

Actes de colloques

## V<sup>e</sup> conférence du DLM-Forum

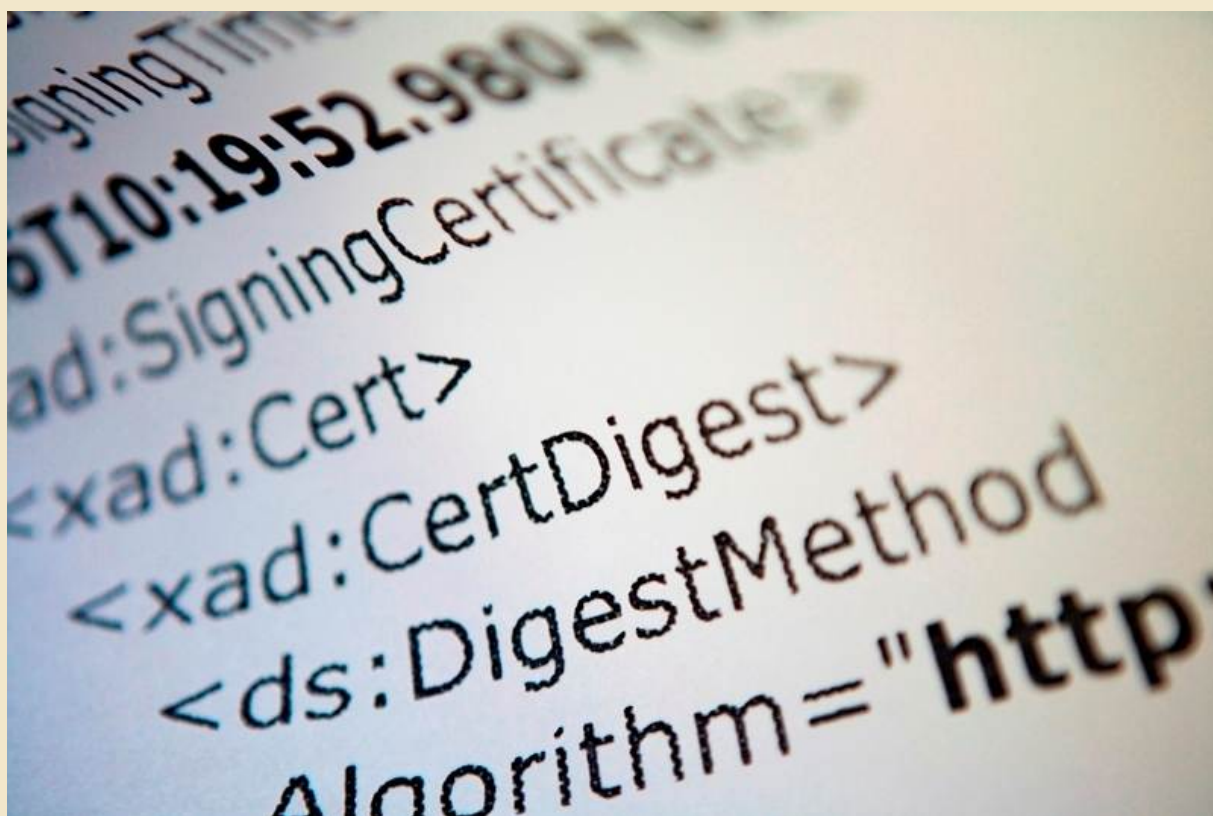
### V<sup>th</sup> DLM-Forum Conference

La gestion de l'information et des archives électroniques  
en Europe : réalisations et nouvelles directions

Toulouse, décembre 2008

*Information and Electronic Records Management in  
Europe: Achievements and new Directions*

*Toulouse, December 2008*



Volume 1

Direction des Archives de France

## **Actes de la V<sup>e</sup> conférence du DLM-Forum**

- La gestion de l'information et des archives électroniques en Europe :  
réalisations et nouvelles directions -

Toulouse, 10, 11 et 12 décembre 2008

## ***V<sup>th</sup> DLM-Forum Conference Proceedings***

- *Information and Electronic Records Management in Europe: Achievements  
and new Directions* -

*Toulouse, 10, 11 & 12 December 2008*

**Volume 1**

**2009**

La préparation à l'édition des actes de ce colloque a été assurée par Jean-François MOUFFLET, conservateur au département de l'innovation technologique et de la normalisation de la direction des Archives de France, avec le concours de Pierre FOURNIÉ, chef du département des publics, et Florence BARRETO, son assistante. Les contributions ont été réunies par Christine MARTINEZ, responsable de la cellule internationale de la direction des Archives de France, Dimitri SARRIS, son assistant, et Jean-François MOUFFLET.

*The preparation of the proceedings of this conference for publication was carried out by Jean-François MOUFFLET, curator in the Département de l'Innovation Technologique et de la Normalisation of the Direction des Archives de France, with the help of Pierre FOURNIÉ, head of the Département des Publics, and Florence BARRETO, his assistant. The papers of the speakers were put together by Christine MARTINEZ, head of the International Relations of the Direction des Archives de France, Dimitri SARRIS, her assistant, and Jean-François MOUFFLET.*

En couverture

Aux Archives municipales de Limoges, archivage des Marchés Publics dématérialisés, technologie de pointe en matière d'horodatage et sécurisation des transferts.

© Direction des Archives de France, photographie François Carlet-Soulages, 2007.

© Direction des Archives de France, Paris, 2009

ISBN : 978-2-86000-339-1

En application du Code de la propriété intellectuelle (articles L 122-4 et L 122-5), toute reproduction ou traduction intégrale ou partielle faite sans le consentement de la direction des Archives de France et de l'auteur ou de ses ayants droit ou ayants cause est illicite. Par conséquent, les copies ou les reproductions du présent ouvrage sont strictement réservées à l'usage privé du copiste. Sous réserve que soient indiqués clairement le nom de l'auteur, celui de la direction des Archives de France et la source, la représentation ou la reproduction d'extraits du présent ouvrage à des fins exclusives d'illustration dans le cadre de l'enseignement et de la recherche sont autorisées dès lors que le public auquel cette représentation ou cette reproduction est destinée est composé majoritairement d'élèves, d'étudiants, d'enseignants ou de chercheurs directement concernés et que l'utilisation de cette représentation ou cette reproduction ne donne lieu à aucune exploitation commerciale.

Publié en ligne le 29 avril 2009 (version 1).

# Note liminaire

## La V<sup>e</sup> conférence du DLM-Forum



Créé à l'initiative de la Commission Européenne, le DLM-Forum est une association regroupant des professionnels issus des archives publiques, des consultants et des vendeurs spécialisés dans la gestion, la conservation et l'archivage des documents électroniques. Son but est favoriser la coopération et l'échange dans ce domaine, principalement au sein de l'Union Européenne. Tous les trois ans se tient une conférence générale sur ces thèmes et la responsabilité de cette organisation repose sur l'état-membre qui préside l'Union Européenne.

La V<sup>e</sup> conférence du DLM-Forum a été organisée par la direction des Archives de France (Ministère de la Culture et de la Communication) en partenariat avec l'Association des Archivistes français dans la cadre de la présidence française de l'Union Européenne. Elle a été accueillie à Toulouse du 10 au 12 décembre 2008. Son objet principal était de mettre en perspective les réalisations, les outils et les nouvelles directions constatées au niveau européen pour la gestion et l'archivage des documents électroniques. Par ailleurs cette rencontre était l'occasion idéale pour présenter le nouveau modèle d'exigences types pour la maîtrise de l'archivage électronique : MoReq2, paru en février 2008, et qui est une mise à jour des précédentes spécifications, supervisée par le DLM-Forum et la Commission européenne.

## Remarques sur le recueil

Chaque article est précédé du résumé envoyé par les auteurs lors de l'appel à communications. Ces résumés sont en majorité disponibles en anglais et en français afin que l'absence de traduction intégrale n'empêche pas le lecteur de prendre facilement et rapidement connaissance du contenu des articles. En outre ces résumés ont leur utilité quand le texte n'a pas été transmis par les auteurs. Les biographies des auteurs sont insérées à la fin de l'ouvrage.

Ce recueil est édité uniquement sous forme électronique : il a donc été conçu pour être le plus interactif possible. Des liens hypertexte permettront au lecteur :

- d'atteindre l'article désiré depuis le sommaire.
- de consulter la biographie de l'auteur en cliquant sur son nom. A l'inverse, il pourra regagner l'article depuis la biographie en cliquant de nouveau sur le nom.
- de télécharger après le résumé des articles le diaporama présenté sur place par les intervenants. Ces diaporamas ont été convertis au format PDF.
- éventuellement de consulter en pleine résolution *via* un navigateur Internet les images de basse qualité insérées dans les articles en cliquant sur celles-ci.

## Version et évolution de ce document

Ce document est la première version de la publication électronique du recueil, publiée le 29 avril 2009 sur le site de la direction des Archives de France.

Cette version peut être amenée à évoluer, notamment au cas où des intervenants viendraient à transmettre le texte complet de leur intervention, ou si une traduction bilingue complète du recueil serait effectuée.

# Notice

## **The V<sup>th</sup> DLM-Forum Conference**



Created on the initiative of the European Commission, the DLM Forum is a community of professionals working in public archives, consultants and commercial providers specialising in the in the area of the management, preservation and archiving of electronic documents. Its aim is to promote co-operation and sharing in this field, mainly throughout the European Union. Every three years, a general conference on these themes is organised by the Member State that holds the presidency of the European Union.

The V<sup>th</sup> DLM-Forum Conference was organised by the Direction des Archives de France (Ministry of Culture and Communication) and the Association des Archivistes Français during the French Presidency of the European Union. It took place in Toulouse from the 10<sup>th</sup> to the 12<sup>th</sup> of December 2008. Its main goal was to focus on European achievements, tools and new directions in the field of digital records and archives management. The event was also the ideal opportunity to present the new Model Requirements for the Management of Electronic Records - MoReq2, published in February 2008 as an update to the previous specifications and overseen by the DLM-Forum and the European Commission.

### **Some comments on this publication**

Each paper is preceded by the abstract sent by the authors during the call for papers. Most of the abstracts are available both in English and French so that, in spite of the lack of a complete translation, the reader can easily and quickly find out about the content of the paper. Furthermore the abstract is useful in those cases where the full paper has not been sent by the author. Short biographies of the authors follow the papers at the end of the publication.

These proceedings are only available as an electronic publication: they were consequently designed to be as interactive as possible. Hypertext links will allow to the reader:

- to access from the summary the selected paper.
- to consult the speaker's biography from the paper by clicking on his name; conversely, it is possible to go back to the paper from the biography by clicking once again on the name.
- to download after the abstract the slides of the presentation as delivered during the conference. These slides have been converted into PDF documents.
- as a possibility, to have with a web browser a full resolution view of low-quality pictures inserted in the articles by clicking on them.

### **Version and further evolutions of the publication**

This document is the first version of the electronic publication of the proceedings which was published on the Direction des Archives de France website on 28<sup>th</sup> April 2009.

This version may evolve, particularly if some speakers provide the full text of the missing papers or if a complete translation of the proceedings in the two languages is undertaken.

# Sommaire / Summary

## Volume 1

### ***Séance plénière d'ouverture*** \_\_\_\_\_ **10**

### ***Plenary opening session*** \_\_\_\_\_ **11**

*Les Archives sont-elles, dans le cadre des stratégies nationales, des partenaires pour affronter les défis du monde numérique ? Quel rôle peuvent-elles jouer ?* \_\_\_\_\_ **12**

Martine de Boisdeffre, Directrice des Archives de France

*The Tortoise and the Hare* \_\_\_\_\_ **17**

Martin Berendse, Director of the Nationaal Archief of the Netherlands

*A View from the Other Side: Digital Continuity in New Zealand - Standards, Shared Services and Open Source Solutions* \_\_\_\_\_ **19**

Evelyn Wareham, Programme Manager Digital Sustainability, Archives New Zealand

### ***Séance 1 - L'administration électronique et les normes*** \_\_\_\_\_ **28**

sous la présidence de Raimo Pohjola, ancien directeur général adjoint des Archives nationales de Finlande

### ***Session 1 - e-Government and Standards*** \_\_\_\_\_ **29**

chairing Raimo Pohjola, former Deputy Director General, of the National Archives of Finland

*Interoperability cross borders and over time - challenges and approaches* \_\_\_\_\_ **30**

Karl Wessbrandt, Strategic Development Officer, Swedish Administrative Development Agency

*Transformation, e-government and electronic records management: a strategic approach to linkage* \_\_\_\_\_ **36**

Stephen Harries, Independent Consultant, Critical Information (UK, Brighton)

*Rôle et applicabilité des normes* \_\_\_\_\_ **43**

Marie-Anne Chabin, Expert en archivage, Archive 17

*Le Costa Rica et l'administration électronique* \_\_\_\_\_ **52**

María Teresa Bermúdez-Muñoz, Professeur à l'Université du Costa Rica

## **Séance 2 - La gestion des archives électroniques et leur accessibilité** 68

sous la présidence de Claes Gränström ancien directeur adjoint des Archives nationales de Suède

### **Session 2 - Managing and accessing e-information** 69

chairing Claes Gränström, former Deputy Director of the Swedish National Archives

*Shooting Ourselves in the Foot: Is Digitization Decaying the Archival Aura?* 70

Adam B. Cowling, MAS/MLIS Candidate, University of British Columbia

*The impact of the UK Freedom of Information Act on digital records management in the public sector* 77

Dr Elizabeth Shepherd, Reader in Archives and Records Management

Dr Alice Stevenson, Research Associate, School of Library, Archive and Information Studies, UCL

*Creating access to electronic records: A two-level approach* 85

Yun Petersen Yoneyama, Special Advisor, Project manager for the Accessibility Project, Danish National Archives

*Traitement et valorisation des flux électroniques, un exemple de Mémoire d'Entreprise.* 96

Marius Delail, Chef du service Gestion de l'Information et de la connaissance, Centre National d'Etudes Spatiales (CNES), Toulouse

## **Séance 3 - Du records management aux archives : développer des stratégies pour transférer les données** 106

Sous la présidence de Ian McFarlane, spécialiste des documents et de l'administration électroniques pour le gouvernement britannique

### **Session 3 - From records management to archives - developing strategies to transfer records** 107

Chairing Ian McFarlane, formerly specialist in electronic records and e-government for the UK Government.

*Transfer and ingest of digital records in national archives: a comparative study of European countries* 108

Kuldar Aas, National Archives of Estonia

Raivo Ruusalepp, Estonian Business Archives Consultancy

Zoltan Szatucsek, National Archives of Hungary

*Revising the Standards for Digital Transfers* 118

Jan Dalsten Sørensen, Principal Consultant, Head of Section for Digital Archiving, Department of Appraisal and Transfer, National Archives of Denmark

*Vers une décentralisation et une démocratisation encadrées du records management au sein des grandes entreprises : le projet ISIS*\_\_\_\_\_124

Sabine Mas, Professeure adjointe à l'EBSI, Université de Montréal (Canada)  
Yves Marleau, Président de Cogniva Information Solutions Inc. (Canada)

*Archives and the architecture of information society*\_\_\_\_\_136

Markku Nenonen, Deputy Director General, National Archives National Archives of Finland

*Preserving European Security and Defence Policy Mission Records*\_\_\_\_\_145

Stephen Howard, Chief Registry and Archives, EULEX-Kosovo

*La problématique de la sauvegarde numérique des e-mails en Belgique : directives et recommandations pour les administrations publiques*\_\_\_\_\_154

Sébastien Soyez, Attaché scientifique, Archives générales du Royaume de Belgique et Archives de l'Etat dans les Provinces

*Archivage de la Banque d'information radiophonique des Deux-Sèvres (BIRDS)*\_\_\_\_\_170

Brigitte Pipon, Directrice des Archives départementales des Deux-Sèvres

*The Nationaal Archief of the Netherlands as a Trusted Digital Repository for record creators and archival institutions in the Netherlands*\_\_\_\_\_179

Jacqueline Slats, Head of Digital Longevity Department, Nationaal Archief - National Archives of the Netherlands

## **Séance 4 - La conservation pérenne des données numériques**\_\_\_\_\_182

sous la présidence de Francesco Barbedo, Instituto dos Arquivos Nacionais, Portugal

## **Session 4 - Long-term preservation**\_\_\_\_\_183

chairing Francisco Barbedo, Instituto dos Arquivos Nacionais, Portugal

*La sélection des documents dans le système numérique*\_\_\_\_\_184

Roberto Guarasci, Anna Rovella, Maria Taverniti  
Università della Calabria, Rende (CS), Italia

*Preservation planning: Why & how to implement? A case study from the Nationaal Archief of the Netherlands*\_\_\_\_\_194

Bart Ballaux, Senior Advisor Digital Longevity, Nationaal Archief - National Archives of the Netherlands

*Archiver et diffuser l'information au sein du système SPAR de pérennisation de l'information numérique*\_\_\_\_\_196

Charlotte Fabre, Consultante EIM, responsable fonctionnelle SPAR, Atos Origin  
Thomas Ledoux, Chef de projet SPAR, BnF  
Laurent Sollier, Architecte, responsable technique SPAR, Atos Origin



*La plate-forme d'archivage électronique du Département des Yvelines : des enjeux stratégiques  
aux choix techniques* \_\_\_\_\_ **203**

Elisabeth Gautier-Desvaux, Directeur des Archives départementales des Yvelines

Gaëlle Mignot, Chargée de mission informatique aux Archives départementales des Yvelines

***Intervenants*** \_\_\_\_\_ **214**

***Speakers*** \_\_\_\_\_ **215**

# Séance plénière d'ouverture

---

---

# ***Plenary opening session***

---

---

# LES ARCHIVES SONT-ELLES, DANS LE CADRE DES STRATÉGIES NATIONALES, DES PARTENAIRES POUR AFFRONTER LES DÉFIS DU MONDE NUMÉRIQUE ? QUEL RÔLE PEUVENT-ELLES JOUER ?

---

Martine de Boisdeffre

Directrice des Archives de France

## ■ Les révolutions numériques en cours

Les révolutions technologiques, pour reprendre le titre d'une table ronde organisée dans le cadre du colloque international de l'Unesco, qui s'est tenu les 18 et 19 novembre derniers sur le thème « Les archives demain », impactent fortement, comme l'ensemble des secteurs d'activité, la profession et la pratique des archives.

La première révolution, c'est la démocratisation de l'accès aux ressources culturelles, illustrée brillamment par l'ouverture il y a trois semaines, du portail Européana ; c'est le développement de la numérisation à grande échelle de fonds archivistiques et leur diffusion à une vitesse prodigieuse sur les réseaux Internet (les 100 millions de pages en ligne sur Internet pour les fonds de l'état civil seront dépassés en France en 2009) tout en s'appuyant sur toutes les potentialités du Web 2.0 en matière de participation collaborative de la part des internautes pour enrichir l'exploitation et la connaissance des fonds.

La deuxième révolution est liée au développement de l'administration électronique et de la dématérialisation des processus métier avec la production d'originaux numériques dont il convient d'organiser la conservation sécurisée, comme on le faisait pour le papier. Cette révolution est induite en France par la loi du 13 mars 2000. Désormais, l'écrit sur support électronique est admis en preuve au même titre que l'écrit sur support papier : « *sous réserve que puisse être dûment identifiée la personne dont il émane et qu'il soit établi et conservé dans des conditions de nature à en garantir l'intégrité* ». De même, la signature électronique est introduite, qui repose en fait sur une technologie particulière, la cryptographie à clé publique, dont la particularité en effet est de pouvoir s'assurer raisonnablement qu'un acte n'a pas été modifié durant la durée de sa transmission entre un expéditeur et un destinataire et que son auteur ne peut pas prétendre ne pas en avoir été l'auteur.

Déjà les administrations ont depuis plus de 20 ans largement informatisé leurs procédures passant ainsi de l'âge des vénérables énormes registres papier et multiples fichiers à entrées multiples, à celui des bases de données qui se transforment pour certaines d'entre elles, en véritables bases de connaissance intégrant des outils d'aides à la production et des outils d'édition. Avec la dématérialisation des documents intimement liés à l'utilisation d'outils de signature électronique, on bascule dans un autre univers avec une exigence absolue de sécurité et de confiance, tant les imperfections existantes dans l'univers papier avec ses difficultés de classement, sa poussière et ses locaux encombrés ne peuvent plus être supportables dans l'univers numérique où toute approximation et erreur se traduisent par des pertes énormes d'informations, des risques vertigineux en matière de falsification de données ou même d'accès non autorisé à des millions d'informations confidentielles. Or la dématérialisation, dans un objectif évident et parfaitement justifié d'efficacité, rapidité des traitements, amélioration des services et économie de moyens, touche tous les secteurs jusqu'aux plus sensibles (les données et documents touchant à l'état des personnes, leurs biens, leur intimité) ou aux plus stratégiques, sans peut-être que tous les risques afférents soient bien mesurés.

Observer les pratiques actuelles, très majoritairement mises en œuvre dans les bureaux en matière de production bureautique et de messageries sans que le travail des agents ait été encadré par des pratiques de

qualité de gestion de l'information, nous fait tous frémir : combien de cas de fichiers électroniques mal nommés, mal gérés, mal classés et par là même inaccessibles, sans gestion de versions et par conséquent sans possibilité de leur accorder la moindre confiance sur la véritable version validée, combien d'enregistrements inexistantes ou lacunaires, de messageries débordantes et inorganisées, de formats devenus avec le temps non interprétables, d'intranets sans procédures rigoureuses d'alimentation, sans fiches de métadonnées structurées associées...

Plus fondamentalement que dire d'un système d'information dont on n'aurait pas pris en compte dès sa conception le cycle de vie des informations produites (durée de conservation, sort final) ? En résultent des ruptures de chaînes, un affaiblissement de la force de preuve voire des risques de pertes de données.

Que dire encore d'un projet de gestion électronique de documents n'intégrant pas cette réflexion sur le cycle de vie des informations scannées ? Ou encore de systèmes d'information développés avec une vision exclusivement verticalisée sans imaginer une fonction d'archivage électronique commune à plusieurs de ces systèmes et permettant par là même des économies d'échelle et des gains substantiels ?

Sans parler de la dématérialisation totale de l'information, visant à ne produire que des originaux numériques. On le voit déjà avec le processus comptable dans les collectivités, les offres de marchés électroniques, les délibérations et arrêtés des collectivités soumis au contrôle de légalité de l'Etat, et bientôt les minutes et les dossiers judiciaires et pourquoi pas les registres de l'état civil, pour lesquels on ne s'appuierait pas sur un mode de conservation sécurisée ?, tant il est vrai que l'information numérique n'est que le résultat d'une combinaison réussie à une date X entre des matériels, des logiciels, des périphériques, des systèmes d'exploitation et que prétendre conserver plus de 10 ans une information de cette nature, nécessite de récupérer et contrôler cette information de la manière la plus autonome par rapport à son environnement technologique, afin de pouvoir la restituer des années plus tard, dans l'environnement alors existant.

Les solutions pour tenter de répondre à ce véritable défi (une information numérique n'est pas faite pour durer) sont connues : une structuration de l'information puissante (métadonnées et classification), une réplique des données sur des sites distants, une surveillance des supports, une migration de ces supports voire des formats dans lesquels l'information est encodée (qui peut lire aujourd'hui un document écrit avec Word 2 ?), une mutualisation des plateformes d'archivage électronique. Encore faut-il avoir la possibilité de les mettre en œuvre !

## ■ Qu'en est-il pour les institutions d'archives face à cette révolution ?

### Le contexte

Il s'agit de vénérables institutions (je rappelle que nous venons de célébrer les 200 ans d'installation des archives nationales à l'hôtel Soubise) qui, bien avant leur rôle en matière de patrimoine, sont les artisans de la conservation de valeur de preuve des activités des individus et des Etats.

En effet, la maîtrise de l'information produite et reçue par les organisations permet, au-delà du rassemblement d'un patrimoine essentiel à l'histoire et plus largement, à l'ensemble des sciences humaines, de faciliter la prise de décision et les arbitrages ainsi que de disposer rapidement de l'ensemble des preuves juridiques nécessaires à la reconnaissance des droits et enfin, de pouvoir répondre à l'obligation de transparence de l'administration vis-à-vis des citoyens.

Or les Archives, même si leur organisation et leur prise en charge sont régulées par des lois, décrets et règlements depuis des dizaines d'années, même si la conservation des archives dites historiques se fait dans des bâtiments de plus en plus performants, devenus en quelques décennies de véritables lieux culturels, même si leur traitement est confié à des professionnels scientifiques et techniques de haut niveau, en amont, dans les administrations et les entreprises, les archives sont une réalité mouvante, difficile à cerner, perçue souvent comme une charge en terme de locaux dans les administrations, peu valorisée, dont la fonction est mal identifiée.

Or, les révolutions numériques sont une chance à saisir par la profession parce que c'est à travers les nouveaux défis qui se posent dans ce domaine que les Archives vont pouvoir prouver qu'elles sont de véritables leviers pour la modernisation de l'Etat. C'est d'ailleurs cette dimension qui a été mise en avant par l'audit sur l'archivage dans les ministères mené en 2007 dans le cadre des différentes vagues d'audits lancés alors pour préparer la modernisation de l'Etat.

### **Une collaboration étroite avec la direction générale de la modernisation de l'Etat**

Et d'abord en travaillant en complémentarité étroite avec les organes de l'Etat chargés de cette modernisation. Déjà en 2004, l'archivage numérique a été inscrit dans le plan stratégique de l'administration électronique par le comité interministériel de la Réforme de l'Etat du 2 février 2004 (action ADELE 103 devenue depuis 2006, une initiative fonctionnelle IF-08) : en effet, « *l'administration électronique, pas plus que l'administration traditionnelle, ne peut être efficace si elle est amnésique. C'est pourtant le risque qu'elle court actuellement, tant en raison de l'absence de prise de conscience de la nécessité de l'archivage électronique que de la faiblesse des moyens et des procédures disponibles pour l'assurer* ». C'est ainsi qu'une collaboration fructueuse s'est nouée avec l'agence pour le développement de l'administration électronique puis la direction générale de la modernisation de l'Etat, dans plusieurs domaines.

*De nombreuses actions de sensibilisation ont ainsi été conduites conjointement* : des matinées dites « ADELE » organisées par la DGME qui permettent d'attirer au-delà des archivistes (producteurs, informaticiens, secteur privé), aux multiples interventions en commun devant le milieu professionnel des archivistes, mais également des services administratifs et des informaticiens, devant les associations représentatives des collectivités territoriales, dans des salons, des séminaires relatifs à la dématérialisation et l'administration électronique.

*Des référentiels ont été élaborés en commun* : par exemple, une politique d'archivage sécurisé dans le secteur public, sous l'égide de la direction centrale de la sécurité des systèmes d'information (DCSSI), qui fixe les conditions minimales d'ordre juridique et technique, mais également fonctionnel et organisationnel, pour qu'une archive conserve sa valeur juridique tout au long de son cycle de vie et notamment détermine les rôles et responsabilités des différents acteurs du processus.

*De même a été publié en 2006 le standard d'échange de données pour l'archivage* qui vise à permettre l'interopérabilité et par conséquent à faciliter les échanges entre des services producteurs et des services d'archives : il définit tant les messages échangés que la structure et le contenu du bordereau de versement qui accompagne les fichiers à transférer pour archivage. A leur arrivée dans le service d'archives, les informations portées dans le bordereau de versement (métadonnées) sont intégrées dans la base de données descriptive des archives (qui permet de retrouver les archives de multiples producteurs de données) et les fichiers sont écrits sur les espaces de stockage sécurisés (soit en ligne, soit hors ligne sur supports amovibles).

Enfin, cette collaboration a abouti à l'actuelle réalisation d'une plate-forme d'archivage électronique pilote (PIL@E) qui sera utilisée aux Archives nationales, à Fontainebleau tandis qu'ont été impulsées des actions dans les collectivités territoriales<sup>1</sup>, le tout dans le contexte de l'émergence d'un marché de tiers-archivistes.

Aujourd'hui cette collaboration va se poursuivre à travers l'action sur l'archivage électronique inscrite dans le plan Economie numérique 2012 du secrétaire d'Etat Eric Besson et plus précisément à travers le plan stratégique pour l'administration électronique dans lequel l'archivage est inséré, dont l'élaboration et la mise en œuvre sont portées par la DGME.

### **Des pratiques de records management clés de la réussite**

Avec la production numérique, ce sont les pratiques de « record management » qui sont mises en avant avec la nécessité impérieuse pour les archivistes et autres professionnels de l'information d'intervenir en amont auprès des producteurs là où se constitue l'information, car si l'on veut qu'un document soit conservé autant que nécessaire, il faut impérativement qu'il soit, dès le début de son cycle de vie, enregistré, classé, tracé, assorti d'attributs relatifs aux droits d'accès et aux durées de conservation.

Ainsi depuis trois à quatre ans, les actions se multiplient d'archivistes qui deviennent de véritables prestataires de services pour des services producteurs et des services informatiques, dans le cadre de projets

<sup>1</sup> Par exemple, la mise en œuvre d'une plate-forme pour les archives départementales des Yvelines.

de gestion électronique de documents qui fleurissent, à travers la prise en compte du cycle de vie et de la structuration de l'information des applications métier qui ne cessent de se complexifier et de voir leur champ s'étendre, et enfin à propos des nouveaux projets de dématérialisation complète avec production de flux XML.

Je citerai l'exemple de l'action remarquable aujourd'hui menée sur les dossiers des maisons départementales des personnes handicapées (MDPH) qui ont pris en charge les anciens dossiers des Cotorep et de la CNES pour les adultes et jeunes handicapés. L'association des MDPH a souhaité créer un groupe de travail sur la numérisation et l'archivage numérique des dossiers à la faveur des multiples projets de GED et de dématérialisation en cours. Le grand intérêt de ce groupe est qu'il est constitué de directeurs des MDPH mais également d'informaticiens et d'archivistes (archives départementales, direction des Archives de France). Cette complémentarité s'est révélée d'une grande richesse. En quelques mois, a été ainsi élaboré un vademecum sur la conduite d'un projet de GED dans une MDPH. Généralement un projet de GED est conduit très rapidement avec un objectif unique de gain de temps de recherche et de gain d'espace (on élimine les dossiers papier après numérisation), avec une numérisation massive externalisée du stock et une étude de coûts trop rapidement menée et limitée car n'intégrant que le très court terme en se limitant aux seuls coûts de la numérisation proprement dite.

Les réflexions menées autour de ce projet ont au contraire prouvé qu'il fallait élargir la réflexion de manière à pouvoir proposer des solutions porteuses d'un véritable gain de productivité et d'efficacité et avant tout bien déterminer le périmètre de la GED. Par exemple il est apparu absurde de numériser le stock inactif (c'est-à-dire des dossiers dont la durée d'utilité administrative est dépassée). C'est ainsi que dans ce département-ci (la Haute-Garonne), cette suggestion de l'archiviste a permis de sortir du périmètre de la GED, 25% du stock et d'ainsi faire économiser 50 000 € et beaucoup plus en fait si on considère la charge des serveurs ainsi allégée sur le long terme.

Encore plus finement, la réflexion a permis de se rendre compte qu'il était inutile de scanner des pièces (arrivant en flux) dont le contenu était intégralement repris dans le système d'information existant. Là encore des économies étaient rendues possibles. Et finalement l'ensemble de ces éléments a permis de se rendre compte qu'on pouvait de fait parfaitement éviter le recours à l'externalisation, parce que d'une part le périmètre était en fait beaucoup plus réduit qu'on ne le pensait et d'autre part que recourir à l'externalisation, avait pour conséquence de renoncer à une structuration des dossiers (travail trop onéreux à faire réaliser en externe) et qu'on avait alors des dossiers scannés dont le mode de recherche était trop fruste (l'agent pour retrouver une information devait parcourir tout le dossier sans possibilité d'interrogations plus fines) alors même qu'un des buts de la GED était de permettre aux agents de retrouver très rapidement la bonne information.

Autre exemple : l'accompagnement en amont (au moment de l'écriture du cahier des charges) du volet « archivage » du projet SOLON 2, à savoir la dématérialisation des projets de lois, décrets, arrêtés, tout au long de leur circuit interministériel, jusqu'à la publication au journal officiel, avec introduction d'outils de signature électronique. Outre les questions centrales relatives au cycle de vie de l'information (à partir de quel événement peut-on considérer qu'un dossier est clos et peut être considéré comme un « record » ?), à la durée d'utilité administrative, aux éliminations et transferts pour archivage définitif avec définition du format d'export, les archivistes ont insisté sur la question des formats des fichiers stockés dans SOLON. Aujourd'hui dans SOLON 1, le choix du format a été Word 97. Ceci signifie concrètement que des dossiers créés et stockés dès 2006 sont à ce format-là et devront être conservés sous ce format durant 10 ans, c'est-à-dire jusqu'en 2016 après reprise des données dans SOLON 2 lorsque celui-ci sera mis en œuvre (à partir de 2010, 2011). Personne ne s'était posé la question de la compatibilité entre Word 97 et le futur format pivot qui sera choisi dans SOLON 2 et si, en 2016, il serait encore possible d'ouvrir ces fichiers avec Word 97 ou de les interpréter correctement avec le logiciel qui sera alors utilisé. Une solution devra donc être trouvée. Nous avons d'ores et déjà quelques pistes.

Bref, à partir de simples exemples, on perçoit mieux l'apport essentiel de ces gestionnaires de l'information que sont les archivistes, aux côtés des producteurs et des informaticiens. En devenant de véritables prestataires de service hautement qualifiés, ils apportent un véritable gain qualitatif aux projets, ils permettent grâce à une vision transversale papier/numérique, une maîtrise raisonnée des flux d'information, une amélioration de la qualité du travail des agents, une conservation raisonnée des documents et données (on élimine tout ce qui peut l'être en respectant strictement le cadre réglementaire et à l'inverse, on conserve tout ce qui doit l'être) et, au final, ils sont des instruments pivots d'amélioration de la qualité du travail de l'administration.

Certes, les révolutions en cours représentent un défi difficile à relever pour le monde des Archives : au-delà des multiples réalisations et de la vitalité du domaine qui transparaît à travers les dizaines de communications que nous allons avoir le plaisir d'écouter dans le cadre de cette conférence, les difficultés sont multiples : inégalités croissantes entre les pays, les territoires ; difficultés techniques encore peu ou mal résolues concernant certains types de formats ; problèmes relatifs à la mise en place de répertoires internationaux de formats, à la signature électronique des documents numériques et à sa conservation dans le temps ; questions concernant la conduite de changement à mener dans les services d'archives avec la mise à niveau et la requalification des personnels, le travail à effectuer avec de nouveaux interlocuteurs au premier plan desquels les informaticiens, et cela dans un contexte où l'effort à accomplir pour continuer à prendre en charge le papier dans un univers qui sera de plus en plus hybride, reste immense et se complexifie à la faveur des réorganisations de l'administration.

Et pourtant si nous relevons ce défi, nous aurons réussi à changer l'image des archives et de nos services : non plus des fonctions support mais des fonctions de prestataires de services ; non plus l'organisation d'un stockage d'informations « mortes » mais une conservation dynamique de documents et données dans un contexte de responsabilisation croissante des différents acteurs ; bref, faire des archives une ressource immatérielle stratégique dans un contexte en constante évolution.



# THE TORTOISE AND THE HARE

---

Martin Berendse

Director of the Nationaal Archief of the Netherlands

*The Tortoise told the Hare that he was willing to bet that he could beat the Hare in a race to the other side of the field. The Hare thought that he had plenty of time, so he lay down to take a nap. When he awoke, he saw that the Tortoise had almost reached the finish line. The Hare ran very fast to catch the Tortoise, but he was too late: the Tortoise won the race.*

*The moral of the story is: slow and steady wins the race. This fable is very old, predating even the great Frenchman Jean de La Fontaine. It was written by Aesop, a Greek storyteller who lived 2500 years ago. The fable and the proverb 'slow and steady wins the race' illustrate the fact that persistence pays off in the end. The metaphor of the Tortoise and the Hare can also be applied to the world of government information. Governments have been producing documents on paper for centuries, but these only seem to be valued as long as they are necessary in the administrative process to provide proof of claims or to defend a position. Today, governments also has to preserve documents as a service to its citizens, but as soon as they are no longer needed governments runs off like the Hare, leaving the stacks of paper behind for the 'Tortoises' - we archivists. It took quite some time before all countries accepted the principle that paper archives should be transferred to an independently managed and administered national archive, and there are still countries where for example military archives or those of the Ministry of Foreign Affairs are kept separate. In the digital age we currently find ourselves in a situation that somewhat resembles the time before there was any archival legislation. Again there are very many government officials who do not realise that the files we create on our computers are archives that are subject to legal restrictions, could provide evidence and could contain information of vital importance for our citizens. We are therefore faced with the challenge of increasing awareness of this issue, and we at the Nationaal Archief have adjusted our strategy accordingly, or rather; we have developed a new strategy to meet this challenge. The key to good governance is not only in developing the correct standards and then imposing them from above, as we archivists would like to do, but in creating awareness through communication, a good administrative environment through politics and the correct judicial framework through legislation. The new informational film that the Nationaal Archief has commissioned for Dutch politicians and leading civil servants best illustrates our new strategy. The purpose of the film is to inform and warn the Dutch population about a new phenomenon in our history: a gap in government memory. The developments in digital technology over the past few decades have come so quickly that we have forgotten to think about storing all of the digital files we read and send.*

*Luckily, we have realised this in time, and the looming collapse has spurred many to take a more innovative approach to access to government information. The Dutch Government started in 2006 the programme 'Information in Order' with the statement: digitalisation makes government more accessible to its citizens and increases their influence on policy, while improving the service provided to the citizens as 'customers'. Dutch Government wishes to take advantage of the opportunities offered by Information and Communications Technology (ICT) to improve the standard of service to the business community and the general public. The use of ICT for such public administration purposes is often referred to as 'the Electronic Government' or simply 'egovernment'.*

*Since the previous congress, the DLM has been working on the development of MoReq2. This standard will replace MoReq1 to become the European standard for Document and Records Management Systems. Over the coming period, the DLM and the EU hope to continue this process by creating a testing and certification organisation. Unfortunately, the Netherlands had never accepted MoReq1 due to its archivist perspective. Our ministries and agencies did not recognise in it their requirements for business and accountability processes. The new MoReq2 is also a technical standard. This situation in which the standards are not generally accepted is undesirable and provides only apparent solutions.*

*At the European level, parties are also making agreements regarding the foundations of e-government and interoperability. This presents many opportunities for imbedding the requirements for good governance in the structure, but as of yet the archival institutions have not been party to the discussions. In the Netherlands, we are working to create ties between the 'Information in Order' programme and the NEN-ISO norms in the field of records management via the Nederlandse Overheid Referentie Architectuur, or NORA (Dutch Government Reference Architecture).*

*Almost all government organisations are busily engaged in contributing to e-government. In so doing it is important to establish principles concerning the way in which it is possible to collaborate, to link processes smoothly to one another and exchange data. NORA was based on the European Interoperability Framework (E.I.F.) 1.0 and adopted its main principles. The challenge for NORA is to achieve interoperability, better*

*chain management, function clarification, the legally mandated re-use of data, and the strive for more uniform definitions and better harmonized laws.*

*The NEN-ISO norms are widely accepted in the Netherlands, in part because ministries and agencies participated in the design phase. This brings us closer to our goal: integrating records management in archive builders' digital operational processes from the very beginning in order to enable the transfer of digital archives in an orderly and accessible manner. This may seem very obvious, but it rarely occurs in reality. When we in the Netherlands ask why our bureaucracy does not stand up and cheer the implementation of our standards, we have to look for the answer in our own actions. We do not gather to discuss the matter, and as you know, we in the Netherlands have a long tradition of discussing matters as long as necessary to find a mutually-acceptable solution (the "polder"-model). When making decisions, the Dutch are the Tortoises of Europe; crawling slowly from one position to the next. We meet, discuss and debate until we are almost ready to collapse, but once we've made a decision then we stick to it. This enables us to beat any headstrong Hare any day of the week. Tortoises of the world, Unite!, I would say.*

*In summary: in the world of e-government and digital preservation I perceive a holy trinity, a trias digitalibus, if you will, consisting of administration, technological standards and legislation. These three elements cannot be considered separately because they are indissolubly connected to one another. European governments are therefore warned: do not try to run away with digital government information like the Hare, because the Tortoise will eventually pass you by in the end.*

Film about the Digital Nationaal Archief shown during the presentation:

[http://www.nationaalarchief.nl/archiefbeheer/DNA/dna\\_eng.asp](http://www.nationaalarchief.nl/archiefbeheer/DNA/dna_eng.asp)

# A VIEW FROM THE OTHER SIDE: DIGITAL CONTINUITY IN NEW ZEALAND – STANDARDS, SHARED SERVICES AND OPEN SOURCE SOLUTIONS

---

Evelyn Wareham

Programme Manager Digital Sustainability, Archives New Zealand

*The Public Records Act 2005 has raised the profile of records across the state sector and local government significantly in New Zealand. Implementation of the Act is supported through mandatory standards for digital information management. Our next challenges are the establishment of an independent audit programme covering over 2,500 public offices, and the expansion of capacity for digital continuity and archiving.*

*Working in cooperation and partnership is the norm in New Zealand's government sector. This applies to our work on standards development for records management, to potential sharing of services, to pooling intellectual resources and to adoption of open source solutions. In the area of digital sustainability, Archives New Zealand has agreed that it cannot make progress without collaborative relationships - with national partner institutions, with government agencies and with archives authorities in Australia, North America and Europe. We particularly value our role as partners in the Australasian Digital Recordkeeping Initiative.*

*This paper will give an overview of current cooperative initiatives in New Zealand, lessons learnt and paths yet to tread. It will discuss the development of a whole-of-government 'Digital Sustainability Strategy' for public sector information. This strategy aims to set a vision, goals and action areas for creation, management and preservation of digital information. Actions proposed include possible shared services for current and semi-current records, development of standards, establishment of a multi-disciplinary community of practice and consideration of public access implications. The strategy also envisages major technology projects for digital preservation. The paper will also talk about Archives New Zealand's experiences with development of an interim digital archive based on open source software (Fedora and Fez). Our aim is to be innovative, cost efficient and to implement solutions based on the continuum model. The Public Records Act 2005 presents a golden opportunity to build momentum and progress in this area across government -- our challenge is not to miss it.*

[See the presentation.](#)

## **La conservation du numérique en Nouvelle-Zélande : normes, services partagés et outils libres**

*La loi de 2005 sur les archives publiques (Public Records Act) a permis aux archives du secteur public et du gouvernement local en Nouvelle-Zélande d'être beaucoup plus accessibles. La mise en application de la loi s'appuie sur des normes obligatoires en matière de gestion des informations numériques. Nos prochains défis seront la mise en place d'un programme d'audit indépendant concernant plus de 2 500 entités publiques et le développement de capacités en matière de pérennité et d'archivage numériques.*

*Le travail en coopération et en partenariat est la règle au sein du secteur public néo-zélandais. Ceci concerne notre travail sur le développement de normes de records management, l'éventuel partage de services, la mise en commun des ressources intellectuelles et l'adoption d'outils libres. En ce qui concerne la pérennité du numérique, les Archives de Nouvelle-Zélande reconnaissent qu'elles ne peuvent évoluer sans la collaboration des institutions nationales partenaires, des agences gouvernementales et des autorités d'archivage australiennes, nord-américaines et européennes. Nous sommes particulièrement fiers de notre rôle en tant que partenaires dans le cadre de l'initiative australasienne de préservation des documents numériques (Australasian Digital Recordkeeping Initiative).*

*Cet article présente les initiatives de coopération menées aujourd'hui en Nouvelle-Zélande, les enseignements tirés et les progrès à réaliser. Il analysera le développement d'une stratégie de pérennité numérique à l'échelle de l'ensemble de l'administration concernant les informations du secteur public. Cette stratégie vise à établir une perspective, des objectifs et des domaines d'action en matière de création, de gestion et de conservation des informations numériques. Les actions proposées comprennent les services partagés possibles pour des archives actives et semi-actives, le développement de normes, la création d'une communauté de pratiques et la prise en compte des conséquences de l'accessibilité publique. La stratégie prévoit également des projets technologiques majeurs en matière de conservation du numérique. Cet article présente également les expériences des Archives de Nouvelle-Zélande pour le développement d'archives numériques provisoires basées sur des logiciels libres (Fedora et Fez). Notre objectif est l'innovation, la rentabilité et la mise en œuvre de solutions qui assurent la pérennité du contenu. La loi de 2005 sur les archives publiques constitue l'occasion idéale pour créer une dynamique et progresser dans ce domaine, notre défi est de ne pas la rater.*

[Consulter la présentation.](#)

You may be asking what is “digital continuity”? Like technology, language is subject to constant innovation, change and redundancy. Developing new words could be called a key characteristic of the archival profession in Australia and New Zealand. Having given the English language “recordkeeping” as a comprehensive term for the full spectrum from records to archives, replacing “records management” and “archives administration”, some key archives authorities in New Zealand and Australia are now moving to the next trend. Experience is showing us that the word “record” is understood too narrowly. If we want government agencies to apply rules and principles to databases, business systems and emerging technologies, the term “information” may be more useful. Likewise, “digital archive” and “digital preservation” limit people’s understanding of our purpose and ambitions to the historical and the static. In the digital environment, we need to influence systems from day one to ensure their longevity, and to continue to intervene actively in order to maintain information for as long as it is needed. This is what we mean by “digital continuity”.

## ■ A View from the Other Side: Australian and New Zealand Approaches

Common approaches to digital recordkeeping characterise the major archives programmes across Australia and New Zealand. In the late 1990s, most major archives authorities in Australia and New Zealand reoriented their programmes to focus on supporting good practice recordkeeping from creation to archiving. The change arose in part from the holistic approach to recordkeeping grounded in the records continuum model, which emphasized the common ground between current records and historical archives. This was part of the same trend that gave us the Australian Standard on Records Management (AS4390) and led to the development of “third generation” archives laws across several Australian states and in New Zealand.

Each of these developments was a response to the challenge of digital information. Previously, the main emphasis had been on appraisal and retention of records of enduring value. Archives had stood as the “ambulance at the bottom of the cliff”, receiving the remnants of office systems. In the digital environment, it had become clear that intervention was needed much earlier in order to ensure that enduring records would result from business practice.

## ■ Continuum Approach

The continuum model, developed by Frank Upward and others at Monash University proposed a unified approach to replace the records life-cycle model’s definite differentiation of records and archives. One of the architects of this model, Sue McKemmish, clarifies that:

“Records continuum thinking focuses on the unifying purposes shared by all recordkeeping professionals, defined as to do with the delivery of frameworks for accountable recordkeeping regimes that enable access to essential, useable evidence of social and business activity in the business, social and cultural domains. Records continuum thinking and practice are underpinned by a concept of records which is inclusive of, not exclusive to, records of continuing value (archives)”<sup>1</sup>

The continuum approach is one of integration - between records and archives, evidence and memory, business and heritage records managers and archivists. It also offers agility - records are not viewed as fixed in one part of the continuum, but free-flowing between dimensions. An archival document is seen as having permanent value even before its creation; this question of status becomes a key to the design phase of systems. Appraisal becomes a process that takes place during the design of systems and focuses on business functions and activities.

---

<sup>1</sup> Sue McKemmish, ‘Yesterday, Today and Tomorrow: A Continuum of Responsibility’, first published in Proceedings of the Records Management Association of Australia 14th National Convention, 15-17 Sept 1997, RMAA Perth 1997. Available online at: <http://www.sims.monash.edu.au/research/rcrg/publications/recordscontinuum/smckp2.html> (accessed 20 November 2008).

## ■ Influencing Records Creators

Over the past decade, archives programmes in Australia and New Zealand have re-focused their energies on influencing records creators. As Adrian Cunningham of the National Archives of Australia stated in 1999: “The National Archives has reinvented itself as a proactive promoter of best practice recordkeeping - our core business now spans the entire records continuum”.<sup>2</sup> Most archives authorities have developed extensive frameworks of standards and guidance for the records creating agencies in their jurisdictions, focusing on digital. Major archives authorities have increased their training and support programmes to advise agencies on how to create and maintain adequate digital records. Much effort has been focused on influencing software developers and vendors to develop products that support the principles and practice of recordkeeping (especially in handling documents and correspondence). Archivists have become (or been replaced by) advisors, with a shift in roles and skills from hands-on technical work with records to communications, change management and consultancy. To persuade decision makers and managers across the public sector of the necessity for change, the language and look of archives programmes have also changed. There has been a move to market archives’ services and key messages through branded programmes like E-Permanence (National Archives of Australia), Continuum (Archives New Zealand) and most recently FutureProof (State Records New South Wales).

## ■ Standards and Influence

Archives New Zealand’s reinvention as an accountability agency with the power to influence records creators had its first major step forward with the establishment of the organisation as a government department (or ministry) reporting directly to a Minister in 2001. Whereas previously, the Chief Archivist had been a lower level manager within a large department, this change made her a Chief Executive who could sit as equals with the heads of foreign affairs, treasury or social development. This was followed in 2003, with the repackaging and launch of Archives New Zealand’s suite of government recordkeeping standards, guidance and services under the brand “Continuum - Create and Maintain”. The Continuum Recordkeeping Resource Kit was launched by the State Services Commissioner in Wellington, and has become a well-regarded source of advice.

Like several other jurisdictions, New Zealand’s standard setting and promotional activity has been backed by a new archives law that empower the archives authority to set mandatory compliance standards. New Zealand’s Public Records Act 2005 was the result of over six years concentrated development effort. It has many similarities with legislation developed a few years earlier in New South Wales and Western Australia. Key developments include a shift from a law that establishing the archival regime, to a law that governs how records are made, managed and disposed of across the state sector.

The Public Records Act, which came into force in 2005, set a new framework for creating and managing information across the New Zealand public sector. The Act’s purpose is to promote government accountability through reliable recordkeeping, enhance public confidence in the integrity of government records and protect New Zealand’s documentary heritage. Core elements of the new framework include:

- a broad and format-neutral definition of a record, which includes digital information of all kinds
- wider coverage – the Act applies to state enterprises and Crown entities such as district health boards, tertiary education institutions, and schools, as well as to the core public service
- a new requirement for organisations to create and maintain full and accurate records in accordance with normal prudent business practice
- a continuation of the existing requirement to dispose of records systematically in accordance with the Chief Archivist’s authorisation
- the authority for the Chief Archivist to set mandatory standards

---

<sup>2</sup> Adrian Cunningham, Sue McKemmish, David Roberts and Barbara Reed, ‘The Australian Records Management Standard’, Society of American Archivists Electronic Records Section Annual Meeting, 1999. Available online at: <http://www.archivists.org/saagroups/ers/AS4390.ppt> (accessed 20 November 2008).

- enhanced compliance mechanisms including inspections, directions to report, and independent audits of recordkeeping practices

A key requirement has been added that requires all public sector agencies to “create and maintain full and accurate records of their activities” and to maintain these in accessible form for as long as they are required. The New Zealand law is also backed by a rigorous compliance framework. It empowers the Chief Archivist to issue mandatory and discretionary standards, and requires her to report annually to Parliament of the state of recordkeeping in the public sector. From 2010, she is also required to commission regular independent audits of recordkeeping practice in all agencies covered by the Act and to report the results of these audits to Parliament.

In 2006, Archives New Zealand worked together with Standards New Zealand to develop a strategy and framework for standards under the Public Records Act. This places heavy emphasis on the role of collaboration and consultation in standard setting. Each standard is developed by an advisory group bringing together representatives of affected agencies and expert interest areas. The mandatory standards issued since 2005 include a standard on electronic recordkeeping metadata and a standard on creation and maintenance of records. Further discretionary standards include a technical specification for electronic recordkeeping systems (primarily targeting electronic document and records management systems), and a standard on digitisation or scanning of analogue records for integration in digital recordkeeping systems. The digitisation standard has been accepted as the basis for development of a new ISO standard. The electronic recordkeeping systems standard, which was based on MoReq1, has been picked up by the International Council on Archives as the basis for module two of its *Principles and Functional Requirements for Records in Electronic Office Environments*.

Audits of recordkeeping in all public offices are a requirement of the Public Records Act. These will commence in 2010. The 2683 organisations covered by the Act will be audited over five years from 2010 to 2015. Archives New Zealand’s approach to audits will be designed to support the desired outcome - good recordkeeping in government that supports good business practice. The audits will be focused on continuous improvement, risk analysis and outcomes.

Since the Act passed into law, Archives New Zealand’s research on the state of recordkeeping across all public offices shows that awareness of the Act is growing throughout the public sector. More and more public offices are assessing their compliance, and beginning projects to improve their recordkeeping. Archives New Zealand’s 2007 government recordkeeping survey indicated that 70 percent of responding public offices now have policies and procedures in place to assess their compliance with the Act. This is a substantial increase from 2006, when only 46 percent of respondents to the survey indicated they had such policies and procedures. There are also other indicators of increasing awareness. The number of state sector staff attending Archives New Zealand’s government recordkeeping forums in Wellington has doubled since 2005. Archives New Zealand is also aware that many public offices and local authorities are seeking to employ specialist recordkeeping staff to manage their records and information.

## ■ Digital Continuity: The Remaining Problem

The Public Records Act defines “records” as any information that is compiled, recorded or stored in any format. Digital information is clearly covered by this definition. This includes emails, websites, databases, business systems, word processing documents, text messages, blogs, wikis, instant messages, video recordings, geographic information systems, and any new technologies introduced into the government work environment. As long as such information is created or received in the course of business, it is covered by the requirements of the Act. Most public offices manage at least some of their records electronically. The most common systems are those used to control financial management and payroll information. In Archives New Zealand’s 2007 government recordkeeping survey, 79 percent of responding public offices indicated that they manage these kinds of records using electronic systems.<sup>3</sup> Comments made by respondents to the survey indicate that many public offices also have specialised digital information systems in place that contain records specific to their business. Examples include systems for managing student records in a tertiary education institution, or clinical records in a district health board. Increasingly, public offices are also recognising the need to actively manage other kinds of digital information such as emails, word processed documents, spreadsheets, and web pages. Increasing numbers are seeking to establish control by implementing document and records management systems or enterprise content management systems. In 2007, 43 percent of public offices indicated that they had such systems -- an increase from 35 percent in 2006.

But despite the increasing awareness of recordkeeping requirements and moves to adopt new systems, digital information poses major challenges across the New Zealand public sector. Digital storage may be increasingly cheap. But digital continuity is a continued cost. Digital continuity is ensuring digital information remains available and useable for as long as it is needed. It means protecting digital files from technological obsolescence. This includes making sure that data is migrated to new formats as appropriate.

67% - that’s two thirds - of New Zealand government agencies report that they hold records that they can no longer access. Of these, 31% have records where controlling index is not available, 20% have records that require computer software or hardware that is no longer available, and 15% have records stored on obsolete storage media. Many are planning steps to deal with this, including records audits to find lost records, migration to new hardware/software/formats and copying or digitising records.

Many issues continue to face recordkeeping across government. Large volumes of information and current recordkeeping systems can make recordkeeping time consuming. Digital recordkeeping methods are often still based on paper paradigms. Solutions are based on technology that is not perfect, such as electronic document and records management systems. Also, advocates for good information management are often not part of senior management teams. There is a lack of awareness that agency culture change is needed in addition to technology responses. And more digital recordkeeping education and training opportunities are needed, especially given that technology is moving very fast - e.g. wikis, blogs, text, Facebook.

## ■ Digital Continuity Strategy

Archives New Zealand is currently leading working to face the challenges of digital continuity through development of a whole of government Digital Continuity Strategy, in support of New Zealand’s Digital Strategy 2.0. This will aim to ensure that agencies can meet the Public Records Act requirement that records remain accessible and useable until they no longer required. The strategy aims to achieve a common vision, approach and understanding of the issues and roles in addressing them. It is being developed with an interdepartmental strategic advisory group, through a consultative process. The New Zealand Cabinet authorised the strategy for release for consultation in August 2008. Since then, the strategy has been available on the web, distributed widely in print and posted online as a collaborative wiki. As a result over 60 written submissions have been received. The aim is to compile, analyse and respond to this feedback and have the strategy back to Cabinet for approval by early 2009.

The key messages in the strategy are as follows:

---

<sup>3</sup> Archives New Zealand, *Government Recordkeeping Survey 2007*, full report available online at: <http://continuum.archives.govt.nz/government-recordkeeping-surveys.html>.



- **There when you need it.** Information will be maintained as long as needed. Some is needed only for a few months, some forever.
- **Authentic and reliable.** Information is tamper-proof and free of technological rights restrictions. It can be trusted to be authentic and reliable.
- **Trusted access.** New Zealanders can be confident that can find and use information that is publicly available, and that their sensitive information will be protected from unauthorised access.
- **Do nothing, lose everything.** If no action is taken, public sector digital information will be lost. We need a proactive approach to maintain information for the future.

A comprehensive programme of goals and actions has been put forward in the strategy. These include infrastructure development for a shared service for storage of government information, and comprehensive digital archiving capacity. As well as these potential technology projects, many actions are proposed to increase understanding and awareness of digital continuity and to enabled sound management of information from day one. The strategy also notes the need for a high level functional appraisal framework and tools to identify information at risk of obsolescence across the public sector. It is expected that the implementation of the strategy will be overseen by a high level governance group, including the Chief Archivist and other relevant departmental chief executives.

## ■ Digital Archive Developments

Given their strong leadership influence at the current records end of the spectrum, where are Australia and New Zealand when it comes to digital preservation programmes? In fact, it could be said that we've been surprisingly slow starters - well behind some European and North American programmes. Why? There is no legacy of digital preservation activity stretching back to the 1960s in our part of the world - in those years most archives authorities in the region were in their infancy and resources would not have stretched to specialist areas. For much of the 1990s, influenced by the ideas of David Bearman and others, there was vociferous debate on a "post-custodial" approach to the management of digital archives. Agencies would retain digital records for the long term and archives authorities would act as controllers, advisors and facilitators of public access. Through strong influence on front end systems in agencies, there would be guarantees of reliable maintenance of archives within business systems. Perhaps in the digital environment of the 1990s, when significant digital systems were complex, proprietary and full of customised transactional data this post-custodial approach may have seemed the only option. However, in today's recordkeeping environment, where much information is created in standardised formats and systems, a centralised approach to at least part of digital archiving has become the accepted approach. The National Archives of Australia quietly moved to a custodial approach to digital archives several years ago, and others have followed suit, although in many cases resources have not been available to make this a reality.

Using Martin Berendse's metaphor, Australian and New Zealand archives are the hare, rather than the tortoise when it comes to digital preservation programmes. The first archives authority in our part of the world to take major steps into digital archives management was the Public Records Office of Victoria, with its Victorian Electronic Records Strategy (VERS). This programme, developed in partnership with the state's scientific and industrial research department, broke new ground for digital preservation. It continues to be exceptional in its integration of office systems in agencies with a flow to archives. The approach is grounded in creation of records in standardised, metadata encapsulated formats within agency systems, enabling smooth and efficient transfer to archives. The digital archive has been operational for some years now, and contains several hundred million records known as "VERS encapsulated objects" or VEOs. In contrast, the other well established digital archive in the region has taken a very different slant. The National Archives of Australia has taken a "format agnostic" approach, allowing most types of records to be transferred to the archives in their original form, with accompanying metadata. Once received, records are normalised to XML using an open source tool, Xena, developed by the National Archives. In its early phases, the system has sat at a distance from the National Archives of Australia's standards and advice programmes for records creators.

The national libraries of Australia and New Zealand have also shown leadership in digital preservation. The PANDORA project at the National Library of Australia was one of the earliest web harvesting initiatives, and it has been followed by early systems development to support mass digitisation and maintenance of born-digital

materials. This year, the National Library of New Zealand launched its National Digital Heritage Archive, a leading edge technology project developed together with a network of commercial and professional partners in Europe, Israel, North America and Asia. Many other archives and libraries in Australia and New Zealand are still starting out or seeking the necessary funding to achieve their goals.

Archives New Zealand is partly in this category - one aim of the Digital Continuity Strategy is to assess the case for further major digital archive development in the New Zealand public sector. However, we are also testing the waters with low-key, open source development. On this basis, we launched an interim digital archive in July 2008, which has enabled four transfers of public archives since its launch. IDA is an opportunity to learn and build capability. From the outset, its intention was to be a minimal solution, with manual supporting process acceptable. It does not yet provide direct online public access.

Following some years of research and requirements development, Archives New Zealand saw the implementation of an interim system as the best way to continue to learn what would be required for a full digital archive solution. It also enables the institution to fulfill its responsibilities under the Public Records Act and protect vulnerable records while a full solution is developed. Rather than adopting a specific digital preservation approach, the interim digital archive enables a "pick and mix" practical assessment of different options, such as migration, normalization and encapsulation. It is based on available open source systems (Fedora / Fez), complemented by specific tools for format recognition and normalization. The analysis and preparation of records for ingest is a manual process, with significant developer input required. Once ingested into the system, records are identified and validated by two open source tools (DROID / PRONOM and JHOVE) and the normalized into XML and other standard formats by the National Archives of Australia tool XENA. Once these ingest processes are complete, the records are stored and preserved in the Fedora repository.

Archives New Zealand expects the interim digital archive to be in use for two to three years. A registration of interest process has been completed and planning is underway for a full digital archive solution, to replace the interim archive. This would provide larger scale and faster capacity, enhanced public access, integration with finding aids systems and integration with agency systems for ingest. The Public Records Act and issues with ongoing accessibility of information should give incentives for government agencies to transfer records to Archives New Zealand. There is a golden window of opportunity to introduce new systems and services during the lead up to introduction of recordkeeping audits in 2010, when the spotlight is on records. However, digital archive systems are only one part of the puzzle. The Digital Continuity Strategy will guide future developments with a focus on creation and maintenance of reliable records from day one.

## ■ Cooperation and Partnership

Cooperation, collaboration and partnerships are the essential modus operandi in New Zealand's government sector - a country where team sport metaphors dominate the business world. This applies to our work on standards development for recordkeeping, to potential sharing of services, to pooling intellectual resources and to adoption of open source solutions. In the area of digital continuity, Archives New Zealand has agreed that it cannot make progress without collaborative relationships - with national partner institutions, with government agencies and with archives authorities in Australia, North America and Europe. We particularly value our role as partners in the Australasian Digital Recordkeeping Initiative.

By reinventing our role as proactive promoters of best practice recordkeeping across the entire records continuum, archives authorities in New Zealand and Australia have become key partners for auditors-general, ombudsmen, information commissioners, government chief information officers and e-government programmes in their jurisdictions. The profile of recordkeeping has risen substantially. And there is a solid relationship between information managers in agencies and their archival advisors.

Consultation and partnership lead to ownership and acceptance of the end result - whether standards, strategies, services or technology solutions. Without these, our services will at best be challenged, and at worst remain unknown or ignored. We need to continue to strengthen these relationships in order to achieve our goals for digital continuity.



# **Séance 1**

# **L'administration électronique et les normes**

---

---

Sous la présidence de Raimo POHJOLA  
Ancien directeur général adjoint des Archives nationales de Finlande

# ***Session 1***

# ***e-Government and***

# ***Standards***

---

---

*Chairing Raimo POHJOLA*  
*former Deputy Director General of the National Archives of Finland*

# INTEROPERABILITY CROSS BORDERS AND OVER TIME – CHALLENGES AND APPROACHES

---

Karl Wessbrandt

Strategic Development Officer, Swedish Administrative Development Agency

**Vision :** *To achieve the EU principles of free flow of goods, persons, capital and services, we still need to develop a lot more of collaborative administrative processes, supported by standardised, or at least harmonised e-Government procedures.*

**Challenges :** *Much effort has been invested in establishing technical standards and developing system solutions supporting those. But often things don't work well, even when the technology is put in place. So that was not enough, and we need to identify problems and develop solutions from different stakeholder's viewpoints. Apart from technical issues on how to produce, transfer and store information in agreed, secure and sustainable ways, we also need to take care of that the meaning of the information is understood as intended. We have many obvious language and cultural barriers in the Pan-European perspective, but also more subtle differences between sectors, organisations and vocations that can lead to severe misunderstandings if we are not aware of and take care of those semantic issues. In other cases, organisational differences must be recognised and resolved, like questions of authority, responsibility and financing. And perhaps most difficult to resolve, we have legal discrepancies, often in connection with data protection, security and integrity. Between EU members, even when legislation is based on implementing the same directives, but also within the member states, where different acts often are incompatible when it comes to implementing cross organisational e-Government processes.*

**Approaches :** *The consciousness of the need to collaborate between stakeholders from different viewpoints is now being more and more recognized. However, this tends to become a very complex task so to be able to do that in a structured and controlled way, we need structured models and methods to apply.*

**Interoperability frameworks and architecture guidelines :** *Frameworks and architectures are tools designed to accomplish those tasks, but there are many different levels and competing schools, so a good question is on what level we have to agree. Do we need a common top down approach or can we work bottom up and harmonise only when the need occurs? Pro's and con's of different approaches within the EU and in member states will be further discussed in the paper.*

**The role of standards :** *Standardisation is a recognised and often preferred way to solve the needs for interoperability and long term sustainability, but history shows that it also has some hurdles, and the competition between standards and standardisation organisations can make the selection of standards a complex task and sometimes even a political issue. To make the right decisions concerning standards can be a critical success factor to get an optimal balance between stability and progress. And that balance level is much depending on the application area concerned.*

**The way forward :** *The paper gives an overview of the current state and further planning to achieve a better interoperability cross borders and over time within the EU and in member states, and gives examples of potential interoperability challenges and possible solution approaches.*

[See the presentation.](#)

## **L'interopérabilité à l'épreuve des frontières et du temps : défis et approches**

***Perspectives :** pour accomplir les principes européens de libre circulation des marchandises, des personnes, des capitaux et des services, nous devons développer des processus administratifs beaucoup plus collaboratifs qui s'appuieraient sur des procédures de services publics en ligne normalisés, ou tout au moins harmonisés.*

***Défis :** des efforts importants ont été investis dans la création de normes techniques et le développement de solutions systèmes pour les supporter. Mais souvent, les choses ne fonctionnent pas correctement, même lorsque la technologie est mise en place. Cette situation ne peut donc plus durer et nous devons identifier les problèmes et développer des solutions à partir des avis des différentes parties prenantes. En dehors des questions techniques concernant les moyens de produire, transférer et stocker les informations de façon certifiée, sûre et pérenne, nous devons également veiller à ce que le sens des informations soit compris comme prévu. Dans une perspective paneuropéenne, nous sommes de toute évidence confrontés à des barrières linguistiques et culturelles, mais également à des disparités plus subtiles entre les secteurs, les organisations et les activités professionnelles qui peuvent conduire à de graves confusions si nous ne sommes pas conscients de ces problèmes sémantiques et si nous les négligeons. Dans d'autres cas, les différences doivent être reconnues et résolues, telles que les questions d'autorité, de responsabilités et de financement. La question la plus difficile à résoudre serait celle des divergences juridiques, souvent en ce qui concerne la protection, la sécurité et l'intégrité des données, entre les États membres de l'UE, même lorsque la législation est fondée sur les mêmes directives, mais également au sein des États membres où les différentes lois sont souvent incompatibles lorsqu'il s'agit de mettre en œuvre des processus inter-organisations d'administration en ligne.*

***Approches :** aujourd'hui, il est de plus en plus reconnu que les parties prenantes d'avis différents doivent collaborer. Cette tâche semble toutefois devenir de plus en plus complexe. De ce fait, pour y parvenir de façon structurée et contrôlée, nous devons appliquer des modèles et des méthodes structurées.*

***Cadres d'interopérabilité et lignes directrices en matière d'architecture :** les cadres et les architectures sont des outils conçus pour accomplir ces tâches mais il existe plusieurs niveaux et écoles différentes. La question est donc de savoir sur quel niveau nous devons nous accorder ? Avons-nous besoin d'une approche descendante ou pouvons-nous travailler de façon ascendante et harmoniser seulement si nécessaire ? Les avis favorables et défavorables sur les différentes approches au sein de l'UE et au sein des États membres seront discutés ultérieurement dans cet article.*

***Le rôle des normes :** la normalisation est le moyen reconnu et souvent privilégié de répondre aux besoins en termes d'interopérabilité et de durabilité à long terme, mais l'histoire montre qu'elle présente des inconvénients, et qu'en raison de la concurrence entre les normes et les organismes de normalisation, le choix des normes peut devenir une affaire complexe et parfois même un enjeu politique. Les normes choisies peuvent avoir un impact majeur sur la capacité à atteindre un équilibre optimal entre la stabilité et le progrès. Ce niveau d'équilibre dépend pour beaucoup du domaine d'application concerné.*

***Marche à suivre :** cet article expose l'état des lieux actuel et les futurs projets visant à améliorer l'interopérabilité transfrontalière et à long terme au sein de l'UE et des États membres, et présente des exemples de défis en matière d'interopérabilité et de solutions possibles.*

[Consulter la présentation.](#)

## ■ Vision

To achieve the EU principles of free flow of goods, persons, capital and services, we still need to develop a lot more of collaborative administrative processes, supported by standardised, or at least harmonised e-Government procedures. Historically, eGovernment has been on the EU policy agenda since 1993 with the Delors paper, and interoperability has been a target at least from the eEurope 2005 initiative that was launched in 2002. The i2010 strategy is the EU policy framework for the information society and media. It promotes the positive contribution that information and communication technologies can make to the economy, society and personal quality of life. The European Commission presented it in June 2005 as the new initiative for the years up to 2010 and it can be regarded as the first significant interoperability-related initiative originating from the European Commission. Further, the ministerial declarations of Manchester 2005 and Lisbon 2007 went into more detail on what is necessary to accomplish; “Reaping the benefits of eGovernment”, was the title of the Lisbon eGovernment conference.

## ■ Challenges

In parallel with the development of the political visions, and partly independent of them, much effort has been invested in establishing technical standards and developing system solutions supporting those standards. But from the interoperability aspect, often things don't work very well together, even when standardised technology solutions are put in place. So standardised technology was obviously not enough, and we need to analyse and identify problems in different dimensions and develop solutions from different stakeholders' viewpoints.

Apart from technical issues on how to produce, transfer and store information in agreed, secure and sustainable ways, we also need to take adequate measures so that the meaning of the information is understood as intended.

We have many obvious language and cultural barriers in the Pan-European perspective, but also more subtle differences between sectors, organisations and vocations that can lead to severe misunderstandings if we are not aware of and take care of those semantic issues. But still, we cannot realistically expect everyone to agree on using unified, standardised terminologies. Traditions and old relations are mostly too strong to make that approach feasible. Instead, we need to start out with admitting and recognising the differences, documenting them and making them visible, as a first step towards harmonisation, or when that is not possible, towards finding practical work arounds that can bridge the gaps where necessary.

In other cases, organisational differences must be recognised and resolved, like questions of authority, responsibility and financing. In terms of organisational structure, the public sector looks very different in various EU countries. Some have a centralised governance, some are federations with limited central responsibilities, and others have a decentralised approach where public services and responsibilities are distributed over hundreds or thousands of agencies, regions and municipalities. To build a pan-European eGovernment service, where responsibilities are differentiated in a way that a central authority in one country may need to interact with hundreds of local authorities in neighbouring countries can lead to lots of organisational and other issues that can be hard to handle.

And perhaps most difficult to resolve, we have legal discrepancies, often in connection with data protection, security and integrity. This is the case between EU member states, even when legislation is based on implementing the same directives. Examples can be found in the areas of electronic signatures, of access to and reuse of public sector information (PSI directive) or the implementations of the service directive, but similar cases can also occur even within member states, where different legal acts may prove to be incompatible when it comes to implementing cross organisational e-Government processes.



## ■ Approaches

The consciousness of the need to collaborate between stakeholders from different viewpoints is steadily growing. However, these extended needs for collaboration tend to become very complex, so to be able to do that in a structured and controlled way, we need an architectural approach and we need to develop new structured models and methods to apply.

We will also need to accept, and even embrace, that not every public organisation can manage or afford to run and control every system or function in-house. Modern approaches and business models, in IT jargon known under different labels such as outsourcing, shared services, SaaS (Software as a Service), Cloud Computing and PPP (Public Private Partnership) may offer the viable solutions at affordable costs that we should be looking for. However, a shift from in-house operation to using external services requires that the public sector can act as informed and professional buyers, negotiating contracts that secure the required service level agreements.

## ■ Interoperability frameworks and architecture guidelines

Frameworks and architectures are tools designed to accomplish those tasks, but there are many different levels and competing schools, so a good question is on what level we have to agree. Do we need a common top down approach or can we work bottom up and harmonise only when the need occurs? Pro's and con's of different approaches within the EU and in member states will need to be balanced. We cannot start from zero with a blank sheet, as many nations and sectors have already invested a lot in architectural work and have published their National Interoperability Frameworks or similar. The same may apply on regional or local level within member states, and cross borders in sectors like health care, defence or justice.

Those approaches, that can be characterized as “bottom-up” or rather “middle-down”, are valuable contributions that must be taken into account, but in some cases they also pose problems that need consideration. This is the case if overlapping architectural approaches or specific recommendations follow different directions and end up with incompatible recommendations. If these recommendations have a purely national or sectoral scope, with no need to reach “out of the box”, that may be quite acceptable, but when it comes to needs for external interactions, such incompatibilities must be handled. Some obvious technical contradictions were discovered between different national interoperability frameworks according to an inventory made by the CEN/ISSS eGovernment Focus Group last year. Examples of this are recommendations to use incompatible web services standards and to use incompatible office document formats. The first step is to discover such discrepancies, followed by discussions if harmonization is possible, and if not, how the differences can be bridged where needed.

Most existing national or sectoral interoperability frameworks are focusing on the technical and, to some extent, semantic interoperability dimensions. But as discussed in the challenges section above, interoperability must deal with several dimensions in parallel. The European Interoperability Framework version 1 (EIF), published by the EC programme IDABC<sup>1</sup> in 2004, was adding the organisational dimension to the technical and semantic ones. In the draft EIF version 2.0 that is currently being finalized, a legal dimension has also been added, and all four dimensions are now held together within a common frame, labelled “political context”.

A central section of the draft EIF 2.0 states eleven important “underlying principles”, including recommendations that should govern the development of pan-European eGovernment Services (PEGS). Principle number ten is about records management and archiving and the headline reads “Preserve information over time”. The related recommendation states that “Public administration should formulate recommendations regarding format and maintenance procedures (including records management and archiving) aimed at long-term information preservation by public administrations, informed by international standards and good practices”, and also contains a reference to the DLM Moreq project.

As an additional aspect of the EIF draft, the concept “Interoperability chain” is introduced, to emphasize that different functions or components are not independent. eGovernment services are process oriented and require collaborating, interoperable services with open interfaces between them. As an example, an

---

<sup>1</sup> IDABC (Interoperable Delivery of European eGovernment Services to public Administrations, Businesses and Citizens) see <http://ec.europa.eu/idabc>.

interoperability chain may consist of a single point of contact like a user portal, featuring identification and records management functions, followed by a secure data transport for information exchange with other stakeholders, including sustainable electronic archives. To accomplish the practical provision of such a chain of independent but interoperable eGovernment services, a service oriented architecture (SOA) model may be the most appropriate approach.

A vendor neutral SOA standard can be found in the OASIS<sup>2</sup> Reference Model for Service Oriented Architecture<sup>3</sup>.

This document states that it expects that any design that adopts the SOA approach will:

- Have entities that can be identified as services as defined by the Reference Model;
- Be able to identify how visibility is established between service providers and consumers;
- Be able to identify how interaction is mediated;
- Be able to identify how the effect of using services is understood;
- Have descriptions associated with services;
- Be able to identify the execution context required to support interaction; and that
- It will be possible to identify how policies are handled and how contracts may be modelled and enforced.

## ■ The role of standards

Standardisation is a recognised and often preferred method to solve the needs for interoperability and long term sustainability of interfaces, but history shows that it also has some hurdles. The competition between standards and standardisation organisations can make the selection of standards a complex task and sometimes even a political issue. To make the right decisions concerning standards can be a critical success factor, to get an optimal balance between stability and progress. And that balance level is much depending on the application area, the stakeholder population and the political environment concerned. There is an on-going debate on the status of different standards and specifications, on which organisations that should be recognised as issuers of standards to be used in public procurements and on the minimum acceptable degree of openness of standards.

The openness can be discussed both in terms of the standardisation process, regarding participation, decision making, maintenance and revision procedures etc. and in terms of the standards documents as deliverables and how they may be used, e.g. in terms of intellectual property rights like patents and licenses. Here the two main directions are to adopt the so called FRAND principles (Fair, reasonable and non-discriminatory) license terms or to require irrevocably royalty free use. To bring some structure into these and other standards selection issues, the IDABC programme is running a project called CAMSS (Common Assessment Method of Standards and Specifications) that defines objective criteria for the selection of eGovernment standards and specifications to be used by public authorities.

These criteria are structured into four categories:

- Suitability - ability to solve needs and promote interoperability
- Potential - scalability, stability, maintenance
- Market conditions - adoption, implementations, support and
- Openness - of deliverables and of standardisation processes.

How those criteria and assessments are valued and prioritised in a specific case is then up to each member state or evaluating organisation, according to the subsidiarity principle and the regulations applicable.

<sup>2</sup> Organization for the Advancement of Structured Information Standards

<sup>3</sup> See <http://docs.oasis-open.org/soa-rm/v1.0/soa-rm.pdf>.

An overall recommendation is to follow what is known as "the more the better" principle - where you recognise that the standardisation world is not black and white and that you may need to make some trade-offs from what appears to be the ideal situation, from your viewpoint.

Today, there is more or less consensus between governments and industry that more openness promotes interoperability. But a better balance is needed between the business models of the IT industry, the standardisation organisations and the consequences of applying different openness criteria, so there is a vivid on-going debate on this, and currently no consensus on what openness means in reality, on the definition of open standards, or even if there is something that should be called open standards at all. Some governmental organisations within the EU and some member states are very active in this debate and we must wait and see if there will be a constructive outcome in the near future.

## ■ The way forward

We live in exciting times, with rapid change and many unknown factors that will influence our future.

Information growth continues at explosive rates and so will the number of cross jurisdiction interactions.

In this emerging new environment, eGovernment in general and archiving practices in particular, need to accept and adopt to some consequences that seem inevitable but are not yet always fully recognised:

1. New eGovernment processes will increasingly cross organisational and jurisdiction borders, and so will the related information assets. This implies that the legal responsibility and the governing standards and processes for the creation, exchange and preservation of eGovernment records will get more complex and not be as self-evident as was often the case in the good old analogue and silo-oriented situation.

This means that the records management issues and responsibilities must get increased attention already in the design phase of new processes.

2. The majority of eGovernment information and related eGovernment process records will be in electronic form and the volumes will outgrow the practical and economical capacity to preserve everything.

Even if system storage is getting cheaper and indexing and search technologies are developing fast, governments (or tax payers) will not be prepared to invest the resources needed to manage it all within the traditional archival organisations, not to mention the cost of migration to new technology generations.

This means that retention periods and destruction dispositions must be critically reviewed and that the long term cultural heritage value of keeping large volumes of similar records must be re-evaluated and in many cases replaced by a strategy where just samples are being kept.

3. Both above points will also lead to that non-governmental organisations and service providers will play a more significant role both for eGovernment and for archiving purposes. For both financial and organisational reasons, governments will not be able to maintain full operational control.

This means that standards, contract terms, authorisations and certifications will get an increased importance as tools to manage and govern external services. And the basis for the necessary service level agreements and quality control systems will be an adequate and consistent set of agreed standards and specifications. The importance of developing and using standards for eGovernment will thus continue to grow, and to ensure long term interoperability, to encourage a fair market competition and to avoid vendor lock-in, those standards should be as open as possible.

However, this will be an ongoing development process, so we must accept that the current world is not perfect, and that what is considered to be a perfect situation is not yet defined or agreed upon.

But what we all can do is to be active in the discussions. Directives, laws, market practices, standardisation procedures and business models can and certainly will continue to change over time, and our ideas and contributions can influence the directions of those changes.

# TRANSFORMATION, E-GOVERNMENT AND ELECTRONIC RECORDS MANAGEMENT: A STRATEGIC APPROACH TO LINKAGE

---

Stephen Harries

Independent Consultant, Critical Information (UK, Brighton)

*How should electronic records management position itself with regard to the major changes taking place in public services and the structure of the public sphere? The traditional emphasis of records management -and ERM -has been on compliance with legislative requirements and corporate governance, and in the electronic environment this has been manifested as attempts to embed compulsory compliance mechanisms in modernising information technology platforms. Yet this approach is not popular with users or senior managers (for different reasons) and records managers have consistently faced an 'uphill struggle' to demonstrate added value.*

*In contrast, the emphasis of modernisation is on agility and flexibility, rapid response to customer demands, and incorporation of government, corporate and third sector efforts within a directed but loose framework. Faced with ever-increasing citizen expectations, and ever-decreasing funding -the challenge of 'doing more with less' -public services are undergoing a process of radical transformation affecting how they work, who they work with, and the fundamental relationship between government and citizen. New web technologies disperse the 'glittering fragments' ever more widely, and call into question any centralising theory and praxis. The traditional approach of records management -a stolid demand for compliance without evident benefits -looks unattractive in this context. How can ERM maintain an emphasis on capturing, managing and sustaining the record, whilst also responding to these challenges?*

*To do so, it must break the link with archival theory, and recognise that the value that electronic records adds to e-government is in the instantiation of formal and explicit knowledge by understanding and working with the knowledge formulation processes. Records in public services run through processes that start with the citizen and involve many different types of service delivery organisations. They give form to high level policy formulation and individualised personal choices. The systems we put in place to manage electronic records create the justification for capturing records in themselves. ERM must move to a much more strategic view, locating itself within the knowledgecentred processes that generate the purposes of records management.*

*To demonstrate the value added of ERM to e-government decision-makers and funders, it is time to move beyond a "recording what happened" ethos, to a much more engaged and interactive style. For example, ERM should use e-records compliance tools to review where apparent gaps in the record indicate a failure of business process, and intervene, not at the compliance, but at the business management level: and so make a value-added contribution to improvement. ERM should understand the varying communities of practice and action in a public service process, and design mechanisms to mediate between them -for example, to facilitate the formalisation of knowledge from deliberative and participative citizen initiatives into the organisation. ERM should re-use its own product to review, analyse and question the knowledge base of the policy-making process. This requires the development of an 'electronic records management with attitude' that is critical, discursive and interventionist.*

*This paper will present the early findings of a project to explore the implications of such a style of ERM in the fields of general practice healthcare, and of adult social care in local administrations. It will suggest, and open a discussion on, the form of a methodology for embedding electronic records management in public service transformation initiatives.*

[See the presentation.](#)

## **Changements, gouvernance électronique et gestion des documents numériques : pour une stratégie d'intégration**

*Comment devrait-on se positionner en matière de gestion des documents numériques par rapport aux changements majeurs qui affectent actuellement les services publics et la structure de la sphère publique ? L'importance que l'on attribue traditionnellement au records management, y compris étendu à la gestion des documents numériques, est en adéquation avec les exigences de la législation et de la gouvernance d'entreprise, ce qui, dans l'environnement électronique, se traduit par des tentatives d'intégration de mécanismes de conformité imposés dans la modernisation des plates-formes de technologie de l'information. Mais une telle approche rencontre une certaine résistance (et ce, pour différentes raisons) de la part des utilisateurs ou des cadres dirigeants et les responsables en gestion documentaire doivent mener un difficile combat de tous les instants pour en démontrer la valeur ajoutée.*

*Par contre, la modernisation privilégie la plasticité et la flexibilité, la rapidité de réponse aux demandes des clients et les efforts intégrés du secteur public, du secteur des entreprises et des services et ce, dans un cadre dirigé mais souple. Confrontés aux attentes de plus en plus fortes des citoyens et à un budget de moins en moins important, donc au défi de « faire plus avec moins de moyens », les services publics connaissent actuellement une transformation radicale qui change leur façon de travailler, leurs partenaires de travail et les rapports fondamentaux entre le gouvernement et ses citoyens. Les nouvelles technologies Internet permettent de disséminer plus largement que jamais des « trésors d'informations », questionnant par là-même toute théorie et toute pratique centralisatrice. L'approche traditionnelle en matière de gestion documentaire, une froide exigence de conformité sans avantages évidents, semble, dans ce contexte, peu attirante. Comment la gestion des documents numériques peut-elle continuer à privilégier la capture, la gestion et la conservation des documents tout en relevant ces défis ?*

*Pour ce faire, elle doit rompre avec la théorie archivistique et reconnaître que la valeur ajoutée par les documents électroniques à l'e-gouvernance réside dans l'instantiation d'un savoir formel explicite et à travers la compréhension et la participation au processus de formulation du savoir. Dans les services publics, les documents font l'objet de traitements qui partent du citoyen et dans lesquels sont impliqués de nombreux types d'organismes de prestation de services différents. Ils permettent de formaliser des politiques de haut niveau et des choix personnels individualisés. La capture de documents justifie à elle seule la mise en place de nos systèmes de gestion des documents électroniques. La gestion documentaire numérique doit évoluer vers une dimension plus stratégique en se plaçant au cœur de processus axés sur le savoir qui déterminent les objectifs de la gestion documentaire.*

*Pour démontrer aux décideurs en matière de gouvernance électronique ainsi qu'aux bailleurs de fonds la valeur qu'ajoute la gestion des documents numériques, il est temps de passer d'un principe d'« enregistrement de ce qui a eu lieu » à une démarche beaucoup plus engagée et interactive. La gestion documentaire électronique devrait, par exemple, utiliser des outils de mise en conformité pour indiquer les lacunes dans des suites d'information provenant d'une faille dans le processus et intervenir non pas au niveau de la mise en conformité, mais au niveau de la gestion globale, apportant par là sa contribution à valeur ajoutée à la modernisation. La gestion des documents numériques devrait prendre la mesure des différentes communautés de pratiques et d'activités dans un processus de service public global et développer des mécanismes d'arbitrage entre ces communautés, par exemple en formalisant les connaissances à partir d'initiatives citoyennes délibératives et participatives pour permettre leur prise en compte par l'organisation. La gestion des documents numériques devrait réutiliser ses propres produits pour étudier, analyser et interroger la base de connaissances du processus décisionnel politique. Ceci implique de développer une gestion des documents électroniques avec une attitude critique, discursive et interventionniste.*

*Cet article présentera les premières conclusions d'un projet d'étude sur les implications de ce type de gestion pour les soins médicaux généraux et les soins sociaux pour adultes dans les administrations locales. Il proposera et initiera une discussion sur une forme de méthodologie dédiée à l'intégration de la gestion documentaire électronique dans les initiatives de réforme des services publics.*

[Consulter la présentation.](#)

## ■ Introduction

The focus of this paper is on the future prospects, of initiatives to embed electronic records management within the UK public sector, drawing on recent experience within government agencies. It considers the changing public policy context of these initiatives, and the challenges which this poses for electronic records; and suggests an approach to engaging with these challenges by a strategic linking of records with knowledge processes within government.

As with that of many other countries, the UK government sector has seen a wave of activity in electronic records management initiatives over the last ten years. A new, modernising, administration came into power intent on applying new technologies, and the philosophy of “new public management”, to the delivery of government services; and espousing a mantra of evidence-based policy-making - typified in the New Labour phrase of doing “what works”. The UK National Archives (at that time the Public Record Office) recognised an opportunity to secure the future of electronic archive material by attaching targets on electronic records lifecycle management to this modernisation process: arguing that successful and sustained delivery of e-government requires effective management of electronic information through managed creation, capture, organisation, appraisal, migration and disposal - and electronic records and document management (EDRM) provides the necessary framework.

Whilst the primary interests of the Archive was to secure its own future as the repository of digital records by extending its influence into the domain of current records, the arguments for working government departments rested on business needs. These arguments were promising: EDRM as an underpinning and enabling function for e-government, which helped to build trust and confidence in the general public; an assurance of accountability and compliance with information legislation through proper management of corporate memory; an organisational development tool for government agencies to promote culture change and a shared understanding of purpose; and organised and accessible records to provide a sound information base for policy-making.

In practice, the results from these initiatives have been mixed. The timing coincided with the first introduction of new Freedom of Information legislation, and the emphasis frequently drifted from modernisation and business change to compliance and ‘business-as-usual’. Technology frequently took too high a profile, and EDRM initiatives were sometimes treated as an IT systems project rather than as a systemic programme concerned with the integration of organisational strategies, people, technology, and meaningful information: when success is measured by the presence of an icon on a desktop, user resistance is highly likely. In addition, the availability of an electronic records base for review and analysis has not had a noticeable impact on the policy-making process itself. Consequently, some report that the achieved benefits are less than anticipated for the cost involved.

At the present time, the wave of early adopters in central government are pausing for reflection, while local government, health and other sectors look to learn the lessons. How well does the model of EDRM applied in this period stand up, and how might electronic records management best position itself for the future?

## ■ From ‘modernisation’ to ‘transformation’

In part, the above narrative reflects a more general experience: that effective e-government turns out to be much more difficult to achieve than was expected in the first flush of enthusiasm. Many broader government programmes have struggled to make an impact in this complex environment. In response, the rhetoric of UK government has moved from ‘modernisation’ to ‘transformation’. As well as giving renewed impetus to programmes of change, this move also marks a development in the overall policy context for the business of government.

The fundamental drivers for ‘transformation’ are two-fold: the need to meet ever-increasing customer expectations by government service users; and a long-term pressure, and incremental reduction, in the resources available to government agencies to deliver these services. In brief: the need to do more with less.

Transformation approaches promote fundamental, radical and systemic, change in structure, process and practice, rather than incremental change to existing business processes. Expensive new infrastructure projects, such as the implementation or upgrade of EDRM, will have to demonstrate value-for-money through delivery of recognised and achievable transformation benefits - they will have to 'earn their keep' - to continue to be promoted in this emerging environment.

Essentially, the transformations that such initiatives aim to introduce are changing the emphasis of much government activity, downplaying the direct delivery of services to citizens, and instead stressing a co-ordination and enabling role; creating and overseeing a policy and operational framework that brings together a range of other actors in order to realise effective service delivery. One strand impacts on the relationship between the service and its delivery: it stresses the use of partnership arrangements between different professions, or with other organisations in the private or not-for profit sectors. For example, the delivery of adult social care may involve social workers working for a local government authority; health professionals working within separately administered National Health Service medical units; and others from voluntary and not-for-profit organisations. Government agencies may also commission the delivery of services through private sector companies, with commercial arrangements that focus on payment by results, but without any direct operational control by the state.

In these circumstances, operationalisation of the the function is spread over several different organisations, whose members possess different beliefs, values and interests, and which have different organisational systems and policies; and consequently with different approaches to recording and organising their information. In the above example, health professionals and social workers have quite different ways of thinking and modes of practice; and yet, successful delivery will depend upon a common understanding between all these disparate groups, or at least the ability to interpret these differences. This will be expressed (amongst other ways) in the records they keep and their use of that information for knowledgeable action.

A second strand of transformation concerns a change in the relationship between the service and the client., characterised as co-production. Here, the aim is to move away from managing the client - i.e. doing things to the client - towards empowering the client to manage themselves, allowing them to resolve their problems directly through their own action, perhaps supported by the direct provision of resources. For example, the introduction of "personal budgets" into adult social care provides the client with the financial resources to buy directly the services that they decide themselves they require, from a range of providers, This strand is mirrored by a trend towards personalisation of services, and by more frequent attempts to engage the public in greater participation in policy development, for example, through citizen panels and juries: these face participants with the kind of resource allocation choices, and options between uncertain outcomes, which policy-makers normally routinely with.

In this context, one kind of question for ERM asks: *To what extent do current models of ERM practice and implementation sufficiently support such radical business change, and at the same time capture valid, reliable and complete records of the full process?* A second, related, set of questions concern the governance issues: *Where accountability and responsibility are diluted across various agencies and actors, does the record adequately reflect the necessary range of inputs to decision-making and the points of appropriate accountability, in a manner compliant with information risk and information policy legislation?* The first kind of question focusses on the contribution of ERM to delivering the business needs of government, and the second to issues of policy-making in a democracy; but both come together on the ground where the production of records, information and knowledge meet in the delivery of e-government.

## ■ Government, records and Web 2.0

In parallel with these movements (or arguably as part of the same broad societal trend) we have seen the ubiquitous rise of Web 2.0 technologies embedded in many peoples' daily lives. While these applications, from an archiving viewpoint, clearly offer technical challenges for the capture, organisation and maintenance of electronic records, their impact on the actual workings of government remains as yet uncertain. The "cyber-utopian" case argues that Web 2.0 makes possible a radical democratisation of the production of new culture and knowledge, facilitating participation at all levels; anyone with access to the Internet can co-operate with others across space and time in ways previously impossible. On the other hand, counter-arguments stress the socially destructive effects on individuals as virtual consumers, and the difference between the availability of

data and its translation into meaningful information and knowledgeable action. In a sense, these opposing arguments are posing different conceptions of collective action, in a way which is very relevant when thinking about the making and delivery of public policy.

Actual examples of innovative Web 2.0 applications by government are not common. Most applications - ministerial blogs, e-petitions, etc - are examples of one-to-many or many-to-one communication, reminiscent of traditional forms of government consultation. Role-playing virtual environments, such as Second Life, could offer exciting opportunities for simulating policy options and studying their practical evolution in a virtual environment (asking “do they really work?”) before policy implementation in the real world; however, to date known examples have typically been limited to discussion between senior civil servants - a very cohesive community of practice. The most telling examples of Web 2.0 applications have come from interventions by civil society groups, aiming to influence government policy from the outside, and specifically wishing to avoid an incorporation in formal policy making and delivery. The important questions for ERM are not only about how to archive Web 2.0 applications, but very fundamentally: *In what circumstances do Web 2.0 type manifestations constitute a valid part of the record, how do they relate to the imperatives of governance and accountability, and should such applications be actively encouraged to improve accountability?*

## ■ Models of ERM and models of policy delivery

The previous section describes a progressive change in the model of policy delivery: how government gets things done. In part, this is not a new thing: the cycle of change in which public services are delivered through a mix of the two basic delivery poles - the state or the market - is well-known. At some times and in some countries, state-owned organisations or explicit government agencies directly manage and deliver services to the client/citizen, funded by taxation; in others, this is handed to private sector companies operating in a market environment, either in a ‘laissez-faire’ mode or via economic devices such as state incentives or subsidies.

This is a hierarchical model of policy delivery. Whether private or public, both types of organisation have a clear structure of control, well-defined objectives and boundaries, and established lines of accountability. It is possible to make a clear mapping between organisational structures, the elements of a function and its constituent transactions, even though some elements may be delegated to another organisation.

This model makes a good fit with classic approaches to ERM implementation, based on functional analysis, legislative remit, or warrant of a community of practitioners. These approaches can produce a well-structured fileplan against which records can be captured, map this plan to transaction processes that produce records, link the creating software to a central capture and management system, develop appraisal and disposal rules for each category of record, and if wished, lock down any undesirable activity outside this framework. This is the basis on which many ERM implementations have been built.

Emerging models of policy delivery are network-based. Benefits are realised by empowering staff and customers to act independently, and the focus is on outcomes rather than process. Central control is diluted, boundaries are permeable, and the co-ordination of activity is fluid; the emphasis is on strategic management of frameworks. In such environments of uncertainty, functional mapping alone does not elicit the ‘big picture’; and the network may span jurisdictional boundaries, as well as professional and practitioner communities.

If we are considering records solely as a repository recording the formal activity of an organisation, then a classical approach to ERM implementation may be sufficient. If, though, we are considering its contribution to the business of transformation - which earlier sections argue is necessary for survival - a multi-layered approach is needed, one which maps the organisational level to the network level. Building on its established role as guardian of corporate memory, ERM has three particular benefits to offer in this area:

- a focus on the inter-relationships between different record-creating communities, and thereby on the knowledge process level: knowledge in the sense of meaningful information put into action;
- disciplinary tools to support accountability across inter-operating communities, capturing implicit, tacit, knowledge into formal and explicit record sets, and thereby the ability to assure good governance by assessing the range and modes of participation in knowledge processes;



- the ability to use explicit organised knowledge, as reflected in the form of captured records sets, to review, analyse and challenge the business processes within and between organisations, identifying issues and areas for improvement.

## ■ Linking records management to e-government via knowledge processes

A shift in focus towards active management at the knowledge process level emphasises a critical assessment of record content and context. From this perspective, ERM initiatives can be seen as an enabler of organisational and social learning, deploying an interventionist approach that links the dynamics of record creation and management with strategic objectives and the production of outcomes. It can assess content with a view to influencing the range and nature of records, and feed this into decision-making and policy-making processes; it can assess the context of record creation to understand and map the various knowledge communities, institutions and interests against business processes and outcomes.

In the transformation scenarios described above, multiple knowledge communities interact with each other through the knowledge processes in order to produce policy outcomes. Three types of communities can be distinguished:

- communities of practice, pursuing a common enterprise, bound together by an implicit understanding of shared beliefs, values and norms, and with well-accepted and embedded routines and activities;
- communities of interest, holding a limited set of shared interests and objectives in common, linked in a temporary coalition but without the stronger glue of a common professional praxis;
- communities of action, those who actually carry through the necessary actions to achieve outcomes.

In any given individual situation, these types of knowledge communities may be distinct, or they may be overlapping. In the direct delivery of a government service by a government agency, the community of practitioners possesses the primary interest in achieving outcomes, and itself takes the necessary actions to realise them. In network-type delivery models, the communities are more often distinct. For example, an urban regeneration programme may involve planners and architects as practitioners working to develop strategic frameworks within which local citizen groups, voluntary organisations, and private development companies act to realise practical objectives; the whole programme sponsored by a political coalition formed temporarily for this purpose.

## ■ Records management, archives management and transformation

Understanding the contribution of ERM through this type of knowledge-based framework begins to reposition it as a respected element of transformation activity, and as an enabler of better governance, providing the intellectual framework for practical information management. An active, interventionist, deployment of ERM will increasingly have to operate at this level of knowledge dynamics, but must also map knowledge processes and communities down to the classic organisational / functional level, and to the technology level, as well. Viewed through such a lens, new technologies such as Web 2.0 can be placed in a coherent management framework that builds on the achievements of electronic records management activity to date, without being distracted: it is clearer to determine how they fit into the ERM remit, and how they can be integrated into policy-making and delivery.

The logic of earlier ERM initiatives led many national Archives towards an engagement with the nexus of business and information issues of government, an area where few countries have clear responsibilities allocated to a central co-ordinating agency. There remained, however, a tension between objectives: an Archives becomes involved in order to safeguard future accession of electronic records, but in order to be able to influence the direction of government information management, must engage with current policy and delivery issues - an unfamiliar area which it may be uncomfortable operating in. This logic appears to challenge

established non-interventionist principles of archives management; and, ultimately, the Archives must decide what kind of an organisation it wishes to develop into: to remain essentially an Archival institution, or to grow into a broader government information agency, with a more strategic and directive focus. This choice will determine how the issues are framed, and that framing will condition the nature and level of problem solutions offered.

The UK National Archives has recently published a discussion paper, asking for contributions from software developers in debating the next wave of ERM systems. While technological solutions are of course a major issue, the problem is essentially a human and organisational one - how do groups and individuals working together in knowledge-based processes to achieve a common purpose, best manage their information to support delivery of their vision, within the imperatives of public information policy? - and not a technologically determined one.

This paper is suggesting an approach which addresses the issues in these terms, arguing that further progress will be achieved by grasping the opportunities offered by the current emphasis, not just on e-government, but on a transformation of the ways in which government is delivered to citizens.

# RÔLE ET APPLICABILITÉ DES NORMES

---

Marie-Anne Chabin

Expert en archivage, Archive 17

*Les dix dernières années ont vu la publication de plusieurs normes et textes de référence pour le records management, l'archivage électronique ou la conservation des données numériques.*

*Parallèlement, les projets se multiplient sur ces mêmes questions, mettant en évidence l'ampleur du chantier, entre expression des besoins et solutions pour y répondre et, de fait, les réalisations sont encore rares.*

*Quelle est la place des normes dans ces projets ? Leur application est-elle optimale ?*

## **Cartographie des textes normatifs**

*Le rôle des normes est d'une part de répandre les bonnes pratiques, d'autre part de cadrer le marché pour la mise en œuvre de solutions pertinentes et interopérantes, mais face à la liste toujours plus longue des normes relatives à l'archivage (ISO-15489, ISO 14721, NF Z42-013, MoReq2, mais aussi ISO 19005, PAIMAS, ISO 23081...), le responsable de projet peut légitimement se demander laquelle ou lesquelles il peut ou doit appliquer : les textes sont-ils complémentaires, disjoints, concurrents ?*

*Cela requiert une cartographie critique des textes normatifs, plus synthétique et plus concrète que la juxtaposition des chapitres introductifs « champ d'application » de chaque norme, car on voit bien que les normes sur le sujet couvrent un spectre assez large de fonctions, de processus voire de démarches, depuis la définition d'une politique jusqu'aux techniques de pérennisation, en passant par la structuration de l'information, les fonctionnalités des systèmes et l'organisation des services.*

## **Comment les utiliser ?**

*On peut diviser le public visé par les normes en deux groupes : les spécialistes (éditeurs, prestataires, administrateurs de systèmes) et ceux qui ne le sont pas ou pas encore, ou qui n'ont pas nécessairement vocation à le devenir (responsables de projets, utilisateurs finaux).*

*Pour la seconde population, il semble que la diffusion et l'utilisation des normes soient quelque peu freinées leur volume (plusieurs centaines de pages pour certaines, et des milliers quand on les additionne) et leur langage (recouvrement des concepts, vocabulaires différents d'une norme à l'autre pour les mêmes notions...).*

*Par ailleurs, les textes existants sont souvent théoriques et abstraits, avec fort peu d'exemples, d'abord parce que le genre normatif ne s'y prête guère, ensuite parce les cas pratiques au niveau international sont assez difficiles à énoncer.*

*D'où la nécessité de guides explicatifs et de recueils de bonnes pratiques. Des groupes de travail ont déjà publié des guides pratiques sur ISO 15489 et sur MoReq. L'effort pédagogique mérite d'être poursuivi et amplifié, en faisant connaître les réalisations abouties, leur coût et leurs bénéfices, en estimant les avantages et les inconvénients à décliner les normes selon son environnement ou à n'en retenir que certains aspects, en décrivant les pré-requis et les conditions d'une bonne utilisation des normes voire les contre-indications.*

## **Conclusion**

*La normalisation, poussée par l'émergence de nouvelles problématiques et de nouveaux défis techniques, doit aussi rechercher un équilibre entre la production de normes et leur application, c'est-à-dire veiller a priori à leur applicabilité et auditer a posteriori leur application. La normalisation ne saurait devenir sa propre finalité ; l'objectif commun est bien que les informations soient tracées, archivées, conservées et accessibles au mieux.*

[Consulter la présentation.](#)

## **Role and applicability of standards**

*During the last ten years, several standards and reference texts have been published in the areas of records management, electronic archiving, and the conservation of digital data.*

*In parallel, projects dealing with these issues have been multiplying, highlighting the size of the task, including expressing needs and finding the solutions to address them, while in fact, there are few achievements.*

*What is the role of standards in these projects? Are they being applied optimally?*

### **An overview of normative texts**

*The role of standards is on the one hand, to disseminate best practice, and on the other, to provide a framework for the market for the implementation of pertinent and interoperational solutions. However, faced with the ever-increasing list of standards for archiving (ISO 15489, ISO 14721, NF Z42-013, MoReq2, and also ISO 19005, PALMAS, ISO 23081, etc.), a project head can reasonably wonder which one or ones he can or should use. Are the texts complementary, disjoint, or in competition?*

*What is needed is a critical overview of normative texts, better summarised and more concrete than simple juxtaposition of the introductory 'field of application' sections for each standard. It is easy to see that the standards in this area cover a large spectrum of functions, of processes, even of methods, from the definition of a policy through to perpetuation techniques, and including information structuring, the functionalities of systems, and the organisation of services.*

### **How should they be used?**

*The public for whom the standards are intended can be divided into two groups - the specialists (publishers, service providers, system administrators) and those that are not, or not yet, or who are not likely to become, specialists (project heads, final users).*

*For this second group, it seems that an obstacle to the diffusion and use of standards is their volume (several hundred pages for some, and thousands when they are all added together) and their language (overlap of concepts, vocabularies that differ from one standard to another for the same concept, etc.).*

*In addition, current texts are often theoretical and abstract, with very few examples. This is because these are difficult to provide in the field of standards, and also because practical cases at the international level are fairly difficult to set out.*

*This indicates the necessity of explanatory guides and collections of best practice. Work groups have already published practical guides on ISO 15489 and MoReq. Efforts to teach should be continued and amplified, by discussing successful projects, their cost and their benefits, estimating the advantages and disadvantages of adopting standards depending on the environment, or of keeping only some aspects, and describing the prerequisites and the conditions for correct use of standards and possibly their contraindications.*

### **Conclusion**

*Standardisation, driven by the emergence of new problems and new technical challenges, must also find a balance between the production of standards and their application, meaning that we must ensure their applicability at the outset and audit their application afterwards. Standardisation is not an end in itself - the common goal, after all, is that information be tracked, archived, conserved and accessed in the best possible conditions.*

[See the presentation.](#)

## ■ Introduction

Les normes se portent bien, particulièrement dans le domaine qui nous occupe aujourd'hui, c'est-à-dire si on s'en tient au titre du DLM : « gestion de l'information et des archives électroniques » en français, « information and records management » en anglais ».

Mon expérience de vingt-cinq années dans ce métier me fait dire que si le mot « information » commun aux deux langues ne soulève pas de difficulté particulière d'interprétation, l'équivalence « gestion des archives électroniques » - « records management » est moins évidente. C'est pourquoi, je vais commencer par préciser, dans les deux langues, les termes avec lesquels je vais décrire le domaine dont je vais parler : l'archivage et la conservation numérique, en français, « records management and digital preservation » en anglais.

Depuis une dizaine d'années, un certain nombre de textes ont été mis à disposition des responsables de projet et d'autres sont annoncés. Ces textes sont de provenances diverses et de contenus variés.

Dans le même temps, les projets se multiplient, dans le public comme dans le privé, avec des responsables de projets qui relèvent aussi bien du management, du système d'information, de la direction juridique ou des fonctions archives et documentation. Les événements déclencheurs varient. Tantôt, c'est la contrainte réglementaire et le risque qui prédomine, tantôt la pérennisation des archives nativement numériques.

Puisque nous sommes en Europe avec le DLM Forum, en France, ici à Toulouse, et dans le monde séculaire des archives, j'ai choisi d'analyser ce constat sous le patronage de deux grandes figures de la République des Lettres de l'Europe du XVII<sup>e</sup> siècle dont nous sommes tous les héritiers : René Descartes et Jean de La Fontaine.

Je retiens pour l'instant de Descartes les conseils suivants extraits de l'iusable *Discours de la méthode* : « ne recevoir jamais aucune chose pour vraie, que je ne la connusse évidemment être telle », « diviser chacune des difficultés que j'examinerais, en autant de parcelles qu'il se pourrait », et « de conduire par ordre mes pensées, en commençant par les objets les plus simples et les plus aisés à connaître, pour monter peu à peu, comme par degrés, jusques à la connaissance des plus composés ».

## ■ Rôle des normes

La première question sera : qu'est-ce qu'une norme ?

La seconde : comment s'articulent les normes du secteur de l'archivage et de la conservation numérique (*records management and digital preservation*) ?

### Qu'est-ce qu'une norme ?

Il existe plusieurs définitions de la notion de norme. Les trois citations suivantes<sup>1</sup> donnent un éclairage assez précis :

- Un document approuvé par un organisme reconnu, qui fournit, pour des usages communs et répétés, des règles, des lignes directrices ou des caractéristiques pour des produits ou des services ou des procédés et des méthodes de production connexes, dont le respect n'est pas obligatoire.
- Spécification technique ou autre document accessible au public, établi avec la coopération et le consensus ou l'approbation générale de toutes les parties intéressées, fondé sur les résultats conjugués de la science, de la technologie et de l'expérience, visant à l'avantage optimal de la communauté dans son ensemble et approuvé par un organisme qualifié sur le plan national, régional ou international.

---

<sup>1</sup> Extraites de <http://www.marche-public.fr/Marches-publics/Definitions/Entrees/Normes.htm> (les deux premières) et de [http://www.numeraladvance.com/Role\\_des\\_Normes/Pourquoi\\_les\\_Normes\\_/Introduction.htm](http://www.numeraladvance.com/Role_des_Normes/Pourquoi_les_Normes_/Introduction.htm) (la troisième)

- Le rôle premier d'une norme est de faire aboutir un consensus entre les différentes parties prenantes, qui se formalise par un texte, et dont l'objectif est de définir les parties communes, techniques et organisationnelles, ouvertes à tous et qui favorisent les échanges, la standardisation et la maîtrise des coûts entre les différents acteurs économiques.

A noter que la langue française distingue la norme du standard « dont les spécifications sont imposées par le marché » mais l'anglais ne fait pas cette distinction.

La typologie des normes distingue les normes « produits » et les normes de « management » qui décrivent les compétences et les savoirs faire pour les mettre en œuvre les technologies ainsi que l'organisation correspondante. Les normes de management se rattachent à ISO 9001, ISO 14001, etc.). Une étude souligne que « les normes produits et les normes de management forment un ensemble qui recherche en permanence la cohérence et offrent le cadre d'un langage commun »<sup>2</sup>.

En partant de ces définitions, on peut dire qu'une norme se caractérise par quatre critères principaux :

1. le domaine professionnel,
2. la nature de la norme (produit ou management),
3. l'instance de normalisation et sa portée géographique (national, régional ou international). L'organisme de normalisation par excellence est l'ISO (International Organization for Standardization), organisé en comité et sous-comités, avec un représentant dans chaque pays, doté d'une structure similaire. A noter que, souvent, une norme ISO provient d'une norme nationale portée au niveau supérieur. Certains textes ISO portent également le label IEC (Commission Électrotechnique Internationale). La normalisation au plan régional, européen notamment, est plus récente.
4. les bénéficiaires, c'est-à-dire, d'après les citations ci-dessus, les parties intéressées, le public et les acteurs économiques.

L'identification d'une norme (celle dont tel utilisateur a besoin) ne devrait donc pas soulever de difficulté, dès lors que le domaine professionnel est bien circonscrit et que l'instance de normalisation est unique ou hiérarchisée.

De fait, si l'on cherche les règles de transposition des caractères chinois en caractères latins, on trouve très rapidement « la » réponse, à savoir ISO 7098. Si on veut connaître les exigences de ténacité pour les matériaux à température cryogénique, on est très vite guidé jusqu'à ISO 21028.

Quand on s'attaque à l'archivage (*archives and records management*), force est de reconnaître que c'est beaucoup moins simple.

### **État des lieux des textes normatifs sur l'archivage et la conservation numérique (records management and digital preservation) ?**

Les normes ISO forment le noyau dur du corpus :

- ISO 15489-1 et 2 : 2001, sur le « Records management » (expression non traduite en français dans la traduction de 2001) ; ISO 15489 a donné lieu à deux normes « filles » : ISO 23081 sur les métadonnées en 2006 et ISO 26122 sur les processus en 2008 ;
- ISO 14721 : 2003, « Modèle de référence pour un système ouvert d'archivage d'information (OAIS) » / *Reference Model for an Open Archival Information System (OAIS)*, avec son « Guide méthodologique associé pour la négociation Producteur / Archive », publié en 2006 sous le numéro ISO 20652 ;
- ISO 19005-1 : 2005 : format de fichier des documents électroniques pour une conservation à long terme, avec les mises à jour (ISO 24517 et ISO 32000).

On peut y ajouter la norme ISO/IEC 27001:2005 qui décrit les exigences relatives aux systèmes de gestion de la sécurité de l'information, dans la mesure où de plus en plus de projets d'archivage et de conservation numérique sont liés à la problématique de sécurité de l'information.

---

<sup>2</sup> <http://www.numeraladvance.com> (déjà cité).

Le titre des normes n'inclut pas le nom du comité qui l'a préparée, de sorte qu'on ne voit pas au premier abord que ces normes sont issues de comités techniques assez différents. Ceci met en évidence le caractère transverse de la question :

- comité n° 46 « Information et documentation » pour ISO 15489,
- comité n° 20 « Aéronautique et espace » pour ISO 14721,
- comité n° 171 « Applications en gestion des documents » pour ISO 19005,
- sous-comité 1/SC 27 « Techniques de sécurité des technologies de l'information » pour ISO/IEC 27001

Ce sont ensuite les textes non ISO de portée internationale, avec notamment deux textes très récents :

1. le premier est celui qui est à l'honneur à cette manifestation : « MoReq2, Exigences types pour la maîtrise de l'archivage électronique » / *Model Requirements for the Management of Electronic Records*, paru en février 2008 et supervisé par le forum DLM et la Commission européenne
2. le second est les « Principes et exigences fonctionnelles pour l'archivage électronique/ *Principles and Functional Requirements for Records in Electronic Office Environments*, publiés par le Conseil international des Archives (CIA) en juillet 2008<sup>3</sup>.

Ces deux textes, assez proches quant au contenu, proposent des recommandations et des spécifications relatives aux Systèmes d'Archivage Électronique (Electronic Records Management Systems) qui préoccupent autant le marché que les entreprises, et qui ne sont pas couverts à ce jour par les normes ISO. L'avenir dira comment les choses évoluent.

Au plan national, on trouve des textes de référence (normes de droit ou normes de fait) dans de nombreux pays, mais on remarque qu'ils n'émanent pas seulement des organismes nationaux de normalisation mais aussi de l'administration ou de communautés professionnelles spécifiques. Par exemple, se rattachent à ISO 15489 même s'ils lui sont chronologiquement antérieurs :

- DoD 5015.2 pour les États-Unis (*Department of Defense - Design criteria standard for electronic records management software applications*) dont bon nombre d'éditeurs se réclament et de fait plusieurs outils sont certifiés au regard de ces exigences américaines.
- TNA (*The national archives*) pour le Royaume-Uni,
- VERS ([The Victorian Electronic Records Strategy](#)) pour l'Australie,
- DOMEA (*Document management and electronic archives*) pour l'Allemagne.

D'autres textes nationaux renvoient davantage aux aspects de conservation numérique. Ainsi pour la France :

- NF Z42-013 révisée en 2008 : Archivage électronique - Spécifications relatives à la conception et à l'exploitation de systèmes informatiques en vue d'assurer la conservation et l'intégrité des documents stockés dans ces systèmes
- Standard d'échange de données pour l'archivage. Transfert - Communication - Élimination - Restitution, publié par la Direction générale et la modernisation de l'État (DGME) et la Direction des Archives de France (2006).

A noter que la certification, notion indissociable de celle de norme, commence à prendre forme, notamment avec la certification du modèle OAIS et celle du modèle MoReq au travers de la batterie de tests qui constitue la grande innovation de MoReq2.

---

<sup>3</sup> Le Conseil international des Archives est par ailleurs connu pour les travaux de normalisation dans le domaine de la description archivistique, avec ISAD(G): Norme générale et internationale de description archivistique (2000), ISAAR(CPF) : Norme internationale sur les notices d'autorité archivistiques relatives aux collectivités, aux personnes et aux familles (2004) et ISDF : Norme internationale pour la description des fonctions (2008) mais ces sujets dépassent le cadre du DLM Forum.

La liste, bien sûr, n'est pas exhaustive mais elle souligne bien deux caractéristiques récurrentes :

- le double enjeu de maîtriser ce qu'on veut archiver et de conserver ce qu'on a archivé, autrement dit les deux composantes de la problématique : le management (identification de l'information à risque, sécurité de la production, contrôle du cycle de vie) et la technique (intégrité, disponibilité, sécurité de la conservation, pérennité) ;
- la pluridisciplinarité des parties prenantes si on en juge par la diversité des positionnements des normes au sein de l'ISO et par la pluralité des initiatives hors ISO (marché, institutions, professionnels ...).

Le schéma ci-après tente de situer une quinzaine de normes dans un plan qui met fait ressortir d'une part l'approche (archivage / *records management* ou conservation numérique / *digital preservation*), d'autre part la distinction management / outil.

Nota Bene : les normes ISO sont sur fond gris foncé ; les autres normes de portée régionale ou internationale sont sur fond moyen et les textes nationaux sur fond gris clair.

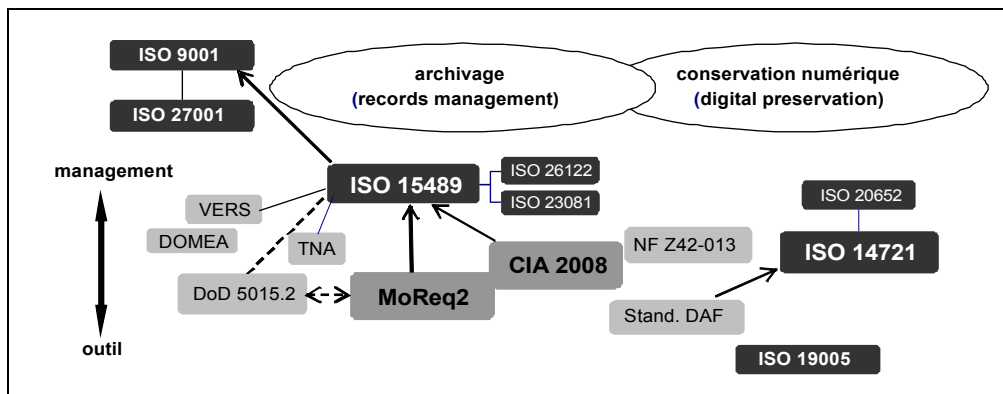


Schéma. Mise en perspective des principaux textes normatifs relatifs à l'archivage/*records management* et à la conservation numérique/*digital preservation*

## ■ Application et applicabilité des normes

Sur l'application des normes, trois questions se présentent :

- les normes sont elles appliquées ?
- qu'est-ce qui fait qu'elles le sont ou ne le sont pas ?
- comment peut-on améliorer la situation ?

### Diffusion et utilisation des normes

Quels sont les indicateurs pour apprécier la diffusion et l'application d'une norme ? Le nombre d'exemplaires vendus ? Les citations de la norme dans la documentation interne des organismes ou des entreprises ? Le chiffre d'affaire réalisé par les produits étiquetés « conforme » à la norme ?

Les chiffres de vente ne peuvent se mesurer que si la norme est payante. De ce point de vue, la France est un exemple intéressant : la norme française NF Z42-013 relative à la « conception et exploitation de systèmes informatiques en vue d'assurer la conservation et l'intégrité des documents » s'est vendue depuis sa première édition en 1999 à plus de 2000 exemplaires. Les résultats sont équivalents pour la version française d'ISO 15489 publiée en 2001<sup>4</sup>. On peut dire que ce sont, pour les critères de l'AFNOR (Association française de normalisation), deux « succès de librairie ».

Pour les normes accessibles gratuitement sur Internet, il faudrait mesurer les téléchargements mais ce serait encore peu significatif car les textes circulent par d'autres voies. Cependant, il apparaît que leur

<sup>4</sup> Merci à Gérard Weisz (AFNOR, CN 171) et à Marie-Martine Guillabert (AFNOR, TC 46/SC 11) pour les chiffres qu'ils m'ont aimablement communiqués.



gratuité n'est pas un facteur décisif de diffusion. Par exemple, MoReq est à ce jour largement moins diffusé en France que NF Z42-013 ou ISO 15489, quelques centaines d'exemplaires sans doute ?

D'autre part, il faut nuancer entre le fait de connaître l'existence d'une norme, son acquisition, sa lecture (et j'ajouterais sa nécessaire relecture réitérée), son utilisation dans la rédaction d'une politique, d'une procédure ou d'un cahier des charges, et son application réelle, c'est-à-dire le fait que ce que dit la norme se traduit concrètement par des faits.

Finalement, le baromètre des réactions des utilisateurs dans les salons, les conférences, les formations, les forums de discussions n'est sans doute pas le moins fiable. Je ne saurais me prononcer sur un plan international mais pour la France, ma connaissance des milieux de l'archivage et ma veille quasi-quotidienne de ces questions me fait dire que :

- le nom des principales normes est assez bien connu des utilisateurs finaux potentiels, d'autant plus que les éditeurs les appellent volontiers à la rescousse de leurs présentations commerciales ;
- la diffusion des normes et leur téléchargement sont encourageants ; la demande du public est là et il faut tenir compte du fait que ces textes sont récents et encore peu enseignés ;
- l'utilisation des normes est en revanche encore embryonnaire ; certes, elles sont exploitées rigoureusement par les équipes de certaines grandes entreprises ou grandes institutions qui s'attaquent à une analyse pointue et comparative du corpus normatif disponible, mais on rencontre surtout de très nombreux responsables de projets d'archivage électronique qui cherchent à « s'y retrouver dans le maquis des normes ». Les retours d'expérience sont maigres et laissent paraître que les normes ne jouent pas le rôle qu'elles sont censées jouer dans le développement du marché.

### Les freins

On lit au début du module 1 des *Principes et exigences fonctionnelles pour l'archivage électronique/ Principles and Functional Requirements for Records in Electronic Office Environments*, publiés en juillet 2008 par le Conseil international des Archives cette phrase paradoxale mais éclairante : « On constate que plus les entreprises développent les nouvelles technologies et introduisent de nouvelles méthodes de travail, plus les méthodes et les procédures d'archivage perdent de leur efficacité »<sup>5</sup>.

Et en effet, on assiste avec l'électronique à une sorte de dérégulation de la gestion de l'information dans le temps, un chassé-croisé des performances technologiques avec les règles de gestion, alors que les normes seraient justement censées réguler ces relations.

Pour ma part, je vois quatre freins plus particuliers au développement des normes d'archivage.

Le premier frein est un certain flou dans les concepts et dans le vocabulaire qui les décrit. Certes, il existe des glossaires mais chaque norme a le sien et, malheureusement pour l'utilisateur, ils ne disent pas la même chose ou, ce qui est plus grave, ils utilisent les mêmes mots pour décrire des réalités différentes, apportant de la complexité là où les choses sont finalement assez simples. Le phénomène est accentué par les difficultés de traduction de l'anglais. Cependant, l'ambiguïté de certains termes n'est pas propre à la France ni aux pays latins mais est assez caractéristique du fait que l'archivage (*archives and records management*) se situe aujourd'hui au carrefour de plusieurs métiers, ceux du documentaire et ceux de la technique, sans oublier l'impact croissant des risques. Le texte de référence du CIA note par exemple dans le module 3 (point 2.3.4) que le mot « archiving » a des sens différents pour les records manager et pour les informaticiens. Autre exemple, l'OAIS utilise un vocabulaire nettement différent des autres textes.

Le deuxième frein tient à la concurrence, apparente ou réelle, entre les normes qui émanent d'instances différentes, sous le regard dubitatif de l'utilisateur. Par ailleurs, devant la profusion normative, les outils qui se prétendent conformes à tous les textes de référence finissent par perdre de leur crédibilité. Il y a plus grave, c'est de voir dans certains cahiers des charges d'organismes publics ou d'entreprises privées la liste systématique de toutes les normes existantes, sans discernement, comme si la compilation de textes hétérogènes étant garante de qualité, quand on ne mélange pas normes, loi et guides utilisateurs...

Troisième frein : la complexité des normes et les coûts de mise en œuvre qui en découlent. Pour ma part, tout en restant solidaire de MoReq2, en tant que membre du comité éditorial, je fais partie de ceux qui

<sup>5</sup> Voir <http://www.ica.org/fr/node/38975> (traduction Marie-Anne Chabin).

dénoncent sa lourdeur excessive (les 330 pages de spécifications ont de quoi rebuter les meilleures volontés). Sans doute est-il nécessaire de pouvoir conduire de grands projets en s'appuyant sur des spécifications complètes mais il faut aussi tenir compte du fait que tous les utilisateurs n'ont pas le même niveau de besoins. Les projets d'archivage ne visent pas tous à maîtriser d'emblée tous les facettes possibles et imaginables d'un système d'archivage, pour des raisons à la fois fonctionnelles et financières bien compréhensibles. Il faudrait que les normes soient une aide à la progressivité et non un facteur d'alourdissement.

Enfin, et c'est en partie du au caractère somme toute récent de la problématique, les normes d'archivage et de conservation numérique paraissent théoriques, éloignées de la réalité de terrain à laquelle sont confrontés les utilisateurs. Comment se laisser séduire par une norme quand on ne peut toucher du doigt des exemples de sa mise en œuvre et du retour sur investissement qu'elle peut procurer ?

### **Des leviers pour demain**

Ces freins ne sont pas fatals. Les réponses existent. Si le diagnostic est partagé, on peut trouver des orientations et des initiatives comme autant de leviers pour une normalisation et une application des normes plus efficaces, plus conformes

Ce sera par exemple :

1. veiller à la cohérence des concepts et des termes, dans les différentes langues, utilisés par les différentes normes ; il serait ridicule et inefficace d'imposer un espéranto de l'archivage mais il faut que les concepteurs et promoteurs des normes donnent à voir à ceux qui peuvent en bénéficier les convergences et les divergences de fond et de forme ;
2. procurer au public destinataire des normes une meilleure visibilité et surtout une visibilité globale de l'ensemble des textes normatifs du secteur, de sorte que l'utilisateur final, sans qui rien ne progressera durablement, puisse apprécier plus facilement l'intérêt ou l'avantage à utiliser telle ou telle norme :
  - est-ce que je retiens la norme NF Z42-013 ou le modèle OASIS ? ou les deux ?
  - dois-je me plonger dans ISO 26122 alors que ma direction générale n'a pas encore signé la politique d'archivage (*records management policy*) ?
  - est-ce que je peux concilier ISO 23081 et le modèle de métadonnées de MoReq2 dans mon projet ?
  - etc.
3. renforcer la compétence et le professionnalisme des maîtrises d'ouvrage, tant dans le domaine de l'archivage (records management) que de la conservation numérique (digital preservation), afin que les équipes projets puissent décrire leurs besoins avec plus d'acuité et dialoguer plus efficacement avec les offreurs et les prestataires, dans un échange utile à tous, et surtout à la réussite des projets ; cela peut notamment se faire au sein de clubs utilisateurs ;
4. encourager le phasage des projets (pour les responsables de projets) et moduler les solutions de mise en œuvre (pour les éditeurs), à la fois pour rassurer les utilisateurs et pour permettre un retour sur investissement crédible pour les sponsors et financeurs des projets ;
5. insister sur l'importance du management pour la réussite des projets, en soulignant davantage les liens entre l'archivage et la conservation numérique d'une part, et la qualité (ISO 9001) et la sécurité de l'information (ISO 27001) de l'autre ;
6. favoriser les publications de guides d'accompagnement des normes, rédigés par les utilisateurs finaux, ou du moins avec eux. Les responsables de projets d'archivage français ont par exemple aujourd'hui à leur disposition sur plusieurs sites Internet :

*Comprendre et pratiquer le records management. Analyse de la norme ISO 15489 au regard des pratiques archivistiques françaises*, élaboré par un groupe de travail commun à l'association des archivistes français (AAF) et l'association des documentalistes (ADBS) en 2005<sup>6</sup> ;

---

<sup>6</sup> [www.adbs.fr](http://www.adbs.fr) et [www.archivistes.org](http://www.archivistes.org)

*Apprivoiser MoReq pour archiver et conserver l'information*, réalisé par un groupe de travail de l'association Ialta France en 2007 et diffusé notamment avec le soutien du Club des responsables de politiques et projets d'archivage (CR2PA)<sup>7</sup>.

Ces deux publications comportent quelques témoignages d'utilisateurs à usage d'autres utilisateurs mais on doit et on peut très certainement aller plus loin, illustrer davantage les bonnes pratiques.

## ■ Conclusion

Pour conclure, je donne la parole à La Fontaine, au travers de la morale de trois fables<sup>8</sup> :

- « Ne faut-il que délibérer, La cour en conseillers foisonne ; Est-il besoin d'exécuter, L'on ne rencontre plus personne » ;
- « Le trop d'expédients peut gêner une affaire ; On perd du temps au choix, on tente, on veut tout faire ; N'en ayons qu'un mais qu'il soit bon » et
- « En toute chose, il faut considérer la fin. ».

Le ton est persifleur ; c'est le propre du fabuliste.

Au-delà de la littérature, le message est qu'il serait peut-être opportun de faire une pause dans la normalisation, en tout cas de ne pas surenchéris dans une normalisation pour la normalisation mais au contraire de recentrer la production des normes sur leur applicabilité pour le bénéfice du plus grand nombre. Devant la complexité réelle des systèmes d'information et de l'impact de l'environnement réglementaire et organisationnel sur la gestion de l'information, il faut se défaire de l'idée qu'une situation complexe requiert une solution complexe et rechercher au contraire des solutions simples, tant techniques que méthodologiques. Enfin, il est capital de ne pas perdre de vue que l'objectif est bien de mesurer de progrès réels dans la maîtrise de l'archivage électronique ; les normes et les outils, aussi beaux soient-ils, ne sont que les instruments et non la finalité de l'action.

---

<sup>7</sup> [www.ialtafrance.org](http://www.ialtafrance.org), [www.cr2pa.fr](http://www.cr2pa.fr), [www.fntc.org](http://www.fntc.org), [www.archive17.fr](http://www.archive17.fr)

<sup>8</sup> Ces trois « morales » sont issues dans l'ordre des fables : *Conseil tenu par les rats* (L. II, 2), *Le chat et le renard* (L. IX, 14) et *Le renard et le bouc* (L. III, 5)

Le site <http://www.eturama.com/edu/lafontaine/fables-english.php> donne une traduction en anglais des fables dont voici les extraits correspondants :

- To argue or refute Wise counsellors abound;
- The man to execute Is harder to be found. (*The Council Held By The Rats*)
- Expedients may be too many, Consuming time to choose and try. On one, but that as good as any, It's best in danger to rely. (*The Cat and the Fox*)
- Whatever way you wend, Consider well the end (*The Fox and the Goat*)

# LE COSTA RICA ET L'ADMINISTRATION ÉLECTRONIQUE

---

María Teresa Bermúdez-Muñoz

Professeur à l'Université du Costa Rica

*Le Costa Rica est un petit pays d'Amérique centrale à peine grand comme la Suisse avec une population de 5 millions d'habitants qui figure en bonne position au classement mondial de l'indice de développement humain (IDH). Son économie est ouverte et basée sur l'agriculture et les services, notamment le tourisme.*

*Dans cette perspective, le propos de la communication est donné un avis sur les points suivants :*

*1. Les efforts régionaux pour donner un nouvel élan à la modernisation de l'État comme élément important du renforcement de la gouvernance démocratique et du bon gouvernement : eLAC 2007, SMSI Conférence régionale de l'Amérique Latine et des Caraïbes, Lettre Iberoaméricaine du e-gov.*

*2. Quelques indicateurs élémentaires sur l'intégration des TICs dans les opérations des gouvernements qui fourniront un premier aperçu du pays : composition de la population, population urbaine, taux d'alphabétisation, PNB-PPA par habitant, lignes de téléphone, téléphones portables, postes de radio, postes de télévision, utilisateurs d'Internet, nombre de fournisseurs d'accès Internet.*

*3. La façon comment le Costa Rica envisage-t-il la normalisation dans le développement de l'administration électronique. Le contexte politique, économique, sociale et technologique ; le niveau éducatif (les demandes de la mondialisation impliquent non seulement la modernisation des procédés d'apprentissage et d'enseignement, mais aussi une offre de modèles éducatifs qui permettront de former pour apprendre tout au long de la vie, pour développer l'esprit critique permettant d'appliquer, de recréer et de créer la connaissance) ; des aspects juridiques, organisationnels et méthodologiques ; les services transactionnels clés ; les services d'information archivistique ; les projets pilotes; l'Internet / le commerce électronique ; le déroulement du travail, la sécurité, l'interopérabilité des systèmes d'information publics, l'archivage électronique, l'accès à l'information publique, le respect des données personnelles.*

*4. Le rôle joué par la Direction des archives nationales dans la normalisation de la gestion des archives. La normative archivistique, le système national des archives, la gestion de documents, les normes nationales de description, les archives électroniques.*

*5. Le rôle joué par la Section d' Archivistique de l'Université du Costa Rica dans la formation professionnelle des archivistes. Antécédents (création du Diplômât archivistique en 1978, profil de technicien pour faire le traitement des archives historiques), évolution de la carrière : création du Baccalauréat, en 1996 (programme orientée toujours vers le traitement des archives historiques, beaucoup de cours d'histoire; peu de cours d'informatique) ; changement du programme pour l'orienter vers les archives courantes: moins cours d'histoire, plus de cours d'informatique, du management des archives, en 2000 ; création de la Licence, en 2004 (programme orienté vers les documents électroniques car tous les métiers liés au "document" ont été affectés par l'entrée dans "l'ère du numérique").*

*La combinaison de ceux aspects permettra une analyse et une réflexion critique sur les processus de l'administration électronique au Costa Rica.*

[Consulter la présentation.](#)

## **Costa Rica, electronic administration and archives**

*Costa Rica is a small country in Central America, not much larger than Switzerland, with a population of 5 million. It is well positioned in world human development index (HDI) rankings, with an open economy based on agriculture and services, tourism in particular.*

*From this perspective, the aim of this paper is to give an opinion on the following points:*

- 1. Regional efforts to boost the modernisation of the State as an important element for reinforcing democratic governance and good government. These include eLAC 2007, the WSIS Regional Conference for Latin America and the Caribbean, and the Ibero-American Letter on e-government.*
- 2. A few basic indicators on the integration of ICTs in government operations which provide a 'snapshot' of the country: composition of the population, urban population, literacy rate, GNP-PPP per inhabitant, telephone lines, mobile telephones, radio sets, TV sets, internet users, number of internet service providers.*
- 3. How Costa Rica views standardisation in the development of electronic administration. The political, economic, social and technological context: the educational level (the demands of globalisation require not only the modernisation of learning and teaching processes, but also the provision of educational models that enable lifelong training for learning, in order to develop the critical thinking needed to apply, recreate and create knowledge); the legal, organisational and methodological aspects; the key transactional services; the archive science information services; pilot projects; internet/e-commerce; workflow, security, interoperability of public information systems, electronic archiving, access to public information, and respect for personal data.*
- 4. The role played by the National Archives Authority in standardising archive management. Standards in records management, the national archive system, document management, national description standards, and electronic archives.*
- 5. The role played by the Archive Science Department at the University of Costa Rica in providing vocational training for archivists. Background (creation of the archive science Diploma in 1978, with a technician profile for the processing of historical archives), career development: creation of the Baccalaureate in 1996 (syllabus still focused on processing of historical archives, many history classes, few computing classes); the change in the syllabus to orient it more towards present-day archives: less history and more computing and archive management classes, in 2000; creation of a degree course in 2004 (syllabus focused on electronic documents as all 'document-related' professions have been influenced by the advent of the digital era).*

*By bringing together all of these aspects, it will be possible to make an analysis and critical reflection on the processes of electronic administration in Costa Rica.*

[See the presentation.](#)

## ■ Introduction

Le Costa Rica est un petit pays de l'Amérique Centrale, de 4,5 millions d'habitants, jouissant d'un indice de développement humain (IDH) important, même si cette position s'est dégradée ces dernières années, passant de 42 à 47. Son économie est basée sur l'agriculture et les services, en particulier le tourisme.

Selon l'indice de compétitivité global (ICG) 2007-2008<sup>1</sup>, le Costa Rica a occupé la 63<sup>ème</sup> place parmi 131 pays étudiés, se situant en 6<sup>ème</sup> position en Amérique latine et dans les Caraïbes. Ce rapport signale, parmi les aspects positifs du Costa Rica, l'indépendance du Pouvoir Judiciaire, la qualité du système éducatif, l'intensité de la concurrence locale, la professionnalisation dans la gestion d'entreprise et la faible fuite de cerveaux, ainsi que la solidité de la banque, l'impact positif de l'investissement étranger, la sophistication du secteur productif et le degré d'innovation. Comme aspects négatifs sont cités particulièrement la bureaucratie et le développement déficient des infrastructures (ports, aéroports, voies ferrées et routes).

En matière économique, pendant l'année 2007, le Costa Rica a présenté une haute croissance du PIB, ce qui a contribué à la production de nouveaux emplois et à une hausse du revenu moyen des foyers ; de même, une décision politique a permis de renforcer le régime non contributif de retraites et le programme « Avançons »<sup>2</sup>, ce qui s'est traduit par une réduction de 3.5% de la pauvreté totale et de 2% de la pauvreté extrême. D'autre part, le pays a continué à accumuler des capacités dans les domaines de la santé et de l'éducation. Toutefois, selon les experts, les facteurs économiques et politiques générateurs de ces bons résultats restent fragiles. En outre, des situations contradictoires se présentent puisque, à côté des progrès sociaux, il y a une augmentation des inégalités<sup>3</sup>.

Tout comme d'autres pays de la région latino-américaine, le Costa Rica s'est aussi proposé de développer l'administration électronique (gouvernement numérique) afin de faire face aux défis de la Société de l'Information et de la Connaissance (SIC).

## ■ La SIC en Amérique Latine et au Costa Rica

Depuis l'année 2000, des initiatives en relation avec la modernisation des États et le développement de la SIC ont vu le jour en Amérique latine. Parmi celles-ci peuvent être signalées les suivantes :

- Lettre Ibéro-américaine de la Fonction Publique de 2003, approuvée par la V<sup>e</sup> Conférence Latino-américaine des Ministres de l'Administration Publique et de la Réforme de l'État, dont le but est de promouvoir la professionnalisation de la fonction publique afin d'améliorer les niveaux de confiance des citoyens en l'administration publique, l'efficacité gouvernementale, la lutte contre la corruption, et la capacité de croissance économique durable des États Ibéro-américains.
- Code Ibéro-américain du Bon Gouvernement dont le but est « de contribuer à rendre digne au vue de la société le service public... » et à ce que « l'activité gouvernementale devienne une pratique de la promotion de l'intérêt général, et s'éloigne de l'abus odieux des ressources publiques à des fins d'intérêt partisan ou particulier. »<sup>4</sup>
- L'École Ibéro-américaine de l'Administration et des Politiques Publiques est l'instrument utilisé par les États américains d'origine ibérique pour mettre en œuvre les engagements acquis dans « la charte ibéro-américaine de la Fonction Publique » et « le Code ibéro-américain du Bon Gouvernement » et a pour but « de contribuer, par la socialisation d'expériences de gouvernement d'Ibéro-Amérique, au renforcement du caractère institutionnel public, à la construction de la citoyenneté, à la consolidation

<sup>1</sup> Monge G. Ricardo y otro (2008). *Competitividad e innovación en Costa Rica. Innovación, Competitividad y Crecimiento. Desempeño de Costa Rica y de su sector de las TICs.* CAATEC. Cartago, CR.

<sup>2</sup> Le programme « Avançons » offre des aides aux familles des collégiens et lycéens pauvres et a pour but l'universalisation de l'éducation pour les moins de 17 ans.

<sup>3</sup> XIV Informe «Estado de la Nación». 2008. Programa «Estado de la Nación». San José.

<sup>4</sup> Código Iberoamericano de Buen Gobierno. 2006. Montevideo.

de la gouvernance démocratique, l'amélioration de la capacité de gestion de l'État et la consolidation des systèmes de la fonction publique professionnelle dans la région... »<sup>5</sup>

- Groupe de Rio. (2001) Déclaration des Chefs d'État et de Gouvernement sur l'Évaluation du Groupe de Rio pendant ses quinze premières années et projection du mécanisme sur le nouveau siècle, « le potentiel des technologies de l'information et des communications pour le développement de la démocratie, l'économie et le progrès social » est reconnu dans ce document. Les signataires se sont accordés également « sur la convenance de partager la connaissance, la technologie et l'information, en profitant de l'infrastructure de connectivité et en promouvant l'accès universel des populations à cette dernière ... » et ils soulignent que « l'utilisation des technologies et des réseaux numériques pour le développement et le progrès social, bien qu'il représente un défi technique, devient avant tout un enjeu politique, institutionnel, social et culturel qui exige une vision et une proposition qui élargisse les espaces de coopération entre nos pays et entre les secteurs public et privé, et doit également s'appuyer sur des institutions financières et d'autres organismes internationaux. »
- Déclaration de Florianópolis<sup>6</sup> dans laquelle on exprime pour la première fois « l'aspiration partagée des pays de l'Amérique latine et des Caraïbes de s'intégrer comme membres pleins de la société de l'information avec efficacité, équité et durabilité, dans le cadre de l'économie globale fondée sur la connaissance ».
- Déclaration d'Itacuraja regroupant des recommandations de l'Amérique latine et des Caraïbes au III<sup>e</sup> Congrès International de l'UNESCO sur les défis moraux, juridiques et sociaux du Cyberspace.
- Déclaration de la Rencontre Indigène Interaméricaine par laquelle des représentants de 45 millions d'indigènes se sont engagés à contribuer à la construction d'une société authentique de l'information et de la communication, encadrée dans les principes d'égalité dans la diversité et du plein respect des droits humains.
- Déclaration de Lima. Sixième Réunion Ministérielle de l'APEC sur l'industrie de l'information et les communications, dans laquelle est souligné, entre autres, « que le progrès technologique accéléré, la convergence des réseaux d'information et de télécommunications et la mondialisation de services et d'applications demande un partage de l'information, avec échange des meilleures pratiques et des expériences visant à l'élaboration des politiques et des régulations entre les économies.»
- Forum sur l'Accès à l'Information en Amérique Latine et aux Caraïbes (LAC) : c'est un espace de dialogue et d'échange, né dans le cadre du Moniteur du caractère privé et de l'accès à l'information dans la région latino-américaine et des Caraïbes.
- Agenda de Connectivité pour les Amériques. Plan d'Action de Quito dont l'objectif est « d'offrir aux citoyens des Amériques des opportunités pour développer et utiliser des connaissances, de sorte qu'ils puissent pleinement profiter des possibilités de renforcer la démocratie, de produire prospérité et d'exploiter leur potentiel humain. »
- Déclaration de Bávaro par laquelle est adopté un ensemble de principes directeurs de la transition à la société informationnelle et sont identifiés les thèmes prioritaires qui doivent être abordés pour que les pays de l'Amérique latine passent du consensus politique à un accord opérationnel décrit dans l'eLAC2007.<sup>7</sup>

## ■ L'eLAC2007

L'eLAC2007 est le Plan d'Action Régionale sur la Société de l'Information pour l'Amérique latine et les Caraïbes pour les années 2005-2007 dont l'objectif est de faciliter la coopération régionale et l'échange de meilleures pratiques pour réduire des coûts et des temps d'apprentissage dans l'acquisition et l'application des technologies de l'information et de la communication (TIC). Il constitue également un instrument de médiation

<sup>5</sup> Escuela Iberoamericana de Administración y Políticas Públicas. 2006. Montevideo.

<sup>6</sup> Mauras Marta y Mariano Ferrero (2007). El Plan de Acción Regional eLAC2007 : una nueva concertación regional para la Sociedad de Información Inclusiva. Proyecto CEPAL/Unión Europea.

<sup>7</sup> Op Cit.

entre les objectifs globaux du Nouveau Millénaire à long terme (2015) et les objectifs spécifiques à court terme répondant aux besoins de l'environnement régional et permettant aussi des évaluations et des retours constants.

L'eLAC2007 est constitué par 30 objectifs et 70 lignes d'action qui comprennent cinq secteurs critiques d'action identifiés par les pays de la région :

1. l'accès et l'inclusion numérique,
2. la création de capacités et de connaissances,
3. l'efficacité et la transparence des contenus et des services publics,
4. les instruments de politique, et
5. l'environnement permettant la mise en place.

Une évaluation effectuée par la CEPALC au Plan d'Action Régional Elac2007<sup>8</sup>, confirme un progrès significatif du développement des sociétés de l'information en Amérique latine, puisque des 27 secteurs d'action évalués, 15 reflètent des avancements, y incluant des avancements importants, alors que 12 présentent des avancements modérés ou insuffisants. Voici quelques résultats.

### **1) Domaine d'accès et d'inclusion numérique**

Un fort progrès a été constaté dans les centres communautaires et dans les gouvernements locaux, alors qu'aucun progrès n'a été constaté dans les centres de santé en ligne.

Pour parler des progrès dans le domaine d'accès et d'inclusion sociale, il faut d'abord connaître les avancements effectués dans l'infrastructure informatique et la connectivité, puisque ce sont les instruments qui permettent le développement des TIC. Dans le cas du Costa Rica, la connectivité a été exclusivement entre les mains de deux entreprises publiques : l'Institut costaricien de l'électricité (ICE) et Radiográfica Costarricense S.A (RACSA). Toutefois, en juillet 2008, a été approuvée une loi qui permet l'ouverture des télécommunications, ceci en réponse aux exigences de l'accord de libre échange entre les pays de l'Amérique Centrale et la République Dominicaine avec les Etats-Unis (CAFTA).

Afin de répondre aux demandes croissantes d'une plus grande connectivité, ces entreprises ont dû réaliser d'importants investissements technologiques. Dans le cas de l'ICE « ses capacités vers des technologies optiques d'avant-garde (...) sont en train de s'étendre pour remplir les futures demandes de tous leurs clients, en vue d'une société totalement numérique. »<sup>9</sup> Toutefois, l'investissement effectué par l'ICE en 2006 a diminué de 20% par rapport à l'année 2005.

Pour sa part, RACSA « a développé des projets innovants en télématique et technologies d'information ». Elle a, en outre, remplacé les connexions par satellite par des liaisons internationales et a investi dans le développement de réseaux nationaux de haut débit et a intégré le service à larges bandes passantes en doublant son investissement afin de fournir les services d'Internet.

Selon les données de l'indice global d'« e-readiness » des Nations Unies, le Costa Rica occupait, en 2005, la 70<sup>ème</sup> position sur 191 pays, alors que selon l'indice du «The Global information Technology Report » 2005-2006, World Economic Forum 2006, le Costa Rica a été classé 69<sup>ème</sup> dans l'Indice de Connectivité, et passant en 2008, au poste 59.<sup>10</sup>

Il reste évident également que le progrès réalisé par Costa Rica dans le développement de services numériques sur Internet a été de type informationnel mais pas dans les services de type transactionnel ou de réseaux, situation qui place le Costa Rica parmi les pays les moins avancés en matière de Gouvernement Numérique.

<sup>8</sup> Observatorio para la Sociedad de la información en América Latina y el Caribe (OSILAC). (2007). Monitoreo del eLAC2007 : avances y estado actual del desarrollo de las Sociedades de la Información en América Latina y el Caribe.

<sup>9</sup> Programa de la Sociedad de la Información y el Conocimiento. (2008) II Informe. Universidad de Costa Rica. Ce rapport couvre l'année 2007.

<sup>10</sup> Plan de Acción del Gobierno Digital en Costa Rica 2008-2010. Sur le site: <http://www.gobiernofacil.go.cr/gobiernodigital/index.html>.



## **Des services numériques de l'Administration Publique**

96% des institutions publiques au Costa Rica ont accès à Internet et 74% disposent d'un site Web public, mais 41% ont l'accès à Internet moyennant une largeur de bande égale ou supérieure à 1Mbps<sup>11</sup>. Toutefois, en faisant la comparaison avec l'étude effectuée par CAATEC (2002), un certain progrès est constaté puisqu'en 2002, seulement 21% des institutions étudiées disposaient d'une largeur de bande supérieure à 1.204 kbps.

Comme il se dégage d'une étude proposée par le Ministère du Plan et effectuée par l'Université de Costa Rica avec le soutien du Programme des Nations Unies pour le Développement (PNUD), 92 % des institutions publiques costariciennes offrent un certain type de service numérique aux citoyens et 67% de ces services sont effectués tant par téléphonie mobile que par téléphonie fixe.<sup>12</sup>

Dans le Gouvernement Central (2006), bien que 95% des ministères aient un site web, ils ont seulement complété la phase de présence ou informationnelle<sup>13</sup> (selon le modèle utilisé par les Nations Unies), laissant un grand vide dans les phases interactive et transactionnelle. Par contre, les banques publiques ont obtenu un développement plein des quatre phases : paiement de services publics comme l'eau, le téléphone, les impôts municipaux, l'électricité. Dans le cas des institutions non financières et de service, 100% sont dans l'étape informationnelle, 17 % ont une information sur les démarches institutionnelles et seulement 10% permettent de faire quelques démarches en ligne.<sup>14</sup>

De la même façon, des 81 gouvernements locaux, seulement 8 municipalités disposent d'un site web avec information institutionnelle et conditions pour les démarches. 62.5% permettent quelques démarches en ligne : autorisations de plans, autorisations de construction et paiement d'impôts municipaux.

Pour l'année 2007, 70 % des services offerts par les institutions publiques, au moyen de l'Internet et du téléphone, sont encore informationnels et seulement 30% sont transactionnels.

L'étude effectuée par le PROSIC (2008) montre le peu de présence dans Internet des municipalités dont seulement un cinquième des 81, ont une page web. En outre, 80% des municipalités possédant une page se trouvent au centre métropolitain du pays, il s'agit des villes de San José, Heredia et d'Alajuela. Un aspect négatif est que les pages ne sont pas périodiquement mises à jour, ce pourquoi il est fréquent de trouver une information désuète.

## **Détention des TIC dans les foyers**

En ce qui concerne la détention des TIC, pendant l'année 2006, le Costa Rica montre les couvertures suivantes :

- Radio : 84.9%
- Téléphonie résidentielle : 93.7%
- Téléphonie mobile : 56.4%
- TV à couleur : 93.7%
- TV par câble : 24.4%
- Ordinateurs : 28.2%
- Internet : 9.8%
- Électricité : 99,15%

Si l'on compare les indices de détention des TIC de 2006 avec celles de 2005, on peut remarquer un accroissement dans les cas suivants : 6.53% dans la téléphonie mobile, 2.26% dans la TV par câble, 1.20% dans

<sup>11</sup> Estudio CONATIC-MICIT (2005): Recursos Informáticos y Tecnológicos de la Administración Pública.

<sup>12</sup> CICAP, Universidad de Costa Rica (2006) Proyecto Diagnóstico sobre el Gobierno Digital en Costa Rica.

<sup>13</sup> Informationnel : les pages ne permettent pas d'interaction avec le citoyen et elles se limitent à offrir une information institutionnelle et à fournir des formulaires qui doivent être présentés personnellement ou remplis à la main.

<sup>14</sup> Programa de la Sociedad de la Información y el Conocimiento. (2007) Informe. Universidad de Costa Rica. Ce rapport couvre l'année 2006.

les ordinateurs, 0.98% dans la TV couleur et 0.09% dans l'électricité. Mais il y a eu aussi une diminution dans les domaines suivants : 1.37% dans la radio, 0.39% dans Internet et 0.19% en téléphonie résidentielle.<sup>15</sup> Il est important de signaler que l'accroissement dans la téléphonie mobile s'est produit dans tous les niveaux économiques des foyers.

D'autre part, si on analyse la détention des TIC dans les foyers des zones urbaines et rurales du Costa Rica, on observe une fracture numérique claire dans l'année 2005, tandis que 75.8% des foyers de la zone urbaine détiennent au moins une ligne de téléphone résidentielle ; dans la zone rurale ce pourcentage descend à 50.5%.

En 2006, la fracture numérique continue entre les zones rurales et les zones urbaines, puisque 75% de la téléphonie résidentielle et 65% de la téléphonie mobile se situent dans les zones urbaines contre 51% de téléphonie résidentielle et 43% de téléphonie mobile dans les zones rurales.

Par rapport à l'utilisation des ordinateurs, 36% se présentent dans les zones urbaines et seulement 15% dans les zones rurales ; concernant l'utilisation de l'Internet, 14% se présentent dans les zones urbaines contre 4% dans les zones rurales. De même, « la moyenne de personnes par foyer utilisant l'Internet est 2.2% et 2.51% respectivement en 2006 et 2007. L'âge moyen du cybernaute était de 28 ans et a augmenté à 29 ans en 2007. »

<sup>16</sup> Une donnée curieuse à signaler, selon l'analyse du PROSIC, est en relation avec le lieu depuis lequel se réalise l'accès à l'Internet : 33% depuis les centres d'étude, 28% depuis le lieu de travail et 33% depuis « les cafés internet ».

Finalement, selon des données de la Commission Nationale des Technologies de l'Information et des Communications (CONATIC)<sup>17</sup>, le Costa Rica occupe la troisième place selon l'indice mondial de consommation individuelle de téléphonie mobile avec une moyenne de 253 minutes par mois, seulement au-dessous des Etats-Unis et de Hong Kong.

## 2) Domaine de Création et des capacités de connaissance

Un progrès important a été enregistré dans les réseaux de la recherche et de l'éducation, mais le secteur des sciences et technologies n'a obtenu aucun progrès.

Dans le cas spécifique du Costa Rica, selon le Rapport de Compétitivité Globale 2007-2008 qui comprend 131 pays, le pays se trouve dans un niveau de transition vers une Économie Basée sur la Connaissance (EBC), ce niveau correspond au revenu par habitant du pays et qui lui permet d'aspirer aux niveaux plus élevés de bien-être économique, s'il continue sa marche avec succès vers une EBC. Parallèlement, bien qu'il montre un certain avantage comparatif dans les secteurs de l'éducation de base et de la santé, il n'atteint pas encore les « standards de pays en transition vers des économies innovatrices, ou bien, économies qui sont déjà dirigées par l'innovation. »<sup>18</sup>

Pendant les dix dernières années, le Costa Rica a assisté à une croissance systématique d'entreprises en rapport avec les TIC, tant des nationales que des multinationales : composants (conception, manufacture, assemblage et ventes de hardware)<sup>19</sup>, développement de software, services de conseil dans le développement, mise en œuvre et maintien de systèmes, entre autres. 59.8% des entreprises costariciennes sont consacrées aux services directs (conseils, développement de software), 21.4% au développement de software, et 6.8% aux composants de hardware.

Quant au niveau d'innovation, les activités les plus significatives ont été en rapport avec les innovations des produits ou du service (lancement d'un nouveau produit ou améliorations dans les services) ; après s'est produite l'innovation de l'entreprise (changements dans l'organisation pour améliorer les processus productifs).

---

<sup>15</sup> Programa de la Sociedad de la Información y el Conocimiento. (2007) I Informe. Universidad de Costa Rica. Ce rapport couvre l'année 2006.

<sup>16</sup> Opus Cit. Página 160.

<sup>17</sup> Comisión Nacional de Tecnologías de Información y Comunicación (CONATIC). Boletín Ciencia y Tecnología N° 53 Dic. 2006. <http://www.conicit.go.cr/boletin/boletin27/conatic.shtml>.

<sup>18</sup> Hewitt John y Ricardo Monge (2008). Los Costarricenses en la Economía basada en el Conocimiento. Infraestructura, destrezas, uso y acceso de las TICS. Serie Costa Rica Digital. Comisión Asesora en Alta Tecnología (CAATEC).

<sup>19</sup> Monge G. Ricardo y otro (2008). Competitividad e innovación en Costa Rica. Innovación, Competitividad y Crecimiento. Desempeño de Costa Rica y de su sector de las TICS. CAATEC. Cartago, CR.

### 3) Transparence et efficacité publiques

Un fort progrès en e-éducation, quelques progrès en e-gouvernement et en information publique et patrimoine culturel, un progrès modéré en e-justice et la protection environnementale et aucun progrès en e-santé.

Selon Monge et Hewitt (2008), le Costa Rica « fait face à d'importants défis, comme la couverture de l'éducation secondaire, l'accès à l'Internet dans les écoles et les collèges, ainsi que la couverture de l'éducation universitaire et la qualité de l'éducation dans les domaines des mathématique et des sciences. »<sup>20</sup>

Toutefois, le Costa Rica a promu l'utilisation des TIC de manière créative, tant dans l'enfance que dans la jeunesse par le Programme National d'Informatique Éducative (PRONIE) du Ministère de l'Éducation Publique (MEP) et la Fondation Omar Dengo (FOD), en passant d'une couverture de 14.5% pendant l'année 1988 à 53.4% en l'année 2006 (éducation primaire) et avec une couverture plus grande à deux parties dans les collèges et lycées de l'éducation secondaire au travers du PRONIE.

Quelques programmes et projets effectués par la FOD<sup>21</sup> sont les suivants :

- Programme Institut Innov@ en rapport avec des expériences informatiques éducatives et qui est composé de plusieurs projets comme Démocratie et Citoyenneté Active, Internet et Apprentissage en Ligne, Innovation Éducative, Robotique et apprentissage par conception, compétences Numériques et Entrepreneurs et Productivité Numérique, entre autres.
- Centre pour la Citoyenneté et la productivité Numérique (CIPRODI)
- Réseau de Télématique Éducative (RTE)
- *Ciber@prendiz* Applications de l'Internet pour l'Apprentissage Éducatif (AIAE)
- Intel® Instruire pour le futur
- Labor@
- Journal Numérique Zone M, et d'autres.

Ainsi, à travers le Programme d'Amélioration de la qualité de l'éducation préscolaire et de la générale de base (PROMECE), un investissement de 23.8 millions \$ a été fait en infrastructures, équipement, formation et ressources didactiques dans des secteurs urbain-marginaux et ruraux, afin de réduire la désertion scolaire et améliorer la qualité académique dans l'éducation secondaire.

Par rapport aux universités publiques, seulement quelques pas naissants ont été faits pour incorporer les TIC dans l'académie ainsi que dans l'administration. Un exemple de cela se présente à l'Université du Costa Rica où des efforts ont été faits pour fournir aux enseignants l'accès à Internet sans fil mais en raison des problèmes dans les locaux et des connexions en mauvais état, ces initiatives n'ont pas encore réussi. De même, même si l'enseignant souhaite incorporer les TIC pour donner ses cours et même en disposant de quelques technologies, la bureaucratie universitaire frustre les enseignants dans l'utilisation des TIC.

### 4) Instruments de politique

Un progrès très important concernant les indicateurs et la mesure, un certain progrès en stratégies nationales et aucun progrès en financement, en politiques d'accès universel et cadre législatif.

Au Costa Rica, les premiers pas du gouvernement numérique sont donnés pendant l'administration de Miguel Angel Rodríguez avec l'initiative appelée « Agenda Numérique » dont l'objectif principal était l'approbation de la Loi sur la signature électronique par l'Assemblée Législative.

Postérieurement, pendant l'Administration d'Abel Pacheco de l'Espriella a été créée la Commission Nationale des Technologies de l'Information et Communication ayant pour but de proposer les politiques et les stratégies en matière de technologies de l'information et la communication adressée aux institutions du

---

<sup>20</sup> *Op. Cit.*

<sup>21</sup> Fundación Omar Dengo : <http://www.fod.ac.cr/>.

secteur public, ainsi que recommander des linéaments techniques et administratifs pour orienter l'action dans le domaine au moyen des activités suivantes :

- Établissement de plans stratégiques en TIC
- Développement de plans de sécurité et continuité de l'affaire
- Adoption de normes et standards ouverts et interopérabilité
- Adoption de normes coût-bénéfice dans les investissements en TICs
- Impulsion des initiatives de Gouvernement Électronique
- Mise en œuvre de la loi sur les certificats, la signature numérique et les documents électroniques
- Réalisation de diagnostics des sites web dans les institutions publiques
- Réduction de la fracture numérique au moyen de programmes comme ACCÈS et Centres communautaires d'Accès à Internet (CCAIs).

Finalement, au début de l'administration d'Oscar Arias Sánchez, a été créé le Secrétariat Technique de Gouvernement Numérique de la Présidence de la République et la Commission Intersectorielle de Gouvernement Numérique, présidée par le Second Vice-président de la République et intégrée par les Ministères de la Présidence, du Plan National et Politique Économique, de l'Agriculture et de l'Élevage, de l'Économie, de l'Industrie et du Commerce, de l'Environnement et l'Énergie, de la Science et la Technologie et celui de la Coordination Interinstitutionnelle.

Bien qu'il ait été annoncé comme une des priorités de la présente administration, le projet de Gouvernement Numérique a traversé une série des hauts et des bas qui ont empêché son exécution de manière convenable: à un an près de sa création, et en raison de la démission du Second Vice-président, la présidence du Secrétariat est resté en mains de la Première Vice-présidente de la République qui, à son tour, a démissionné en octobre 2008 pour se consacrer à la préparation de sa pré-candidature aux élections présidentielles de 2010. Afin que le projet d'e-gouvernement ne soit pas paralysé, le président de la République a décidé de transférer son administration à l'Institut costaricien de l'électricité (ICE) le 13 novembre 2008.

Selon une étude récente de l'Université du Costa Rica, parmi les principales faiblesses institutionnelles du gouvernement numérique ou e-gouvernement, se trouvent l'inexistence de bases de données adéquates, des connexions à Internet inexistantes ou lentes, la faible inter-connectivité numérique interinstitutionnelle et avec les bureaux régionaux. Un autre aspect très important est l'ignorance des besoins et des préférences des citoyens. À ceci s'ajoute une vision institutionnelle fermée, qui pourrait neutraliser les initiatives de développement institutionnel, et les restrictions budgétaires qui conditionnent la vitesse et la portée du développement de cette nouvelle façon d'offrir des services publics.

Afin de faire avancer le développement du Gouvernement Numérique et ne disposant pas de ressources suffisantes, le Secrétariat Technique a effectué quelques alliances stratégiques avec des institutions publiques ayant mis en œuvre les TIC dans leur travail quotidien. C'est le cas, par exemple, de l'alliance avec la Banque de Costa Rica (BCR). Grâce à cette alliance, le service pour la livraison de passeports et de permis de conduire est offert par les agences de la BCR.

Les résultats obtenus sont positifs en 2007 ; entre juillet et décembre, 160.000 demandes de passeport ont été traitées dans tout le pays avec une durée de réponse de 15 minutes. Le Secrétariat Technique considère que l'économie en temps et en argent est très significative, la couverture du service s'est étendue d'un point à 38 points dans tout le pays, ce qui a permis la diminution du temps de réponse aux utilisateurs, d'une moyenne de 4 heures à 15 minutes par démarche et la livraison d'un passeport de 72 heures à 48 heures.

D'autres expériences réussies du Gouvernement Numérique ont été la numérisation du Registre National, du système national de la législation, les achats numériques au moyen de COPRARED, le paiement des services publics et des impôts au moyen de l'Internet, l'autorisation des plans constructifs par le Collège fédéré d'ingénieurs et d'architectes, entre autres.

Par rapport aux plans et aux linéaments pour favoriser l'utilisation des TIC dans le pays, des projets prioritaires à court et à moyen terme, et d'autres à long terme ont été prévus<sup>22</sup> :

### **Projets prioritaires**

- Infrastructure de connectivité de l'Etat costaricien (ICEC). Selon des estimations des experts, cela se traduit par une économie d'approximative de 0.5% du PIB (100 MUSD).
- Système électronique d'achats de l'Etat costaricien avec une économie estimée d'environ 2% du PIB (US\$ 500 MUSD par an).
- Contribution Numérique. Une croissance des revenus fiscaux de 2% du PIB (500 MUSD) est prévue.

### **Projets à terme court et moyen**

- Automatisation de l'encaissement de péages
- Autorisations d'immigration
- Système unique de bénéficiaires
- Portail et centre des données
- Interopérabilité du registre de citoyens
- Programme en reliant des idées
- Développement et certification des cafés Internet
- Villes Numériques
- Portail de commercialisation d'enchaînement de MIPYMES (micro, petites et moyennes entreprises)
- Télétravail
- Formation en ligne
- AyA Numérique (Institut de l'eau)
- Croix Rouge Numérique
- Bibliothèques Virtuelles
- Publication d'actes administratifs

### **Projets à long terme**

- Système intégré d'identification, agences et rendez-vous
- Système intégré de dossiers médicaux
- Modernisation de cartes d'identité
- Vote Électronique

Par rapport au cadre normatif, bien que la loi sur les certificats, les signatures électroniques et les documents électroniques a été approuvée depuis 2005, elle n'a pas encore été mise en vigueur au Costa Rica, dû au fait que les conditions de sécurité informatique établies dans le règlement de la Loi s'avèrent assez onéreuses et les entreprises intéressées ont considéré que le marché est trop petit pour un investissement aussi important. Il existe aussi une grande quantité de lois en rapport avec l'utilisation et la conservation des documents électroniques.

---

<sup>22</sup> Secretaría Técnica de Gobierno Electrónico. Plan de Acción Gobierno Digital 2008-2010. Enero 2008, San José.

## 5) Environnement pour la mise en œuvre

Un fort progrès du Suivi du Sommet Mondial de la Société de l'Information et l'exécution eLAC 2007. Toutefois, malgré la reconnaissance des progrès importants réalisés afin de développer une Politique Nationale de l'Information, il faut encore la mise en place des mesures pour renforcer et réviser les politiques déjà existantes afin qu'elles répondent aux défis et aux demandes de l'environnement global et permettent le développement humain, social, politique, économique et culturel dans chaque pays.<sup>23</sup>

Parmi les contributions internationales à l'Amérique latine pour promouvoir le développement d'un Système National d'Information (SANS), la plus significative a été celle de l'UNESCO, à travers le Programme de la Société de l'Information pour l'Amérique latine (INFOLAC) dont l'objectif est de « Renforcer l'autonomie individuelle de l'ensemble des institutions nationales de l'Amérique latine et des Caraïbes pour créer et opérer des mécanismes d'organisation et d'accès à la ressource informative et intensifier son utilisation dans la formulation, l'exécution, l'évaluation et l'administration de leurs plans, leurs programmes, leurs politiques et leurs actions de développement économique, social, scientifique, technologique et culturel. »<sup>24</sup>

Dans le cas du Costa Rica, selon une étude effectuée par Molino, en 1990, plusieurs initiatives ont été faites afin de créer un système national d'information. La première de celles-ci a été la création du Conseil National des Recherches Scientifiques et Technologiques (CONICIT) qui opère depuis 1972.

Par la loi 7099 du 26 octobre 1988, a été établi un prêt avec la Banque Interaméricaine de Développement pour contribuer au développement scientifique et technologique au moyen du financement de projets de recherche, de construction et d'équipement de centres nationaux et universitaires de recherche et pour la formation de ressources humaines spécialisées.

En 1990, par la loi de la promotion du développement scientifique et technologique, il est établi de manière explicite l'objectif de favoriser les activités d'appui, y compris la documentation et l'information scientifique et technologique, et les activités stratégiques y ont été définies.

En 1993, par le décret exécutif 22282 du 15 juillet 1993, a été créé le Système National d'Information Scientifique et Technologique (SINYCIT) qui coordonnera la production d'information technique afin de la fournir aux secteurs productifs du pays pour qu'ils puissent se moderniser et soutenir le développement économique. Ses principaux objectifs sont les suivants :

1. « Procurer l'utilisation nationale des ressources d'information nationale par la spécialisation de celles-ci, avec une organisation autonome et autochtone, mais avec un accès disponible aux utilisateurs des différents secteurs.
2. Promouvoir l'utilisation du réseau téléphonique commuté national (RTC), du réseau de transmission de données publiques et de l'interconnexion aux réseaux de recherche et au réseau des réseaux Internet, ainsi qu'aux réseaux locaux.
3. Pousser l'organisation interinstitutionnelle pour un développement coordonné et planifié des systèmes et des services d'information, dans les différents secteurs nationaux et en complément de l'information du Registre scientifique et technologique ».<sup>25</sup>

D'autres organismes en rapport avec les politiques d'information sont le Ministère des sciences et de la Technologie qui définit la politique générale ; le Conseil National des Recteurs (CONARE), qui coordonne les aspects académiques ; l'entreprise Radiográfica Costarricense, S.A (RACSA) qui offre les services de réseaux publics ; et la Commission Consultative sur les hautes technologies du Costa Rica (CAATEC) qui réalise un suivi permanent du développement des TIC dans le pays.

Les principales difficultés identifiées dans notre système national d'information sont les suivantes :

- L'absence des politiques et des plans nationaux intégrés.
- Le manque d'union et de coordination entre les différentes unités d'information existantes dans le pays.

<sup>23</sup> Pájaro Rosaura y otra. (2007). *Objetivos y áreas de acción de política nacional de información*. Dans: *Hacia la construcción de políticas nacionales de información: la experiencia de América Latina*. UNESCO.

<sup>24</sup> <http://www.infolac.org>

<sup>25</sup> <http://www.sinalevi.pgr.cr>.

- Le déficit de ressources économiques. La plupart d'unités d'information ne disposent pas de budget fixe suffisant leur permettant d'offrir des services et des matériaux optimaux.
- Nécessité de formation de ressources humaines.
- Normalisation des procédures, software, outils, etc.
- Meilleure connaissance des nécessités des utilisateurs et la formation de ces derniers.
- Mécanismes pour l'encaissement des services.
- Continuité des services après arrêt d'appuis internationaux.
- Nécessité de maintenir et de promouvoir des relations avec des organismes d'appui.
- Nécessité de conseiller des techniques sur les orientations pour l'acquisition des biens et services informatiques.
- Des volumes de production de matériaux éditoriaux limités.
- L'information existante dans le pays n'est pas exploitée de manière exhaustive.
- La non inclusion des archives et des musées comme composantes du Système, malgré l'insistance de l'UNESCO qui les considère comme unités fondamentales de la fourniture d'information.

Malgré ce qui est précédemment exposé, quelques politiques<sup>26</sup> en rapport avec l'information sont identifiées :

- Politiques d'information scientifique et technologique. Conseil National pour les Recherches Scientifiques et Technologiques (CONICIT), Registre scientifique et technologique.
- Politiques d'information statistique et cartographique. Institut national des statistiques et des recensements : recensements de la population et enquêtes de foyers
- Politiques d'Accès à l'information publique des administrations. Constitution Politique : accès à l'information publique et droit de demande et de réponse rapide. Conventions : Déclaration Universelle des Droits de l'Homme, Convention Américaine sur les Droits de l'Homme, Pacte International des Droits Civils et Politiques.

## ■ La « Carrière d'Archivistique »

Les études d'Archivistique font leur apparition au Costa Rica avec l'ouverture d'un Diplôme en archives administratives (1978) dont le but était d'offrir une formation très élémentaire aux fonctionnaires de l'Administration Publique responsables des archives de l'État, mais surtout pour former les fonctionnaires des Archives Nationales. Cette première étape de la Carrière s'est nettement caractérisée par une formation empirique et par une analyse d'archive finale ou historique. Formation empirique parce que la plupart des professeurs étaient des historiens qui avaient suivi quelques « cours de formation, des ateliers et des stages de courte durée » en Argentine et en Espagne, et une approche d'archive finale parce qu'une grande partie des professeurs étaient des fonctionnaires des Archives Nationales dont l'expérience se limitait strictement aux documents historiques. Le Diplôme a fonctionné pendant douze ans, sans changements excepté deux modifications substantielles : l'inclusion de deux cours d'Informatique (1985) et l'introduction du cours Atelier d'archivistique générale (1990).<sup>27</sup>

En 1992, est créée le Bachillerato<sup>28</sup> en Archivistique afin de répondre aux besoins du marché de travail du pays, en raison de l'existence d'un cadre normatif pour la création et le développement du Système National d'Archives. Malgré l'intention d'introduire dans cette deuxième étape l'analyse de « gestion de documents » et

<sup>26</sup> Bustamante, José (2007). Situación comparada en América Latina y El Caribe. Dans: Hacia la construcción de políticas nacionales de información: la experiencia de América Latina. UNESCO.

<sup>27</sup> Echavarría Solís, Ana Lorena y José Bernal Rivas Fernández (2005). Reseña de la formación profesional de archivistas en Costa Rica. Ciudad Universitaria Rodrigo Facio.

<sup>28</sup> Bac+4, correspondant à l'ancienne licence française ou à un master 1.

celle « d'Archivistique intégrée », l'objectif n'est pas atteint, en particulier, parce que le corps d'enseignants reste presque le même que celui qui offrait les cours du Diplôme : les cours ayant un rapport avec les processus techniques d'archivistique restent les mêmes que pour le Diplôme, suivant une approche historique, sans pratique archivistique et avec une théorie très élémentaire, pour ne pas dire de base. En outre, une grande quantité de cours d'Histoire continuent d'être donnés.

En 1999, une restructuration du plan d'études est entreprise afin que les étudiants soient mieux formés pour exercer leur travail au sein des archives administratives tant publiques que privées, et en prévision de l'apparition du document électronique ; un cours d'Histoire est éliminé et remplacé par un cours de Sciences de l'information. Toutefois, le problème des cours ayant une relation avec les processus archivistiques continue, dû au fait que les enseignants ne se mettent pas à jour et continuent d'enseigner l'Archivistique de manière traditionnelle.

En 2004 et 2005, le plan d'études du Baccalauréat est modifié partiellement et est créée la maîtrise en Archivistique afin de former des professionnels en Sciences de l'Information capables de concevoir et d'administrer des systèmes de gestion de l'information, de développer des recherches de haut niveau et de promouvoir une culture d'information institutionnelle. Bien que les problèmes avec les processus archivistiques cités précédemment continuent, un changement est perçu chez les étudiants qui cessent d'avoir un rôle passif et adoptent une position plus proactive.

De même, depuis 2006, le cours « Analyse et Conception d'Archives I et II » est complètement changé. Ce cours est le dernier cours d'Archivistique obligatoire avant l'obtention du baccalauréat. Le changement consiste à ce que les étudiants doivent appliquer la Norme ISO 15489 dans un organisme public ou dans une entreprise privée. Pendant le premier cycle de l'année (mars-juillet), ils développent les étapes A, B, C et D de la Norme, tandis que pendant le deuxième cycle sont appliquées les étapes E, F, G et H. Le travail a été à la fois très enrichissant pour l'enseignant et pour les élèves, mais épuisant, en raison de la nécessité de reprendre les processus archivistiques dans leur totalité.

D'autre part, bien que nombre de cours, conférences et ateliers en rapport avec l'Archivistique ont été offerts, l'un des plus grands vides détectés dans la Carrière d'Archivistique est l'absence d'un programme de formation continue qui permette la mise à jour constante de professeurs et anciens élèves.

Au début de l'année 2008, un processus d'auto-évaluation des études d'Archivistique est entrepris afin de détecter des forces, faiblesses, menaces et opportunités, et de chercher l'amélioration continue dans tous les processus internes (académiques, administratifs et estudiantins) et de répondre ainsi efficacement aux nécessités du marché du travail archivistique. Ce processus d'auto-évaluation se terminera durant l'année 2009 et l'idée est d'aboutir sur une ré-ingénierie du cursus.

Finalement, les promotions conclues jusqu'à ce jour sont les suivantes : 227 *Diplomado* (Diplôme), 60 *Bachillerato* (licence ou Master 1) et 8 *Licenciatura* (Master 2).

## ■ Les Archives Nationales et le « Système National d'Archives »

La création des Archives Nationales date de 1881 quand l'historien Leon Fernández a réussi à réunir des documents de la période coloniale et de la période postérieure à l'Indépendance du pays. Jusqu'à la décennie soixante-dix, les Archives Nationales ont eu un rôle de garde des documents qui formaient le patrimoine documentaire du pays, sans prendre soin des documents administratifs qui étaient en train de se produire (et dans beaucoup de cas éliminés) au sein des institutions de l'État.

À partir de l'année 1972, le Costa Rica est choisi par l'Organisation d'États Américains (OEA) pour développer un plan pilote en matière d'archives. Grâce à ce Plan Pilote, les Archives Nationales reçoivent la visite d'experts archivistes du Conseil International des Archives comme M. Tanodi, Mesdames Bodil Ulate et Zelia Zaher, M. Bruno Delmas et Mme Vicenta Courtois Alonso qui orientent les Archives Nationales vers des meilleures pratiques archivistiques. De même, une grande quantité de bourses pour des fonctionnaires des Archives Nationales sont accordées et ils vont suivre une formation non réglée, principalement en Argentine et en Espagne, mais aussi en France, en Suède et en République Dominicaine.



Tandis que les Archives Nationales menaient une révision profonde et se restructuraient, « un développement de beaucoup d'autres éléments fondamentaux d'un véritable système d'archives est entamé : les archives des institutions publiques; la formation professionnelle des archivistes ; les ressources économiques nécessaires et indispensables ; l'infrastructure adéquate, etc. »<sup>29</sup>

En 1974, est créé le Conseil d'administration des Archives Nationales dont le principal objectif était de doter d'un bâtiment propre à la Direction Générale des Archives Nationales (DGAN), objectif qui se matérialise durant l'année 1993 avec l'inauguration de la première étape (un bâtiment moderne sur une surface de 6.600 m<sup>2</sup>) et durant l'année 1998 avec la deuxième étape (bâtiment pour réservoirs, le secteur de conservation, salles de formation et cantine pour les fonctionnaires, sur une surface de 4.000 m<sup>2</sup>).

Pendant la décennie 80, la DGAN entame une série d'activités qui allaient préparer le terrain pour la consolidation du Système National d'Archives : cours de formation non réglée pour les fonctionnaires publics responsables des archives administratives; visites inspection et assessorat aux institutions de l'État; rencontres archivistiques nationales et bancaires; circulaires à l'Administration Publique sur les soins des documents.

La décennie des 90 débute avec l'approbation de la Loi 7202 du Système National d'Archives, réglementation d'avant-garde pour cette époque-là et qui a servi de base pour d'autres lois archivistiques de la région, comme par exemple le cas de la Colombie. Parmi les articles les plus significatifs de cette Loi, on peut souligner les suivants :

- Création du Système National d'Archives dont le recteur est la DGAN.
- Toutes les archives de l'État, excepté (par une erreur de rédaction) celles de l'Organe Électoral, sont obligées d'observer les linéaments en matière archivistique qu'émet la DGAN.
- Parmi ses fonctions, la DGAN doit donner des recommandations en matière de gestion de documents.
- Création de la Commission nationale de sélection et d'élimination de documents et interdiction, sous peine de prison, d'éliminer des documents de l'État sans l'approbation préalable de cette Commission.
- Toutefois, cette loi est restée désuète dans beaucoup d'aspects et il faut une prompt modification ou l'élaboration d'une nouvelle loi qui réponde aux défis archivistiques du XXI<sup>ème</sup> siècle. Concernant les ressources humaines, le nombre de fonctionnaires est augmenté de 36 en 1980 à plus de 100 en 2008.

Une autre réalisation importante de la DGAN a été la création de plusieurs commissions : Commission de la description documentaire, Commission de normalisation de la gestion de documents et d'autres processus archivistiques et, finalement, Commission de gestion du document électronique. Toutefois, de ces trois commissions, seulement la Commission de description documentaire fonctionne aujourd'hui et malgré le grand travail effectué pendant les cinq dernières années, elle n'est pas parvenue à élaborer des normes nationales de description qui servent à orienter la description documentaire depuis les archives de gestion. Il a été question, à un moment donné, de combiner la connaissance académique de la Carrière Archivistique avec la partie pratique de la DGAN pour travailler ensemble dans cette Commission, mais cette proposition a été rejetée par la DGAN.

Au niveau international, la DGAN a réussi à obtenir la Présidence d'Association latino-américaine d'archives; être le siège de la première réunion du Comité Exécutif du CIA effectuée en Amérique et participer dans plusieurs forums internationaux.

Par rapport à l'utilisation des TICs, les Archives Nationales sont en retard dans leur application à cause de deux facteurs fondamentaux : les restrictions budgétaires et la résistance des hautes autorités à les utiliser pour offrir un meilleur service à l'utilisateur. La politique de la DGAN a été, jusqu'à ce jour, de microfilmer les documents les plus anciens et ceux se trouvant très détériorés. L'investissement le plus grand effectué par la DGAN est celui d'avoir converti presque la totalité de son inventaire papier en données de base microis, base de données conçue et créée par les fonctionnaires de cette direction lors d'un cours de formation. En outre, la vision de la présence physique de l'utilisateur dans la salle de recherches reste encore, et la possibilité d'offrir un service virtuel à l'utilisateur a été négligée. Cette vision limitée des besoins de l'utilisateur actuel empêche l'accès des internautes locaux et étrangers à l'information si précieuse que gardent les Archives Nationales.

---

<sup>29</sup> Chacón Arias, Virginia (2000). El ayer y el hoy del desarrollo archivístico. <http://www.archivonacional.go.cr/articulos.htm>.

La DGAN dispose d'un site web qui offre des informations institutionnelles : réglementation, structure organique, services, etc., mais n'offre pas l'option de réaliser des consultations en ligne. Sur ce thème, les Archives Nationales du Costa Rica montrent du retard par rapport aux archives nationales dans la région latino-américaine, tel est le cas des Archives Nationales du Chili dont le site met à disposition de l'internaute un système de consultations et de certification électronique de documents.

En ce qui concerne la consolidation du Système National d'Archives pendant les premières années du XXI<sup>e</sup> siècle, selon une analyse effectuée par Bermúdez Muñoz (2008)<sup>30</sup>, -basée sur un rapport des Archives Nationales de l'année 2004 où les institutions de l'État sont classées par A, B et C, où le A correspond à un organisme ayant une archive centrale, un archiviste et un comité de sélection et d'élimination de documents ; le B à un organisme où il manque un des trois paramètres susmentionnés, et le C à un organisme n'ayant qu'un seul ou aucun de ces paramètres-, seulement 19% des institutions ou d'organismes publics remplissent la réglementation en vigueur en relation avec les archives : tel est le cas du Pouvoir Législatif, de quelques ministères, institutions autonomes et municipalités. 13 % ont fait un certain effort pour administrer adéquatement ses documents et 32 % n'ont pas montré de souci pour développer une culture archivistique au sein de leurs institutions. Parmi ce dernier groupe, classé dans la catégorie C, on peut trouver la Présidence de la République, le Ministère de la Présidence, le Ministère des Finances, le Ministère l'Environnement et de l'Énergie, le Ministère du Logement et des Etablissements Humains et le Ministère des Sciences et de la Technologie.

Toutefois, si on compare la quantité d'institutions publiques qu'avaient des archives organisées pendant l'année 1990 et celles en 2004, on peut constater qu'il y a eu un progrès dans le processus de consolidation du Système National d'Archives, bien que très lent.

Institutions publiques	1990 <sup>31</sup>	2004 <sup>32</sup>
Pouvoir législatif	2	3
Pouvoir judiciaire	3	3
Pouvoir exécutif	3	12
Administration décentralisée	15	27
Municipalités	3	35

Les Archives Nationales ont émis quelques directives en rapport avec la gestion de documents, à savoir :

- Utilisation du papier et de l'encre dans la production des documents ayant une valeur scientifique culturelle (2003)
- Loi N° 7202 du Système National d'Archives et du patrimoine documentaire (2005)
- Importance de l'accomplissement de la Loi N° 7202 du Système (2006)
- Règlements techniques sur l'administration des documents produits par moyens automatiques (2004 et 2008)

Malheureusement, pendant ces deux dernières années, les Archives Nationales ont montré un recul dans leur rôle de recteur du système : des places à pourvoir dans des postes de responsabilité clé pour la consolidation du Système National et pour la conservation du patrimoine documentaire ont été couverts par des historiens et non par des archivistes. De même, sa projection dans le système a diminué en limitant les conseils et les inspections seulement aux organismes ne disposant d'aucun type d'archive organisée.

<sup>30</sup> Bermúdez Muñoz, María Teresa (2008). Los archivos y la brecha digital. I Conferencia Internacional sobre Brecha Digital e Inclusión Social. PROSIC. San José.

<sup>31</sup> Chacón Arias, Virginia (2000). El ayer y el hoy del desarrollo archivístico. Dans : <http://www.archivonacional.go.cr/articulos.htm>

<sup>32</sup> Bermúdez Muñoz, MT. Op. Cit.

## ■ Des groupes organisés

Les archivistes se sont organisés depuis de nombreuses années. Dans la plupart des cas, ces organisations sont nées à la demande des autorités de la DGAN. Certains de ces groupes sont les suivants :

- Association d'Archivistes Costariciens (ACA). 1983.
- Commission d'Archives Universitaires (CAU). 2000.
- Commission Archivistique Interbancaire (CAI). 1980
- Commission Interinstitutionnelle de Chefs ou de Représentants d'Archives centrales du secteur public (CIAP). 1989.
- Commission d'Archives Municipales (CAM). 2001.

## ■ Conclusions

Bien que le Costa Rica montre encore des indices acceptables, selon l'Indice de Développement Humain et selon d'autres indices comme celui de la compétitivité et de l'innovation, ce qui est vrai c'est que les indices socio-économiques, d'éducation, santé et environnement se sont détériorés pendant les dernières années.

D'autre part, plusieurs initiatives ont été entamées, au niveau national, pour faire avancer le projet de gouvernement numérique afin de répondre aux engagements acquis avec la région latino-américaine, mais la plupart d'entre elles se sont avérées infructueuses par manque d'une politique nationale d'information et de développement des TICs, bien que des réalisations comme la numérisation du Ministère des Finances, le Système National de Paiements Électroniques et les démarches de demande de passeport et de permis de conduire, par le biais des agences de la Banque de Costa Rica, entre autres, peuvent être signalées. Concernant le thème de l'éducation, des progrès ont pu être faits grâce en particulier, à l'action de la Fondation Omar Dengo et pas à celle du Ministère de l'Éducation Publique.

En ce qui concerne les études d'Archivistique au Costa Rica, des efforts ont été faits afin de les intégrer dans le secteur des sciences de l'information, mais elles restent en retard par rapport au développement des bibliothèques, faute d'une plus large vision dans la coordination de la carrière, d'une part, et de l'autre, d'une résistance manifeste de quelques enseignants à se mettre à jour et modifier l'analyse traditionnelle de l'Archivistique par une approche plus intégrale.

Finalement, quant au Système National d'Archives et au rôle de la Direction Générale des Archives Nationales, il semblerait que le modèle avant-gardiste des années 90 se soit épuisé et il s'avère urgent de réformer la Loi 7202 pour que le Système soit consolidé, ainsi que de modifier le service à l'utilisateur, pour répondre aux nécessités des utilisateurs virtuels.

# **Séance 2**

## **La gestion des archives électroniques et leur accessibilité**

---

---

Sous la présidence de Claes GRÄNSTRÖM  
Ancien directeur adjoint des Archives nationales de Suède

# ***Session 2***

# ***Managing and***

# ***accessing***

# ***e-information***

---

---

*Chairing Claes GRÄNSTRÖM*  
*Former Deputy Director of the Swedish National Archives*

# SHOOTING OURSELVES IN THE FOOT: IS DIGITIZATION DECAYING THE ARCHIVAL AURA?

---

Adam B. Cowling

MAS/MLIS Candidate, University of British Columbia

*The real, physical spaces of records management and archival services are collapsing inward towards an informational singularity, the virtual Web. Here, digital representations displace real world tangibility and claims to knowledge are perpetuated in greater mass than ever before. Authority is threatened and so too is archival, epistemological, and ethical credibility. The proposed paper aims to give light to some of the issues associated with the incorporation of digital information onto new technological platforms and the widespread proliferation of digital information on the Web. It will discuss perceived oversights of European and international e-information initiatives which fail to take into account critical aspects of traditional archival theory. First, there is an ever-increasing consensus that digitization initiatives pander to the technological whims of a largely unknown user through online organisational schema which are understood to provide increased access through the availability of reproductions, but which only provide a potentially 'undemocratic' perception of access. Second, current initiatives continue to erode record context and disregard the concept of authenticity. The influence of increasingly multi-functional Web tools and the opportunity for public exposure has encouraged records repositories to rush to digitize their holdings without much consideration for what effect this has on the continued representation of contextual materials. In some cases, the integration of Web functionality has led to an entirely new reinterpretation, where foundational archival concepts like original order and provenance fall by the wayside; eliminated is context, in favour of the juxtaposition of selected materials in the interest of creating attractive, but abstracted narratives. Virtual or "invented" records repositories, like The Parallel History Project on Cooperative Security (PHP)<sup>1</sup> pick and choose archival items from any number of contexts and combine them to serve such purposes. The paper will also address a concern over misuse of e-initiative resources, which are often directed towards short-lived, description starved websites. If records managers and archivists are to resign themselves to the fact that most future sources of funding will be directed towards e-information, then they must not be satisfied with the requests of sponsoring agencies for quantity over quality, for access over contextualization and authenticity. Resource sustainability is identified as a critical issue in regard to maintaining access to and preserving e-information. Managers must take into account the costs, of both time and money, and the upkeep of a website and digital repository. Initial funding must provide for maintenance of the Web presence over its expected duration, and must account for the possible time and money needed for system migration of data and the possible re-digitization of materials as improved means become available. Taking into account the current online misconstructions, the proposed paper will conclude by expanding upon the increasingly positive possibilities for management and accessibility to e-information for those who maintain the intrinsic archival elements necessary for users to understand and interpret digital (and digitized) information in the correct way.*

[See the presentation.](#)

---

<sup>1</sup> Swiss Federal Institute of Technology at Zurich, *The Parallel History Project on Cooperative Security*, (2007), <http://www.php.isn.ethz.ch/index.cfm> (12 November 2007).

## **Un cas d'auto-sabordage ? La numérisation porte-t-elle atteinte à l'aura des archivistes ?**

Actuellement, les espaces physiques réels dédiés au records management et aux services d'archives sont en train d'imploser et de converger vers une singularité informationnelle: le Web virtuel. Les représentations numériques y supplantent la matérialité du monde réel et le droit à la connaissance est maintenu avec des volumes de données plus importants que jamais. L'autorité s'en trouve menacée et, avec elle, la crédibilité archivistique, épistémologique et éthique. L'article proposé aura pour objectif d'éclaircir certains des problèmes liés à l'intégration des données numériques sur les nouvelles plates-formes technologiques et à la prolifération sous tous azimuts des données numériques sur Internet. Il traitera des erreurs que l'on a pu constater sur des initiatives de numérisation européennes et internationales qui n'ont pas su prendre en compte certains aspects critiques de la théorie archivistique traditionnelle. Tout d'abord, il existe un consensus toujours plus large selon lequel les initiatives de numérisation sont mises en œuvre pour satisfaire aux caprices technologiques d'utilisateurs, inconnus pour la plupart, à travers la mise en ligne de systèmes d'organisation censés fournir un droit d'accès plus large aux informations grâce aux reproductions disponibles, mais qui ne font que donner l'illusion d'un droit d'accès élargi potentiellement « non-démocratique ». Ensuite, les initiatives actuelles elles-mêmes continuent à saper le contexte documentaire en négligeant l'aspect d'authenticité des documents. L'influence des outils Internet dont les fonctionnalités sont de plus en plus nombreuses ainsi que la perspective de l'exposition ont poussé les dépôts d'archives à numériser leurs fonds à toute allure sans prêter une grande attention à l'effet produit sur la représentation permanente de documents contextuels. Dans certains cas, l'intégration des fonctionnalités Internet a conduit à une réinterprétation entièrement nouvelle où des concepts archivistiques fondateurs comme le respect de l'ordre primitif et de la provenance sont abandonnés en chemin ; le contexte est éliminé au profit de la juxtaposition de documents sélectionnés pour élaborer des récits attrayants, mais extraits de leur contexte. Les dépôts d'archives virtuels ou "inventés" comme le Projet d'Histoire parallèle sur la sécurité coopérative (PHP)<sup>2</sup> choisissent un peu arbitrairement des pièces d'archives issues de nombreux contextes qu'ils combinent pour servir ces objectifs. Cet article pointera aussi une certaine inquiétude liée au mauvais usage qui est fait des ressources des e-initiatives qui s'orientent souvent vers des sites éphémères, peu fournis en description. Si les records managers et les archivistes doivent accepter le fait que la plupart des sources de financement futur iront aux documents électroniques, ils ne doivent pas pour autant accepter les demandes des organismes clients qui font primer la quantité sur la qualité et l'accessibilité sur la contextualisation et l'authenticité. La viabilité des ressources est identifiée comme un point critique pour ce qui est du maintien de l'accès aux données électroniques et de leur conservation. Les responsables doivent tenir compte des coûts en temps et en argent d'un site Internet et d'un dépôt numérique ainsi que de leur entretien. Le financement initial doit pouvoir assurer l'entretien du site Internet sur sa durée de vie attendue et prendre en compte le temps et l'argent nécessaires à la migration des systèmes de données et à la renumérisation éventuelle de documents lorsque des moyens plus performants apparaissent. Tout en considérant les erreurs d'interprétation que l'on trouve sur Internet, l'article proposé se terminera par un développement sur les possibilités de gestion et d'accessibilité aux données électroniques de plus en plus prometteuses qui s'offrent à ceux qui maintiennent les exigences archivistiques intrinsèques nécessaires à une bonne compréhension et à une bonne interprétation par les utilisateurs des données numériques (et numérisées).

[Consulter la présentation.](#)

---

<sup>2</sup> Swiss Federal Institute of Technology, Zurich, *The Parallel History Project on Cooperative Security*, (2007) <http://www.php.isn.ethz.ch/index.cfm> (12 novembre 2007).

## ■ Introduction

The very definition of 'archives' is a subject of great contemporary debate. Much of the significance of the word's traditional meaning has been lost in its modern usage. The fundamental concepts of the real, physical space, and archives' meaning as derived from the interconnectivity of the whole *fonds*, are collapsing inward towards an informational singularity, the virtual Web. Here, it is possible for information to be accessed without a physical presence, and for electronic representations to take the place of tangible interactions. In such a place, record authority is threatened and so too is archival, epistemological, and ethical credibility. This paper aims to give light to some of the issues associated with the wide-spread proliferation of digitized information on the Web. It will explore the perceived threats of online digitization initiatives to the ever-relevant principles of archival theory, and to the future sustainability of the archival profession as a whole. This paper will focus on four identifiable threats. These are the ever-increasing consensus that digitization initiatives must pander to the technological whims of a largely unknown user; the continual erosion of context and disregard for authenticity; the misrepresentation of both archives and the archivist; and the misdirection of critical resources.

One must ask if the push for a reinterpretation of the archival theoretical structure, brought on by the challenge of digitization, is based on any proven and established fundamental change; or, have we just been taken in by the potential of the user's insatiable appetite for content, and afraid of being left behind? This paper aims to encourage archivists and records managers to take heed of the potentially jeopardizing effects of wide-spread digitization and accessibility via the Web.

## ■ Pandering to the Unknown 'User'

There are a number of reasons for which institutions choose to digitize primary source materials. These include, but are not limited to, pressure from users and the general public, pressure from sponsoring agencies or public authorities, a desire to participate in the idea that the Web is a universal deposit of all human knowledge, and the pressure or intention to remain up-to-date. These reasons all relate to the perception that archives should satisfy the needs of the user. But who is this user? One would think that the hysteria surrounding the development of digitization initiatives was based on some influx of clear quantitative information, but considering how important this information of user behaviour would seem to be, few studies have directly addressed the issue.<sup>3</sup> The seemingly unfounded assumption seems to be that users want easier, faster access to information; that they will be satisfied if they can view a wide assortment of digital reproductions online.

There seems to be a misconception about the access that is provided through digital reproduction and its capacity of opening up archives and digital repositories to a greater breadth of users. What is in fact provided through digitization is a mere perception of access. The satisfaction of users towards a reproduction in place of an original may be associated with what Walter Benjamin identified as that mode of human sense perception which shifts over time.<sup>4</sup> The increasing functionality of the Web has further shifted user perception away from what is real, towards a satisfaction with what is perceived as real because it is easily accessible.

There is a destructive element to the promise of liberal access which provides effortless exposure to vast amounts of digital information. This statement may seem in direct opposition to the popularized notion of the 'democratization' of knowledge which new technologies allow. However, there is something to be said for labour. There is something to be said for time. If all is immediately at our finger tips, the value of deep investigation will be lost. The serendipitous nature of knowledge, which has drawn generations of citizens to archives, will mean nothing if we lose that connection to the real thing. One might say that digitization maintains a connection to the real for researchers of today, but that is because those of our time still possess a remnant of the conception that these materials represent the real thing to be found in a physical space.

<sup>3</sup> Ian G. Anderson, "Are You Being Served? Historians and the Search for Primary Sources," *Archivaria* 58 (Fall 2004), 86.

<sup>4</sup> Walter Benjamin, "The Work of Art in the Age of Mechanical Reproduction," in *Illuminations: Essays and Reflections*, ed. Hannah Arendt, translated by Harry Zohn, (New York: Harcourt, Brace & World, 1968. Reprint, New York: Schocken Books, 1988. Originally published as *Illuminationen*. Frankfurt a.M., Germany: Suhrkamp Verlag, 1955), 222.



However, as time passes, it would seem likely that people will become accustomed to digitized materials so easily accessible, and will lose the connection to the tangibility of the unique item. There will be no challenge of investigation. A search will instantaneously get the result you are looking for, and you will expect it every time. There will be no more surprises, no more joy of coming across something that few people have seen before, because everyone will have already found it with the very same ease with which you now find it in front of you. When everything is so easy, there is no effort required, and with no effort, there is no reward, and with no reward, no joy, and with no joy, no point. There is something to be said for the labour and the time of it all.

## ■ Erosion of Context and Disregard of Authenticity

Digitization is a major task in selection. It is difficult, or even impossible to digitize everything. Thus, how are certain materials selected so as to represent the context of the whole? The influence of increasingly multi-functional Web tools and the opportunity for exposure has encouraged archivists to rush to digitize their holdings with little consideration for what effect this has on the continued representation of bonded information. In some cases, the integration of Web functionality has led to a whole new reinterpretation of the archives, where foundational archival concepts like original order and provenance fall by the wayside; eliminating archival context in favour of the juxtaposition of selected information in the interest of creating a new stage from which to present a 'historical' narrative. Removal of a record from the context of its interrelationships with other documents threatens its ability to serve as evidence. "Virtual archives" or "invented archives", which pick and choose items from any number of digital repositories and combine them in this way, represent projects of this nature. *The Parallel History Project on Cooperative Security (PHP)*<sup>5</sup>, for example, combines in a single narrative, materials from the Public Record Office (PRO) at Kew/London, the US Department of Defense at Washington, DC, and materials cited as simply a CIA Historical Declassification Release.<sup>6</sup> *Montreal, 500 Years of History in Archives*<sup>7</sup> reproduces single items selected from a group of twenty-three Montreal repositories and places them within the text of a historical narrative with little to no descriptive or contextual information. While the project may present an introductory experience for the casual user, the initiative's creators would have done greater justice to their holdings had they represented them using a "classic" archival approach. Being faithful to this approach involves the reproduction of archival sources in the same way researchers access them in a paper environment, avoiding any bias which might be introduced by a different arrangement. Archives best represent themselves and their holdings when *fonds* and series digitized are organized and presented in their original context, preserving their original structure and organization.

"Invented archives" are not archives, and it should be made absolutely clear that such projects do not represent an archival effort, but rather reflect institutionally generated exhibitionism. Even if these projects were to identify themselves as such, they still contribute to a misrepresentation of archival materials on the Web. They represent a blatant violation of the principle of archival impartiality, and pose a significant threat to the authenticity of records.<sup>8</sup> Referring to one of the largest image digitization projects in the world, that of Corbis Corporation, Geoffrey Batchem writes that its efforts have, "erased the aura of authenticity...and replaced it once and for all with the simulated glitter of reproducibility."<sup>9</sup> While rarely on such a scale, the combined effort of over-enthusiastic archivists to provide access to digitized reproductions is not unlike Corbis Corp. What we are experiencing is desensitization to the real. It is the transfer of value from the original to the perception of value of the reproduction. It is a loss to what is real, and a ignorant satisfaction of 'access' to what is not.

<sup>5</sup> Swiss Federal Institute of Technology at Zurich, *The Parallel History Project on Cooperative Security*, (2007) <http://www.php.isn.ethz.ch/index.cfm> (12 November 2007).

<sup>6</sup> William Burr and Robert A. Wampler, eds. "Lifting the Veil on Cosmic: Declassified US and British Documents on NATO Military Planning and Threat Assessments of the Warsaw Pact," *The Parallel History Project on Cooperative Security*, (21 May 2002) <http://www.php.isn.ethz.ch/collections/colltopic.cfm?lng=en&id=14968> (12 November 2007).

<sup>7</sup> City of Montreal and GARM, *Montreal, 500 Years of History in Archives*, [http://www2.ville.montreal.qc.ca/archives/500ans/portail\\_archives\\_en/accueil.html](http://www2.ville.montreal.qc.ca/archives/500ans/portail_archives_en/accueil.html) (12 November 2007).

<sup>8</sup> Authenticity, as we know, is the trustworthiness of a record as a record; the quality of a record that is what it purports to be, free from tampering or corruption. See InterPARES 2 Project, "Authenticity," InterPARES Glossary, [http://www.interpares.org/ip2/ip2\\_terminology\\_db.cfm](http://www.interpares.org/ip2/ip2_terminology_db.cfm) (12 November 2007).

<sup>9</sup> Geoffrey Batchem, "The Art of Archiving," in Ingrid Schaffner and Matthias Winzen, eds. *Deep Storage: Collecting, Storing, and Archiving in Art*, (Munich and New York: Prestel, 1998), 49.

## ■ Misconceptions about Archives and Archivists

The continued accommodation of user whims by digitization initiatives threatens the position of responsibility traditionally entrusted to archives and archivists. By creating the impression that digitization initiatives are another form of entertainment, an additional source of online distraction, archivists run the chance of damaging their professional profile in the eyes of sponsors and researchers. One is reminded of Georges Duhamel, who in expressing his clear detestation for film, lamented that it represented nothing but a distraction, whereas art, demanded a viewer's concentration.<sup>10</sup> The same sentiment can apply to the changing perceptions of archives. The shift from the concentration on robust archival quality to the consumption of shallow quantity should be identified as a destructive force proliferated by blind enthusiasm for digitization. Archives demand more from their user. They can only be fully understood by use and by perception. A tactile appreciation of archives cannot be accomplished solely through visual means, but must rather be accompanied by the enablement of habit. Digitization hardly affords the required means to comprehensively understand and experience archival information. While the multi-faceted functionality of digitization initiatives can be seen to provide a new informational experience, such functionality destroys what has traditionally been a thought and use driven process of knowledge building. Archives, through digitization initiatives, become just another form of mass media, like any other; and the archivist is relegated to the role of director, forced to concede to popular appetites.

## ■ Misappropriation of Resources

Digitization initiatives, many of which fail due to a lack in strategic planning, divert funding and resources away from archival processing and even endanger the capacity for institutions to create and maintain online finding aids that more properly address the functions and responsibilities of archives. Furthermore, if institutions become accustomed to the funding of digitization initiatives, which only provide for shallow description, it is likely that future funding will treat limited description as the norm, eliminating the need for rich description and further diluting the contextual essence of archives. Archivists must not be satisfied with the requests of sponsoring agencies that only fund digitization initiatives. The importance of maintaining the physical repository must be communicated to sponsors, so that funding is not entirely siphoned away into the virtualization of information.

Sustainability is a critical issue in regard to the management of resources and the funding of digitization initiatives. Archivists must take into account the costs, of both time and money, for the upkeep of digitization initiatives. Initial funding must provide for maintenance of the Web presence over its expected duration, and must account for the possible time and money needed for system migration of data and the possible re-digitization of materials as better tools become available. Further costs will always be incurred, even if no changes are made to content, due to servicing of servers, backup creation and logging, user requests, error correction, and the maintenance of digital objects. These financial requirements can be a continuing burden for institutions with irregular or unstable funding sources. While it may not be possible for those engaged in the management of archival materials to ignore the Web, they must approach it with caution and with a view to long-term sustainability, so as to avoid wasting more money on mediocre efforts.

As more information is digitized, a preference to provide access solely to digital reproductions will become more prevalent. The capacity of archival institutions to maintain continued access to original materials is thus threatened by the ease by which access to electronic information can be refused. While archives as a place can be physically appropriated in order to gain access when citizens are denied, as was done during the French Revolution<sup>11</sup>, wouldn't the new 'democratized' system simply require the flick of a switch to remove access? Web publication of digitized information also provides sponsoring agencies the opportunity to justify cost cutting measures such as mergers and reduction of reading room services. It opens the way for sponsors to make claims that, since digital copies are available online, there is little reason to maintain current hours of

---

<sup>10</sup> Benjamin, "The Work of Art in the Age of Mechanical Reproduction," 239.

<sup>11</sup> After French revolutionaries took control of the French archives, Article 37 of the Messidor decree proclaimed that "Every citizen is entitled to ask in every depository...for the production of the documents it contains." See Ernst Posner, "Some Aspects of Archival Development Since the French Revolution," *A Modern Archives Reader: Basic Readings on Archival Theory and Practice*, edited by Maygene F. Daniels and Timothy Walch, (Washington D.C: National Archives and Records Service, 1984), 5-6.

operation for access to originals. According to *The Vancouver Sun*, the merger of the Provincial Archives of British Columbia with the Royal B.C. Museum resulted in a “new ideological mantra” of cost recovery. New reading room access hours were implemented as a result of the measures. Full service hours were made available only when the average working person would be unable to visit, resulting in the archives being inaccessible 87.7 per cent of available time each week.<sup>12</sup> The article’s author asked how this could satisfy the province’s statutory mandate to ‘provide access’ to its citizens. A response from the province to such a question would likely include a URL to a digital repository. In this case, tax dollars have funded a merger and a digitization initiative which are now endangering citizens’ rights to original materials. Here, we see not democratic access provided to the citizen, but a mere eluding gesture providing the perception of access.

## ■ Past the Point of No Return? Suggestions for the ‘Future’ of Digitization

Here, we will not rescind the arguments discussed above, but will nevertheless give recognition to some potential benefits of archival digitization. These include the provision of access to those who might not normally be capable of accessing archival holdings, due to geographic isolation, physical disability, etc; the potential for small lesser funded archives to build awareness of their holdings; the potential use of online initiatives as educational aids; and digitization’s contribution to preservation by reducing the physical handling of original materials. It may seem naïve to believe that institutions will abandon digitization initiatives for the few reasons elucidated in the paragraphs above, therefore, the following suggestions will be proposed so as to contribute to a sort of re-conceptualization of the digitization process.

Institutions should ensure users are provided with a means to acquaint themselves with archival terms and arrangement, through the availability of tutorials, glossaries, etc. If digitization initiatives are to be used for research or as educational resources, the first lesson should be on how to use archives. An institution’s digitized Web presence should at all times maintain archival arrangement, even when full text searching of the online finding aids is enabled. Institutions should ensure that users are able to browse hierarchal levels if they so choose. If institutions are fearful that the common user will be turned away by large amounts of contextual information, the Web presence should provide a means by which deep, robust contextual description can still be viewed. This could take the form of an abstract view of a contextual summary and a full text view of the entire description.

The purpose and target audience of online digitization initiatives should be made evident to users, as should the institution’s selection and presentation methodologies. Archival networks and archivist associations should consider a grading or ranking system of some kind to rank the comprehensiveness of digitization initiatives. In this way, users can determine the likely usefulness of the site depending on their particular needs.

Archivists should set as a priority the need to make sponsoring agencies aware of the fact that funding can be better directed and utilized towards more worthwhile projects and preservation efforts. Archives should consider, and weigh carefully, the pros and cons of putting only finding aids online versus both finding aids and digitized materials. If archives find it necessary to place digitized materials online, they should take steps to ensure confidence in the digital reproduction. Archival networks and archivist associations should emphasize, among their members, the importance of utilizing agreed upon standards for all aspects of an institution’s digitization initiative. For aspects of the digital online representation that do not fall under an existing standard (e.g., new web development or finding aid encoding tools), archives should be cautious of their implementation, and should become involved with other archival bodies to research these areas, and look towards the possibility of developing standards for their use. Archival networks and archivist associations should promote knowledge sharing among the information professions, so as to develop the most useful and effective means by which to establish a healthy and cooperative Web presence. Competitive behaviour by renegade information organizations (e.g., to implement the newest technological gizmo before anyone else), should be highly discouraged within these knowledge sharing networks.

Archives intending to develop an archival digitization presence would likely do well to consider the theoretical concerns over what an archive is and what in the future it may be as a digital resource, while

---

<sup>12</sup> Stephen Hume, “Cost recovery, limited access subvert the archive’s role,” In *The Vancouver Sun*, (31 October 2007), 1. <http://www.canada.com/vancouver/columnists/story.html?id=20eed08f-a368-4cb7-8e0b-8095de588787&p=1> (10 November 2007).

additionally taking into account the more practical questions of archival management. Those capable of doing this work must be capable of thinking critically about the long term effects and sustainability of electronic archival information. Digitization initiatives may have some positive attributes; however, these can only remain positive so long as the digitization initiative remains supplemental to the physical archives.

Information will inevitably be sourced by users from the myriad options that the Web will continue to provide, and it does make sense that archives have representation on this medium; but, we must not get ourselves caught up in the transitory technological zeitgeists of the day. We should not let the current popularity of a technology in its early stages of development dictate and decay our present and future work. In these times, we must think critically about what digitization means to long term sustainability, and should identify those threats to the valuable physicality of the archival space. This paper has attempted to bring a few of these threats to light so that they can be further discussed; including the questionable accommodation of general user behaviour; the weakening of context and neglect of authenticity; the misrepresentation of archives and the archival profession; and the misallocation of resources. Archival knowledge has stood the test of time, and it is something archivists should have the faith to stand confidently by through changing times and societal shifts.

## ■ BIBLIOGRAPHY

Anderson, Ian G. (Fall 2004). Are You Being Served? Historians and the Search for Primary Sources. *Archivaria* 58, 81-129.

Benjamin, Walter. 1988. The Work of Art in the Age of Mechanical Reproduction. In *Illuminations: Essays and Reflection*. Edited by Hannah Arendt. Translated by Harry Zohn. New York: Harcourt, Brace & World, 1968. Reprint, New York: Schocken Books, 1988. Originally published as *Illuminationen*. Frankfurt a.M., Germany: Suhrkamp Verlag, 1955. 217-251.

Burr, William, and Robert A. Wampler, eds. 21 May 2002. Lifting the Veil on Cosmic: Declassified US and British Documents on NATO Military Planning and Threat Assessments of the Warsaw Pact. *The Parallel History Project on Cooperative Security*. <http://www.php.isn.ethz.ch/collections/colltopic.cfm?lng=en&id=14968> (12 November 2007).

City of Montreal and GARM. *Montreal, 500 Years of History in Archives*, [http://www2.ville.montreal.qc.ca/archives/500ans/portail\\_archives\\_en/accueil.html](http://www2.ville.montreal.qc.ca/archives/500ans/portail_archives_en/accueil.html) (12 November 2007).

Hume, Stephen. 31 October 2007. Cost recovery, limited access subvert the archive's role. In *The Vancouver Sun*, 1-2. <http://www.canada.com/vancouver/columnists/story.html?id=20eed08f-a368-4cb7-8e0b-8095de588787&p=1> (10 November 2007).

InterPARES 2 Project. 2004. Authenticity. InterPARES Glossary. [http://www.interpares.org/ip2/ip2\\_terminology\\_db.cfm](http://www.interpares.org/ip2/ip2_terminology_db.cfm) (12 November 2007).

Posner, Ernst. 1984. Some Aspects of Archival Development Since the French Revolution. *A Modern Archives Reader: Basic Readings on Archival Theory and Practice*. Edited by Maygene F. Daniels and Timothy Walch. Washington D.C: National Archives and Records Service. 3-14.

Schaffner, Ingrid, and Matthias Winzen, eds. 1998. *Deep Storage: Collecting, Storing, and Archiving in Art*. Munich and New York: Prestel.

Swiss Federal Institute of Technology at Zurich. 2007. *The Parallel History Project on Cooperative Security*. <http://www.php.isn.ethz.ch/index.cfm> (12 November)

# THE IMPACT OF THE UK FREEDOM OF INFORMATION ACT ON DIGITAL RECORDS MANAGEMENT IN THE PUBLIC SECTOR

---

Dr Elizabeth Shepherd

Reader in Archives and Records Management

Dr Alice Stevenson

Research Associate, School of Library, Archive and Information Studies, UCL

*The UK Freedom of Information (FOI) Act 2000 came into force in January 2005. Much of the published research into the impact of this legislation in the UK has focused on compliance, rather than the management of and access to e-information which underpins compliance. This oversight is surprising as the Lord Chancellor's Code of Practice on the management of records, published in compliance with FOI, emphasized how effective records management enables authorities to meet the obligations of the FOI. University College London's School of Library, Archive and Information Studies (UCL, SLAIS) recently received Arts and Humanities Research Council (AHRC) funding to address this gap in research. A project analysing the impact of the FOI Act 2000 on digital records management in the public sector is therefore now underway. In this paper we shall present the preliminary key results of our investigation. Although the research focuses on the UK, the issues identified will have resonances in other European nation states facing questions about access to public sector records in an era of e-government and transparency.*

*Given that the FOI Act applies to over 115,000 public authorities, the current project has chosen to undertake a detailed study of local authority councils in the South East of England, including London. Whilst this will not provide generalization across the whole public sector, this research has produced rich data that can be compared with other sectoral studies. The local authority sector is of particular interest as weaknesses in the management of records by local authorities have been noted in high profile reports and following recent tribunals, Practice Recommendations from the Information Commissioner's Office against two city councils have exposed the risk of not meeting sectoral compliance advice. Moreover, despite the high volume of FOI requests on a wide range of topics to local authorities, evidence to the Constitutional Affairs Committee reported that local authorities have less assistance than other sectors in complying with FOI and coping with its demands. By concentrating, therefore, on the study of the FOI experience in local authorities we will be able to offer an opportunity for significant improvement in this area.*

*Our investigation has been conducted with four specific research objectives in mind:*

- 1. how well digital records management services prepared for and coped with the first three years of FOI implementation.*
- 2. what contribution digital records management services make to the ability of public authorities to comply with the FOI Act.*
- 3. how the user experience of FOI is affected by the management of records*
- 4. what the implications are of FOI so far for good practice in records management.*

*In presenting these results we shall focus on the perspectives of three different groups:*

- 1. records managers*
- 2. institutional FOI policy managers*
- 3. FOI requesters and user communities.*

*From these perspectives we shall demonstrate how each group contributes to and benefits from the inter-relationship between digital records management and FOI. We shall also identify examples of good practice that can be shared within the records management and public sector communities both in the UK and the EU nation states.*

[See the presentation.](#)

## **Les conséquences sur la gestion des documents électroniques dans le secteur public de la loi britannique sur la liberté d'information (UK Freedom of Information Act)**

*La loi sur la liberté d'information (FOI - UK Freedom of Information Act 2000) est entrée en vigueur au Royaume-Uni en janvier 2005. La plupart des recherches publiées sur les conséquences de cette loi au Royaume-Uni se sont intéressées à la conformité et non à la gestion et à l'accessibilité des informations électroniques qui sont associées à la conformité. Cette omission est surprenante puisque le code des pratiques du Lord Chancelier sur le records management, publié conformément à la FOI, insiste sur le fait qu'un records management efficace permet aux autorités de satisfaire aux obligations de la loi sur la liberté d'information. L'école des bibliothèques, archives et sciences de l'information de la University College (University College London's School of Library, Archive and Information Studies - UCL, SLAIS) a reçu récemment un financement de la part du conseil de recherche pour les arts et les sciences humaines (Arts and Humanities Research Council - AHRC) pour combler cette lacune dans la recherche. Un projet visant à analyser les conséquences de la loi sur la liberté d'information sur la gestion des documents électroniques dans le secteur public est donc en cours. Nous présenterons dans cet article les résultats clés préliminaires de nos recherches. Même si la recherche porte sur le Royaume-Uni, les problèmes identifiés auront des répercussions sur d'autres pays européens confrontés au problème de l'accessibilité des archives du secteur public à l'ère des services publics en ligne et de la transparence.*

*La loi sur la liberté d'information concernant plus de 115 000 autorités publiques, le projet actuel a choisi de mener une étude détaillée des conseils municipaux dans le sud-est de l'Angleterre, y compris à Londres. Même si les résultats ne se généralisent pas à l'ensemble du secteur public, cette recherche a révélé des données substantielles pouvant être comparées à d'autres études sectorielles. Le secteur des autorités locales est particulièrement intéressant car des faiblesses au niveau du records management mené par les autorités locales ont été révélées par des rapports largement diffusés. Suite à des décisions de justice récentes, des recommandations pratiques du Bureau du commissaire à l'information ont été émises à l'encontre de deux conseils municipaux pour leur exposer le risque de ne pas répondre aux avis de conformité du secteur. Par ailleurs, malgré la multitude de demandes relatives à la liberté d'information adressées aux autorités locales sur une grande variété de domaines, il a été prouvé au Comité des affaires constitutionnelles que les autorités locales bénéficient de moins d'assistance que d'autres secteurs leur permettant d'être conformes à la loi sur la liberté d'information et de répondre à ses exigences. De ce fait, en nous concentrant sur l'étude de l'expérience de la loi sur la liberté d'information dans les autorités locales, nous serons en mesure de proposer des solutions pour améliorer ce secteur de façon significative.*

*Notre étude a été réalisée en gardant à l'esprit quatre objectifs de recherche spécifiques :*

- 1. dans quelle mesure les services de records management se sont-ils bien préparés à la loi sur la liberté d'information et de quelle manière ont-ils fait face à la mise en œuvre de la loi durant les trois premières années ?*
- 2. dans quelle mesure les services de records management ont-ils permis aux autorités publiques d'être conformes à la loi sur la liberté d'information ?*
- 3. dans quelle mesure l'expérience de l'usager de la loi sur la liberté d'information est-elle influencée par le records management ?*
- 4. quelles sont à ce jour les implications de la loi sur la liberté d'information sur les bonnes pratiques en matière de records management ?*

*Nos résultats concerneront en particulier les perspectives de trois groupes différents :*

- 1. les responsables d'archives*
- 2. les responsables institutionnels de la loi sur la liberté d'information*
- 3. les demandeurs d'informations relatives à la loi sur la liberté d'information, et les communautés d'utilisateurs.*

*A partir de ces perspectives, nous démontrerons dans quelle mesure chaque groupe participe et bénéficie des interactions entre le records management et la loi sur la liberté d'information. Nous identifierons également des exemples de bonnes pratiques pouvant être partagées dans les domaines du records management et du secteur public, entre le Royaume-Uni et les pays membres de l'UE.*

[Consulter la présentation.](#)

## ■ Research context

Political rhetoric in the UK, as in many democratic states, makes much of the need to increase public participation in the political process; improving public understanding of, and re-establishing public trust and confidence in, government decision making; increasing transparency, openness and accountability of public authorities; and, ultimately, better government decision-making on behalf of citizens. Many national governments accept the concept of access to information or freedom of information, and have introduced legislation: the UK is no exception. The UK Freedom of Information (FOI) Act 2000 came fully into force in January 2005 and represents the most important UK legislation to date to include records management guidance for all public authorities. The Act imposes significant duties and responsibilities on public authorities to give access to information. To achieve this, authorities need to know what information they hold, manage and retrieve information effectively, deal expeditiously (within 20 days) with FOI requests, and disseminate information through a publication scheme. As such, records management practices are deemed to underpin a public authority's ability to comply with the Act. This relationship between FOI legislation and the practice of corporate records management was formally recognised by Parliament in the form of a Code of Practice on Records Management, which states that:

“Freedom of information legislation is only as good as the quality of the records and other information to which it provides access. Such rights are of little use if reliable records are not created in the first place, if they cannot be found when needed or if the arrangements for their eventual destruction or transfer to an archives service are inadequate.”

During the lead up to the enactment of the full provisions of the Act in 2005, the theme of good records management practice was advocated by the Information Commissioner, who has responsibility to promote the observance of the Code. The Code was accompanied by detailed guidance from The National Archives, and from some professional associations. The Code itself is not compulsory, but it is held to be best practice. Nevertheless, the Information Commissioner can intervene in cases where public authorities are found to be failing to meet expected standards of good practice by providing advice, carrying out assessments, issuing Practice Recommendations and Enforcement Notices.

At UCL we are running a research project this year, funded by the Arts and Humanities Research Council, which is examining what the impact of the UK Freedom of Information Act has been on records management services in public authorities, especially local government. We are examining four specific research questions. First, how well records management services prepared for and coped with the first three years of FOI implementation. Secondly, what contribution records management services make to the ability of public authorities to comply with the Act. Thirdly, how the user experience of FOI is affected by the management of records and, fourthly, what the implications are of FOI for good practice in records management. We are looking at these questions from three perspectives:

- records managers
- FOI policy managers
- FOI requestors and user communities

We hope to demonstrate how each group contributes to and benefits from the inter-relationship between records management and FOI, and to examine the co-operation and partnerships needed in order to maximise the benefits of FOI. Our study cohort includes local government authorities in the South-East of England and London, at county, unitary, city and London borough levels. We have approached 50 authorities seeking interviews with those responsible for records management and for freedom of information. Later in the year we will hold two focus groups with requestors to elicit their views. The team of researchers on this project comprises academics Drs Elizabeth Shepherd and Andrew Flinn, and Research Associate, Dr Alice Stevenson, who is carrying out the research full-time until March 2009.

## ■ Literature review

Although records management as a distinct activity began in local government by the 1960s, there was little perception that the records manager should have a strategic role within the organisation until well into the 1990s. Records management has developed more quickly in the past decade, perhaps as a result of the changing organisational structure of authorities, the improvement in records management training on university courses, the development of performance indicators and quality initiatives, and the evolution of information technology systems. The Labour administration's Modernising Government agenda and e-government target of making all government services available electronically by 2005, led to modernisation of public services. The Local Government Act 1999, for example, committed local authorities to develop long-term forward looking policies; to consult the communities they serve in order to deliver services; to deliver efficient services to high standards; and to make full use of information technology in delivering those services. This had clear implications for electronic document and records management systems, and records management was thought by some to be significant in effecting such changes. Even so, in 2002 The National Archives (TNA, 2004) noted "concern that at present there was little or no coherent records management provision within many authorities" and the Audit Commission had "evidence from 25 audits that records management is widely undervalued in a range of local authorities". Such inadequacies were especially noticeable in the area of electronic records management, in which local authorities "had yet to make any significant progress" (Barata 2002; Barata 2004).

The Audit Commission's review of e-government implementation in England found too much emphasis on the implementation of front-end systems and not enough emphasis on re-engineering the entire range of processes and systems in local government. Key questions were also raised over the sustainability of the electronic government agenda in terms of both the physical (people, finance and technical infrastructure) and non-physical resources (skill, plans and strategies) available to authorities. Similarly, Benyon-Davies noted in Welsh local authorities the over-emphasis of e-government initiatives on front end systems, such as providing access to information via the website, rather than a deeper consideration of social dimensions, institutional context and organisational structures .

Compliance with freedom of information was linked closely to records management from the outset. Several challenges were envisaged. One potential problem was the contradiction between the remit of the Act which covers information, and the emphasis on records management in the guidance. Another was that a whole public authority has to be accountable, yet implementing an organisation-wide programme would be challenging. As Bailey (2005) noted:

"It is easy to create a records management policy. It is comparatively easy to agree a corporate file plan and retention schedule. It is far harder to convince staff that they should cease filing their information locally and managing it according to their own ways and should embrace new, shared ways of working"

Since the implementation of the FOI Act, the focus in local authorities has been on compliance, rather than the records management that underpins compliance. Nevertheless, surveys conducted by the Constitution Unit at UCL (Amos et al, 2006, 2007, 2008) have reported that improvements in records management were a positive effect of FOI compliance. A quarter of local authorities surveyed in 2007 felt that FOI had had a positive impact on records management, including better filing and naming of electronic records, streamlining of information procedures and improved records management systems. Yet, the same surveys report records management issues (such as bad record keeping, difficulties in locating data and determining whether it was held) as among the main barriers to compliance with the Act.

Research carried out in Scotland by Burt and Taylor reported that while FOI "is perceived to contribute to change, it is not a catalyst for change per se". They noted that FOI is seen as an administrative task, rather than "a mechanism through which to generate strategic organisational change". However, their research suggested that FOI was one factor among many which are generating "a growing awareness of the importance of information as a corporate resource" and it thereby lends "new importance to the development of formal records and information management policies, systems, processes and procedures within local government". The "FOIA has led to a surge in interest in the maintenance of electronic and paper records" and public authorities had shown "a marked increase in interest in records management as a general 'business issue". And yet, the Scottish research found little evidence of real strategic change in the way information was managed within authorities, instead observing a partial and patchy administrative shift and, perhaps, small changes in



the ways in which people recorded information. For example, it was noted “one thing people don’t do now is record personal opinion. This is probably good as it removes potential for prejudicial views”. It was against this context that we have been undertaking our research at UCL.

## ■ Research methodology

Given the complexity and diversity of the organisation of the local government sector in the UK we decided that it would not easily be represented by quantitative measures, rather we suspected that the different arrangements of FOI and records management activities in varying institutional contexts may have an impact on the relationship between the two. We adopted a qualitative approach in order to explore the issues from the perspective of the respondents based on their work context (Gorman and Clayton 1995). Semi structured interviewing was the main data collection method selected, which although time consuming has worked well in other similar research (Shepherd and Ennion 2007). Relevant topics were initially identified and the possible relationships between these topics and the issues became the basis for more specific questions, allowing both the interviewer and the person being interviewed the flexibility to probe for details or discuss issues (Seidman 1998). In designing the areas to be covered, existing toolkits were reviewed, including the Lord Chancellor’s Records Management Code, the Model Action Plan and The National Archive’s Evaluation Workbook and Methodology. Work by UCL’s Constitution Unit which addresses preparedness for FOI and by the Information Commissioner’s Office also informed the interview preparation.

Potential interview participants were identified with the assistance of The National Archives’s FOI contacts list, as well as through approaching councils directly via their generic FOI mailbox and by posting invitations to participate on the records management and FOI listservs co-ordinated by JISC, as well as Records Management Society local group email lists and the London Information Rights Forum. Initially, it was envisaged that ten interviews would be conducted with records managers and ten with FOI policy managers. In reality we found that in many local authorities individuals held dual responsibilities, whilst in others the roles of FOI and records management were seen in the broader context of information governance with individuals fulfilling a wide job specification. Thus such a neat division was not possible. A total of 22 interviews, typically of 40-60 minutes duration, were held and each was fully transcribed, with any identifying information redacted to preserve anonymity of the participants.

The rich data collected was analysed using the qualitative computer software NVIVO, version 7. NVIVO supports qualitative research in five main ways: managing data, managing ideas, querying data, graphically modelling concepts and reporting from the data. Whilst the programme does not represent a methodological advance, it does facilitate an increase in the complexity and detail possible in coding. This is the stage at which the research is currently, but already several issues are emerging.

## ■ Emerging Issues

One of the assumptions made prior to undertaking the interviews for this project was that FOI has been a major driver for improvements in records management, a not unreasonable assumption based upon the literature review. However, whilst most interviewees agreed that FOI had given records management a higher profile and increased awareness of it as a corporate function in a local council setting, there were few tangible examples of the way in which records management had been directly transformed as result of FOI:

“It’s more high profile, but the actual way we do things hasn’t changed”

*Records Manager, County Council*

The lack of any actual transformation in practices is particularly evident with regard to digital records and the comments recorded on issues such as control of email, shared drives and version control of documents echoed the conclusions of Barata concerning the lack of improvements to back office systems such as these. Whilst these digital areas were identified as problematic by every interviewee, few related the problems associated with managing these with their ability to comply with the FOIA. Rather, as long as some information

could be provided to the customer, regardless of whether it was all the information potentially available or the definitive final version of a document, that was deemed to be sufficient to comply with FOIA:

“We haven’t had, or at least to my knowledge we haven’t had something which is dependent upon version control. It only needs one or two people who understand the FOI game probably to cause havoc and mayhem. Luckily we haven’t actually encountered that. Clearly, you can see sometimes, particularly with journalists they clearly have got a standard template request letter but certainly with journalists they will take what they’re given because they’ve got a copy deadline to produce some article”.

*Archivist, County Council*

“At the moment we are relying a lot on trust and people say ‘oh I wrote this, it is the only version’, which we provide and send out, and we just don’t have the resources to call their bluff on it. But if there is proper version control and everything can be found then we are going to have some hard choices to make. I know the Information Commissioner has said the final draft is the final draft, it’s information not documents. You get an FOI request worded in the right way then you’ve got to consider all the drafts that you have.”

*Data Protection and FOI officer, London Borough*

“My guess is that the quality of what we are providing - we might answer a request - but the quality of what we are providing on the whole is not good because of RM. If we had better RM the quality would be better because you will be sure the documents you are providing will be the latest one.”

*Corporate Information Manger, London Borough*

This raises questions about the role of digital records management in being able not only to facilitate locating information, but also potentially permitting the quality of the information retrieved to be improved. It also suggests perhaps that references to records management improvements may be with paper records in mind rather than digital systems. So, for instance, whilst the introduction of retention and disposal schedules seems to have occurred in the last few years in many of our case studies, partly linked to FOI, this rarely extends to digital equivalents. Therefore, if a hard copy document is disposed of in line with council policy the digital equivalent is very often not attended to, but may exist in an unstructured space either on a shared drive or personal computer area, and remain disclosable if located or specifically asked for. As commented in the above cases, this has the potential to be a problem, especially as several interviewees noted that not only has the number of FOI requests increased in the last year, but also that they have increased in their complexity.

Another issue emerging is the question of to what extent records management supports FOI compliance. Records management systems do not appear to be the primary point of consultation in order to locate information. Rather, when a request is received in most of the institutions interviewed requests are sent out to specific individuals. In the quotation above from ‘Data Protection and FOI officer, London Borough’, for example, it is clear that central FOI officers are very much reliant on individuals supplying information and they themselves rarely have central access to check systems. This was echoed by several other individuals, for example:

“a lot is not stored in a records management document manager. It is still stored on servers that are unique to each directorate or unit or service so we have to chase humans to get access to the information.”

*Senior Information Officer, County Council*

“You still get things where things do run late and we have things that run late sometimes because of capacity because it is simply complicated, the person isn’t there or the person who was dealing with it left and there’s no a new person in post. It is always going to happen. But

just occasionally you have things that run late because somebody can't be bothered and that's just frustrating, but it is going to happen with anything in any authority because in any authority you have somebody who doesn't really think 'it's my job' or 'I haven't got the time' or 'I'll do that later', and it just doesn't get done."

*Information Management Officer, County Council*

Thus reports of delays in finding information within the Act's 20-day limit are partly a result of knowledge management and work prioritisation rather than simply a reflection of poor records management per se, as is perhaps inferred in some reports.

It also emerged that, as reported by Benyon-Davies (2003), FOI was only one of a series of factors driving records management, and interviewees often had difficulty pinpointing any one reason for developments in the area. These factors included the changing working structures in local authorities to more flexible working and hot-desking and the need to reduce the amount of (hard copy) material in storage. In particular the UK government's emphasis upon electronic government was seen as a particular impetus for the introduction of Electronic Document and Records Management systems (EDRMs), which was often driven from IT departments. Thus many organisations that had adopted records management and FOI situated these functions in IT to align with the technological advances being pushed by other initiatives and legislation. In some cases this led to tensions between an IT department responsible for EDRMs and central records management functions.

"I sort of got pushed aside because of the fact that we are going to get an EDRMs... the information governance team which is led by an IT person rather than a records manager and so there is a bit of tension about who is responsible for what and who has the final say for what."

*Records Manager, London Borough*

These are just a few of the issues that are emerging and it should be emphasized that this is very much a work in progress. Nevertheless, our research is beginning to suggest that in local government the relationship between records management and FOI is not as straightforward as implied by the Lord Chancellor's Code of Practice.

## ■ Future Research

We will continue to explore the interview data generated using NVIVO, but the project also aims to engage with the requestor experience of FOI and whether this is affected by records management. Therefore, two requestor focus groups are being assembled currently, most likely with participants drawn from the two of the primary users of FOI in the UK identified in the interviews: journalists and local campaign groups. Such groups often have extensive experience of making repeated requests and are thus the best source of information on current practices and issues. One particularly interesting forum and point of contact for such individuals and groups is the website *whatdotheyknow.com*, which allows people to make requests online and anybody can view the requests and the responses. In this way the full communication process is revealed and managed online.

We hope to have more complete results to report next year. Our work so far has thrown some light on the complex FOI-records management relationship and it is clear that although the questions might seem to be about compliance with legislation, social and organisational relationships are as likely to influence the outcome of an FOI request as formal systems and standards for compliance.

## ■ References

Amos, J., S. Holsen, et al. (2006). *Freedom of Information Act 2000: The first year. The experience of local authorities in England*. London, Constitution Unit, UCL.

Amos, J. and S. Holsen (2007). *FOIA 2000 and local government in 2006. The experience of authorities in England, Northern Ireland and Wales*. London, The Constitution Unit, UCL.

Amos, J, B. Worthy, et al (2008). *FOIA 2000 and local government in 2007. The experience of local authorities in England*. London, The Constitution Unit, UCL.

Bailey, S. (2005). "Assessing the Impact of the Freedom of Information Act on the FE and HE Sectors." *Ariadne* 42.

Barata, K. (2002). *Archives in the Digital Age. A study for Resource*. London, Resource.

Barata, K. (2004). "Archives in the digital age." *Journal of the Society of Archivists* 25(1): 63-70.

Bazeley, P. (2007). *Qualitative Data Analysis with NVivo*. London, Sage Publications.

Benyon-Davies, P. and M. D. Williams (2003). "Evaluating electronic local government in the UK." *Journal of Information Technology* 18(2): 137-149.

Burt, E. and J. Taylor (2007). *The Freedom of Information (Scotland) Act 2002. New Modes of Information Management in Scottish Public Bodies? Report to the Scottish Information Commissioner*. St Andrews, Scottish Information Commissioner.

Burt, E. and J. Taylor (2008). "Managing Freedom of Information in Scottish public bodies: administrative imperatives and political settings." *Open Government* 4(1).

Gorman, G. and P. Clayton (2005). *Qualitative Research for the Information Professional: A Practical Handbook*. London, Facet.

NAO (2002). *Better Public Services Through E-Government*. London, National Audit Office.

McLoughlin, I. and J. Cornford (2006). "Transformational change in the local state? Enacting e-government in English local authorities." *Journal of Management and Organization* 12: 195-208.

Seidman, I. (1998). *Interviewing in qualitative research: A guide for researchers in the education and the social sciences*. New York, Teachers College Press.

Shepherd, E. and E. Ennion (2007). "How has the implementation of the UK Freedom of Information Act 2000 affected archives and records management services?" *Records Management Journal* 17(1): 32-51.

TNA (2004). *Report on responses to the National Archives' consultation paper CP03/01: Proposed National Records and Archives Legislation. Proposals to change the current legislative provision for records management and archives*. London, The National Archives.

# CREATING ACCESS TO ELECTRONIC RECORDS: A TWO-LEVEL APPROACH

---

Yun Petersen Yoneyama

Special Advisor, Project manager for the Accessibility Project, Danish National Archives

*The Danish National Archives has collected electronic records from public administration, mostly from governmental entities since the early 70s. There are now more than 3000 extracts from systems in preservation. All records have structural information describing how to re-create a database with the same content.*

*The challenge for the Access Project was to create a system to access archival database records that is both easy to use for non-professionals and at the same time gives access to the full structure and the power of query facilities in database systems. This article describes how this challenge was met through application of a two-tiered design.*

*In 2007-2008 The Access Project created a modular archival access application for electronic records called SOFIA. Two of the modules each exemplify an approach to access:*

- *The Journalizing Systems Module uses a model-based approach. This module gives users access to a query interface with a high level of immediate usability for most users.*
- *The Register Module uses a tool approach. This module gives users access to the full database record including all structural information. With this module, users can create their own semantic ordering<sup>1</sup> of the record, and use this to generate queries according to interest.*

*It should be noted, that both modules are general programs, which can be used for any applicable record. Journalizing Systems records are all databases, so they can be used though the Register Module as well. The modules are therefore general in the sense, that there is no need to reprogram modules for specific records.*

*This dual approach rests on a theoretical assumption: To understand and retrieve information from databases is first of all is a question of analyzing the semantic order in the electronic record. This is typically done through studies of documentation and the database structure.*

*The goal for the Register Module is to create a tool, for the selection and aspectation<sup>2</sup> according to actual interests. With this module, users can create and execute relevant queries without knowledge of SQL. Some basic knowledge of database structure is however necessary here. The tool approach makes it feasible for the Archive to create access with minimal preparation work.*

*A second assumption is that for classes of IT-systems, there are recurring semantic orders. For the model-based approach, the Access Project developed a general model for journalizing systems containing the most common kinds of fields, and their structural relationships according to Danish administration practice.*

*The goal for the Journalizing Systems Module was to create an easy to use search system. Journalizing Systems can be searched using the same field names regardless of the nomenclature used in the original system. The same search system and user interface can therefore be used across the collection of journalizing systems records in the archive (more than 600.)*

*The search user interface is optimized for usability, so experience in searching one record may be re-used on another. Preparation for access is done by an archivist, and usually takes few days. The whole procedure including preparation, transfer of the record from the preservation system and quality control usually takes a week.*

*The experiences we have gained through internal test production of SOFIA shows that it is possible to create much use value for users with minimum spending of archival resources.*

[See the presentation.](#)

---

<sup>1</sup> Semantic ordering is here understood as the order relevant for a given interest. For a given approach, some part of the record are central and other irrelevant. For other approaches this may be the other way around. For databases, this means that some tables are central and queries should use certain relationships.

<sup>2</sup> "Aspectation" is used in its semantic meaning here: To attribute meaning or order on selected materials.

## La création d'applications pour accéder aux documents électroniques : deux approches

*Les Archives nationales danoises ont constitué, depuis le début des années 1970, des fichiers électroniques provenant des organismes publics, dépendant principalement du gouvernement. Il en existe à ce jour plus de 3 000, extraits des systèmes d'archives. Tous les fichiers contiennent des données structurelles décrivant la procédure pour reconstituer une base de données à partir du même contenu.*

*L'objectif du projet Access était de créer un système pour accéder aux bases de données des archives, qui soit à la fois facile à utiliser pour le profane et, en même temps, assez sophistiqué pour permettre une exploitation exhaustive de la structure avec toute la puissance des fonctions de requêtes habituelles des systèmes de base de données.*

*Cet article explique comment ce défi a été relevé par une conception à deux niveaux.*

*En 2007-2008, le projet Access a créé SOFIA, une application modulaire d'accès aux fichiers électroniques des archives. Chacun des modules de SOFIA représente un type d'accès particulier :*

- *Le module conçu pour les systèmes de journalisation utilise une approche basée sur un modèle. Ce module offre une interface d'interrogation ergonomique très facile à mettre en œuvre par la plupart des utilisateurs.*
- *Le module à registre est conçu comme un outil à la disposition des utilisateurs. Ce module permet aux utilisateurs un accès complet au fichier extrait de la base de données incluant toutes les informations structurelles, leur permettant de créer leur propre hiérarchie sémantique<sup>3</sup> pour le fichier et de l'utiliser pour générer des requêtes selon leurs besoins.*

*Il est à noter que les deux modules sont des programmes d'ordre général, aptes à être utilisés pour accéder à n'importe quel fichier. Les fichiers des systèmes de journalisation sont tous des bases de données et sont donc accessibles également à partir du module à registre. Les modules sont donc dits d'ordre général dans le sens où il n'est pas nécessaire de reprogrammer des modules pour des enregistrements spécifiques.*

*Cette double approche repose sur une hypothèse théorique : comprendre et extraire des données des bases de données nécessite d'abord une analyse de la hiérarchie sémantique du fichier électronique. Ceci est généralement effectué à travers une étude de la documentation et de la structure de la base de données.*

*L'objectif du module à registre est de créer un outil permettant la sélection et l'aspectation<sup>4</sup> selon les données recherchées. Ce module permet aux utilisateurs de créer et d'exécuter des requêtes pertinentes sans aucune notion préalable de SQL – bien qu'une connaissance des principes fondamentaux de la structure des bases de données soit quand même requise. L'approche utilitaire permet aux responsables des archives de créer des accès aux fichiers avec un minimum de travail préliminaire.*

*Une deuxième hypothèse est que certaines catégories de systèmes informatiques opèrent suivant des hiérarchies sémantiques récurrentes. En ce qui concerne l'approche basée sur un modèle, le projet Access développa un modèle général de système de journalisation comprenant les types de champ les plus courants, et leurs relations structurelles suivant la pratique dans l'administration danoise.*

*L'objectif ici était de créer un système de recherche facile à utiliser. Les systèmes de journalisation peuvent être interrogés à partir des mêmes noms de champ quelle que soit la nomenclature du système utilisée à l'origine. Le même système de recherche et interface utilisateur peut, de ce fait, être utilisé à travers toute la série des enregistrements des systèmes de journalisation des archives (plus de 600).*

*L'interface de recherche est optimisée pour faciliter l'utilisation, ce qui fait que l'expérience acquise lors d'une recherche d'un enregistrement peut être réutilisée pour un autre. Le travail préliminaire requis pour la création d'un accès est effectué par un archiviste et demande généralement quelques jours de préparation. La procédure complète, y compris la préparation, le transfert du fichier du système d'archivage et le contrôle de qualité prend généralement une semaine.*

*L'expérience acquise pendant les tests de validation internes de SOFIA a démontré qu'il est possible de créer beaucoup de valeur ajoutée pour les utilisateurs, avec un minimum d'impact sur les ressources des archives.*

[Consulter la présentation.](#)

---

<sup>3</sup> Hiérarchie sémantique est compris ici comme étant la hiérarchie pertinente pour un intérêt établi. Pour une approche donnée, quelques parties du fichier sont considérées comme étant centrales et d'autres non pertinentes. Pour d'autres approches, cet aspect de la recherche peut être inversé. Pour les bases de données, cela signifie que quelques tables sont centrales et les recherches doivent utiliser certaines relations.

<sup>4</sup> « Aspectation » est utilisé ici dans sa signification sémantique : attribuer un sens ou un ordre à des matériaux sélectionnés.

## ■ Introduction

The Danish National Archives have collected electronic records from public administration, mostly from governmental entities since the early 70'ies. There are now more than 3000 extracts from system in preservation in a form that facilitates the re-creation of their content as databases. This article describes how Danish National Archive has created an access system to the digital archival records. Two search modules were developed, with different but complementary approaches for meeting a useful balance between ease of use and requirement for technical knowledge among users.

## ■ The Access Project

The Access Project (in Danish shortened to TGP) started as a preliminary examination of the problem area in 2005, and was turned into a full project from the beginning of 2006. The purpose of the project was to investigate and create an access system for digitally created archival records.

In Danish legislation, the format of digital records is described in great detail through an executive order (the current version is the ministerial order 342 of March 11. 2004). The ministerial order standardizes how archival records should be delivered to public archives. As a standard, it specifies the structure, acceptable file formats, character set etc. for digital records. Data are delivered as flat files and metadata are described in a few auxiliary files. One metadata file contains structural information describing how to re-create a database with the same content.

The starting point for The Access Project was therefore in some ways quite clear. We should create a system that can re-create the databases, and make digital archives available to archive users and internal users.

The project was limited in two ways: The access should be made from our reading rooms (i.e. not through the Internet), and the system should only be developed for current the version of the standard. The last limitation was possible, as the Format and Structure Conversion project (FSK) was conducted in parallel with The Access Project. The FSK project reformat and convert older archival records into the current format, character sets, etc.

### The challenge

The ambition for the Access Project was to create a system to access archival database records, that is both easy to use for non-professionals and at the same time gives access to the full structure and the power of query facilities in database systems. But databases are by their nature not applicable for access by users especially without knowledge about the way data are structured in databases. Some form of IT-system is needed to access the data.

The Danish National archive uses a conversion strategy, so only data and structure are preserved - not the original system. With a conversion strategy for conservation, it is the practice to preserve the "content" of archival records i.e., the data and not the IT-systems the data were created by.

## ■ The Access System

The Access Project has created a general system for accessing digital archival records. The security part of the system is capable of differentiate between readily accessible records, as well as record that can only be accessed by archivists or through a dispensation. Dispensation may even differentiate between access to the whole record, partial access, access through predefined search facilities, or access to documentation only.

The user part of the system called Sofia is a modularized framework with currently 5 user accessible modules and another two for administrative purposes. Beside the two search modules, we developed modules to access the original documentation, to access metadata, and to access notes made during preparation of the archival record for access. Sofia is accessible from all internal PCs and through thin clients in most of our reading rooms. Beside Sofia the project has created supporting programs for user administration, database loading of archival records, and others.

All in all, the system can create access to any archival record that can be re-created as a database. That covers most of the more than 3.000 digital archival records currently collected by the Danish National Archives. The current exception is digitized records including audio and video records.

## ■ The Case Management Systems Module

The first search module we developed was the Case Management Systems Module. This module gives users access to a simple query interface which we designed to be easily usable for most users.

The Case Management Systems Module is designed to supports the known, current uses for a particular genre of systems. The data was authentic but the search facilities created was not intended in any way to emulate the original system. On the contrary, the system was designed as a general application for the whole genre of systems. Case management systems was chosen as genre, as the Danish National Archive has a large collection, and clear use cases for these systems. The primary use is for finding a particular case file where the case file ID is unknown. A secondary use is to support appraisal of paper documents registered in these systems, as these archival records will be appraised in these years.

Using this approach, it was possible to create a single program that can access any case management system record, and support the search for case file identifications and other well known search needs.

Case management systems were operational from the mid 1980ies and were developed as standard systems from a limited number of companies. There are relatively few (9 major) systems, that encompass most of the digital archival records from these systems. In case management systems, documents are registered and organized in case files. Early systems retain only registrations of documents, whereas newer systems can also be used as a repository for electronic documents stored or scanned into the system.

The Case Management Systems Module offers a standardized set of search fields, so all case management system records can be searched using the same field names regardless of the nomenclature used in the original system. In the same way result fields was normalized into standardized fields. The module can therefore be used across the collection of case management records in the archive (more than 600.) The standardization of fields create ease of use, and experience in searching one record may be re-used on others.

### Using the case management Systems Module

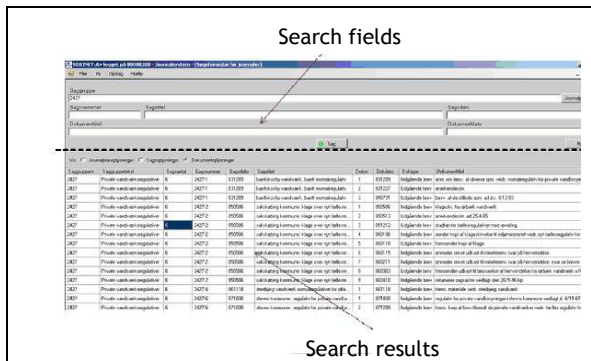


Fig. 1: Case Management Systems Module user interface

(Click on the picture to see it in full resolution)

The user interface of the Case Management Systems Module in Sofia is a simple search system (fig. 1). In order to be useful in Danish public administration, case management systems must provide the same basic



relationships between filing plan and case files, and between case files and documents. The standardization of fields is possible because of the ubiquity of these basic relationships. Technically, all fields are described in a XML file for each archival record. Each field points out a field in the model, and have a data type, that describes which search operation the field supports. E.g. text fields support wildcard searches and date fields support date range searches.

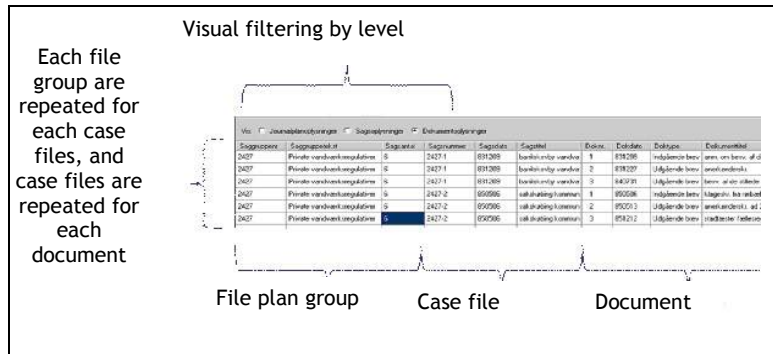


Fig. 2: Sofia result table

(Click on the picture to see it in full resolution)

The result table (fig. 2.) presents search results across the involved entities in the model. Therefore each document has a separate line, but a case file with several documents spans a set of lines. A filter can be applied, that reduces the result table visually, so only distinct case files are shown or even filing plan numbers entries.

Archivist has used this module since the first release of Sofia in March 2008. They found it to be useful in the activities it was designed for, e.g. finding specific case files and appraisal plan for paper documents. The simplicity of the user interface is an advantage here, as many users have reported that, Sofia is so easy to use, that they could retrieve useful information right from the start.

### Model-based access

To handle the preparation of the archival record, we have developed what could be called a model-based approach to creating access. As a top-level model, we developed a general archival model for Case Management Systems.

The model contains the most common kinds of fields, and their structural relationships according to Danish administration. In a sense, archival models reflect what archivist expect can be found in an archival record belonging to the genre regardless of the actual structuring and naming in the preserved digital archival records.

The general model was further refined into system specific models - one for each type of actual system used in creation of the archival records. The system specific model describes where the actual data usually can be found, if this particular system was known to be used.

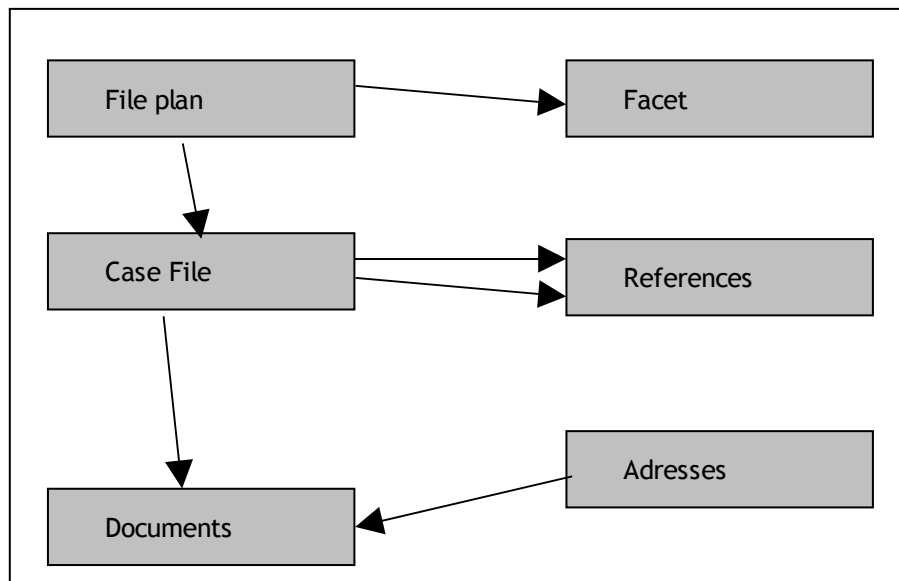


Fig. 3: General Case Management Systems Model

The model for case management systems is a very simple model with 3 main entities (filing plan, case file and documents) supported by three supplementary entities (See fig. 3). All data fields that are useful for retrieving case files and documents can be attributed to these six entities. Useful here means that these data fields are valuable as field for specifying the search or as part of the result<sup>5</sup>.

### **Preparation of case management system archival records for access**

Case management systems records cannot be used directly in this module without some preparation work. In an ideal world, system-specific models should be sufficient to connect the archival record and the access system. In practice, the archivist preparing the record, use a system specific model more as a template and not simply as an answer, to where data actually are.

Preparation work consists of connecting the fields in the model with fields in the actual case management system record. But the archivist preparing the record must also look into the actual data preserved. For example, useful data are sometimes distributed in several fields, or sometimes in a altogether different field. Textual descriptions of a case file are often contained in a "title" field, but sometimes much better data for search can be found in a "comment" field. Even with two agencies using the same version of a particular system, data can be found in different places, because of differences in actual use patterns in the agencies.

All in all models are not solutions but more like templates or suggestions for the archivist preparing the records.

The actual result of the preparation work is a set of database views, where each view describes where the actual data for an entity in the model can be found. This is formulated in SQL as individual searches, that the module knows how to combine into complex searches.

A database, where a system specific model is applicable, can usually be prepared in two-three days. The whole procedure including transfer of the record from the preservation system, preparation and quality control usually takes a week.

<sup>5</sup> The model shown here is the simple model used, when the system does not contain actual digitized documents. "Document" in the model above should be read as "document registration".

## Technical architecture

Technically, the connections between archival record and database are created with database views. The module is programmed to retrieve data through these views, and how it access data is described only here. (see fig. 3). In practice, this is the biggest gain on the model-based approach: There is no need to program specific adjustments for each archival record.

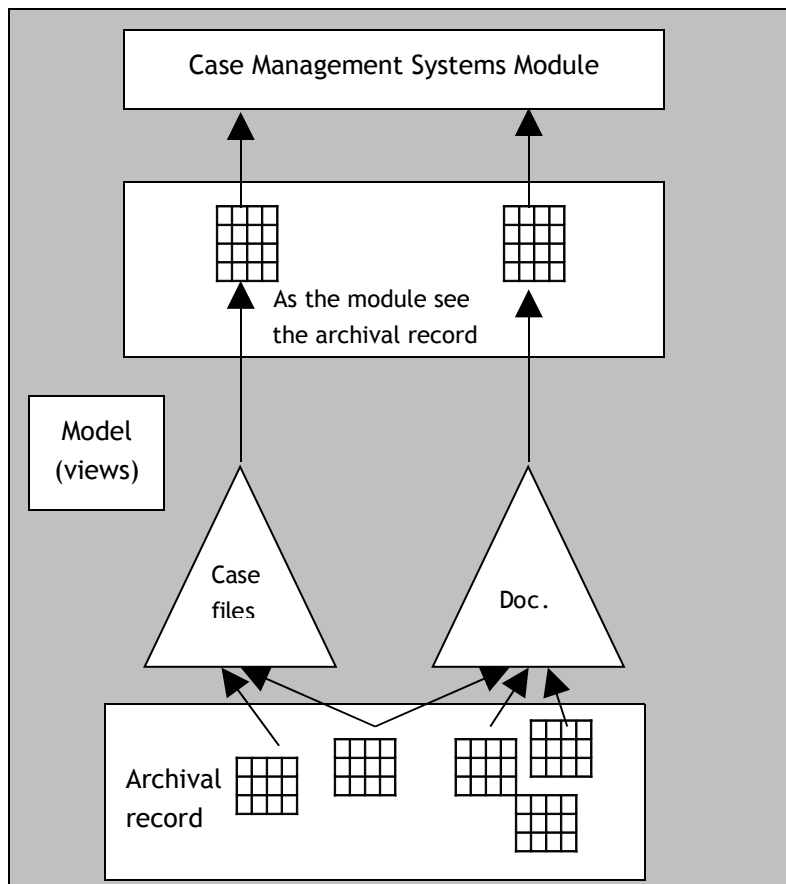


Fig. 4: Access with views

By using views, the program only need to know about the general model. Each view translates between the fields and entities in the model and the actual structure and fields in the re-created database.

### Advantages of the model-based approach.

- The user interface only gives access to the standardized fields of the model, and can therefore be kept simple, compared to the real structure in the re-created databases. Simplicity creates ease-of-use and standardization makes it much easier, to transfer experience in use may across the entire collection.
- In programming the module, only the general model is considered thereby simplifying the program. All actual details are handled outside the program through views or by forms. Therefore the module itself is a general module for the whole genre, and can be further developed and maintained as such.
- The model isolates system development from archival preparation work from each other. They can be carried on in parallel and be further developed independent of each other (unless the general model is changed.)

## **Disadvantages of a model-based approach**

- Models "filters" the archival records so only the fields described in the model are searchable. Some users may prefer conducting their own investigation into the archival record. Possibly using other kinds of queries than those supported by the model.
- Delivering digital archival records require preparation work, and therefore cost some working hours per record.

## **■ The Register Module**

To address some of the disadvantages of the model-based Case Management Systems Module, a new search module, the Register Module was created as part of Sofia version 1.1 (Nov/Dec 2008).

The ambition for the Register Module was to enable users to conduct their own research of an archival record including creating their own queries without requiring knowledge of SQL<sup>6</sup>. Even though ease of use still is important in the design, it is necessary to demand a basic knowledge of databases, and how information is structured in relational databases. The Register Module is more complex and for most people, it will not be as easy to use as the Case Management Systems Module.

### **Designing the Register Module**

To design the register module, we assembled an expert group among our colleagues with experience in analyzing digital archival records. Initially users seeks to understand the archival record, usually using documentation and structural metadata. Often this is supplemented by simple queries to see data as well as testing the users understanding. Later the user actively construct queries to retrieve relevant data. Often supplemented by visits back to documentation and metadata. Studying documentation and querying data are therefore used in both modes, but the emphasis changes from one to the other.

To support this the design should always give access to both documentation and queries without emphasizing one or the other. This makes it possible for users to control the emphasis, and to seamlessly change from one to the other.

### **Documentation and metadata**

In practical terms, we did not need to create access to the documentation, as the existing module was sufficient. But a better access to structural metadata was needed. The user needs easy access to an overview of the tables contained in the database, and select relevant tables to be included in queries. In Sofia V 1.0, metadata was accessible through a simple text viewer showing the XML structure description. In the Register Module this was changed into a more navigational tree-view containing a list of all tables in the record. Selecting a table, on pane shows the known metadata for the table, and a list of fields in the table (fig. 5). Tree-view is a convenient way of displaying structured information and should be usable even for an archival record containing 30-40 tables.

---

<sup>6</sup> SQL - Structured Query Language is the standard native query data manipulation language for relational databases. SQL has very powerful search facilities, but complex queries can be very challenging to write.

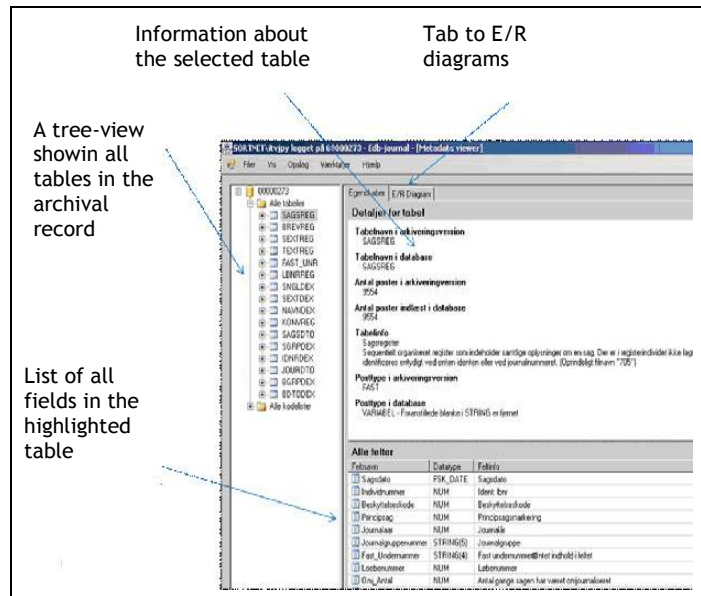


Fig. 5: The Register Module  
(Click on the picture to see it in full resolution)

### Selecting tables and creating queries

The user can select and drag tables into the pane E/R diagram pane. If there are relationships between the tables these are automatically drawn on the diagram. By selecting relevant tables, the user controls which tables and relationships the query uses.

In the lower left corner an area has been set off for search fields. These can be designated simply by dragging a field from a table into the area (Fig. 6).

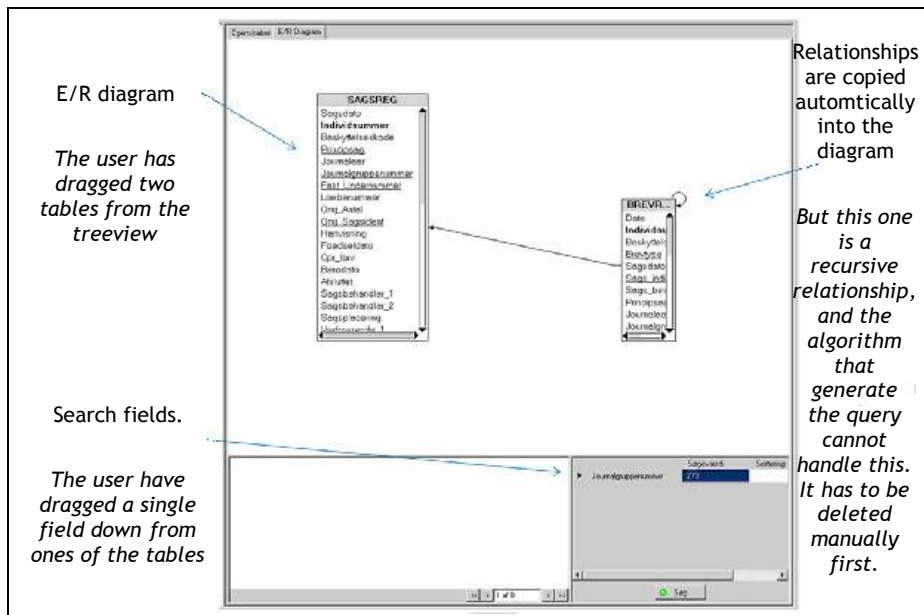


Fig. 6: Selecting tables and creating search fields  
(Click on the picture to see it in full resolution)

After search fields have been selected, the module can be used as any other search form. Filling in the search field and executing the query, will result in a dataset containing the search result (Fig. 7). This can be

manipulated in more or less the same way as a result set obtained through the Case Management System Module.

The screenshot displays a search interface with three main sections:

- E/R diagram:** Located at the top left, it shows a diagram of database relationships.
- Search results:** A table of results with columns: Sigsigle, Pri, Juv, Jour, Lin, Vidensmed\_1, Vidensmed\_2, Date, and Developp. The table contains 15 rows of data.
- Fields list:** A vertical list of fields on the right side, including: Indivoteurname, Beskyttelseskode, Brucodtag, Avsnr, Juvnoyordnummer, Fast\_Lindefnummer, Lantbeskrivning, Omr\_Artid, Omr\_Sigstakt, Havsning, Foodvaldsko, Qtr\_ker, Sveridato, Aftuter, SigtBehandlar\_1, SigtBehandlar\_2, Sigtglaesning, Vidensmed\_1, Vidensmed\_2, Vidensmed\_3, Ertmsord\_1, Ertmsord\_2, Beskyttelseskode, Sigsigle, Sigsigle\_indidat, Sigsigle\_beskyttelseskode, Farsning, Juvnoyord, Juvnoyordkode, Juvnoyordgruppe, Fast\_indivoteurname, Lovbeskrivning, Brevstat, Sveridato, Sveridatobeskrivning, News, Ntch\_2, Aftuter, Aftuter\_2, Adress, Adressbeskrivning, Land, and Farsning.

Annotations on the screenshot include:

- "E/R diagram" pointing to the top left section.
- "Search results" pointing to the table of results.
- "All fields from all tables are retrieved, but the user can hide irrelevant ones." pointing to the fields list.
- "Number of results found" pointing to the bottom of the results table.

Fig. 7: Search result

(Click on the picture to see it in full resolution)

## Tool approach to access

The screen-shots above are taken from our current version of the Register Module, which will be released in early December. It is still an advanced prototype, as some supplementary features need to be developed.

Compared to the Case Management Systems Module, the Register Module uses a quite different access approach. The module is designed as tool for the user, rather than a search form. The user should be able to go back and forth between seeking information in the documentation and test out queries. It should support the user's perspective and not an archival pre-conception of what might be useful, that governs the use. We think we have achieved to support this, even though it is a bit rudimentary in the current prototype.

When the user have access to the full database structure, the user can develop their own queries, and therefore need facilities for calling up information, test queries, etc.. Even though there still are some limits on the generated searches, it can generate queries without requiring the user to know how to formulate them in SQL. We hypothesize that it is much easier for a user to learn about basic relational database structures through E/R diagrams compared to learning SQL. The module can show the SQL generated, and a facility to edit the SQL directly will be designed into the module in the near future. Even though most users probably will ignore this, at least some of our archival colleagues can benefit from this.

## Advantages

- The user have full access to all available information about the archival record, the full structure, and thereby all data.
- Queries can be designed according to the users interest and needs
- Delivering digital archival records through the Register Module, do not require preparation work.

## Disadvantages

- The Register Module require some technical knowledge on databases
- The register module is more complex to use

## ■ A two-level access strategy

The two modules shown has each its strengths and weaknesses:

	Model based approach	Tool approach
Ease of use of the user interface	+ Can be kept very simple, easy to use and standardized across the collection	- Complex
Requirement of user knowledge	+ Only requires a basic understanding of Case Management	- Requires at least a basic understanding of database structures and database query. Knowledge of SQL is an advantage, but not necessary.
Preparation work	- Necessary, but can be made more efficient by creating system specific model	+ Not needed
Access to all data	- Not possible. Access depends on the fields chosen in the models	+ All data are accessible

In principle the ideal access system for databases is easy to use, do not require (much) special knowledge, and can query, sort, combine and filter all data available in any way the user may find worthwhile. We could not design a single access module that contains all these qualities. However, by designing two separate modules, we could distribute the advantages into two modules that complement each other.

The Case Management Systems Module can be generalized into a range of built for purpose access systems, whenever actual recurring usage needs emerges. The technology and method are both generalized, and we have experimented with more simple, singular forms based on single SQL queries as a supplement to the full model based method. Forms with single SQL queries are very easy to create using the same technological foundation as the Case Management Systems Module itself, and did not require modification of the programming. Simple search forms can be distributed at an even lower cost of preparation, though it will also typically involve more analysis before preparation of the record. The reason for this is, that there are no system specific models available for now outside the collection of case management systems..

Creating Access with the Register Module require no preparation work at all. And users can access data through the query generation in the module.. Every year new archival records will be available to the public, but none of these do have known recurring uses (yet). Delivering archival records with the Register Module is therefore the most economical way of deliver archival records available in The Access System.

## ■ Conclusion

The Access Project is reaching the end and will close as a development project by the end of 2008. After four years we can look back on many challenges and see the results. First of all we have created a general, user-friendly access system - Sofia - that can handle all digitally created records as long as the archival record is structured according to the Danish archival standard for digital records. Also, a simple and effective method for preparing digital archival records has been created, to support the continued work of creating access for specific needs. And last, a general query building system for archival records have been developed to support unknown needs for retrieving data.

# TRAITEMENT ET VALORISATION DES FLUX ÉLECTRONIQUES, UN EXEMPLE DE MÉMOIRE D'ENTREPRISE.

---

Marius Delail

Chef du service Gestion de l'Information et de la connaissance,  
Centre National d'Etudes Spatiales (CNES), Toulouse

*Le service Gestion de l'Information et de la connaissance du Centre Spatial de Toulouse a mis en place une chaîne de traitement des documents produits dans le cadre des activités du CNES. Ceci avec deux objectifs : d'une part, en accord avec les archives départementales, respecter la loi de 1979 relative aux archives publiques ce qui nous amène à verser périodiquement cette documentation aux archives, d'autre part, mutualiser et valoriser ce travail d'archivage à travers des outils de gestion des connaissances afin de permettre à tous les agents du CNES de profiter de ces informations.*

*En effet, le travail très important réalisé au niveau archivage : tri, numérisation, reconnaissance de caractères (papier), formatage, recherche de doublons, indexation, inclusion des métadonnées, contrôle qualité, versement, est valorisé en interne via la Mémoire d'Entreprise accessible en Intranet par les agents des quatre centres. La mutualisation du flux des documents issus des archives avec les documents, connaissances et outils issus des travaux de GC a permis la création du fonds documentaire de la Mémoire d'Entreprise (100 000 documents, 40 000 références d'articles, liens directs vers les spécialistes...).*

*Ce fonds est traité avec des outils de classement automatique et est accessible à travers un moteur de recherche sémantique. Ces derniers travaillent avec une Base de Connaissance qui regroupe les modélisations des métiers CNES.*

*Cette approche permet :*

- d'éviter de contextualiser les documents en amont ; i.e les auteurs de documents ne se préoccupent pas de métadonnées, ni de mots clés, résumé...*
- de standardiser, contrôler et gérer en configuration, la contextualisation des documents via la Base de Connaissance et les filtres de catégorisation utilisés pour le classement automatique*
- de réutiliser la Base de Connaissance dans des contextes différents : maintien du "sens" dans le temps, classement par métier de lots importants de documents (revues de projet, messagerie...).*

[Consulter la présentation.](#)



## **Processing and exploiting electronic flow: an example of Corporate Memory**

*The Information and Knowledge Management Department of the Toulouse Space Centre has set up a processing chain for documents produced as a part of CNES activities. It has two goals: firstly, to apply the French law of 1979 concerning the archiving of public documents by regularly sending such documents to the archives of the Haute-Garonne département, and secondly, making these archived documents available to all CNES collaborators who may need to access them, using knowledge management tools.*

*The considerable work involved in the archiving process – sorting, digitising, character recognition (for paper documents), formatting, seeking out duplicates, indexing, integrating metadata, quality control and delivery – is exploited internally via the Corporate Memory accessible over the Intranet by staff at all four CNES centres. By pooling the flow of archived documents with the documents, knowledge and other tools resulting from Knowledge Management, we have been able to set up the database of the Corporate Memory (100,000 documents, 40,000 references to articles, direct links to specialists etc).*

*The repository is processed with automatic sorting tools and can be accessed via a semantic search engine. The sorting tools work with a Knowledge Base containing models of all professions represented at CNES.*

*This approach:*

- *avoids the need to contextualise the documents upstream; i.e. the authors of the documents do not need to supply metadata, keywords or abstracts, etc;*
- *means that document contextualisation is standardised, monitored and configuration managed, via the Knowledge Base and the classification filters used for the automatic classification;*
- *allows the Knowledge Base to be used in many different contexts: ensuring that its significance remains clear despite the passing of time, classifying large batches of documents by professions (project reviews, electronic messaging and others) etc;*

*The proposed article could:*

- *explain how paper and electronic flows are processed: standardisation, types of processing, interface with the French National Archives, interface with the Information System, interface with the Corporate memory, feedback etc.*
- *explain the exploitation aspect: Corporate Memory interface, synergy with the Knowledge Base and the text-mining tools, methodology, resources deployed, results etc.*
- *or give an overview of each of these aspects.*

[See the presentation.](#)

## ■ Contexte

### Le CNES

Etablissement Public à Caractère Industriel et Commercial (EPIC), le Centre National d'Etudes Spatiales (CNES) a été créé en 1961. Son rôle est de proposer au gouvernement la politique spatiale de la France au sein de l'Europe et de la mettre en oeuvre.

Il rassemble 2500 agents répartis sur quatre sites, trois en métropole : Paris, Evry et Toulouse, et un en Guyane.

Au sein de la Direction du Système d'Information (DSI), dans la sous Direction Support et Développement (SD), le service Gestion de l'Information et de la connaissance (GI) est structuré en 6 pôles d'activité :

- Archivage des documents techniques, scientifiques et patrimoniaux
- Gestion des Connaissances (GC)
- Veille et Intelligence
- Documentation, bibliothèque (prêt d'ouvrages, revues...)
- Traductions, interprétariat et terminologie
- Mémoire d'Entreprise (ME)

### Les archives

Le CNES en tant qu'EPIC est tenu de verser aux Archives de France, selon la loi N°79-18 du 3 janvier 1979 (renforcée en 2008), un exemplaire des documents produits dans le cadre de ses activités.

Cette obligation de versement aux Archives de France nécessite d'établir une politique d'archivage avec des règles de gestion et de traitement des archives qui se concrétisent par une charte négociée avec les AdF via les Archives départementales de la Haute Garonne (AdHG).

La charte précise : les règles de tri applicables (Tableau de gestion), les supports (CD-ROM professionnels), le contenu des lots (fichier d'indexation, historique, acronymes...), le format (électronique, PDF), les règles de sécurité (conservation d'un double dans un coffre au CNES), le contrôle qualité...

Cette charte, élaborée dans le cadre de l'archivage de flux de documents "papier", doit être mise à jour afin de prendre en compte les flux électroniques.

Depuis 2002 la cellule archives du service GI est responsable des versements, plus de 60 000 documents ont été versés aux AdHG.

## ■ Traitement des archives

En plus du respect de la loi, les objectifs du service GI, pour ce qui concerne les archives sont :

1. réduire le stock de documents "anciens" sur support papier qui encombre les armoires et les caves du CNES (plus de 1km linéaire) et ainsi libérer de la place
2. mettre en place les traitements relatifs à l'archivage des documents qui arrivent directement sur supports électroniques ("flux électroniques")
3. valoriser (rentabiliser ?) le travail réalisé pour les archives

4. améliorer la collecte des documents à archiver
  - a) sensibilisation des intermédiaires : agents, secrétaires, hiérarchie
  - b) établir des interfaces avec les systèmes de Gestion Electronique des Documents (GED) qui existent au sein du CNES
  - c) optimiser la collecte au sein du Système d'Information (SI), en particulier en automatisant le transfert des documents référencés du SI vers les archives et la ME
5. participer à des communautés de pratiques dans le domaine de l'archivage et maintenir nos relations avec les AdHG

Depuis 2001, nous avons traité principalement des documents sur support papier ("flux papier") : tri, numérisation, reconnaissance de caractères, formatage, indexation, inclusion des métadonnées, finalisation (historique, acronymes, jaquette...), contrôle qualité. Puis versements : aux archives CNES, aux archives départementales, en Mémoire d'Entreprise (ME).

Depuis 2005, nous réalisons des études et nous développons des chaînes de traitement (collecte, tri, mise en forme, etc.) pour ce qui concerne les flux électroniques.

En 2006, nous avons mis en ligne, sur l'Intranet, la Mémoire d'Entreprise. Elle regroupe des connaissances explicites (formalisées) et tacites (non formalisées) capitalisées au sein de l'entreprise par les individus au cours de leurs expériences professionnelles. La capitalisation des connaissances permet de transformer les expériences passées en nouveaux moyens de production, de s'inspirer de l'existant pour les projets à venir, elle sert de terreau et est l'un des éléments source d'innovation. La ME est composée d'un fonds documentaire (plus de 100 000 documents), d'une Base de Connaissances (BC), et d'outils tels qu'un moteur de recherche sémantique, un éditeur graphique de réseaux sémantiques, un éditeur de concepts, un indexeur, un outil de classement automatique, un analyseur de requêtes, une interface Web... Elle nous permet de valoriser le travail réalisé pour les archives à travers la mise à disposition, de manière efficace, de l'ensemble des documents scientifiques, techniques et patrimoniaux archivés.

Nous allons détailler plus particulièrement les aspects traitement des flux électroniques et Mémoire d'Entreprise dans la suite du document.

## **Les flux électroniques**

Paradoxalement le traitement des flux électroniques s'avère beaucoup plus délicat que le traitement des flux papiers. Pour ces derniers, le choix est simple ils sont soit lisibles soit illisibles.

Pour les flux électroniques les principaux problèmes sont : le nombre d'interfaces et de supports (GED, serveur, CD-ROM, DVD, disquettes, bandes...), des formats très hétéroclites (Word, Texte, Latex, Powerpoint, Excel, Tiff, PDF image, PDF texte sur image...). Même le PDF de "base" pose problème car les agents récupèrent sur Internet des convertisseurs "exotiques" qui donnent parfois des résultats surprenants.

On doit aussi résoudre le problème de la diversité des droits et attributs appliqués aux documents numériques : document protégé par un mot de passe, document en lecture seule...

Puis apparaît la nécessité d'un identifiant unique et cohérent pour chaque document, alors que l'on peut trouver des documents de même nom dans des répertoires différents...

Mais le problème capital lorsque l'on traite des flux électroniques est celui des "doublons".

En effet, les sources des flux électroniques sont très nombreuses, les documents électroniques peuvent se trouver simultanément : sur le portail d'entreprise, sur une ou plusieurs GED, sur le disque dur d'un ou de plusieurs salariés, sur des clés USB, des CD-ROM, sous forme "papier" suite à une impression, sur un ou plusieurs serveurs de structure, sur un lot qui nous revient de l'extérieur via l'un de nos partenaires...

Compte tenu de la difficulté pour transmettre, dupliquer, sauvegarder, stocker un document électronique on retrouve le même document de très nombreuses fois dans les différents flux que l'on doit traiter, et ceci, même si en "amont" on a instrumenté l'unicité des documents, par exemple en ayant défini un support de stockage unique pour les documents de référence. De plus, du fait de la plus grande "transversalité" des organisations les flux nous arrivent de moins en moins "homogènes", c'est-à-dire non triés.

Les éléments présentés ci-dessus, l'augmentation du nombre de documents traités suite à la prise en compte des flux électroniques, les flux de "Tout venant", la taille du fonds, renforcent la nécessité de traiter les doublons.

Voici les étapes que nous avons mises en place pour résoudre ces problèmes :

1. Même lorsque cela est possible (serveur, clé USB...), il serait très dangereux de travailler directement sur le document original fourni, car une erreur malencontreuse pourrait le détruire ou le rendre inexploitable. La première étape consiste à copier intégralement les documents électroniques tels qu'ils sont, c'est-à-dire avec toute leur structure (dossiers et sous-dossiers compris) dans un espace de travail. Cet espace de travail est structuré en deux parties : une partie générique où l'on va conserver les éléments indispensables (paramètres) pour enchaîner les différents traitements et une partie "zone de travail" structurée en fonction du lot traité. La seule action spécifique en plus de la recopie est l'expansion automatique des fichiers ZIP que l'on est susceptible de trouver dans le flux d'entrée.
2. Il n'est pas possible de gérer tous les formats de fichiers de la même manière. Les fichiers contenus dans les différents dossiers et sous-dossiers sont triés et regroupés dans la structure de travail en fonction de leur type (.DOC, .PDF, .PPT, .PPTX, .MSG,...). Nous en profitons pour homogénéiser les attributs (lecture/écriture) et les accès. Certains formats (.DOC, .TXT, PDF Image...) sont convertis automatiquement en "PDF image sur texte". Pour les autres une action manuelle est nécessaire avant la conversion (mise en forme des fichiers Excel, formats exotiques qui demandent le passage à un format intermédiaire avant la conversion en PDF...). Nous réalisons aussi le dédoublement du flux entrant, les fichiers de même nom, issus de répertoires différents sont soumis au choix de l'opérateur afin de déterminer celui (ceux) que l'on conserve. A la fin de cette étape tous les fichiers du flux d'entrée sont au même format, dans la même zone de travail, avec les mêmes attributs et avec un nom différent.
3. La troisième étape est centrée sur la réalisation du (des) fichiers d'indexation. Ce fichier regroupe un certain nombre d'informations pour chaque document (Titre, Auteur(s), nom du fichier source, confidentialité...) et un lien hypertexte (générique) vers la zone de stockage finale du document. A ce niveau nous distinguons trois types de flux différents :
  1. le flux archives, décrit dans le chapitre 4, ce sont des CD-ROM structurés avec un fichier d'indexation de type archives,
  2. le flux GED, issu de l'une des GED du CNES, c'est en général un flux structuré avec une organisation spécifique. Suivant le cas il faut reprendre : le fichier d'indexation, la nomenclature des noms de fichier, la structure du flux (dossiers, sous-dossiers), le format des fichiers... Par rapport au flux précédent (N° 1) il faut aussi créer la structure de type Archives (conforme au CD-ROM qui sera fourni aux AdHG),
  3. le flux "tout venant", fourni par une structure, ou collecté au cours d'une action de Gestion des Connaissances (capitalisation des connaissances au moment d'un départ en retraite). Dans ce cas il faut aussi créer la structure de type Archives (conforme au CD-ROM qui sera fourni aux AdHG).

Les flux 1 et 2 sont traités automatiquement, le flux trois "tout venant", le plus complexe est traité de manière semi automatique :

1. l'opérateur réalise, via une Interface Homme Machine spécifique : un tri au vue du contenu du document (si le document n'est pas éligible il passe au suivant), ce tri se fait non seulement en fonction du tableau de gestion mais aussi en fonction de lots prés définis (projetX, projetY, R&T, brevets, rapports de stage, divers...). Certains lots sont spécifiques au fonds documentaire de la ME : annonces de communautés de pratiques, sommaires de revues ou d'ouvrages, sommaires de collections... Puis l'opérateur réalise un copier/coller des informations d'indexation (Titre, Auteur(s), date...), le logiciel crée alors les fichiers d'indexation (l'un pour les archives, l'autre pour la ME), formate le flux archives et le flux ME (nomenclature des noms de fichier...) et copie les fichiers dans les zones adéquates,
2. le logiciel détermine s'il existe déjà une structure d'accueil en local (par rapport au flux que l'on est en train de traiter) ou en archives. S'il n'y a pas de structure

locale (pas de document déjà traité appartenant à ce lot) il faut créer cette structure et mettre à jour les paramètres génériques (nouveau document, type tout venant...), sinon il faut utiliser la structure existante puis mettre à jour le fichier d'indexation,

3. le traitement suivant permet de dédoubler les documents par rapport au flux entrant (y a t-il un autre document de même titre dans le flux entrant ?...)
4. La quatrième étape, commune à l'ensemble des flux, met en place le dédoublement des documents par rapport aux archives et au fonds documentaire de la ME (un document du flux entrant possède-t-il le même titre qu'un document déjà en archives ou en ME ?). Elle réalise aussi l'insertion des méta données : titre, auteur, date dans les documents et la fusion des fichiers d'indexation pour les flux déjà existants en archives et en ME (concaténation du fichier d'indexation du flux entrant et des fichiers d'indexation déjà existants...). Pour l'instant, afin de contrôler la bonne exécution des traitements les versements sont réalisés (copier/coller) manuellement, à terme ils seront automatisés.

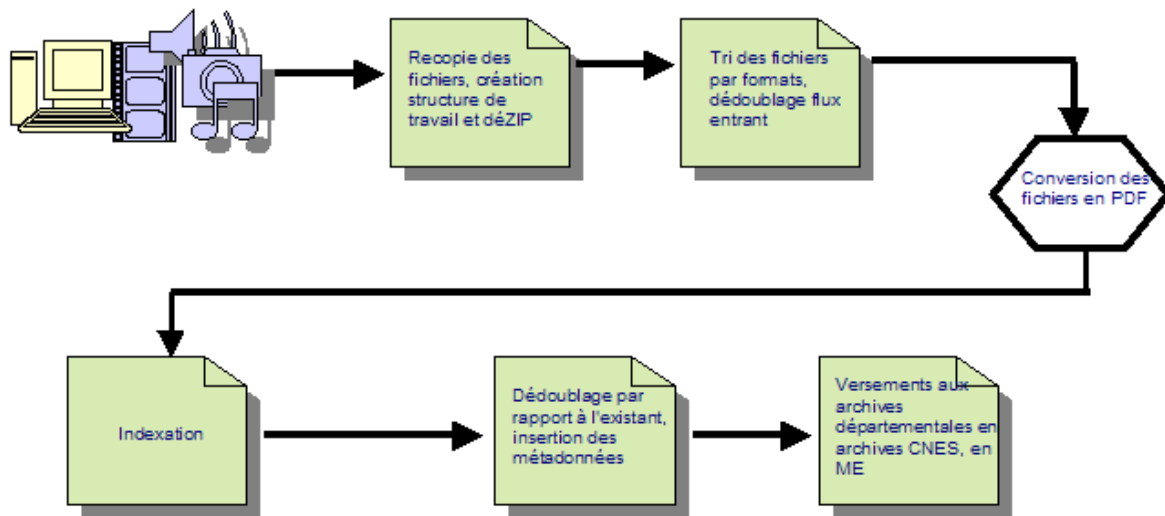


Figure 1 : Schéma des traitements

## ■ Valorisation des travaux relatifs aux archives

Notre deuxième centre d'intérêt concerne la valorisation des travaux relatifs à l'archivage, cette valorisation passe par la mise à disposition des documents d'archives dans un contexte qui permet leur exploitation par tous les agents du CNES. Ce contexte regroupe des documents techniques, des documents patrimoniaux, de la connaissance, du savoir et des services tels que : le téléchargement des documents CNES, des liens vers les spécialistes, est concrétisé par la Mémoire d'Entreprise.

### Présentation générale de la ME (mémoire d'entreprise)

La ME est composée d'un fonds documentaire de plus de 100 000 documents qui comprend les documents versés en archives :

- Les publications ou documents produits par le CNES (Qualité Espace, la Recherche Spatiales, CST-Info...)
- Les documents de projets. (ARSENE, DORIS, SPOT, TOPEX, INTEGRAL...)
- Les documents de structures (activités de R&T, ingénierie et études, support....)

- Les documents d'organismes extérieurs réalisés pour le compte du CNES (études, rapports de contrats...)
- Les brevets CNES
- Les documents issus des travaux de GC (livre de connaissances, interviews patrimoniaux...)
- Les thèses, mémoires, rapports de stages
- Les sujets et rapports de veilles et intelligence (non confidentiels)
- Les pré-prints d'articles d'auteurs CNES
- Les cours d'auteur CNES
- Les notes techniques
- Les documents de travail personnel
- Les documents propriétés du CNES ou avec accord de Copyright

Et des documents non éligibles pour les archives :

- Les sommaires des annonces des Centre de Compétences Techniques (pour trouver les spécialistes et les manifestations organisées par les CCT). Il s'agit des communautés de pratiques du CNES.
- Les index des collections d'ouvrages (Les techniques de l'ingénieur, ... pour lesquels nous avons eu l'autorisation de mise en ligne)
- Les sommaires des principales revues métiers (plus de 40 revues du spatial), qui sont conservées en salle des collections ainsi GI peut répondre à une demande en ligne d'un article de ce fonds.
- Les références des ouvrages des bibliothèques du CNES (Paris, Evry, Kourou et Toulouse.) ce qui permet la réservation en ligne quelque soit le centre où l'ouvrage est stocké

Ce fonds documentaire est en constante évolution en fonction : des versements en archives, de la réception et du traitement des revues, de l'achat et du traitement des ouvrages, des travaux de GC...

La mise à disposition de ces documents, de ces services et de ces liens se fait en utilisant le moteur de recherche de la ME qui est accessible via le portail des CDI du CNES.

Afin d'optimiser la recherche et obtenir un bon niveau de pertinence ce moteur effectue les recherches en sémantique et en plein texte. La recherche sémantique s'effectue à partir d'une Base de Connaissances (BC) qui sert :

1. A l'indexation des termes contenus dans les documents en fonction des concepts contenus dans la BC.
2. Au classement automatique des documents en fonction de leur contenu, dans des catégories pré déterminées (arbres de classement).
3. A effectuer les recherches en tenant compte du contexte d'un terme : catégories, domaines et termes associés en plus des synonymes (lexicalisation) et des tournures grammaticales : singulier, pluriel (lemmatisation). De plus, les concepts sont reliés entre eux par des relations qui permettent d'affiner la contextualisation et d'affecter un niveau de pertinence aux documents trouvés. A travers ces relations les concepts forment des réseaux sémantiques. L'un des objectifs de la cellule GC est de modéliser les métiers du CNES en travaillant avec les spécialistes et les experts. Ce travail consiste à faire le recueil des concepts pour chacun des métiers puis d'établir les liens sémantiques entre concepts, c'est-à-dire des liens de sens décrivant la relation entre deux idées. L'ensemble de ces liens pour un métier forme un ou des réseaux sémantiques.

A ce niveau nous trouvons une forte synergie avec l'activité Traduction/Interprétation/Terminologie, cette activité développe une base terminologique de termes métiers (plus de 20 000) traduits en anglais et en russe. Cette base permet d'initialiser le recueil des concepts par métiers.

## ■ Conclusion

A ce jour nous traitons les flux électroniques suivants :

- Les flux papiers suite à leur transformation en flux électroniques respectant la charte d'archivage
- Certains flux issus de GED (la GED projets du CNES, la GED des clauses techniques des contrats...)
- Les sommaires des annonces des Centre de Compétences Techniques (CCT). Ces sommaires permettent de trouver les spécialistes et les manifestations organisées au sein des communautés de pratiques du CNES.
- Les index des collections d'ouvrages (Les techniques de l'ingénieur, ...) pour lesquels nous avons eu l'autorisation de mise en ligne.
- Les sommaires des principales revues métiers (plus de 40 revues du spatial), ces dernières sont conservées en salle des collections ainsi GI peut répondre à une demande en ligne d'un article.
- Certains flux spécifiques (nanotechnologies, ouvrages des CDI...)

Le traitement "Tout Venant" est en cours de test, il doit être opérationnel fin 2008.

La mise en place de ces chaînes de traitement augmente considérablement le nombre de documents versés en archives : en moyenne 20 000 au cours des années précédentes, 30 000 en 2008 et plus de 40 000 par an à terme.

Cette augmentation a deux conséquences ; l'obligation de traiter les doublons et la sophistication des moyens mis en œuvre pour valoriser ce fonds documentaire.

Nous traitons les doublons dans différentes étapes : en premier lieu dans le flux "entrant", puis pour éliminer les doublons entre le flux entrant et l'existant (archives, ME). Il reste du travail dans la mesure ou la notion de "doublon" doit être affinée : vrai/faux doublons, faux/vrai doublons...

La ME est notre outil de valorisation du flux archives, compte tenu de l'augmentation continue de la taille du fonds les moyens de recherche doivent eux aussi évoluer. Nous avons choisi d'utiliser une approche "sémantique" via une base de connaissances et un moteur de recherche sémantique. Cette approche nous permet, entre autres :

1. d'éviter de mettre en place, auprès des auteurs de documents, des règles et des contrôles relatifs à la contextualisation de leur création ; i.e les auteurs de documents ne se préoccupent pas de métadonnées, ni de mots clés, résumé...
2. de rendre la gestion de la contextualisation indépendante de la création des documents, c'est-à-dire de standardiser, contrôler et gérer en configuration, la contextualisation des documents via la Base de Connaissances. Cette base de connaissances est développée et maintenue par la cellule GC avec l'aide des experts, elle est l'image à une période donnée du contexte de travail du CNES. En effet, elle est basée sur la modélisation des métiers du spatial, cette modélisation suit les évolutions du domaine : nouveaux métiers ou activités, modification d'un métier... La BC évolue dans le temps au rythme des transformations du domaine spatial.
3. de classer automatiquement les documents, via la BC, ce qui permet de remettre chaque document dans son (ses) contexte pour une période donnée. L'analyse des documents non classés participe au contrôle qualité.
4. de réutiliser la Base de Connaissance dans des contextes différents, par exemple : la veille et intelligence, le maintien du "sens" dans le temps (thèse en cours), ou encore l'archivage de la BC pour recontextualiser les documents archivés au cours d'une période donnée. La réactivation et l'étude de la BC, plusieurs années après son archivage, est un élément qui contribue à la connaissance de l'environnement dont sont issus les documents archivés à la même période.

Bien qu'il soit très difficile d'établir le "Return Of Investissement" des efforts liés à l'archivage et la mise en place de la ME, plusieurs "success story" nous autorisent à dire que :

- la consultation des documents archivés évite de réinventer la "roue",
- la présentation "transverse" des documents (la réponse à une requête comprend des documents de tous types : thèses, brevets, rapports de R&T...), de même que les filtres d'affinage de requêtes (sélection des métiers les plus représentatifs de la requête) élargissent le champ de réflexion et sont donc un facteur d'innovation.





**Séance 3**

**Du *records***

***management* aux**

**archives : développer**

**des stratégies pour**

**transférer les données**

---

---

Sous la présidence de Ian MCFARLANE

Spécialiste des documents et de l'administration électroniques pour le gouvernement britannique

***Session 3***

***From records  
management to  
archives - developing  
strategies to transfer  
records***

---

---

*Chairing Ian MCFARLANE*

*Formerly specialist in electronic records and e-government for the UK*

# TRANSFER AND INGEST OF DIGITAL RECORDS IN NATIONAL ARCHIVES: A COMPARATIVE STUDY OF EUROPEAN COUNTRIES

---

**Kuldar Aas**

**National Archives of Estonia**

**Raivo Ruusalepp**

**Estonian Business Archives Consultancy**

**Zoltan Szatucsek**

**National Archives of Hungary**

*This paper will present first results from a co-operation project that started at a DLM Forum Members meeting and involves national archives of most European countries. The project explores current practices with transfer and ingest of electronic records into national archives and looks into new ways of informing the development of tools and standards that build on existing solutions.*

## The problem

*All governmental agencies in Europe create and use electronic records and the volume of electronic records managed and retained in ERM systems is constantly growing. To prepare for an increase in transfer of electronic records from agencies, most national archives in Europe are, at least to some degree, involved in defining and developing solutions for ingest of electronic records from ERM and other business information systems to archival preservation repositories. This has led to significant overlap and duplication of effort between countries that all seem to strive towards the same goal - electronic records safely ingested into national archives.*

## The idea

*To explore ways of learning from each others' experiences and solutions, rather than develop competing solutions over and over again, the National Archives of Estonia invited representatives from most European national archives to join a round-table discussion in late 2007 during the DLM Members Meeting in Lisbon. Scope of the meeting was to invite the archives to join their forces in defining common criteria for the ingest processes, to provide a forum for discussion and learning from each others experience, and to help overcome the duplication of effort by defining practical pre-ingest and ingest solutions for electronic records.*

## The Ingest Co-operation Wiki

*As a result of this first meeting a Wiki was set up in December 2007 to provide a forum space for uploading information on best-practices and standards in the countries involved. As of May 2008, 15 European countries have joined the network and are contributing to the Wiki.*

*The contributors have uploaded descriptions of their current practices on pre-ingest, transfer and ingest of electronic records and databases as well as background information on data creation and legislation. This information has been analysed by the Wiki editors to create "Thematic Pages" which represent a comparative analysis of tools, standards and practices in different areas of interest - pre-ingest, transfer and ingest of electronic records and databases.*

## Potential Areas of European Co-operation

*Based on the information provided by the participating countries it is possible to outline some areas where international cooperation would benefit all national archives:*

- **Generalised tools for creation of packages that are transferred to the national archives.** *Up till now different national archives are developing tools for compiling packages of records and their metadata based strictly on the national rules and requirements. At the same time all archives are moving towards using XML technologies, and most package structures include some XML components, e.g., packaging. A transfer package creation tool that could understand different XML package structures and metadata standards could be used by most agencies in Europe and could even comply with changing guidelines and legislation in different countries.*
- **Transfer of electronic records.** *National archives that have experimented with online transfer of electronic records have relied on existing solutions for data transfer between information systems. These tools are not always suited for secure, bulk data transfer. A standardised transfer protocol and supporting software tools to allow reliable and secure bulk data transfer and validation between agencies and archives are of interest for many national archives in Europe.*

- **Transfer and preservation of specific information systems.** Some types of standard information systems (e.g., GIS, medical systems etc.) are widely used around the world, but not all national archives have the expertise to deal with the whole range of disparate systems that create records. Sharing the knowledge and know-how on ingesting records from standard and bespoke information systems would benefit many archives.
- Other, more technical and detailed issues have also been identified.

#### The future

This collaboration of national archives will continue with collecting further information on emerging standards, tools and practices; analysing the state-of-the-art in digital preservation; highlighting problem areas for partners and proposing solutions in areas where the existing application could have most effect on automating and simplifying the transfer of digital records. The consortium is open and welcomes more members to widen its scope and to be able to offer more to its partners. Further analyses of current practices at national archives in Europe will also be produced.

[See the presentation.](#)

## **Le transfert et l'intégration de documents numériques dans les archives nationales : étude comparative des pratiques européennes**

Cette communication présentera les premiers résultats d'un projet de coopération commencé à l'occasion d'une conférence du DLM-Forum auquel participent des membres des Archives nationales de la plupart des pays européens. Ce projet examine les pratiques actuelles de transfert et d'intégration des documents électroniques dans les archives nationales et étudie de nouveaux modes d'information sur le développement d'outils et de normes basés sur des solutions existantes.

#### **Problématique**

Le volume des documents électroniques gérés et conservés dans les systèmes de gestion des ressources électroniques (ERM) augmente en permanence. Pour se préparer à l'augmentation du volume des documents électroniques transférés par les agences, la plupart des archives nationales européennes se sont engagées, à des échelles différentes, dans la définition et le développement de solutions pour l'intégration des documents électroniques depuis les systèmes ERM et autres systèmes informatiques d'entreprise vers les centres de conservation d'archives. Ce qui a donné lieu à des recoupements et à des redondances importantes des actions des différents pays qui visent néanmoins un même objectif: l'intégration sécurisée des documents électroniques dans les archives nationales.

#### **Idée**

Afin d'étudier les différentes façons de mettre à profit les expériences et les solutions des autres plutôt que de développer encore et toujours des solutions concurrentes, les Archives nationales d'Estonie ont invité des représentants de la plupart des archives nationales européennes à participer à une table ronde organisée à l'occasion de la conférence du DLM-Forum à Lisbonne en fin 2007. L'objectif de cette réunion était d'inviter les différentes archives à conjuguer leurs efforts pour définir des critères d'intégration communs, d'offrir un forum de discussion et de partage d'expériences et d'essayer d'éviter les redondances entre les activités en définissant des solutions pratiques de pré-intégration et d'intégration des documents électroniques.

#### **Coopération Wiki en matière d'intégration**

En décembre 2007, un Wiki a été mis en place afin de proposer un espace de forum dédié au téléchargement des informations sur les bonnes pratiques et les normes des pays participants. En mai 2008, quinze pays européens ont rejoint le réseau et sont devenus collaborateurs du Wiki. Ces collaborateurs ont téléchargé des descriptions des pratiques en cours. Ces informations ont ensuite été analysées par les rédacteurs Wiki pour créer des « Pages Thématiques » qui présentent une analyse comparative des outils, des normes et des pratiques en vigueur dans les différents domaines: pré-intégration, transfert et intégration des documents électroniques et des bases de données.

#### **Domaines potentiels de coopération européenne**

A partir des renseignements fournis par les pays participants on a pu dégager quelques domaines dans lesquels une coopération internationale s'avérerait utile pour toutes les archives nationales:

1. Outils généralisés pour la création des paquets transférés aux archives nationales.
2. Transfert des documents électroniques.
3. **Transfert et conservation des systèmes d'informations spécialisés** (systèmes d'informations géographiques, systèmes médicaux etc.)
4. D'autres thèmes plus techniques et plus précis ont également été identifiés.

### L'avenir

*Cette collaboration des archives nationales va se poursuivre avec la collecte d'informations sur les normes, les outils et les pratiques émergents, l'analyse de l'état de l'art en matière de conservation numérique, la mise en lumière des domaines problématiques pour les différents partenaires et l'offre de solutions là où les applications existantes pourraient s'avérer plus efficaces en matière d'automatisation et de simplification du transfert des documents numériques. Notre consortium est ouvert et prêt à accueillir de nouveaux membres pour élargir son champ d'action et pouvoir offrir un support plus important à ses partenaires. De nouvelles analyses des pratiques actuellement en cours dans les archives nationales européennes seront également conduites.*

[See the presentation.](#)

## ■ Introduction

All governmental agencies in Europe create and use electronic records and the volume of records managed and retained in EDRM systems is continuously growing. Most national archives in Europe are, at least to some degree, involved in defining and developing solutions to prepare for the future transfers of these records to archives. There is a significant overlap and duplication of effort in many European countries in developing tools for ingest of electronic records from ERM and other business information systems to digital archives for long-term preservation.

For these reasons the National Archives of Estonia and Estonian Business Archives decided to put forward a proposal to several national archives for collaboration in gathering and sharing the information on current practices of transfer and ingest of electronic records.

A first meeting of representatives from 10 national archives took place in December 2008 at a DLM Members meeting in Lisbon. The main outcome of this meeting was establishment of the *Archives Ingest Cooperation Wiki* (<http://www.eha.ee/wiki>) to allow specialists from national archives to share information on electronic records ingest and transfer practices and outstanding issues in their countries. The longer term aim of this portal is to make this information available for the general public and to identify areas for direct cooperation.

This paper gives a brief insight into the objectives, development and first outcomes of the collaborative information source.

## ■ Objectives

The main objective of the collaboration is to collect information on transfer and ingest processes of electronic records that both already in use, or planned to be implemented by national archives. The gathered information itself would form a basis for:

- defining common criteria for the ingest processes;
- providing a forum for discussion and learning from each others' experience;
- providing participants with regular analyses based on the information on the Wiki;
- defining recommendations for co-operation in developing practical pre-ingest, transfer and ingest solutions for electronic records.

It should be noted that the goal of the collaboration is to serve as a shared knowledge base, but it has been set up as an ongoing voluntary effort. Detailed practical projects should be spin-offs of the knowledge base rather than part of it. So far, the collaboration has been sustained via voluntary contributions with no direct funding.

## ■ Why a Wiki?

There are many good reasons why a group-work software like Wiki should be used for an international collaboration project. The main argument, however, is that the topics of transfer and ingest are of high priority for most national archives in Europe and research and development in this area has a high pace and new knowledge is changing the methods fast. Traditional information collection methods like surveys and questionnaires would yield results that will be obsolete shortly after the survey was conducted. A dynamic environment like Wiki, on the other hand, enables users to create and edit up-to-date information whenever and wherever suitable. The benefits of using a wiki for information collection and exchange have already become apparent within the first months of the project.

## ■ Scope of the analysis

The main information in the Wiki is structured according to so-called Country Profiles - every participating country has its Country Profile, where information can be created (see Figure 1).



Figure 1: A Country Profile on the Wiki

*(Click on the picture to see it in full resolution)*

Each Country Profile consists of nine chapters that are grouped into four major areas of information. Mostly the chapters provide a general description of the main activities and provide links to more detailed resources on the issue. The main areas each country is providing information on are:

- **Background information.** While developing solutions for pre-ingest, transfer and ingest of electronic records national archives have to take into account several factors. In the Wiki such background information is described in three chapters: legislative and organisational background; management of databases and management of electronic document and records management systems (EDRMSs). The contents of these chapters include short descriptions and links that help the readers understand how electronic records are created and managed, what is the division of responsibilities between agencies and archives and what legislation and other regulations have to be followed. This information is mainly used for a comparative analysis - based on this it is possible to see what the countries have most in common and therefore have the most potential of working together towards practical solutions.
- **Pre-ingest of electronic records and databases.** To have good quality electronic records prepared and validated in the agency in such a way that they can be easily transferred and prepared for preservation in the national archives is a necessity. A country profile in the Wiki has three chapters devoted to this issue: a general chapter that describes general concepts and methods and two chapters to describe both the preparation of data from databases and records from EDRMSs. The short descriptions in these chapters give an overview of the basic standards, regulations and practices while links to more detailed

descriptions enable to dig deeper into the technical issues like SIP structures and metadata standards used in a country.

- **Transfer to archives.** Transferring electronic records and datasets to the national archives is a task with many potentially inherent dangers. A national archives has to ensure that records arrive complete, unaltered and securely at the national archives and that possible errors are discovered as early as possible. The Wiki has one chapter to describe transfer characteristics like transfer media, methods of transfer, procedures and quality assurance methods applied to records.
- **Ingest of electronic records and databases.** With the help of ingest procedures SIPs transferred to the national archives are checked for their quality and made available through the archives' information systems. There are two chapters in the Wiki that describe the procedures, guidelines and tools used to fulfil and automate the tasks. A distinction is made between ingest of electronic records and ingest of databases.

## ■ Analysis of ingest activities

### Analysis in the Wiki

While the Country Profiles can be regarded as being a good way to input information by the participants, they alone do not provide an efficient way for getting a comparative overview of the contents. For that reason the Wiki includes Thematic Pages that bring together all country profiles on a thematic basis. The Thematic Pages are structured similarly to the Country Profiles - five different pages to deliver a comparative overview of legislative and organizational background, creation and management of electronic records, pre-ingest, transfer and ingest. The Thematic Pages are updated whenever new information is input to the Country Profiles. This paper reflects the state of contents of the wiki at the end of October 2008.

On the technical side some simple tools are being developed in the Wiki to automatically collate the information from Country Profiles into Thematic Pages and therefore further simplify the analysis and keep an eye on changes in the Country Profiles.



Figure 2: Thematic Pages of the Wiki

*(Click on the picture to see it in full resolution)*

### Participants to date

The initial group of project participants included 10 national archives: Denmark, Estonia, Finland, Germany, Hungary, Netherlands, Norway, Portugal and Sweden. During the following months contact was made with six more national archives: UK, France, Latvia, Lithuania, Slovenia and Spain.

Information provided by different partners is on different level of detail. Also due to lack of time and resources it has not been possible for all partners to fill their Country Pages. Therefore the analysis delivered in



this chapter is preliminary and not comprehensive of whole Europe. Notably the Nordic countries (Denmark, Norway and Sweden) as well as Estonia and Slovenia were able to input all relevant information to their Country Profiles.

### **Analysis of the Country Profiles**

The current section provides the information available on the Country Profiles at the end of October 2008. The information is presented as comparative tables organised into three major areas (pre-ingest, transfer and ingest).<sup>1</sup> For every area also some preliminary trends are highlighted.

---

<sup>1</sup> Note that the information highlighted in this paper is preliminary. For a more detailed overview and textual descriptions consult the *Country Profiles* and *Thematic Pages* at <http://www.eha.ee/wiki>.

**Pre-ingest**

	Denmark	Estonia	Germany	France	Norway	Slovenia	Spain	Sweden
Who is responsible for preparing records for transfer to public archives?	Agency	Agency		Agency		Agency		
Metadata standards applied	Same as for databases applies	Estonian Records Management Metadata Standard (implemented as a XML Schema)	Several metadata-standards for the exchange of electronic records between ERMS are developed (XJustiz, XFinanzen etc.). An exchange-standard XDOMEA 2.0 is currently under construction.		NOARK-5 metadata and appendix			All systems must be able to export and migrate electronic records. There has to be a preservation plan in place. A generally described amount of documentation and metadata has to be available
RM functional requirements applied	FESD-standard (Derived from NOARK)	Estonian Functional Requirements for Electronic Records Management Systems	DOMEA Concept (Includes a certification program)		NOARK-5 core and NOARK-5 complete	Universal Technical Requirements (based on MoReq, includes a certification program)	A working group is currently preparing guidelines to regulate management and transfer of electronic records	All systems must be able to export and migrate electronic records. There has to be a preservation plan in place. A generally described amount of documentation and metadata has to be available
Standards regulating creation and use of databases	A formal appraisal is done for every new database. If the information in the database is of archival value, a transfer date is set and technical requirements about transfer are sent to the agency	National datasets have to be described according to a national set of requirements specifying necessary technical documentation and semantic descriptions of data and services			Databases have to be reported to the National Archives. If the data is appraised as being of archival value, the database has to be able to produce an extract of the data and to document it in a standardised way	Guidelines on how to create and maintain databases from an archival perspective are currently developed in a public-private partnership		All systems must be able to export and migrate electronic records. There has to be a preservation plan in place. A generally described amount of documentation and metadata has to be available)
Regulations regarding file formats in active systems	No regulations	A non-mandatory framework exists, which recommends to use open-source file formats in public administration.						
EDRMS transfer requirements	Same applies as for DB transfers	A subset of the elements in the Estonian RM Metadata Standard is required			NOARK-5, implemented as a XML Schema	A pilot project is currently defining the necessary metadata standard		Same applies as for DB transfers
DB transfer requirements	A national-level description (metadata) standardised approach is in place with requirements for describing content and structure of the system as well as general information about the system	No description standard is available. Future plans include usage of a subset of the descriptions created while developing a dataset (see above) for transfer and preservation purposes			ADDML (Archives Data Description Markup Language), available as DTD, development of a XML Schema in progress	A pilot project is currently defining the necessary metadata standard		A national-level standardised approach is available on how to describe databases (data and accompanying documentation)
File format limitations for computer files transferred to public archives	TIFF, SQL, flat file	XML, PDF/A, CSV, SQL, TIFF, PNG, WAV, MPEG2		A file format list will be available as part of French general interoperability framework (Référentiel général d'interopérabilité, under processing)	plaintext (UTF-8 or ISO 8859-1:1998, Latin 1), TIFF version 6, XML (also ODF and OOXML), PDF/A, XML, TIFF v6, JPEG, MP3, PCM and PCM-based Wave, MPEG-2, SOSI v2.2 or later	Instead of listing strict archival file formats general requirements for suitable formats are in place. A list of formats which meet these requirements is also available		

Table 1: Comparison of pre-ingest details in European countries

Based on the information delivered in the table it is possible to see that:

- In most of the countries the responsibility of preparing electronic records for transfer lies in the hands of the agencies that created the records;
- There is a strong trend towards restricting the number of file formats allowed by the national archives (list of archival file formats). There is a significant overlap of archival formats in different countries;
- XML is the most popular technology both for implementing metadata sets that describe records and databases as well as for packages used for transfer.

### **Transfer**

	Denmark	Estonia	France	Norway	Slovenia	Sweden
EDRMS transfer package structure	Same applies as for DB transfers	Two XML Schemas are available for	An exchange standard which defines the XML transfer form for digital records is available	NOARK-5 includes specifications of the transfer structure (also as XML Schemas)	A pilot project is currently defining the necessary structure	Same applies as for DB transfers
Database transfer package structure	A standardised structure is available describing how to organise data, documentation and metadata on transfer media	A structure is available describing a folder structure to organise data, documentation and metadata on transfer media			A pilot project is currently defining the necessary structure	A standardised structure is available describing how to organise data, documentation and metadata on transfer media
Transfer possibilities (online / offline)	Offline	Until now only offline, online transfer processes are being developed		Offline	Offline	
Transfer media	CD-R, DVD-R, DVD+R, USB hard disks	until now only CD-R, future plans involve starting to use other off-line media and online transfer		CD-R, DVD-R or DVD+R, hard disk with USB 2.0 connector, memory stick with USB 2.0 connector	CD-R, DVD	
Protocols and standards in online transfer			An exchange standard which defines the XML transfer form for digital records is available			
Software tools used in agencies to prepare transfers		A tool which helps the agencies archivist to extract records and their metadata from EDRM systems, validate them and create transfer packages is available		Arkadukt: a program which produces technical metadata about databases according to ADDML		

Table 2: Comparison of transfer details in European countries

Based on the information in the table it is possible to see that:

- Most national archives transfer digital records using off-line media like CD-R, DVD-R or USB hard disks;
- Most common approach is to define guidelines for structuring information on storage media and for the transfer process, incl. description of media;
- Multiple national archives are conducting research on secure online bulk data transfer possibilities;
- An established online transfer methodology is already available in France.

### ***Ingest***

	Denmark	Estonia	Norway
Outline of ingest tasks	Extensive routines and tools for testing and validating transfers, as well as for preparing data for long-term preservation	Manual check of metadata and viruses, input of archival descriptions into AIS	Routines for validating and testing of the structure and formats exist
Tools used for automating ingest processes			
- DB ingest	Tools are available	Until now no tools are available	Arkade (program to check accordance of datasets to ADDML)
- RM ingest	Tools are available	Until now no tools are available	ArkN3; ArkN4 (programs to check extracts out of NOARK-3 and NOARK-4 conformant ERM systems)
Database of archival descriptions	DAISY	AIS	Asta
Archival media used	LTO, hard disks, DVD	CD-R	CD-R
Number of archival copies	4	2	2

Table 3: Comparison of ingest details in European countries

Based on the three answers it is unfortunately not possible to deliver any generalised trends.

## ■ Potential areas of cooperation between European national archives

Based on information provided in the previous sections it is possible to outline some areas where international cooperation would benefit all national archives. The following recommendations are structured into three areas - pre-ingest, transfer and ingest.

### **Pre-ingest in European countries**

It is reasonable to assume that there is a general need for tools which help record creators to prepare transfer packages in XML format according to the structure defined by national archives. Until now such tools are available only in a few countries but those are strictly bound to their national specific transfer structures. It is recommended to explore possibilities to generalize the functionality of those tools to be able to use them flexibly with different structures and metadata sets.

Quite a large number of different systems that are highly standardised themselves (GIS, medical systems etc.), but need special effort to transfer them to archives, are not dealt with right now by national archives. The main reason for this could be the lack of experience and resources at individual national archives. As the structure and description of these systems can be assumed to be rather similar across Europe it is recommended to explore possibilities for a collaborative effort in preparing guidelines and methods to process these systems.

## **Transfer of digital records to national archives**

As online transfer methods are not yet widely adopted among national archives it is recommended to evaluate already available transfer standards like the French data exchange standard and based on these create a European standard for secure and reliable bulk data exchange between record creators and archives. Also developing general application guidelines to complement this standard would be useful.

## **Ingest of electronic records and databases**

As some tools for automatic checking of the quality of SIPs have already been developed by some national archives (for example Norway and Denmark) it is recommended to analyse whether these tools can be generalised to be supporting multiple information package structures and metadata standards;

The functionality of those automatic quality-check tools is quite similar to the tools recommended above for pre-ingest purposes. Therefore a second possibility is to analyse if it is possible to create a single set of tools which help both to create the transfer packages at the agency as well as to validate them both before and after transferring them to the archives.

## **■ The future**

Results presented in this paper are preliminary and further input is needed to provide a comprehensive set of recommendations for the transfer and ingest procedures in European national archives. While it was possible to extract first recommendations based on the available information there will be a continuous effort to gather information from available partners as is warmly welcoming new members to join the collaboration. Our aim is to continue with the collaboration and to be a primary knowledge base for existing and emerging co-operations and projects among stakeholders in digital preservation.

The main lesson learned from the first year of this collaboration is that the experts working in this field are very much engaged in their everyday activities and therefore finding time to share information on the Wiki can be a hard task for many of them.

A potential funding opportunity has been found to support further analysis of the Wiki content and deliver more detailed and better quality results. The National Archives of Hungary has started an effort to define the main requirements for their transfer and ingest processes. As part of it the information on the Wiki will be analysed and the results of this analysis will be freely available to all participants.

Our hope is that this approach will be adopted by other institutions and projects - the information on the Wiki will be available for general public free of charge as long as all analysis and reports that use the information are made available to all participants!

# REVISING THE STANDARDS FOR DIGITAL TRANSFERS

---

Jan Dalsten Sørensen

Principal Consultant, Head of Section for Digital Archiving, Department of Appraisal and Transfer, National Archives of Denmark

*The National Archives of Denmark has a long history of setting up regulations for the transfer of born digital information and records. The archival legislation enables the National Archives to require the agencies to convert data and documents to archival standards before transfer. In OAIS-terms this means that the SIP (Submission Information Package) that is received by the archives is equivalent to the AIP (Archival Information Package) that is stored in the archive. Of course, for this set-up to work all transfers of data and documents must be meticulously tested. The testing involves automated test of data formats, document formats, and structure, as well as a manual test of the contents of metadata and other parts of the transfer, that can only meaningfully be tested by the human eye.*

*The current regulations for transfer of data and documents in the form of system independent archival versions were implemented in 2000, although with minor revisions in 2002 and 2004. Since then, more than 1300 transfers of data and, if applicable, documents from IT-systems have been approved of and stored in the National Archives. The practical handling of transfer and test has provided us with valuable experiences. Based on those experiences, the Danish National Archives is now in the process of revising the regulations for the transfer of born digital material to public archives. The goal for the project is to formulate a revised set of regulations that*

- *is based on the principle of system independence*
- *is based on standards (ISO, W3C, etc.)*
- *takes into consideration the technological development that has taken place since the current regulations were formulated*
- *takes into consideration the experiences we have made so far*
- *takes into consideration the results that have been achieved in other countries regarding digital transfers*
- *takes into consideration the costs of transfer, for the archives as well for the transferring agencies*
- *will improve the quality of the digital transfers*
- *takes into consideration that the regulations must be adequately supported by current technology*

*The planned launch of the new regulations will be in the beginning of 2010.*

*This paper will sum up the main considerations of the revision of regulations for transfer, and also describe the plans for the implementation.*

[See the presentation.](#)

## Révision des normes pour les transferts d'objets numériques

Les Archives nationales du Danemark ont une longue histoire en ce qui concerne la mise en place de réglementations relatives au transfert d'informations et d'archives numériques natives. La législation relative à l'archivage permet aux Archives nationales d'exiger des agences de convertir les données et les documents aux normes d'archivage avant leur transfert. En termes de normes OAIS, cela signifie que le SIP (paquet d'information à verser) qui est reçu par les archives est équivalent à l'AIP (paquet d'information archivé) qui est stocké dans les archives. Bien entendu, pour que cela fonctionne, tous les transferts de données et de documents doivent être méticuleusement testés. Ce qui suppose un test informatisé des formats de données, des formats et structures des documents, ainsi qu'un test manuel des contenus des métadonnées et autres parties du transfert que seul l'œil humain ne peut tester de façon formelle.

Les réglementations en vigueur relatives au transfert de données et de documents sous la forme de versions d'archivage indépendantes du système ont été mises en œuvre en 2000, avec quelques révisions mineures en 2002 et 2004. Depuis, plus de 1 300 transferts de données et, le cas échéant, de documents issus des systèmes informatiques ont été approuvés et stockés dans les Archives nationales. La manipulation des transferts et des tests nous a fourni des expériences utiles. Basées sur ces expériences, les Archives nationales danoises sont en train de réviser les réglementations relatives au transfert de documents numériques natifs vers les archives publiques. L'objectif du projet est d'élaborer un ensemble de réglementations révisées qui :

- sont basées sur le principe d'indépendance du système
- sont basées sur des normes (ISO, W3C, etc.)
- prennent en considération le développement technologique depuis les réglementations en vigueur
- prennent en considération les expériences que nous avons réalisées à ce jour
- prennent en considération les résultats obtenus dans d'autres pays en ce qui concerne les transferts numériques
- prennent en considération les coûts de transfert, à la fois pour les archives et pour les agences de transfert
- amélioreront la qualité des transferts numériques
- prennent en considération le fait que les réglementations doivent être supportées de façon adéquate par la technologie actuelle

Les nouvelles réglementations devraient entrer en vigueur début 2010.

Le présent article résumera les principales considérations concernant la révision des réglementations relatives au transfert et décrira le programme de mise en œuvre.

[Consulter la présentation.](#)

When transferring data to an archive that has based its preservation on the conversion strategy, data and documents must be converted to formats and standards that are suitable for long term preservation.

The National Archives of Denmark has a long history of setting up regulations for the transfer of born digital data and records. The archival legislation enables the National Archives to require the public agencies to convert data and documents to archival standards before transfer. In OAIS-terms, this means that the SIP (Submission Information Package) that is received by the archives is almost equivalent to the AIP (Archival Information Package) that is stored in the archive.

The current Danish regulations for transfer of data and documents in the form of system independent archival versions were implemented in 2000, although with minor revisions in 2002 and 2004.

Since then, more than 1300 transfers of data and, if applicable, documents from IT-systems have been approved of and stored in the National Archives. The practical handling of transfer and test has provided us with valuable experiences. Based on those experiences, the Danish National Archives is now in the process of revising the regulations for the transfer of born digital material to public archives. This article will sum up the ideas behind the project and present some of the issues we take into consideration when we revise the standards for transfers of electronic, or digital, records.

## ■ Goal and principles

The goal for the project is to formulate a revised set of regulations that...

- is based on the principle of system independence
- is based on standards (ISO, W3C, etc.)
- takes into consideration the technological development that has taken place since the current regulations were formulated
- takes into consideration the experiences we have made so far
- takes into consideration the results that have been achieved in other countries regarding digital transfers
- takes into consideration the costs of transfer, for the archives as well for the transferring agencies
- will improve the quality of the digital transfers
- takes into consideration that the regulations must be adequately supported by current technology

In the following, I will go into more details with the thoughts behind these dots.

Generally speaking, you can say that digital archiving is no easy business. And it can be extremely hard to fully understand the consequences of your choices. If you choose to rely on standards that are not open enough, reliable, well-described etc. you risk that you will not have the ability to access the digital records in the long term. Conservatism is therefore a key word.

### **System independence**

It is crucial for the long term preservation and the ability to convert to new formats as well as making the electronic records accessible for future users that what we accept into the archives is not in any way dependent on the system where data was generated or dependent on a particular, proprietary technology. System independence is thus a key concept for the long term preservation of data and digital documents.



## **Standards**

The requirement that our specifications for transfer are based on standards is, of course, in very close connection with the requirement of system independence. For an archive it is important that the standard is open, and that all documentation is available.

## **Technological development**

Obviously, the it-systems that are used in the public administration have changed and evolved since 1999 when our present regulations were defined. There has also been quite a bit of standardization going on, opening up for possible alternative formats that we did not consider the last time. Of course, the archives must be conservative and refrain from jumping on any bandwagon. On the other hand, we need to adjust our requirements to the it-systems and the technology that is actually in use out there. For example, why continue to demand text delivered in the character set that was standard 10 years ago, when almost every system now uses Unicode?

## **Experiences**

Working with the regulations in practice has given us a lot of valuable experiences both seen from the point of view of the public agencies and their vendors, as well as the archives. A couple of years ago we even had a special "gray area task force" that worked with examining all the different questions and problems that we had encountered and where the solution was not evident from the regulations. It is one thing to define some standards, but it is another to actually work with them in practice. You learn a lot from "taking the plunge" and we want to make sure that those experiences are taken into account.

## **Results and experiences from other countries**

We want to take this opportunity to look around us to see if other countries have developed new or alternative methods for transfers that we can learn from. We realize that the challenge of preserving digital information is very much on the agenda in many other countries, so we want to use the project about revising our standards as an opportunity to look abroad and see if there is any inspiration or experiences that we can learn from.

## **Costs**

You could probably define a standard that would be fairly easy for the agencies to comply with, e.g. if you only required them to transfer a dump from their databases in any form or format, and all their digital documents in whatever format they might be found in the systems of the agencies. Thus you could reduce the costs of transfer to a minimum - for the transferring agency. That would, however, make the test, long term preservation and access extremely complicated and expensive for the archive that receives the transfers. On the other hand, you could also define a standard that - in order to make it as inexpensive as possible for the archives - would make the cost unacceptably high for the public agencies.

## **Quality**

Quality can be an ambiguous term, as you could measure quality by several different parameters. One parameter could be that more of the transferred archival versions of electronic records can be approved of by the receiving archive already after the first test. Another could be that the archival versions are seen as better to use for the end-user, e.g. because it has better and more structured documentation.

The demand for higher quality must of course be seen directly in connection with the requirement that we must take the costs into consideration. It is fairly easy for the archivists to come up with an extensive wish list for improved quality, among other things in terms of documentation but this needs to be balanced against the cost. Even though we have a legislation that enables us to set up demands and regulations that the public agencies have to follow, we need to make a set of regulations that are not impossible for the agencies to comply with.

## **Adequately supported**

We also need to be practical. If we are to be able to work with the standards in an efficient way, we need to be sure that the standards that we have chosen are widely supported so that the archives as well as the producers of the archival versions have a range of tools available for the production, test and quality assurance of the digital records.

### **■ Status**

In the project, we are working on these main subjects:

- document formats
- data formats and character sets
- structure of the archival version
- documentation and metadata
- terminology
- requirements for notification on and approval of new it-systems

Thus, the project has a technical side to it, e.g. when looking at the long term preservation qualities of document formats, data formats etc. It also includes elements that require more classical archival skills, such as defining the requirements for documentation and metadata. And finally, it has an administrative side to it, since the regulations must be issued as a ministerial order.

The project right now can be seen both as a consolidation project, a knowledge project and a development project.

Consolidation, because the project provides a chance for us to go through our original arguments for the selection of formats, standards, structures etc. and see if they are still valid and sensible.

Knowledge, because this project is a framework for structured harvest of information about standards for digital preservation in the widest sense (formats, methods, metadata standards etc.) and a chance to learn more about standards for digital preservation in the world around us.

And then of course also a development project, because we will develop a new standard for digital transfers, as well as the necessary tools, procedures and guidelines to work with this new standard.

The project also includes a terminology part, because we have realized that many terms that we use about digital records, digital archiving and long term preservation can be ambiguous, contradictory or different from the terminology used elsewhere. So, before we finish writing the new set of regulations with guidelines and procedures, we need to go through our terminology to make sure that at least we have a clear and defined description of the terms that we use.

It is also our ambition to have the new regulations translated into English in order to facilitate the exchange of knowledge, ideas and inspiration with colleagues abroad.

### **■ Implementation**

When we have agreed upon the final content of the revised standard, we have to go through a process of implementation. The standard for digital transfers must be issued as a ministerial order, so obviously it takes some time to formulate a text that is precise enough to meet that criterion.

More time consuming, however, is to develop a new test tool that will enable us to test future transfers and make sure that they comply with the revised standard. Other tools are probably necessary or would at least be appreciated by the producers of the archival versions.

It is also necessary to write new guidelines to be used with the new regulations. The text of a ministerial order must necessarily be short and precise. However, to use the regulations in practice you often need a richer text with more examples. Just as we have done with the current regulations, we expect to make one or two examples of archival versions with dummy data available for download on our website because we know out of experience that the concept of an archival version can be very alien to the agencies and especially the vendors or consultants that are to produce it, and a hands-on, downloadable example that they can use as a template is a good help.

Last, but not least, all staff within the National Archives of Denmark that works with digital archiving must receive the necessary introduction to the revised regulations and how to use them in practice. Our procedures and internal guidelines must be updated or revised accordingly.

Revising the standards for transfer of digital records is a project that involves many members of our staff as well as a number of different skills and talents. We feel, however, that the effort is necessary in order to make sure that transfer of digital records continuously can take place in a way that makes long term preservation and access feasible while taking both the quality of the records as well as the costs for agencies/records creators and the archives into consideration.

The expected outcome is both measurable (e.g. costs, time and effort needed to produce and process archival versions) and less measurable (e.g. the improved knowledge and competences of the staff taking part in the project).

After the revised regulations take effect in 2010, we plan to start preparing for a coming revision right away by setting up a framework for monitoring the developments in document formats, data formats, preservation methods etc. We hope to be able to constantly monitor and evaluate the maturity of formats, standards and methods so that the next revision of the standards and regulations for transfer will require a little less effort.

# **VERS UNE DÉCENTRALISATION ET UNE DÉMOCRATISATION ENCADRÉES DU RECORDS MANAGEMENT AU SEIN DES GRANDES ENTREPRISES : LE PROJET ISIS**

---

Sabine Mas

Professeure adjointe à l'EBSI, Université de Montréal (Canada)

Yves Marleau

Président de Cogniva Information Solutions Inc. (Canada)

*Nous proposons de présenter les éléments contextuels, les caractéristiques conceptuelles et l'implantation technologique d'un projet de services d'infrastructure sémantique pour l'information (ISIS) conçus et utilisés pour encadrer la gestion et l'accès aux documents et à l'information corporative dans un environnement électronique décentralisé.*

*Parallèlement aux recherches conduites en Europe dans le domaine de l'archivage électronique plusieurs travaux sont conduits au Canada dans un contexte universitaire et gouvernemental pour résoudre les difficultés rencontrées avec l'usage des systèmes de records management traditionnels dans l'environnement électronique. Les moyens utilisés par les entreprises pour gérer leurs documents reposent largement sur la mise à disposition de moteurs de recherche, de langages documentaires et d'instruments de gestion tels que les schémas de classification, les thésaurus et les calendriers de conservation (Sabourin, 2001; Van der Walt, 2004). Malheureusement beaucoup de ces outils, qui ont été conçus pour faciliter des activités de gestion documentaire spécifiques, sont le plus souvent redondants, insatisfaisants et difficiles à gérer dans un contexte de décentralisation et de démocratisation des activités de création, d'organisation, de conservation et de diffusion des documents électroniques (Tough & Moss, 2003). Alors qu'il appartenait traditionnellement au responsable d'un poste de classement de décrire et d'organiser les documents dans le but de mieux les repérer, les partager, les exploiter et ultimement les conserver ou les détruire, c'est désormais à chaque employé que revient la responsabilité de gérer les documents électroniques créés ou reçus à partir de son poste de travail informatique. Les documents sont "auto-gérés" généralement "sans règle ni discipline" (Skupsky, 1999) par les créateurs et utilisateurs qui n'ont souvent ni mandat, ni formation, ni incitation à se préoccuper de la gestion documentaire. En l'absence de contrôle et d'une gestion normalisée et intégrée des documents, il devient de plus en plus difficile pour les grandes entreprises de se conformer aux lois portant sur l'accès à l'information et sur la transparence du processus décisionnel, d'assurer la défense de leurs droits et la conservation de la trace de leurs activités (Skupsky, 1999; Guercio, 2002).*

*Pour surmonter ces difficultés des chercheurs canadiens ont développé le projet ISIS basé sur une classification à facettes qui articule et intègre de manière transparente les normes et les fonctionnalités des outils traditionnels de gestion documentaire. Conformément à la norme ISO 15489 qui privilégie une organisation des archives courantes et intermédiaires basée sur les fonctions et activités administratives et opérationnelles et au modèle d'exigences pour l'organisation de l'archivage électronique (MoReq), ISIS vise à encadrer l'employé dans la description qu'il fait à la fois du sujet d'un document d'archives (facettes de contenu) et du contexte administratif ou d'affaires dans lequel ce document a été créé (facettes de contexte). Il vise aussi à améliorer la précision de la recherche et à favoriser une plus grande automatisation du processus de gestion du cycle de vie des documents dans le but de délester l'employé d'une tâche qu'il accomplit rarement très bien (Gagnon-Arguin & Bannouri, 1998; Mas, 2007).*

*Dans le cadre de notre exposé, nous commencerons par présenter le contexte de ce projet et les questions qui le sous-tendent. Nous caractériserons ensuite l'infrastructure sémantique et les services qui lui sont associés (e.g. service de thésaurus, service de métadonnées, service archivistique). Pour terminer, nous discuterons des multiples retombées du projet proposé sur les plans pratique (e.g. gestion intégrée et transparente des documents), méthodologique (e.g. conception d'une méthode d'identification et de sélection des facettes) et théorique (e.g. validation empirique du modèle à facettes proposé grâce au développement et à l'implantation d'une solution logicielle) en nous appuyant sur des expériences de déploiement dans différents organismes canadiens (e.g. Bibliothèque et Archives Canada) et internationaux (e.g. Fonds monétaire international).*

[Consulter la présentation : partie 1 ; partie 2 ; partie 3.](#)

## **Towards the decentralisation and supervised democratisation of records management within large companies: the ISIS project**

*We plan to present the contextual elements, conceptual characteristics and technological implantation for a project of Information Semantic Infrastructure Services (ISIS) designed and used to supervise the management of and access to documents and corporate information in a decentralized electronic environment.*

*In parallel to research carried out in Europe in the field of electronic archiving, much work has been carried out in Canada by universities and government in order to resolve problems encountered in the use of traditional records management systems in the electronic environment. The methods used by companies to manage their documents depend largely upon the availability of search engines, documentary languages and management tools such as classification themes, thesauruses and preservation calendars (Sabourin, 2001; Van der Walt, 2004). Unfortunately, many of these tools, which were created to make specific document management activities easier, are often redundant, unsatisfactory and difficult to manage in the context of decentralisation and democratisation of the tasks of creating, organising, preserving and distributing electronic documents (Tough & Moss, 2003).*

*While it was traditionally the role of a classifications manager to describe and organise documents with the aim of more easily finding, sharing and using them, and ultimately preserving or destroying them, it is now up to each employee to manage electronic documents created or received by his own computer terminal. The documents are generally "self-managed ... without any rules or discipline" (Skupsky, 1999) by creators and users who often have neither the authority, the training nor the incentive to deal with document management. In the absence of any control-system or standardized and integrated management of documents, it becomes increasingly difficult for big companies to respect the law regarding access to information and transparency of the decision-making process, or to defend their rights and preserve a trace of their business activities (Skupsky, 1999; Guercio, 2002).*

*To overcome these difficulties, Canadian researchers have developed the ISIS project based on classification by facets which articulate and integrate the norms and functions of traditional tools of document management. The ISO 15489 standard favours an organisation of current and intermediary archives based on administrative and operational functions and activities. In accordance with this norm, and the Model Requirements for the Management of Electronic Documents and Records (MoReq), ISIS aims to guide employees when describing both the subject of a document for archiving (content facets) and the administrative or business context in which the document was created (context facets). It also aims to improve the precision of the search process and favour greater autonomy in the process of managing the life cycle of these documents, with the aim of freeing employees from a task that they rarely accomplish very well.*

*In our presentation, we will start by outlining the context of this project and the questions it raises. We will then describe the semantic infrastructure and the services associated with it (e.g. thesaurus, metadata and archive services). We will conclude by discussing the multiple repercussions of the proposed project on a practical level (e.g. integrated and transparent document management), on methodology, (e.g. creation of a method of identification and selection of facets) and on a theoretical level (e.g. empirical validation of the proposed facets model thanks to the development and implantation of a software solution) by referring to the experience of deploying this in various Canadian organisations (e.g. Canada Library and Archives) and international (e.g. International Monetary Fund).*

[See the presentation: part 1 ; part 2 ; part 3.](#)

L'utilisation répandue des micro-ordinateurs depuis le milieu des années 1980 a favorisé une décentralisation et une démocratisation des fonctions de création, d'organisation, de contrôle, de manipulation et de diffusion des documents en format numérique. Alors qu'il appartenait traditionnellement au responsable d'un poste de classement de décrire et d'organiser les documents dans le but de mieux les repérer, les partager, les exploiter et ultimement les conserver ou les détruire, c'est désormais à chaque employé que revient la responsabilité de gérer les documents électroniques créés ou reçus à partir de son poste de travail informatique. Les documents sont gérés par les créateurs et utilisateurs qui n'ont souvent ni mandat, ni formation, ni incitation à se préoccuper de la gestion documentaire. En l'absence d'encadrement, de contrôle et d'une gestion normalisée et intégrée des documents, il devient de plus en plus difficile pour les grandes entreprises de se conformer aux lois portant sur l'accès à l'information et sur la transparence du processus décisionnel, d'assurer la défense de leurs droits et la conservation de la trace de leurs activités.

Parallèlement aux recherches conduites en Europe dans le domaine de l'archivage électronique plusieurs travaux sont conduits au Canada dans un contexte universitaire et gouvernemental pour résoudre les difficultés rencontrées avec l'usage des systèmes de *records management* traditionnels dans un environnement électronique décentralisé. Dans le cadre de ces travaux, l'équipe canadienne Cogniva à laquelle nous sommes associés a développé le projet ISIS (Integrated Semantic Information Services) basé sur une classification à facettes qui articule et intègre de manière transparente les normes et les fonctionnalités des outils traditionnels de gestion documentaire. Conformément à la norme ISO 15489 qui privilégie une organisation des archives courantes et intermédiaires basée sur les fonctions et activités administratives et opérationnelles et au modèle d'exigences pour l'organisation de l'archivage électronique (MoReq), ISIS vise à encadrer l'employé dans la description qu'il fait à la fois du sujet d'un document d'archives et du contexte administratif ou opérationnel dans lequel ce document a été créé. Il vise aussi à améliorer la précision de la recherche et à favoriser une plus grande automatisation du processus de gestion du cycle de vie des documents dans le but de délester l'employé d'une tâche qu'il accomplit rarement très bien<sup>1</sup>.

Dans le cadre de notre exposé, nous commencerons par présenter le contexte de ce projet et les questions qui le sous-tendent. Nous caractériserons ensuite l'infrastructure sémantique et les services qui lui sont associés. Pour terminer, nous discuterons des multiples retombées du projet proposé sur les plans pratique, méthodologique et théorique en nous appuyant sur des expériences de déploiement dans différents organismes canadiens et internationaux.

## ■ Éléments contextuels

Le développement des technologies de l'information et des télécommunications a entraîné la création et l'utilisation d'un nombre toujours croissant de documents électroniques pour la réalisation des activités quotidiennes au sein des organismes. De plus en plus de documents, qui existent uniquement sous forme électronique, possèdent une valeur administrative, légale, financière et sont essentiels à la réalisation des activités ou à la défense des droits des organismes. Parmi ces documents, certains ont également une valeur historique et font alors partie de la mémoire corporative.

Pour répondre au besoin de gérer, de retrouver et d'utiliser ces documents électroniques, les archivistes et gestionnaires de documents préconisent l'utilisation du schéma de classification institutionnel déjà conçu pour l'organisation des documents sur support papier<sup>2</sup> (Voir Figure 1). Cet outil d'organisation basé sur une structure hiérarchique de catégories descriptives relatives aux fonctions et activités de l'institution permet de regrouper

<sup>1</sup> Gagnon-Arguin, L., & Bannouri, R. (1998). *La classification des documents institutionnels dans les ministères du gouvernement du Québec et le défi informatique : rapport de recherche*. Montréal, QC : Université de Montréal, École de bibliothéconomie et des sciences de l'information.

<http://www.msg.gouv.qc.ca/fr/publications/enligne/administration/ingenierie/classification.pdf>

<sup>2</sup> Conférence des recteurs et des principaux des universités du Québec (CRÉPUQ). (2004). *La gestion des documents numériques des établissements universitaires du Québec : état de situation et planification stratégique*. <http://www.crepuq.qc.ca/documents/arch/Rapport-GGDN.htm>

Guercio, M. (2002). Records classification and content management: Old functions and new requirements in the legislations and standards for electronic record-keeping systems. In *Proceedings of the DLM Forum 2002. Access and preservation of electronic information : Best practices and solutions, May 6-8, 2002, Barcelona, Spain*. Luxembourg: Office for Official Publications of the European Communities, 432-442.

Dhérent, C., & collaborateurs. (2002). *Les archives électroniques : manuel pratique*. [Paris] : Direction des archives de France. <http://www.archivesdefrance.culture.gouv.fr/fr/archivistique/>

logiquement et physiquement tous les documents relatifs à une même activité ou à un même dossier. Il est établi que le schéma de classification institutionnel permet l'organisation des documents à plus ou moins long terme et facilite entre autres la localisation et le repérage des documents par l'ensemble des employés.



Figure 1. Exemple d'utilisation d'un même schéma de classification institutionnel dans l'environnement analogique et dans l'environnement électronique

Malgré les recommandations à cet effet, il est démontré que le schéma de classification institutionnel est rarement utilisé par les employés pour organiser et retrouver les documents électroniques se trouvant sur leur poste de travail (voir Figure 2). Les employés privilégient des schémas de classification plus « personnels » des documents électroniques qui répondent davantage à leurs besoins pour la réalisation des activités quotidiennes qu'à une vision institutionnelle.

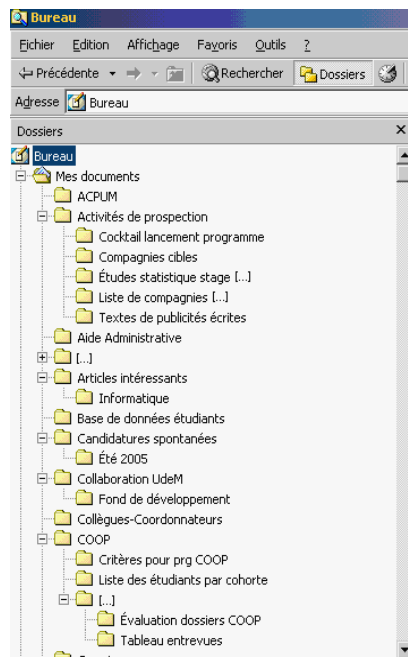


Figure 2. Extrait d'un schéma de classification personnel

Les limites inhérentes à une telle structure hiérarchique obligent les individus à devoir privilégier comme base de classification et selon des critères très personnels, tantôt le thème du document, tantôt le type de document ou encore l'activité à laquelle il se rattache, alors que le document pourrait de fait être classifié à plusieurs endroits. En l'absence d'un cadre normalisé, un organisme se retrouve avec autant de façons de faire qu'il y a d'individus.

L'individualisme qui caractérise l'organisation des documents électroniques dans un contexte de gestion décentralisée pose des problèmes multiples pour les organismes. Il est reconnu que l'absence d'organisation systématique et uniforme des documents électroniques conduit à plusieurs problèmes dont les plus critiques sont l'éparpillement des documents relatifs à une même affaire, l'impossibilité d'assurer la défense des droits et la conservation de la trace des activités de l'organisme en raison de la perte possible d'un nombre important de documents (même ceux qui n'ont pas d'équivalents sur papier), et les difficultés d'interprétation, de repérage, de partage et de réutilisation de l'information consignée liées aux modes d'organisation individuels basés sur des critères très personnels qui s'avèrent le plus souvent incompréhensibles aux autres utilisateurs.

Toutes ces considérations ont conduit les archivistes et gestionnaires de documents à revoir leurs méthodes et à proposer des solutions en vue d'uniformiser l'organisation des documents électroniques à caractère institutionnel. Depuis les années 1990, plusieurs organismes, groupes de travail et professionnels ont mené des réflexions, formulé des recommandations et des normes en vue d'améliorer l'organisation des documents électroniques sur un micro-ordinateur. Ainsi diverses propositions ont été énoncées pour simplifier et faciliter l'implantation et l'utilisation du schéma de classification institutionnel dans l'environnement électronique : formulation des règles de dénomination de fichiers et répertoires, définition de métadonnées, diminution du nombre de niveaux hiérarchiques de la structure de classification et personnalisation des titres des catégories. Malgré les adaptations effectuées, les employés qui ont eu recours au schéma de classification institutionnel pour organiser leurs documents électroniques se disent souvent insatisfaits en raison de la « logique de la classification » qui n'est pas la même que la leur<sup>3</sup>. Les employés interrogés par Gagnon-Arguin et Bannouri préfèrent utiliser soit un schéma de classification « spécifique » issu du schéma de classification institutionnel, soit un schéma de classification « parallèle », plus approprié à l'organisation des documents électroniques se trouvant sur leur poste de travail informatique.

L'exploitation de solutions technologiques a également été envisagée pour aider les individus à organiser leurs documents. Plusieurs méthodes statistiques (par exemple, fréquences des mots, pondération des mots), linguistiques (par exemple, sémantique, syntaxique) ou d'intelligence artificielle (par exemple, systèmes expert, réseaux neuronaux) ont été développées pour organiser automatiquement des données ou des documents électroniques. Toutefois, la classification automatique basée sur le traitement de la langue naturelle et la technologie des réseaux neuronaux n'est pas encore implantée dans les organismes et son efficacité reste à démontrer pour l'organisation des documents électroniques institutionnels.

Dans le domaine des sciences de l'information on explore, par exemple, la possibilité de remplacer les systèmes de classification traditionnellement hiérarchiques par une infrastructure à facettes permettant d'indexer et de naviguer à travers une collection de ressources documentaires référencées sur le Web ou plus modestement dans l'espace personnel d'un individu. La classification à facettes se présente comme une combinaison de groupes conceptuels plutôt que comme une liste structurée de sujets. L'application d'une classification à facettes permet d'éliminer certaines contraintes liées au fait que le schéma de classification traditionnel soit un langage de représentation pré-coordonné. L'impact positif des approches de classification à facettes pour l'organisation et le repérage de l'information a déjà été confirmé dans un environnement réseauté et nous est apparu prometteur pour l'organisation des documents administratifs électroniques<sup>4</sup>.

Dans le cadre de notre recherche doctorale<sup>5</sup>, l'analyse des concepts représentés dans plus d'une vingtaine de structures arborescentes de répertoires créés par des employés et l'observation d'un usage répandu de classes qui représentent plusieurs concepts (par exemple, "Dossiers étudiants A-2003") (voir Figure 3) allait dans le sens d'un transfert souhaitable dans ce domaine d'une classification hiérarchique unidimensionnelle vers une classification à facettes.

---

<sup>3</sup> Gagnon-Arguin, L., & Bannouri, R. (1998). *La classification des documents institutionnels dans les ministères du gouvernement du Québec et le défi informatique : rapport de recherche*. Montréal, QC : Université de Montréal, École de bibliothéconomie et des sciences de l'information, p.23.

<http://www.msg.gouv.qc.ca/fr/publications/enligne/administration/ingenierie/classification.pdf>

<sup>4</sup> Hudon, M., & Mas, S. (2001). *Analyse des facettes pour la classification des documents institutionnels au gouvernement du Québec. Rapport présenté pour le Groupe de travail en classification et indexation*. Montréal: École de bibliothéconomie et des sciences de l'information, Université de Montréal. (Collection en ingénierie documentaire; 13). [http://www.msg.gouv.qc.ca/fr/publications/enligne/administration/ingenierie/classification\\_analyse.pdf](http://www.msg.gouv.qc.ca/fr/publications/enligne/administration/ingenierie/classification_analyse.pdf)

<sup>5</sup> Mas, S. (2007). *Schémas de classification et repérage des documents administratifs électroniques dans un contexte de gestion décentralisée des ressources informationnelles*. Thèse de doctorat en sciences de l'information. Montréal : Université de Montréal. <http://hdl.handle.net/1866/1433>



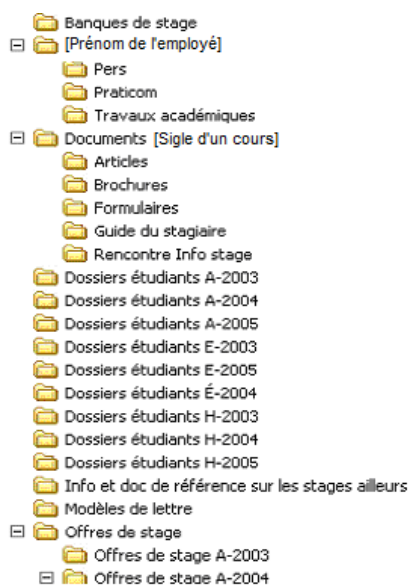


Figure 3. Extrait de la structure classificatoire arborescente du coordonnateur de stage participant à l'étude.

Parallèlement à ces travaux, l'équipe de Cogniva travaillait sur un modèle à facettes et une infrastructure logicielle permettant d'indexer n'importe quelle ressource documentaire numérisée d'une organisation. Ce modèle à facettes a été mis en œuvre après plusieurs années de réflexion et d'expérience auprès de grandes organisations en permettant de regrouper et d'intégrer plusieurs fonctionnalités des outils traditionnels de gestion documentaire : service de thésaurus, service de métadonnées, service archivistique, service de point de vue, carte de savoir, gouvernance. Nous caractérisons ci-après l'infrastructure sémantique et quelques uns des services qui lui sont associés.

## ■ L'infrastructure sémantique ISIS et les services associés

Le métamodèle ISIS repose sur un ensemble de facettes, de valeurs et de relations entre les valeurs des facettes. Il s'inspire de la théorie des facettes de S.R. Ranganathan<sup>6</sup> visant à décomposer le sujet d'un document en facettes. Selon cette théorie, une facette est un critère de division (ou caractéristique commune) non hiérarchique. À la différence d'une classification hiérarchique d'organisation des documents, l'efficacité des classifications à facettes tient à leur capacité à intégrer différentes dimensions d'analyse sur les objets informationnels, à caractériser et à rendre l'accès à l'information plus facile en offrant de multiples voies de navigation vers n'importe quel document.

L'infrastructure sémantique que nous proposons permet de coordonner deux espaces de classification des documents, l'un calé sur les critères institutionnels de l'entreprise (facettes de contexte), l'autre sur des critères métiers plus spécifiques (facettes de contenu). Les concepts relatifs au contexte administratif ou opérationnel d'utilisation et au contenu informationnel des documents, représentés par des termes, sont énumérés à l'intérieur de chacune des facettes, conceptuellement indépendantes les unes des autres mais combinées dans une description quand la situation le demande.

Les facettes constitutives du métamodèle ISIS décrivent le contenu d'un document et l'environnement administratif ou opérationnel dans lequel il est créé, reçu ou utilisé. Comme l'indique la Figure 4, ces facettes peuvent être divisées en trois catégories : (1) *les facettes descriptives du contenu d'un document*, c'est-à-dire les concepts qui décrivent le sujet principal d'un document (i.e. lieu, temps, thème, personne, etc.), (2) *les facettes contextuelles*, c'est-à-dire les concepts qui décrivent le contexte et la forme d'un document (e.g.

<sup>6</sup> Ranganathan, S. R. (1965). *The Colon classification*. New Brunswick, NJ : Rutgers University Press. (Rutgers Series on Systems for the Intellectual Organization of Information; v. 4).

fonction et activité administrative, type de contenu, rôle, position, structure organisationnelle) et (3) *une facette hybride* qui représente le concept d'organisation du travail auquel est associé le document (e.g. projet, dossier, programme, comité).

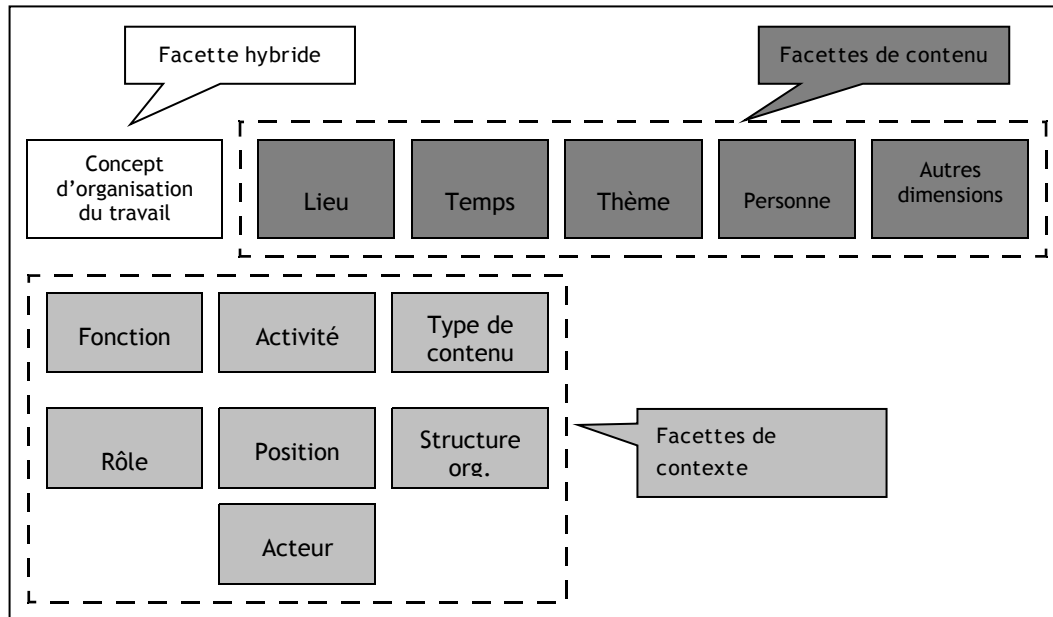


Figure 4. Les facettes constitutives du métamodèle ISIS

**Les facettes de contenu :** Dans le traitement des archives courantes et intermédiaires, le contenu ou le sujet d'un document a traditionnellement été considéré comme secondaire à sa forme et à sa provenance. Aujourd'hui, la fréquence des recherches par sujet justifie l'introduction d'index thématiques contrôlés tant aux niveaux terminologique que conceptuel en vue de rendre plus efficace le repérage de documents pertinents dans les entreprises. Le contenu même d'un document se décline en plusieurs parties : un titre, un auteur, le cas échéant un destinataire, une date, des références au contexte de production de l'information (programme, procédure, décision, etc.); le corps même de l'information, l'exposé ou les données recueillies par l'opération ou la transaction; des informations complémentaires telles que des documents joints ou liés, des illustrations. Ces éléments décrivent "de quoi parle le document", ce qui s'est passé ou ce qui a été dit ou fait. Ils constituent la "mémoire de l'action" et sont capitalisés en fonction de leur importance pour les utilisateurs<sup>7</sup>.

Une partie du contenu d'un document électronique peut être extraite automatiquement (e.g. auteur ou date de création) selon le niveau de structuration du document. L'autre partie est analysée "manuellement" (e.g. le corps du texte ou du message). Selon la théorie des facettes à la base de la *Colon Classification* de Ranganathan, le contenu de tout document touche à cinq aspects de la réalité : la personnalité (groupe, catégorie de personnes, thème ou objet duquel il est question), la matière (substance, propriété, qualité), l'énergie (opération principale, action que l'on décrit en rapport avec l'objet), l'espace (localisation du sujet dans l'espace) et le temps (place du sujet dans le temps). Si la théorie des facettes de Ranganathan offre une syntaxe idéale pour l'expression et la représentation d'un sujet, la liste des facettes de contenu ne peut être fournie *a priori* dans une formule immuable, elle doit plutôt être recherchée de façon pragmatique par un examen de la documentation relevant du domaine à traiter, l'analyse conceptuelle permettant de ramener le nombre théoriquement illimité des termes d'un domaine à un nombre restreint de facettes. On peut toutefois s'attendre à retrouver des facettes de contenu fondamentales telles que le Temps (e.g. Quand a eu lieu l'incident relaté dans le document?), le Lieu (e.g. Où a eu lieu l'incident relaté dans le document?), le Thème (e.g. Quel est le type d'incident?), la Personne (e.g. Qui sont les intervenants?). À ces quatre facettes peuvent s'ajouter d'autres dimensions suivant les besoins exprimés par l'entreprise, par exemple la langue du contenu (particulièrement utile pour les états fédérés ou les organismes internationaux tels que le Fonds monétaire international) ou l'entité ou agent (individu ou organisme) impliqué dans l'opération ou la transaction (Voir Tableau 1).

<sup>7</sup> Rietsch, J.-M., Chabin, M.-A. & Caprioli, E. (2006). *Dématérialisation et archivage électronique: Mise en oeuvre de l'ILM (Information Lifecycle Management)*. Paris : Dunod.

Facette	Description
Lieu	Lieu de l'action, de la transaction, de la production, de l'événement ou de tout autre sujet contenu dans le document (e.g. <i>France - Site X</i> )
Temps	Période de l'action, de la transaction, de la production, de l'événement ou de tout autre sujet contenu dans le document (e.g. <i>2006</i> )
Personne	Personne physique ou morale ou entité impliquée dans l'action, la transaction, la production, l'événement ou de tout autre sujet contenu dans le document (e.g. <i>Entreprise X</i> )
Thème	Sujet de discussion, de description, de représentation ou matière à traiter ou à régler contenu dans le document (e.g. <i>Fraude</i> ).
Autres dimensions	Autres dimensions de contenu qui vont être déterminées par le contexte (e.g. Facette de contenu <i>Fournisseur</i> pour le Service des achats).

Tableau 1. Les facettes de contenu du métamodèle ISIS

Les facettes de contenu sont identifiées à partir de techniques d'entrevue et sélectionnées par une méthode d'analyse documentaire. Cette méthode d'analyse ascendante apparentée au processus d'indexation (analyse de contenu et traduction en langage documentaire) permet d'identifier les facettes de contenu utilisées pour la représentation multidimensionnelle et l'indexation du sujet d'un document en fonction des besoins particuliers de l'entreprise.

**Les facettes de contexte** décrivent les liens entre le contenu d'un document et son environnement administratif ou opérationnel soit les fonctions et activités de gestion et d'exploitation. Les facettes de contexte retenues sont : (1) Fonction, (2) Activité, (3) Type de contenu, (4) Rôle, (5) Poste, (6) Structure organisationnelle, (7) Acteur (Voir Tableau 2).

Facette	Description
Fonction	Ensemble coordonné d'actions poursuivant un même objectif et fournissant un résultat final bien identifié. On distingue généralement les fonctions de gestion administrative communes à l'ensemble des organisations (e.g. <i>Gestion des ressources humaines</i> ) des fonctions d'exploitation propres à la mission d'une organisation (e.g. <i>Gestion de la production</i> ).
Activité	L'activité représente le niveau le plus fin de décomposition des fonctions (e.g. <i>Recrutement du personnel enseignant</i> )
Type de contenu	Informations spécifiées et consignées sur un support physique (e.g. <i>Avis d'absence</i> )
Structure organisationnelle	Unité administrative ou organisationnelle de l'organisme (e.g. <i>Direction des ressources humaines</i> )
Poste	Ensemble de rôles et de responsabilités joués par un même acteur (e.g. <i>Directeur des ressources humaines</i> )
Rôle	Ensemble de responsabilités confiées à un ou plusieurs acteurs dans le déroulement d'une procédure (manière spécifiée d'accomplir une activité) (e.g. <i>Chef de projet</i> ).
Acteur	Personne physique qui remplit un ou plusieurs rôles (e.g. <i>Marc Tremblay</i> )

Tableau 2. Les facettes de contexte du métamodèle ISIS

Même si on peut y retrouver des fonctions et activités communes à tous les organismes (par exemple, la gestion des ressources financières ou la gestion des ressources humaines), une instance du métamodèle ISIS est spécifique à chaque organisme et reflète l'ensemble de ses activités administratives et opérationnelles. De plus, comme tout schéma de classification, ISIS doit correspondre à la réalité documentaire de l'organisme. Par conséquent, ISIS est élaboré à partir de la connaissance la plus complète possible des *activités* et des *documents* de l'organisme. Cette construction est basée sur la combinaison de deux approches : l'approche déductive (connaissances des fonctions et activités administratives) et l'approche inductive (étude de l'ensemble documentaire à organiser). Ces deux approches permettent de recueillir des données sur le fonds documentaire de l'organisme et d'identifier les classes et sous-classes qui formeront la structure interne du schéma de classification.

Les facettes de contexte sont identifiées et sélectionnées par une méthode de représentation et de décomposition fonctionnelle et structurelle. Cette méthode d'analyse descendante permet d'identifier les valeurs pertinentes aux facettes de contexte utilisées pour replacer l'ensemble des documents d'une entreprise dans son contexte de production. Les principes et règles classificatoires théoriques (par exemple, critères d'exhaustivité et d'exclusivité, garantie d'existence documentaire), les méthodes de modélisation des processus métiers (par exemple, OSSAD, UML) et les normes documentaires (par exemple, ISO 15489, ISDF) sont utilisés pour sélectionner les valeurs pertinentes et identifier les niveaux de granularité adéquats pour chacune des facettes.

Le type de contenu (ou mode de représentation conventionnelle de l'information) est une des facettes pivot au sein du métamodèle ISIS car c'est elle qui permet le plus souvent de faire le lien entre les autres facettes de

contexte et les facettes de contenu. Le type de contenu, crée des attentes chez le lecteur et impose une lecture et une utilisation canoniques, c'est-à-dire celles prévues lors de la création du document. Les types de contenu conventionnels permettent de communiquer et repérer l'information plus rapidement. Ils sont nécessaires à l'identification du document et parfois même à sa validité. Il s'agit, par exemple, d'un contrat, des données satellites capturées en application d'un programme scientifique, d'un formulaire d'enquête rempli par un répondant, d'un plan d'avant-projet de construction d'un collège, d'un relevé de compte bancaire, d'un compte rendu de réunion stratégique, etc. Le métamodèle ISIS incorpore inévitablement ce concept familier de type de contenu avec lequel les employés ont un rapport quasi-quotidien et qui apparaît comme critère de regroupement des documents dans les espaces de travail personnel.

L'utilisation d'une facette "hybride", représentant à la fois des éléments de contexte et de contenu du document, intervient, par exemple, au moment de la mise en place d'un nouveau projet, d'un nouveau comité, d'un nouveau programme ou au commencement d'une année budgétaire. Représentée par un concept d'organisation du travail apparenté à la notion de dossier en archivistique<sup>8</sup>, cette facette est utilisée comme moyen de dériver automatiquement au moment de la classification du document les valeurs des facettes de contenu associées à ce projet, comité, programme, affaire, etc. Cette facette permet d'associer et de retrouver les documents relatifs à une même affaire, à un même projet, à un même dossier.

Un des aspects les plus novateurs du métamodèle ISIS repose dans la création de relations sémantiques entre les valeurs des facettes avant même l'implantation opérationnelle du modèle dans l'interface utilisée par l'employé. Plutôt que d'avoir des listes de termes indépendantes les unes des autres, les valeurs des facettes composant le modèle sont reliées sémantiquement les unes aux autres. Ce faisant, on obtient un modèle qui se construit dynamiquement au fur et à mesure de la sélection des valeurs, on élimine dans les listes de termes proposées à l'utilisateur les valeurs non pertinentes, ce qui restreint le nombre de valeurs présentes sous chaque facette, facilite et accélère le processus de description et de classification des documents par les employés. La gestion du cycle de vie des documents devient aussi plus transparente. Par exemple, la connaissance du poste d'un employé (e.g. Directeur technique) qui est dérivée automatiquement à partir du login ou nom d'utilisateur entré dans le système et qui est associé à un acteur (e.g. John Doe) permet de présager des rôles (e.g. Chef de projet, Administrateur), des activités (e.g. Gestion des appels d'offres, Suivi et contrôle) qui sont propres à son rôle et des types de contenu (e.g. Contrat d'achat, Appel d'offres) reliés à ces activités. L'attribution automatiquement d'une règle de conservation devient également possible suite au choix des valeurs des facettes Fonction, Activité, Type de contenu et Thème, par exemple.

En mode recherche, l'utilisation des facettes de contenu permet de répondre à des besoins de partage du savoir et de retour d'expérience de type: Qui avait-on contacté pour l'achat des pièces? Qu'est-ce qui a été fait sur les systèmes d'aspiration? Quels sont les acteurs qui ont participé à la réalisation de cette étude?

La terminologie souvent trop générique utilisée dans la structure d'un schéma de classification institutionnel ne correspond pas toujours à celle qu'utilisent les employés dans leur travail. Ces derniers nomment leurs répertoires et leurs documents électroniques selon une sémantique personnelle, souvent reliée au domaine spécialisé dans lequel ils évoluent. Par conséquent, sous chaque facette le vocabulaire est contrôlé par un service de thésaurus monolingue ou multilingue permettant à la fois un contrôle et une flexibilité terminologiques. Selon leur nombre, les termes peuvent être regroupés à l'intérieur d'une petite taxinomie sous chaque facette.

La maintenance et la mise à jour du modèle à facettes sont assurées par un service de gouvernance. L'ajout de nouvelles valeurs associées aux facettes ou de nouvelles relations entre les valeurs des facettes est réalisé grâce à un mode semi-participatif de mise à jour du modèle. Autrement dit, la proposition de nouvelles valeurs et de nouvelles relations par l'employé lui-même est supervisée par un service centralisé de gouvernance assuré par exemple, par un architecte de l'information. La figure 5 illustre les différentes composantes de l'architecture ISIS.

---

<sup>8</sup> Mas, S. & Gagnon-Arguin, L. (2003-2004). Pour un approfondissement de la « notion » de dossier dans la gestion de l'information organique et consignée d'une organisation. *Archives*, 35 (1-2), p.29-48.

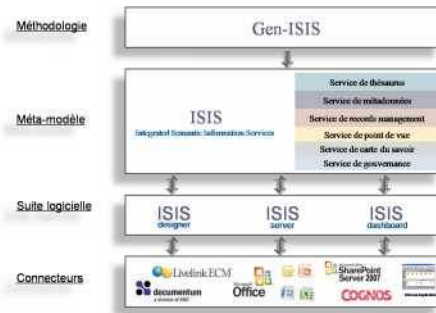


Figure 5. Architecture ISIS

(Cliquez sur l'image pour la voir en pleine taille)

Le métamodèle ISIS constitue une source de métadonnées descriptives d'un document puisqu'il en décrit le contenu conceptuel et documente son contexte de production et d'utilisation. Ces métadonnées sont encapsulées dans le document et dans la solution logicielle, indépendamment de toute plateforme utilisée pour la GED (e.g. Documentum, Livelink, Hummingbird). Le but d'une telle démarche est de proposer un langage classificatoire post-coordonné logique, référentiel et stable à travers le temps, comportant un nombre limité de facettes prédéfinies, et permettant une classification à la fois multidimensionnelle mais aussi consensuelle et standardisée de l'ensemble des documents au sein d'une même entreprise par des acteurs appartenant à un même métier.

## ■ Retombées attendues

Les moyens utilisés par les entreprises pour gérer leurs documents reposent largement sur la mise à disposition de moteurs de recherche, de langages documentaires et d'instruments de gestion tels que les schémas de classification, les thésaurus et les calendriers de conservation. Malheureusement beaucoup de ces outils, qui ont été conçus pour faciliter des activités de gestion documentaire spécifiques, sont le plus souvent redondants, insatisfaisants et difficiles à gérer dans un contexte de décentralisation et de démocratisation des activités de création, d'organisation, de conservation et de diffusion des documents électroniques. ISIS propose un nouveau modèle intégré basé sur le principe de classification par facettes permettant de créer une infrastructure sémantique unique qui regroupe et intègre la fonctionnalité de plusieurs outils de gestion documentaire tels que les plans de classification, les thésaurus, les normes de métadonnées ainsi que les taxinomies. Avec ce nouveau modèle il est possible de mieux décrire à la fois le contenu l'objet informationnel et le contexte dans lequel il a été créé, d'améliorer la précision de la recherche, de gérer l'information à travers son cycle de vie et aussi d'améliorer l'automatisation du processus de classification et d'organisation de l'information, délestant ainsi l'utilisateur d'une tâche qu'il accomplit rarement très bien.

Sur le plan méthodologique, un des enjeux les plus cruciaux pour l'exploitation des modèles à facettes tient à la phase de conception des facettes. Les classifications à facettes sont conçues lors de la phase initiale d'un projet par des professionnels de l'information s'appuyant à la fois sur des bonnes pratiques reconnues (facettes dites de contenu) et sur des analyses, représentations et décompositions des processus métiers et administratifs de l'organisation (facettes dites de contexte). Un des points forts du modèle ISIS est, quant à lui, de permettre la co-construction par les employés eux-mêmes (i.e. sans passer par les professionnels de l'information qui jouent plutôt un rôle d'encadrement, de consolidation et de gouvernance), de schémas de classification adaptés à leurs problématiques. Pour la conception d'une classification institutionnelle multi-facettes, les méthodes de *record management* et de modélisation des processus d'affaires qui sont envisagées sont principalement de trois types :

1. Méthode de représentation et de décomposition fonctionnelle et structurelle : cette méthode d'analyse descendante permet d'identifier les valeurs pertinentes aux facettes de contexte (par exemple, Fonction, Activités, Position, Rôle) utilisées pour replacer l'ensemble des documents d'une entreprise dans son contexte de production. Les principes et règles classificatoires théoriques (par exemple, critères d'exhaustivité et d'exclusivité, garantie d'existence documentaire), les méthodes de modélisation des processus métiers (par exemple, OSSAD, UML) et les normes documentaires (par exemple, ISO 15489, ISDF) seront utilisés pour

sélectionner les valeurs pertinentes et identifier les niveaux de granularité adéquats pour chacune des facettes.

2. Méthode d'analyse documentaire : cette méthode d'analyse ascendante apparentée au processus d'indexation (analyse de contenu et traduction en langage documentaire) permet d'identifier les facettes de contenu (par exemple, Lieu, Entité, Thème) utilisées pour la représentation multidimensionnelle et l'indexation du sujet d'un document en fonction des besoins particuliers de l'entreprise.

3. Méthode de contrôle terminologique : cette méthode permet d'identifier et de valider les termes retenus comme valeurs de facettes et les termes rejetés, équivalents et associés. Les relations sémantiques (par exemple, hiérarchique, associative) entre les termes sont également décrites à l'intérieur d'un langage ou vocabulaire contrôlé de type Thesaurus qui permettra de gérer avec flexibilité le vocabulaire.

En ce qui concerne les retombées théoriques, le projet ISIS permet d'explorer l'utilisation d'un modèle classificatoire à facettes par des employés dans un contexte de travail et de valider empiriquement le modèle à facettes proposé grâce au développement et à l'implantation d'une solution logicielle.

C'est dans ce contexte que Bibliothèque et Archives Canada a entrepris un projet pilote qui utilise l'approche ISIS pour établir les bases d'une approche plus efficace pour la gestion des documents et des dossiers au sein du Gouvernement Canadien. Ce projet permettra à la fois de faciliter le travail des utilisateurs en leur proposant des services de classement et de création de métadonnées semi-automatiques ainsi que celui des archivistes qui bénéficieront ultimement d'une meilleure description du contexte et du contenu des objets informationnels. Plusieurs autres ministères du gouvernement Canadien en sont à différentes phases de l'implantation du modèle ISIS dans leur organisation.

# ARCHIVES AND THE ARCHITECTURE OF INFORMATION SOCIETY

---

Markku Nenonen

Deputy Director General, National Archives National Archives of Finland

*Records schedules are the source of the essential metadata which guides the life cycle management and handling rules belonging to tasks and workflows of an organization. Its structure is based on classification framework of the tasks, handling processes and record types of the organization. Records schedules and long-term preservation are the possible tools to achieve the goals of Information Society*

*Records management has to be seen from the viewpoint of architecture of public administration and it is a tool for the better data management in producing public services. The Draft of the Architecture launched by the ministry of Finance consists of four different interoperable levels: tasks, data, information systems and technology. National Archives of Finland has developed centralized architecture of classification framework and records schedules system which serves the whole administration in Finland. Architecture consists of systems which provide records schedules and handling rules and operative systems which provide services for the citizens and digital preservation system. From the Archives perspective the most important levels of the architecture are the levels of data and information systems. To define the data level, however, you have to identify the needs of the first level, tasks.*

*The importance of records management and long-term preservation in the field of eGovernment is to boost co-operation and reform public administration structures in order to ensure the availability and quality of public services. One big question is how records management and long-term preservation can influence on the productivity of the public sector. From this point the municipal sector of Finland has a key role at the moment. The classification framework of the tasks, handling processes and record types of the organization, could have a significant role. There is right now a huge possibility to have a joint electronic archives system for the whole municipal sector.*

*As a result of this framework development National Archives of Finland can present effective and seamless life cycle of electronic records which guarantees trustworthy and reliable long-term preservation in electronic format. And we can have more efficient data management to improve the service processes so that the customer can have service required in much shorter time and have better services.*

[See the presentation.](#)



## **Les archives et l'architecture de la société de l'information**

*Les tableaux de gestion (T.D.G.) sont la source des métadonnées essentielles qui guident la gestion du cycle de vie et les règles de manipulation afférentes aux tâches et aux flux de travail d'une entité. Leur structure est basée sur une classification des tâches, des processus de manipulation et des modes d'archivage de l'entité. Les T.D.G. et la conservation à long terme sont les outils permettant d'atteindre les objectifs de la société de l'information.*

*Le records management doit tenir compte de l'architecture de l'administration publique. Il constitue un outil permettant une meilleure gestion des données au sein des services publics producteurs. Les Archives nationales de Finlande ont développé une architecture centralisée pour un cadre de classification et un système de T.D.G. qui est appliquée à l'ensemble de l'administration en Finlande. Cette architecture est composée de systèmes fournissant des échéanciers et des règles de manipulation de l'archivage, et de systèmes opérationnels offrant des services destinés aux citoyens et au système de conservation numérique.*

*En matière de records management et de conservation à long terme dans le domaine des services publics en ligne, il est important d'encourager la coopération et de réformer les structures de l'administration publique afin de garantir la disponibilité et la qualité des services publics. La question majeure est de savoir dans quelle mesure le records management et la conservation à long terme peuvent avoir un impact sur la productivité du secteur public.*

*Grâce au développement de ce cadre, les Archives nationales de Finlande peuvent présenter un cycle de vie efficace et sans coupure des archives électroniques garantissant une conservation fiable et à long terme sous un format électronique. La gestion des données est alors plus efficace pour améliorer les processus de service. Le client se voit ainsi offrir des services beaucoup plus rapides et de meilleure qualité.*

[Consulter la présentation.](#)

## ■ Strategic approach

The position and tasks of the Finnish National Archives as part of the information society must be approached from the viewpoint of the strategic objectives of the society and the information society in particular. They determine the operating environment and the objectives that the National Archives must be able to achieve.

In Finland, the information society has been specified, and objectives have been set in three major strategies. Only the main outlines of these strategies are mentioned here, particularly from the viewpoint of the National Archives.

### **National Knowledge Society Strategy 2007–2015**

Two key aspects can be highlighted in the strategy: the development of service provision, and the improvement of productivity and the quality of customer service. The objectives are pursued by making the operations of the public administration, in particular, more effective. The key means and method here are the opportunities offered by information communications technology, that is, a more efficient utilisation of ICT.

### **State administration IT strategy 2006**

The aim is to identify the state administration's shared needs and operations areas that can be efficiently supported with IT. Based on these needs, shared IT systems and services will be produced and adopted in the state administration, and all governmental units will be able to use them.

### **State administration Productivity Plan**

The aim is to cut approximately 15,000 state administration jobs by 2015. This can be achieved by utilising the retirement of baby boomers, which is very rapid in Finland: the most rapid in Europe.

In its strategy and implementation plan, the National Archives has to analyse the strategies mentioned above and take them into account so that the National Archives is able to allocate its own tasks so that it can contribute to the well-being of Finnish society. The situation is favourable for the National Archives, since the new strategy work by the National Archives has been launched, and a new strategy will be completed in the autumn of 2009.

## ■ Architecture of the information society

Comprehensive architecture should be regarded as a good tool in the development of organisation operations and information systems. The architecture enables the establishment of a shared operating environment for both those responsible for the operations and for IT developers. In Finland, comprehensive architecture has been seen as a good method of combining IT tasks in the state administration. It has been regarded as appropriate to build its own, light, but adequately comprehensive and efficient architecture to support operations development.

In Finland, the corporate sector in particular has used the architecture framework to a great extent, but there are not many experiences regarding the use of it as a tool in the development of public administration. From the viewpoint of the information society, architecture can be divided into four different levels: operations and processes, information, information systems, and technology. The relationships between the levels can be seen as follows: Operations and the processes supporting their objectives constitute the highest level. All organisations have their own objectives (ultimately specified in applicable laws) for the implementation of which the organisations have developed processes of their own. The task of operations and related support processes is to create results and output for their own customers and target groups. For the

follow-up of the results and output, indicators are built, showing how efficient and profitable the organisation's operations are.

The information level, information architecture, specifies the information required by the operations and processes. Information is the most essential raw material for operations and processes. In Finland, the specification of the information architecture is very much at a start-up stage, and the development owner, for example, has not been clearly determined. It is also obvious that the evaluation of the significance of the information architecture still needs a reasonable amount of work.

The basic task of information systems is to help operations and processes to, as efficiently as possible, obtain and process the information required by the processes. The most essential objective of information systems is to support the organisation's information processing and information management.

First and foremost, technology shows as standards. Standards are used when building and implementing information systems. The starting point is to use primarily international and, of course, national standards as extensively as possible.

The comprehensive architecture of the information society can thus be described as follows: when it comes to achieving the operational objectives of an organisation, information is the essential resource, and its processing is supported with information systems based on general standards. The efficiency and profitability of the operations are determined by the organisation's operational results and outputs which are evaluated using indicators approved of by the customer.

## ■ The information architecture levels and archives – information level

As mentioned above, information is the essential raw material and method when it comes to achieving operational objectives. It is also a level for which Finland at least does not yet have an owner. The National Archives has traditionally been a strong information possessor and administrator. Traditionally, information is in records, which, however, should be reviewed on the basis of their content instead of their format. The National Archives has very long-standing traditions in administration and service provision information management in particular. It would be fair to say that in the last twenty years, information administration and information technology have come increasingly closer to the traditional area of operations of the National Archives.

Of course, it is important to understand how information architecture is placed in the comprehensive architecture, that is, who and what set requirements for whom and for what. Operations and results are the goal of everything, and this clearly sets requirements for information architecture which is supported by information systems and sets information management/processing requirements for them.

The question is what essential requirements the information architecture (information level) sets for the IT level. At least the three following requirements can be regarded as the starting point: records' life-cycle management, task and process classification, and metadata. In connection with the implementation of the information architecture, it has to be determined how specific the requirements mentioned above should be, and, of course, what the requirements signify in practice, in the implementation of the information systems.

## ■ SÄHKE 1 and related applications

In Finland, the National Archives has addressed information architecture requirement specifications with the SÄHKE standard, which has been in use since the beginning of 2006. SÄHKE 1 was valid for two years until the end of 2008. The National Archives approved the new SÄHKE2 standard at the end of December 2008, which took effect at the beginning of 2009.

SÄHKE 1 set requirements for the information systems in the records administration. SÄHKE sets requirements for those organisations which want to maintain their permanently stored records in electronic

format. The National Archives grants permissions for storing electronic material. In 2008, the permission was granted to two government agencies: The Finnish Rail Agency and the National Land Survey of Finland. Several new applications will be received in 2009. From now on, meeting SÄHKE requirements enables records transfer to the National Archives, that is, at the stage when the National Archives is able to provide an information system for permanent electronic storage. According to the existing plans, the information system (VAPA) will be ready for operation by the end of 2010. At the moment, organisations store their electronic records in their own information systems. From now on, it has to be separately evaluated whether the organisations store their records in their own information systems, or whether they are transferred to VAPA. The key objective is that the organisation's information can be used and utilised as extensively as possible, which means that the information system for archival has to be as open as possible.

The National Archives is renewing the permission for electronic archival so that the SÄHKE2 requirements will be taken into account. It is clear that also the SÄHKE1 requirements are valid during the transition period, that is, until the end of 2011. Naturally, the VAPA system will be based on the SÄHKE2 standard.

## ■ SÄHKE2 basics

The SÄHKE2 standard is based on the electronic management of the entire life-cycle of records, and the standard also covers other than information systems for records management. An electronic archival establishment plan is the steering method for electronic life-cycle management, and the Finnish municipal sector in particular refers to the plan as an information steering plan. Life-cycle management starts off with the creation of records in the information system producing them, that is, at the initial stage of the organisation's processes. Life-cycle management extends to permanent electronic archival of the records, or their termination. The plan covers and particularly emphasises the reusability of records in electronic archives for organisations' service operations and, of course, for research activities.

The electronic archives establishment plan includes task (process) classification which has been taken into account in SÄHKE2 and which is a precondition for utilising the VAPA service.

The archives establishment plan steers the establishment of the metadata concerning the information required by the organisation's processes so that the metadata flows throughout the document life-cycle and is supplemented when necessary. Metadata enables the reuse of records and information. It is essential to ensure that the metadata linked to documents is consistent and is linked to the documents in the document creation process, and that metadata can be automatically linked to the documents, that is, the information systems automatically produce metadata. In Finland, this package has been described with the concept of data steering which is based on information planning (electronic archive establishment plan). The starting point is that the National Archives specifies the metadata to be used, and the information system providers have taken into account the features of data steering in their respective information systems.

SÄHKE2 and the VAPA system are based on the OAIS model. However, the aim has been to take the model closer to the start of the process through data steering planning and a data steering system. This could perhaps be described in a better way by using the OAIS model terminology so that the National Archives aims to specify the pre-ingest stage and tries to actively impact the establishment of the SIP package. These are themes which many presentations at the DLM Forum in Toulouse addressed.

An electronic data steering system supporting records' life-cycle management also provides a view on records stored for a fixed-period, and not only on records archived on a permanent basis. The requirement specifications of the SÄHKE standard and the VAPA have also taken into account the entire document life-cycle management, covering also records stored for a fixed period. At this stage, however, it has not yet been outlined and decided how the electronic management of records stored for a fixed period will be carried out in Finland. In a simplified manner, it can be said that the archival period of a document is only one piece of metadata; that is, metadata management can enable the electronic screening and termination of records.

The electronic archival of databases is a further core challenge for the VAPA system, and this theme was strongly highlighted in Toulouse, too. In Finland, investigations concerning the theme have been started in a National Archives' project launched in the beginning of 2009.

From the viewpoint of the information society architecture, SÄHKE2 addresses information management questions by creating principles on metadata management. This is to say that SÄHKE2 clearly involves questions highlighted primarily in information architecture, but has approached the level of operations and processes in a clearer way than SÄHKE1. This is particularly emphasised in data steering requirements. In addition, SÄHKE2, from the viewpoint of metadata management, clearly sets requirements for information systems. However, SÄHKE2 does not set as high requirements for information systems as Moreq2 does. Generally speaking, it could be said that SÄHKE2 is closer to Moreq2 than SÄHKE1 was to Moreq1. It is likely that in Finland, the SÄHKE2-Moreq2 relationship will be evaluated and described further in the Moreq2 0 chapter, whose implementation has not yet been decided.

SÄHKE2 thus acts as a basis for electronic archival (VAPA). In this connection, it is also important to note VAPA's links to European-level development through other channels besides Moreq2. The objective of the European Digital Library Project, currently known as Europeana, is to make the collections of European libraries, archives, and museums more accessible for citizens. The first version of Europeana was published in late 2008. In Finland, the objectives of the Europeana are implemented in the National Library Project launched in the summer of 2008 by the Finnish Ministry of Education. The project involves all of the essential memory organisations in Finland. The project has two parts: building a customer interface, and long-term archival (LTA). The National Archives has been assigned to carry out the requirement specifications for the latter part.

Eventually, all electronic data produced by memory organisations, intended for long-term archival, will be transferred to the long-term archival data system. Every data set has its special features, but it could be justified to say that the entire data ultimately consists of binary digits. The data must have adequate metadata so that the data can be found and reused.

Electronic records constitute one part of the electronic long-term archival, which is why it is clear that the VAPA and the LTA have obvious linkages. In fact, it can be seen that the National Archives' SÄHKE specifications, the electronic archives establishment plan and the VAPA have also a more extensive significance in building the Finnish information society.

In the implementation of VAPA, there still remain many matters requiring outlines and decisions, but the aim is to address the matters during the spring of 2009. The unanswered questions include: Is the deployment of VAPA based on legislation? At the moment, an electronic archive concerning healthcare is being implemented, based on the legislation which says that healthcare units must hand over their data to said electronic archive. Does the VAPA also cover the municipal sector; that is, would it be an electronic data resource shared by the entire public sector in Finland? In the middle of the first decade of the 21<sup>st</sup> century, strong definitions of policy have been made in Finland concerning the cooperation of the entire public sector and shared IT solutions. VAPA could be the flagship of this development. Records' transfer cycle to the electronic archives must be specified. In Finland, printed material is transferred to the National Archives when the material is forty years old. It is clear that the same cycle cannot be used in the electronic field. It also has to be solved how the administration and maintenance of the information system and related technology will be carried out. It is obvious that the National Archives is not an IT house. It does not have the necessary know-how, and it is not worthwhile for it to acquire the know-how. Instead, IT solutions implemented in the state administration have to be utilised. Also the service offering in the private sector has to be taken into account. Here, as in other cases, the ordering expertise of the owner of the matter (National Archives) is emphasised.

## ■ The question of certification

At the moment, the National Archives grants permissions for the electronic archival of data stored on a permanent basis. As mentioned above, the National Archives granted two such permissions in 2008. The permission is expressly granted to an organisation, not to an individual information system or to an IT provider, for example. It could then be said that the National Archives particularly evaluates the operations of the organisation, including how the organisation utilises the information system it has. As such, the use of a particular information system does not mean that the system would be used in an appropriate way. From the viewpoint of the information society architecture, it is also about the topmost level, that is, the level of operations and processes, not about the third level, that is, the information system. As far as I know, Moreq has set out from information systems certification in particular. It is important to note that in the National

Archives' certification, the second level -- information -- is the level that connects operations with the information system.

I am not saying that said certification system of the National Archives would be comprehensive or watertight. Operations has many gaps and matters and questions to be answered. Of course, the basic question deals with how matters in general should be certified. One starting point - which is perhaps the easiest - would be information-level certification based on a shared information architecture in the public administration. Of course, it is obvious that the National Archives cannot - and could anyone - certify the operations and processes of a certain organisation; and if it is possible, what is the level? On the other hand, the question is about whether the certification of information systems is adequate, which means that it is essential how the information system supports information management/processing and ultimately the operations of the organisation. It would be justified to say that the lowest level of the architecture, that is, the technologies, and the standards in particular, have already been certified, which means that there already is a certification system for the lowest level. It could be assumed that the next certification level would be the information system, which would mean that the progress would take place from the easiest level towards the most difficult one.

It is a different thing who certifies what and how much resources can and should be allocated for certification. In Finland, this will soon be topical in connection with the implementation of a national archive in healthcare. A related discussion deals with patient data systems that deliver information to the centralised archive.

In Finland, the role and tasks of the National Archives are entirely undefined in the certification dialogue. The first question to be answered is to determine the architecture level of the National Archives, and what is its core know-how. Or is it that certification and the architecture levels are topics that do not belong to the National Archives. Personally, I find that the starting point would be what has already been said earlier: The National Archives is an experienced and skilled possessor and manager of information. Experience obtained from records administration is a good basis for information architecture development.

## ■ Challenges

I have above described the strategic policy definitions and objectives of the Finnish information society, and the possible role and tasks of the National Archives in the implementation of the strategy. From the viewpoint of the National Archives, the objective and the operating model are still mainly theoretical, and there are yet many unanswered questions and challenges ahead both within the National Archives and in the operating environment.

In Finland, SÄHKE1 is widely approved, but mainly from the perspective of long-term archival. However, the challenges of the information society provide SÄHKE2 with opportunities: the Finnish information society is also looking for new solutions with substance and effectiveness. Records administration and decision-making processes are the basic pillars of Finnish public administration, and the objectives are shared by all. Document life-cycle management constitutes a major challenge: as mentioned above, data steering (archives establishment planning) extends to the root of the process. At least in Finland, the common view has been that the task of the archives sector is to manage and administer data which is stored on a permanent basis, that is, to focus on the end part of document life-cycle management and nothing else. If and when the National Archives, for a justified reason, goes closer to the start of the process, it will be evident that it steps on somebody else's toes. This is what has happened to the National Archives in Finland.

The biggest challenge for the National Archives is to find strategic partners who understand said package. The document management package, document life-cycle management, is clearly a part of administration development. In Finland, the development of not only the state administration, but the entire public administration, has been centralised to the Ministry of Finance. From the viewpoint of electronic document management, it is clear that the Ministry of Finance is the most important strategic partner for the National Archives. A viewpoint of its own to the whole is introduced by the fact that administratively the National Archives is under the steerage of the Ministry of Education.

The municipal sector, which is very highly autonomous in Finland, is a question of its own. However, the municipal sector has clearly communicated that shared IT systems are not a threat to the autonomy of the

municipalities, and the situation is the opposite if anything. By carrying out cooperation and acquiring shared IT systems, municipalities will have more time to take care of their core operations, that is, to provide services for the municipalities. The municipal sector has been very actively involved in the development of electronic data steering planning, and is adopting a shared task classification. The starting point for the implementation of the VAPA project has been that it would serve as a shared solution for the entire public administration. But, official decisions and policy definitions are missing at the moment.

## ■ The information society and the National Archives

In Finland, the National Archives has the opportunity to be an active and important operator when securing the well-being of Finnish society. The operations of all organisations are based on applicable legislation, and this is the case for the National Archives as well. Finland has a separate Act on Archives, but also a number of other laws which have an impact on the implementation of the information society. A feature shared by almost all of the laws is that they have been enacted in a situation where an approach such as the information society has not existed, and where such an approach has not been possible. On a more general level, it could be said that the legislation has been prepared in a situation where the essential policy definition of the information society, that is, the state administration, has not been seen, and the entire public administration can and should be regarded as a more extensive package where boundaries disappear. At the moment, the Finnish administration is facing a massive change and transition, and the previous boundaries are being removed at an increasingly quicker pace: the transition has been shown and felt in both the state administration and the municipal sector. At the moment, the sectors are examined from their respective starting points, but it is presumable that the next stage will involve a comprehensive review of the public administration, which means that changes are to be expected.

A legislation reform and change is a big and time-consuming job, and it would be justified to ask whether there is time for this, particularly due to the retirement of baby boomers. Surely the quicker way is to interpret the existing laws in a new way. The interpretation of the existing Archives Act, for example, enables the provision of the VAPA service for the municipal sector if this is desired.

The National Archives has started to consider matters from the viewpoint of the information society architecture, particularly from the viewpoint of information architecture. There are ideas and plans concerning the expansion of the task classification, and the archives establishment plan to the direction of ontologies and a semantic Web. Information life-cycle management and related metadata obviously provide a reasonably good basis for specifying the information architecture.

As I see it, the National Archives can contribute to the information society, but it has to find its role and core expertise. It is often argued that with the existing resources, the National Archives is totally unable to carry out even part of the tasks described above. This surely is the case. But, the starting point has to be the specification of a new electronic package, dividing it into sub-areas and looking for solution models and operators for them. The Finnish public administration has shown its agility in many matters, including here, which means that it will point out the necessary resources for an operator that is able to produce the required solutions.

## ■ Advantages

Finally, it is good to note and remember why all of the development and operations described above should be carried out, what the intended advantages are, and what the achievements will be.

In the Finnish administration, an open decision-making process is an essential means for a well-functioning democracy, and the process is used in every public administration organisation. First and foremost, the decision-making process is based on information, its processing and management. The question is how to manage this package as efficiently as possible with the primary aim of providing citizens with high-quality services within a reasonable time. Behind the decision-making processes, there is an administrative system which has to function as efficiently as possible in order to achieve the objective in question. The essential

objective in Finland is that "helping hands", that is, human labour, could from now on be allocated for healthcare services above all.

Information processing and management is key in meeting the objectives described above. The National Archives has some tools for achieving these objectives! However, they still need to be developed further, and, first and foremost, it must be ensured that they are deployed!



# PRESERVING EUROPEAN SECURITY AND DEFENCE POLICY MISSION RECORDS

---

Stephen Howard

Chief Registry and Archives, EULEX-Kosovo

The European Union has recently launched an ambitious Rule of Law mission in Kosovo (EULEX). Over 3000 staff will be deployed, making it by far the largest European Security & Defence Policy (ESDP) mission to date. For further information about ESDP missions, see [http://en.wikipedia.org/wiki/Overseas\\_interventions\\_of\\_the\\_European\\_Union](http://en.wikipedia.org/wiki/Overseas_interventions_of_the_European_Union)

The rapid build-up of the EULEX mission, its unique role, and its limited duration present many information management challenges. A Registry & Archives Expert was thus recruited by the European Union Planning Team in Kosovo at an early stage to assist in resolving these issues.

This paper provides an update on initiatives to capture and preserve mission records six months after the formal launch of EULEX on June 15<sup>th</sup> 2008. It also raises the visibility of a new type of European organization that until now has not benefitted from the archival and records management wisdom of the participating member states. It will cover the following items over the allocated thirty minute slot.

- **Background to the ESDP mission in Kosovo** (mission mandate, deployment timetable, exit plan & organizational structure)
- **Establishing Information Governance in a Rule of Law mission** (identifying the regulatory framework, security classification, relevant records and information standards, obtaining legal advice, networking within EU archival institutions, developing Standard Operating Procedures, borrowing from established UN mission RIM standards)
- **Results of initial information survey** (record legacy from previous missions, paper records, initial electronic records user behaviour)
- **Development of an information management strategy** (in line with mission objectives, recognising unique factors such as transfer of UNMIK records, in collaboration with ICT and Public Information functions)
- **Recruitment of Archive & Registry Office team** (hiring international & local records staff to former conflict zones)
- **Design of Archive & Registry Service Level Agreement** (electronic and manual services, managing shared servers, identification of suitable records storage areas)
- **Development and delivery of training programme** (gaining a slot on the mandatory mission induction programme, identifying document management officers within each component for further training)
- **Building and implementing a Mission File Plan** (with a finite mission life, organisational structure vs. functional approach?)
- **Exploiting ICT opportunities** (Office 2003, Sharepoint vs. Xerox DocuShare)
- **Lessons learned** (what we would do differently, what future missions need)

This paper is in accordance with the conference topic of developing strategies for records management and archival services and demonstrates the work undertaken to provide a records management template for future ESDP missions.

[See the presentation.](#)

## **Préservation des documents d'une mission dans le cadre de la politique européenne de sécurité et de défense**

*L'Union Européenne a lancé récemment une ambitieuse mission « Etat de droit » au Kosovo (EULEX). Plus de 3000 hommes y seront déployés, ce qui en fait de loin la plus importante mission de politique européenne de sécurité et de défense (PESD) à ce jour. Pour plus de renseignements sur les missions PESD voir le site: [http://fr.wikipedia.org/wiki/Politique\\_européenne\\_de\\_sécurité\\_et\\_de\\_défense](http://fr.wikipedia.org/wiki/Politique_européenne_de_sécurité_et_de_défense)*

*La rapidité de la mise sur pied, le rôle unique et la durée limitée de la mission EULEX génèrent un certain nombre de défis en matière de gestion des informations. Dans un premier temps, un expert des registres et archives a donc été recruté par l'Équipe de Planification de l'Union Européenne pour le Kosovo pour travailler à la résolution de ces problèmes.*

*Cet article dresse un état des lieux des initiatives réalisées en matière de capture et de préservation des informations de la mission EULEX six mois après son lancement officiel le 15 juin 2008. Il apporte aussi une plus grande visibilité sur une organisation européenne d'un type nouveau qui n'avait pas jusque-là bénéficié des connaissances des Etats membres participants en gestion documentaire et archivistique. Il traitera dans les trente minutes imparties des thèmes suivants :*

- **Contexte de la mission ESDP au Kosovo** (mandat de mission, calendrier de déploiement, plan de sortie et structure organisationnelle)
- **Mise en place de la gouvernance informatique dans une mission « Etat de droit »** (identification du cadre réglementaire, classification de sécurité, normes applicables aux informations et aux documents, conseil juridique, mise en réseau des institutions d'archives de l'UE, développement de procédures opératoires standard, emprunts de normes de gestion documentaire et archivistique en vigueur dans les missions de l'ONU)
- **Résultats d'une première enquête d'information** (legs documentaire des missions précédentes, documents papier, comportement des utilisateurs des premiers documents numériques)
- **Développement d'une stratégie de gestion des informations** (en adéquation avec les objectifs de la mission, reconnaissant la singularité de certains éléments comme le transfert de documents de la MINUK et utilisant les TIC et les services d'information publique)
- **Recrutement d'une équipe administrative pour la gestion des archives et des registres** (embauche de personnel d'archives international et local dans les anciennes zones de conflit)
- **Elaboration d'un contrat de niveau de service pour les registres et archives** (services électroniques et manuels, gestion de serveurs partagés, identification de zones de stockage adaptées)
- **Conception et conduite d'un programme de formation** (obtention d'un créneau horaire sur le programme obligatoire de formation à la mission, identification des agents de gestion documentaire sur chaque élément du programme à des fins de formation complémentaire)
- **Création et mise en oeuvre d'un Plan de fichiers mission** (mission limitée dans le temps, comparaison structure organisationnelle/approche fonctionnelle?)
- **Exploitation des potentiels des TIC** (Office 2003, comparaison Sharepoint / Xerox DocuShare)
- **Leçons à tirer de cette expérience** (ce que nous pourrions changer, besoins pour les futures missions).

*L'article proposé se rattache au thème de la conférence consacré au développement de stratégies pour les services de gestion documentaire et archivistique et présente le travail réalisé pour développer un modèle de gestion documentaire pour les futures missions ESDP.*

[Consulter la présentation.](#)

The Council of the European Union has recently launched an ambitious rule of law mission in Kosovo (EULEX). Over 3000 staff will soon be deployed, making it by far the largest European Security and Defence Policy (ESDP) mission to date. The rapid build-up of the EULEX mission, its unique role, and its limited duration present many information management challenges. This paper provides a brief update on initiatives to establish an information governance framework and tactical measures to capture and preserve mission records since the formal launch of EULEX-Kosovo in June 2008. This paper also raises the visibility of a new type of European organization that until now has not benefitted from the archival and records management wisdom of the participating member states and demonstrates the work undertaken to provide a records and information management template for future ESDP missions.

## ■ Background to the ESDP mission in Kosovo

*“A majority of European citizens, and an even larger majority of people elsewhere in the world, have probably never even heard of ESDP”*, Jolyon Howorth<sup>1</sup>.

### Context

Since 2003, the EU member states have increasingly coordinated and integrated their security and defence strategies in the form of overseas missions of a civilian and military nature. The Civilian Planning and Conduct Capability (CPCC) is the permanent structure with operational responsibility for the planning and activities of civilian ESDP missions. The CPCC is under the political control and strategic direction of the Political and Security Committee and the overall authority of Javier Solana, the High Representative of the EU Common Foreign and Security Policy.<sup>2</sup>

The primary purpose of these missions has been to mitigate and contain risk in crisis situations within the sphere of European interest. More than twenty missions have been launched to date over a surprisingly large geographical area, within Africa, the Balkans and south Caucuses, the Middle East, and Asia. These can be categorized as police missions; border control and military/technical assistance missions; peace monitoring and judicial training missions; and military missions. Most of these missions have been relatively small, with between ten and several hundred staff.

### Mandate and objectives

The EU has been active in Kosovo for many years. The EU Pillar was a founding component of the United Nations Interim Administration Mission in Kosovo (UNMIK)<sup>3</sup>. After facilitating basic reconstruction following the 1999 conflict, the EU Pillar supported the economic development of Kosovo in line with its European perspective.<sup>4</sup> The Council decided in April 2006 to establish a European Union Planning Team (EUPT) of approximately one hundred staff to prepare for a possible future crisis management operation in the field of rule of law in Kosovo.

In February 2008 the EU Council decided to launch EULEX with an initial mandate of two years. It will achieve initial operational capability in December 2008. Meanwhile UNMIK will continue to exercise its executive authority under UN Security Council Resolution 1244. EULEX will not replace UNMIK but rather support, mentor, monitor and advise the Kosovo authorities on all areas relating to the rule of law, in particular the police, judiciary, customs and correctional services. EULEX will also exercise limited executive responsibilities in specific areas of competence, for example in the investigation and prosecution of serious and sensitive crimes.<sup>5</sup>

---

<sup>1</sup> Jolyon Howorth; “Security and Defence Policy in the European Union”; 2007; Palgrave Macmillan; p2.

<sup>2</sup> More information on ESDP is available via <http://www.consilium.europa.eu/esdp>

<sup>3</sup> More information on UNMIK is available via <http://www.unmikonline.org>

<sup>4</sup> More information on the EU Pillar is available via <http://www.euinkosovo.org/>

<sup>5</sup> More information on EULEX-KOSOVO is available via <http://www.eulex-kosovo.eu>

## ■ Benefits of managing ESDP mission records

The generic benefits of more effective records and information management are well known and apply also to ESDP missions.

- Supporting the transformation to electronic delivery of more economic, efficient and effective functions
- Providing reliable data for the management of property, finance, human resources, organisational performance and risk
- Ensuring compliance with legal requirements for maintaining privacy, security, confidentiality, authenticity and integrity of information and protecting the mission and its employees against litigation
- Increasing the mission's accountability by making decisions transparent to stakeholders through effective record-keeping and improved communications
- Maintaining and exploiting corporate knowledge over time, establishing lessons learned and assisting with research, development and innovation
- Facilitating Europe-wide partnership working through the appropriate sharing of information to agreed standards
- Safeguarding the mission's vital records and supporting business continuity
- Supporting the long-term preservation of the mission's archival records and the Council's institutional memory
- Reducing administrative costs by streamlining the handling of information, reducing time lost in retrieval and reworking
- Minimising accommodation and information storage costs by ensuring the timely and secure disposal of expired manual and electronic records

## ■ Information management challenges

### Local conditions

The infrastructure, climate, economy, and general levels of education within Kosovo present particular challenges to the information manager. The domestic generation of electricity currently relies upon outdated equipment, resulting in frequent power failures. Even when organizations have deployed uninterruptible power supplies at each level of their IT infrastructure, a power failure suffered by a third party (e.g. Internet Service Provider) can interrupt business continuity and cause the loss of electronic records. Telecommunications cables are frequently disturbed during road improvement and construction projects, resulting in further network outages.

The climate of Kosovo is predominantly continental, resulting in warm, dry summers and cold, snowy winters. The average temperature ranges from +30 °C during the summer to -10 °C in winter, although extremes of +37 °C and -26 °C are possible. The availability of quality office and warehouse premises is low in Kosovo and international missions occupy temporary accommodation with low thermal efficiency. It is thus a challenge to maintain BS5454<sup>6</sup> environmental recommendations for archival storage in these circumstances, particularly when power failures regularly interrupt air conditioning and heating systems.

In these circumstances records storage boxes that are frequently handled can fail after a few seasons of extreme weather, and finding replacement stock can be difficult. EULEX recently had to import boxes from the United Kingdom due to the poor board quality of local alternatives, despite an extensive search of the local market. The local recruitment market is also developing, and when hiring local records and information

<sup>6</sup> A temperature range of 16-19 °C and a relative humidity level of 45-60%

management staff it can be difficult to find candidates with relevant professional qualifications. The strongest EULEX candidates had good records management experience and English language skills from previous assignments with international organizations in Kosovo.

### **General organizational character**

Crisis management missions are typically born from actual or potential conflict, and a very high degree of political uncertainty surrounds their launch and deployment. The initial mission duration is limited according to the relevant Council Joint Action, creating a short-term management focus. Attention in the early stages of mission deployment is naturally upon the activities on the “critical path”: the recruitment of international and local staff; the acquisition of assets; and the provision of general services such as medical, IT, building management, finance, logistics, transport and training to the functional departments of the mission.

Once EULEX reaches full operational capability, the mission will have around two thousand international staff (rule of law experts, police officers, judges, prosecutors, customs officials, professional administrators and managers) and approximately one thousand local staff deployed across Kosovo. Despite the rapid build-up of mission strength, and the best “can-do” attitude, there are resource gaps and departments generally have a low capacity for administrative support functions in the initial stages.

This applies also to the shortage of professional records management input on the ground. The organizational chart allows for a central Registry and Archives team of two professional international staff, and two local support staff. There are no further dedicated staff to perform recordkeeping tasks within the EULEX departments, and whilst there is general agreement amongst senior management that information governance is vital for efficiency and accountability, there are many competing demands on resources.

### **Emerging regulatory framework**

At the beginning of 2008 there was no obvious records and information management regulatory framework for EULEX and ESDP missions in general. The EULEX “concept of operations” and “operational plan” did not refer explicitly to records and information management obligations, nor did the Standing Operating Procedure on Communications and Information Technology. The EULEX Registry and Archives team contacted the EU and European Commission archives services for background information on strategies, policies and standards relating to archiving, records and information management within the EU institutions in general.

The EULEX Registry and Archives team was directed to the E-DOMEC<sup>7</sup> team at the Commission, and their excellent information governance resources which are available online. E-DOMEC standards are compulsory for the Commission’s Directorate Generals and services, but any other European institution, agency, mission or task force is welcome to use them as source of inspiration. EULEX subsequently received the advice from the Council that records of ESDP missions once closed will be stored in the Central Archives of the Council in Brussels. In September 2008 the Council’s Working Party on Information debated the public right of access to ESDP records and made a decision about responsibilities and archiving instructions for ESDP missions that will be included in a future CPCC Administrative Handbook.

### **Registration requirements**

The Central Archives of the Council recommend that ESDP missions should adopt a centralised registration system. The registration of documents in registry filing systems remains the “world’s most predominant recordkeeping system”<sup>8</sup>, however the approach is not without its critics. In Europe there has been a long tradition of maintaining document-level registration systems, and this applies also to the bodies of the European institutions. The stated purpose of registration at the Commission, for example, is to a) identify the document in question with certainty and b) certify that the document meets minimum standards and has been sent, received or created on a certain date.

Registration ensures that the document gets onto the relevant file and that appropriate staff are allocated to action a document. The process also provides an audit trail to track the document, and the resulting register

---

<sup>7</sup> More information on E-DOMEC is available via [http://ec.europa.eu/transparency/edoc\\_management/index\\_en.htm](http://ec.europa.eu/transparency/edoc_management/index_en.htm)

<sup>8</sup> Stephens, David O; “Registry: the world’s most predominant recordkeeping system”; available via [http://findarticles.com/p/articles/mi\\_qa3691/is\\_199501/ai\\_n8725164](http://findarticles.com/p/articles/mi_qa3691/is_199501/ai_n8725164)

helps the Commission comply with legal obligations regarding public access to documents. Furthermore, there are strict registration and security requirements for sensitive documents whose unauthorized disclosure would prejudice the interests of the European Union.

Document-level registration can prove difficult to scale to fit larger organizations, may introduce unsatisfactory delays or backlogs in document routing and retrieval, and can be seen as an excess of document control and a waste of recordkeeping resources better spent on other tasks. In many organizations without a strict regulatory environment, and without dedicated filing staff, pragmatic and decentralized filing systems are developed instead. An ESDP mission of 3000 staff deployed in less than optimum conditions is likely to find it challenging to manage a central registry with a small records team of four staff.

Given that up to ninety per cent of records creation, transmission and receipt activities across the mission are done by electronic means, the effective management of email and other electronic office documents is essential to support any registry proposals for hardcopy correspondence and records. If EULEX fails to manage electronic records in a disciplined manner using Mo-Req2 compliant systems, the mission will only fulfil the archiving requirements of the Council by printing such records and retaining them in physical folders.

### **Information technology resources**

ESDP missions can be launched at short notice into theatres of operation, and ideally require a standard set of core IT applications to facilitate the management of personnel, financial, logistics, procurement and reporting functions. At present, missions are deployed with limited information technology architecture and must develop many of their systems in-house. EULEX staff are provided with web-enabled email accounts and granted access to shared Windows folders, but there is currently a lack of electronic records and document management or enterprise content management systems to facilitate records capture and information sharing.

In practice, the initial intermittent and slow connectivity to shared drives within EULEX has resulted in an excessive reliance upon the email system to store and share documents and a proliferation of PST archive files that is unsustainable in the mid-term. The resulting ad hoc use of local PCs and removable storage devices to store data also represents an information security risk. It is anticipated that some of these technical barriers will be overcome through the replacement of congested serial lines with fibre optic cable.

## **■ Information management opportunities**

### **Raising senior management awareness**

In March 2008, the Registry and Archives team was introduced to the EUPT Senior Management Team and reported that in the absence of standardisation, centralisation, and clear roles and responsibilities in the management of records and information, there was a significant risk of record loss. The structure and content of shared drives remained unmanaged and at risk of deterioration. Email management, it was agreed, was likely to be a growing problem in these circumstances. The team outlined a vision of comprehensive information governance that included a network of Document Management Officers; electronic and - where necessary - manual recordkeeping systems; compliance with relevant EU standards, including security; consistency with internal audit requirements; joined up content management; and the development of information management best practices for ESDP missions. Through the mechanism of weekly and monthly reports, the Registry and Archives team has attempted to highlight any shortfalls in the achievement of this vision on a regular basis.

### **Creating the Document Management Officer (DMO) network**

In July 2008, the Administrative Head of Technical Services formally requested from departmental heads a nomination for the post of Strategic Document Management Officer. This Strategic DMO had to be an international staff member, and would have strategic oversight for all information governance issues within their department, including formal registration of electronic and manual correspondence; development and implementation of classification schemes; promotion of email best practice and monitoring the use of shared electronic folders; coordination of access to and possible handover of UNMIK records; definition of

requirements for document management systems; management of intranet content; preservation of document security; compliance with the European Data Protection directive in the processing of personal data; and the co-ordination of mail and courier services.

The activities of the department Document Management Officers would be coordinated and supported by professional input from the Registry and Archives team. The latter would also manage and facilitate access to the inactive electronic and manual archives created by mission departments. In practice the departments have found it very difficult to allocate staff to these tasks and progress has been limited. At the first meeting of the Strategic DMO's, it was agreed to alert senior management to the urgent need for dedicated international and local staffing resources allocated to each department to manage the wide range of essential records and information management activities.

### **Training staff in records and information management**

The mandatory mission induction training from April 2008 onwards provided an opportunity for the Registry and Archives team to raise awareness of the fact that EULEX staff have a duty to make proper arrangements for the records they create; to introduce the policy commitment that EULEX will create full and accurate records as evidence of its business activities and transactions; that information created or received by EULEX is owned by the mission, not the staff member; that staff must adopt shared standards and handle their information consistently; that all documents are not created equal - staff must know how to identify a mission record.

The Registry and Archives team helps teams comply with the "rules of the road" by writing policies/procedures and providing training, advice, and archival services. In the absence of any existing policy framework, and pending the completion of the CPCC Administrative Handbook, the following "ten rules" were included in a handout issued during mission induction.

1. Nominate a Document Management Officer for each section to co-ordinate electronic and manual recordkeeping, and liaise with the Registry and Archives team.
2. Set up shared electronic and manual filings systems, using an approved vocabulary to name files and folders. This may involve formal registry systems and centralised mail-handling services
3. Record details of mentoring functions and maintain appropriate records of correspondence between EULEX and external contacts.
4. Ensure that significant decisions (e.g. relating to purchasing, tendering, property management and recruitment) are fully documented and based on fact and evidence-based opinion.
5. Ensure that project leaders maintain records of all key phases of the project.
6. Apply appropriate security and access controls to records to prevent unauthorised disclosure of EULEX information.
7. Process personal data in accordance with Directive 95/46/EC of the European Parliament and of the Council of 24 October 1995 and subsequent regulations.
8. Identify records that are essential for business continuity and ensure tested back-up procedures are in place to promptly restore these vital records in the event of a disaster.
9. Dispose of records according to the EULEX Records Retention Policy e.g. promptly destroy drafts, working papers and correspondence, including emails, when they are no longer needed by the organisation.
10. Transfer records worthy of permanent preservation to the Central Archives of the Council of the European Union.

Staff were further advised not to over-use email; to give records a meaningful title in the subject field; to regularly use the delete key to remove non-records; to save email records alongside related business documents in folders on shared drives; to manage email within 500MB mailbox limits and not to create PST archive files; not to send attachments with internal emails, but to use hyperlinks; and not to send messages to everyone in the address book without approval.

## ■ Information management architecture

### The EULEX file plan and retention schedule

An early draft of the EULEX file plan was issued for comment in April 2008. It was based initially on the United Nations (UN) Archives and Records Management Service Taxonomy for Recordkeeping in Field Missions of UN Peacekeeping Operations. The Registry & Archives team continues to hold workshops with the Strategic DMO's to revise their part of the file plan tree so that it fits the business better. This approach has been endorsed by the Council Central Archives. It is hoped that a model ESDP file plan will be developed, inclusive of standard retention periods for each records series. A project team will be set up by the Council to address this, but for the time being relevant Commission and Council retention rules will be consulted alongside best practice suggested by the UN Peacekeeping Operational Retention Schedule<sup>9</sup>.

### Outline of general registration proposals

In the absence of robust Electronic Document and Records Management Systems (EDRMS) within ESDP missions, the Council Central Archives has recommended that paper is the preferred archival media. However a parallel system of electronic recordkeeping will be maintained to facilitate quick retrieval, concurrent usage and the re-purposing of content. The following tactical proposals will be implemented without prejudice to future strategic IT developments e.g. Mo-Req2 compliant EDRMS and email archiving/vault applications. Documents and metadata registered in this manner remain exportable for use in future systems.

The strategic goal is to permit powerful searching across registry indexes and departmental record collections, to retain audit trails, to maintain security permissions as appropriate and to delete expired records. In this way EULEX aspires to handover a comprehensive multimedia archive to the Council following liquidation. The use of preservation file formats, such as Adobe's Portable Document Format (PDF) enhances interoperability and minimizes the risk of digital obsolescence.

In general, all incoming hardcopy correspondence to the appropriate departments will be date-stamped, and marked by the primary decision-makers with the name of the responsible officer, the action required, and the task deadline. Documents shall then be routed to the responsible officer for action. Following completion of this action, the original incoming document and a copy of any outgoing correspondence relating to this transaction shall be forwarded to the department registry. Once received by the department registry, all incoming and outgoing mail shall be marked with the creation date of the document, the file classification and a unique registration ID. Each incoming and outgoing document shall then be registered on an electronic index managed by the department registry.

The original document shall be prepared for imaging and processed via optical character recognition. The resulting image/PDF shall be re-titled with its Registration ID and then stored on a network server in appropriately named secure folders managed by the department registry. A URL or hyperlink can be added to the electronic index to permit direct access to the scanned document.

The original manual documents shall be filed in a physical folder titled according to the relevant classification. Note also that internally-generated documents or correspondence that are for internal consumption only will be registered in a similar manner, and filed alongside related documents in the appropriate physical and electronic folders.

Due to the registry index and the electronic version available, it should not be necessary or desirable to create a manual chronological file e.g. IN/OUT. If appropriate the physical folder can be checked-out to the responsible officer for easy reference, or it can be stored in the local department registry. A simple tracking system will be provided to department registries for this purpose.

### Outline of email registration proposals

The most basic approach to the capture of email records would be to print them out and handle them as outlined above. However, each department registry shall be provided with a dedicated mailbox for official

---

<sup>9</sup> More information on the UN taxonomy and retention schedule can be found via <http://archives.un.org/unarms/>



email records (e.g. JusticeRegistry@Eulex.eu). In general, all incoming email correspondence to the appropriate departments shall be date-stamped, and marked by the primary decision-makers with the name of the responsible officer, the action required, and the task deadline. Where not captured automatically, this information can be easily added to the email title or message body.

Emails shall then be routed to the responsible officer for action. The use of Outlook tasking tools may facilitate the allocation and tracking of outstanding tasks. Following completion of this action, the original incoming email and a copy of any outgoing correspondence relating to this transaction shall be forwarded to the department registry. This could be in the form of cc or bcc where the responsible officer has responded directly to the original sender. Only key mission records need be routed in this manner, and the responsible officer might delay submission to the department registry mailbox until a particular email thread is completed.

Each incoming and outgoing email record (and attachment) shall then be registered on an electronic index managed by the department registry. If the volume of email records to register is low, it will be possible to use the same index outlined above. If, as is likely, this method is unsustainable with high email volumes, it will be sufficient to use the department registry mailbox itself as the supplementary index. Additional property fields may be selected within Outlook to add metadata to individual emails so that the index is comparable to the one for hardcopy correspondence.

In order to control the size of the department registry mailbox and facilitate shared access to its content, groups of email can be periodically harvested (for example by Adobe Acrobat Professional) to create a PDF bundle of email that can be saved to a shared drive and searched by the Adobe Reader installed on users' computers. These PDF bundles of email could usefully sit alongside related hardcopy documents that have been scanned as PDFs and born-digital records to permit their consistent management. The original emails located in the responsible officer's mailbox can safely be deleted once this process is completed. Indeed, this will be necessary if users' mailboxes are to remain manageable and capped at 500MB.

## ■ Conclusion: Strategic IT and information management

ESDP missions can be seen to represent the triumph of "brain" over "brawn". The success of mentoring, monitoring and advising activities in support of the rule of law depends upon the collection of data, the sharing of information and the transmission of knowledge in the field. Future missions will benefit from the ESDP recordkeeping standards emerging from Brussels, and the developing professional network of records managers and archivists within the EU institutions. They would also benefit from the early recognition of the value of records and information management, the appointment of strategic information managers, the provision for sufficient records staff within the organisational structure, and a closer relationship between IT development and information management strategy. In this way, despite the considerable challenges of deploying in extreme environments, the records of ESDP missions are likely to be captured and preserved in a satisfactory manner.

# LA PROBLÉMATIQUE DE LA SAUVEGARDE NUMÉRIQUE DES E-MAILS EN BELGIQUE : DIRECTIVES ET RECOMMANDATIONS POUR LES ADMINISTRATIONS PUBLIQUES

---

Sébastien Soyez

Attaché scientifique, Archives générales du Royaume de Belgique et Archives de l'Etat dans les Provinces

*Une mise en perspective du travail réalisé à l'échelle belge sera brièvement exposée. En outre, nous présenterons le traitement de cette problématique par les Archives de l'Etat. Celle-ci étant abordée en deux temps :*

- définir un cadre global pour l'archivage numérique des e-mails par le biais d'une directive générale ;*
- travailler en partenariat avec les institutions publiques, dont les administrations fédérales, afin d'une part de diffuser les best practices et d'autre part de développer des outils facilitant les étapes conduisant à un archivage pérenne et ad hoc de leurs e-mails.*

*L'essentiel de la présentation portera sur la présentation du résultat de la première étape de notre travail, à savoir la publication et la diffusion de la directive sur l'archivage numérique des e-mails. Après un bref rappel du cadre général et légal de l'archivage des e-mails, nous présenterons les trois axes principaux de la mise en œuvre des recommandations de cette directive : quand archiver ?, qui est responsable ? et comment archiver ?*

*Nous y aborderons les responsabilités de chacun en présentant les délimitations en la matière tant au niveau de la sélection qu'au niveau de l'enregistrement et de la conservation. Nous présenterons également un bref aperçu des plans d'action à court terme que les administrations publiques peuvent élaborer en vue d'améliorer la gestion et la conservation des e-mails. Nous aborderons enfin les étapes essentielles d'un archivage conforme des e-mails, qui passent notamment par la sélection de métadonnées et par l'organisation structurée des messages et leur enregistrement dans des formats d'archivage pérenne.*

*En conclusion, nous mettrons en perspective le résultat de cette première étape avec les objectifs de la seconde étape à savoir la diffusion des bonnes pratiques et la mise sur pied de partenariat étroit avec les acteurs du secteur public en vue d'appliquer les recommandations pour un archivage approprié des e-mails.*

[Consulter la présentation.](#)

## **Issues surrounding digital storage of emails in Belgium: directives and recommendations for public bodies**

*The presentation will begin with a brief overview of the work accomplished across Belgium. In addition, we shall describe how the State Archives handle these issues, in two parts:*

- by defining an overall framework for the digital archiving of emails via a general directive;*
- by working together with public institutions, including the federal administrations, with a view to spreading best practice and developing tools to facilitate the stages leading to the long-term and ad hoc archiving of their emails.*

*Most of the presentation will be devoted to the result of the first phase of our work, the publication and dissemination of the directive on digital archiving of emails. After a brief review of the general and legal context concerning archiving emails, we shall look at the three main questions addressed by the directive: When to archive? Who is responsible? and How to archive?*

*We shall consider the responsibilities of all those concerned by describing the limits involved, whether concerning selection, recording or preservation. We shall also look briefly at the short-term action plans that public administrations can adopt to improve the management and storage of emails. Finally, we shall describe the essential phases for standardised archiving of emails, which require such elements as the selection of metadata, the structured organisation of the messages and saving them in long-term archiving formats.*

*In conclusion, we shall compare the results of this first phase with the objectives of the second phase, i.e. disseminating good practice and setting up close partnerships with public sector stakeholders to help them apply the recommendations for the suitable archiving of emails.*

[See the presentation.](#)

## ■ Introduction

Dans l'environnement numérique tout comme pour les supports conventionnels d'information, le processus d'archivage (de sélection, d'enregistrement et de conservation) des documents produits par toute institution publique peut se résumer, d'une part, par l'organisation, la gestion et la conservation des documents qui ont une valeur probante et/ou culturelle, historique et, d'autre part, leur garantir authenticité, intégrité et une parfaite lisibilité (l'absence de celle-ci étant l'équivalent d'une perte, d'une destruction ou d'une altération des documents). La communication de ces documents aux usagers selon des moyens rapides, précis, conviviaux, en fonction de la demande et d'éventuels niveaux d'accès étant un des objectifs finaux de l'archivage.

L'archivage de contenus électroniques est l'ensemble des actions, outils et méthodes mis en œuvre pour les réunir, les identifier, les sélectionner, les classer et les conserver, sur un support sécurisé, dans le but de les exploiter et de les rendre accessibles dans le temps, que ce soit à titre de preuve (en cas d'obligations légales notamment ou de litiges) ou à titre informatif. Le contenu archivé est considéré comme figé et ne peut donc être modifié. La durée de l'archivage est fonction de la valeur du contenu et porte le plus souvent sur le moyen voire le long terme. La conservation est l'ensemble des moyens mis en œuvre pour stocker, sécuriser, pérenniser, restituer, tracer, transférer voire détruire, les contenus électroniques archivés. Pour mettre en œuvre un archivage numérique offrant toutes les garanties de conservation pérenne, il est essentiel de pouvoir sélectionner les documents à conserver sur le long terme, choisir des supports et des formats de conservation adéquats, conserver les métadonnées (descriptives, administratives et techniques) tout au long du cycle de vie des documents et prévoir, le cas échéant, des outils de recherche adaptés aux usagers (directs et futurs) en vue de faciliter la recherche, la communication et l'utilisation des documents. L'archivage numérique peut donc se résumer à l'activité de gestion et de diffusion d'un document depuis sa création afin d'assurer qu'il est en accord avec son temps et disponible pour son éventuelle réutilisation (durant sa phase dynamique) ou sa future consultation (durant sa phase semi-dynamique ou statique).

Dans le cadre de la rédaction de directives et de recommandations sur l'archivage numérique, les Archives de l'Etat en Belgique se sont attaqués à la problématique de l'archivage des e-mails. Les différentes étapes du cycle de vie d'un e-mail y sont présentés, de sa création jusqu'à sa conservation, déterminant de manière univoque les éléments indispensables à leur préservation sur le long terme. Les réponses aux questions suivantes y ont été traitées : Quelles sont les obligations légales? Qu'entend-on par un e-mail? Quand utilise-t-on /doit-on utiliser un e-mail? Quels sont les e-mails qui doivent être archivés? Qui est responsable de l'archivage des e-mails? Comment archiver ses e-mails? Cette première directive tente de jeter les bases pour la mise en œuvre d'une politique cohérente de conservation des documents électroniques au sein des institutions publiques qui sont sous la surveillance des Archives de l'Etat, à savoir les administrations de l'Etat (les services publics fédéraux), le Conseil d'Etat, les tribunaux de l'ordre judiciaire, les provinces, les communes et les organismes d'intérêt publics.

## ■ Obligation légale de l'archivage des e-mails

La loi sur les archives du 24 juin 1955<sup>1</sup> constitue le fondement de l'obligation de conservation des documents administratifs par les institutions publiques belges. Cette loi précise d'ailleurs les personnes et les organisations qui sont concernées par cette obligation<sup>2</sup>. Le législateur de l'époque s'est montré clair sur l'obligation qu'ont ces institutions publiques de verser leurs archives, et par extension tout type de document conservé sur quelque support que ce soit. La législation en vigueur en Belgique couvre donc également l'ensemble des documents numériques et *a fortiori* les e-mails.

A ce titre, l'ensemble des institutions publiques concernées par la loi de 1955 ont l'obligation de conserver leurs e-mails qui ont un statut de document administratif. Du point de vue juridique, la notion de document administratif peut se définir comme étant tout document produit par un service administratif en application

---

<sup>1</sup> *Moniteur belge* du 12 août 1955 (<http://www.moniteur.be>).

<sup>2</sup> Pour le détail de ces institutions, cf. *supra* (Avant-propos).

d'une procédure réglementaire. Par ailleurs, la loi belge du 11 avril 1994<sup>3</sup>, relative à la publicité des actes administratifs (concernant les autorités administratives fédérales), renforce l'obligation qu'ont les institutions publiques, dans un souci de transparence, de conserver leurs documents administratifs.

Cette obligation de conservation des documents par les institutions publiques est néanmoins asservie de limites fixées par la législation sur la protection de la vie privée. Le cadre légal fixant ces limites est réglé selon deux niveaux : l'un européen, avec l'article 8 de la Convention européenne des droits de l'homme de 1954<sup>4</sup> traitant de la protection des données à caractère personnel et étendue récemment à la protection de la vie privée, de la liberté de communication et du secret de télécommunication<sup>5</sup> ; l'autre belge, avec la loi du 8 décembre 1992 sur la protection de la vie privée à l'égard des traitements de données à caractère personnel<sup>6</sup>. Au regard de cette législation, il apparaît que tout document à caractère personnel, c'est-à-dire tout document administratif comportant une appréciation ou un jugement de valeur relatif à une personne physique nommément désignée ou aisément identifiable, ou la description d'un comportement dont la divulgation pourrait causer un préjudice à cette personne, doit être protégé et ne peut en aucun cas être conservé par l'employeur à d'autres fins que celui de l'archivage à long terme. Il est important de signaler que l'automatisation totale et systématique de l'archivage des documents administratifs tels que les e-mails par l'employeur entre en conflit avec cette législation sur la protection de la vie privée de l'utilisateur. Un moyen pour éviter ce conflit avec la législation en vigueur en cas d'automatisation du processus d'archivage des e-mails est celui de l'utilisation de messagerie professionnelle<sup>7</sup>. Ce choix permet de canaliser le flux de la correspondance officielle - tout en évitant les utilisations abusives - et facilite la sélection ultérieure des e-mails à conserver. Quelque soit la solution envisagée, il conviendra pour l'employeur de respecter cinq principes de base : la transparence (tout les personnes impliquées doivent être informées qu'il y a un système d'archivage en place, le cas échéant automatisé), la finalité (doit être conforme à la législation sur les archives), la proportionnalité (seuls certains e-mails seront conservés et pour une durée déterminée), la sécurité et la confidentialité (les accès aux données doivent être justifiés) ainsi que le droit d'accès et de modification (toute personne a le droit de savoir ce qui est conservé à son sujet, de corriger toute information erronée voire de supprimer l'information illégitime).

Les avantages directs pour une institution publique de conserver ses documents, et en particulier ses e-mails professionnels, qu'ils soient produits ou reçus, sont nombreux et indéniables. Lorsqu'une plainte ou une contestation survient, la conservation appropriée de ces documents numériques permettra de fournir des preuves tangibles notamment devant les tribunaux. De plus, une institution ayant investi dans un système permettant un archivage probant des documents numériques assure aux utilisateurs, tant présents que futurs, un accès et une consultation optimales. Pour atteindre cet objectif, il conviendra que le système mis en place (quel qu'il soit) puisse garantir - lors de la conservation des documents d'archives - leur authenticité, leur intégrité, leur fiabilité, une complète sécurité, leur intégrité ainsi qu'une parfaite lisibilité. Nous reviendrons en détail sur ces notions de base dans la partie relative à la conservation des e-mails.

L'archivage électronique des documents offre donc de nombreux avantages dans la centralisation de la conservation à court, moyen ou long terme suivant les contraintes définies, la gestion des cycles de vie et de la sécurité dans une perspective d'archivage légal, la traçabilité des documents archivés et des accès des différents utilisateurs, une ouverture aux différents standards et une définition des formats des documents à archiver.

---

<sup>3</sup> *Moniteur belge* du 30 juin 1994 (cf. par ailleurs les lois et décrets des entités fédérées sur la publicité des actes administratifs).

<sup>4</sup> Convention de sauvegarde des Droits de l'Homme et des Libertés fondamentales (pour le texte complet de la convention, cf. site internet du Conseil de l'Europe <http://www.coe.int/>).

<sup>5</sup> Les Directives européennes du 24 octobre 1995 relative à la protection des personnes physiques à l'égard du traitement des données à caractère personnel et à la libre circulation de ces données (Journal officiel n° L 281 du 23 novembre 1995) et du 12 juillet 2002 concernant le traitement des données à caractère personnel et la protection de la vie privée dans le secteur des communications électroniques, communément appelée directive "vie privée et communications électroniques" (Journal officiel n° L 201 du 31 juillet 2002).

<sup>6</sup> Loi du 8 décembre 1992, adaptée par la loi du 11 décembre 1998 sur base de la directive européenne du 24 octobre 1995 et mise en application par arrêté royal le 13 février 2001 (*Moniteur belge* du 13 mars 2001).

<sup>7</sup> Ce boîte aux lettres professionnelle peut le cas échéant prendre le nom de l'entité pour laquelle l'agent est en fonction.

## ■ La structure et le format des e-mails

Afin de préciser le champ d'activité l'archivage des e-mails, il est indispensable de présenter une définition précise du courrier électronique. Suivant la directive européenne « vie privée et communications électroniques » du 12 juillet 2002 transposée dans la législation belge le 11 mars 2003<sup>8</sup>, le courrier électronique peut être défini comme « *tout message sous forme de texte, de voix, de son ou d'image envoyé par un réseau public de communications qui peut être stocké dans le réseau ou dans l'équipement terminal du destinataire jusqu'à ce que ce dernier le récupère* »<sup>9</sup>. La directive européenne précise à cet égard que : « *les services actuellement compris dans la définition du courrier électronique comprennent : le courrier Simple Mail Transport Protocol (SMTP), c'est-à-dire le courrier électronique « classique », le service de messages courts (SMS), le service de messages multimédias (MMS), les messages laissés sur répondeurs, les systèmes de messagerie vocale y compris sur les services mobiles, les communications « net send » adressées directement à une adresse IP. Les bulletins d'information envoyés par courrier électronique tombent également dans le champ d'application de cette définition* ». Cette liste non limitative connaîtra certainement des extensions futures en fonction des prochains développements technologiques. A la lumière de cette définition, il est important de préciser que la directive préparée par les Archives de l'Etat concerne exclusivement le courrier électronique classique, appelé communément par les anglophones « e-mail »<sup>10</sup> et par les francophones « courriel »<sup>11</sup>. La définition commune d'un e-mail désigne le service de transfert de messages envoyés par un système de messagerie électronique par l'intermédiaire d'un réseau informatique vers la boîte aux lettres électronique d'un destinataire choisi par l'expéditeur.

Il est à noter que les messages électroniques envoyés et reçus sont généralement collectés au sein de deux entités distinctes d'une boîte d'e-mail<sup>12</sup> : les répertoires Réception (*Inbox*) et Envoi (*Send*). Seuls les messages maintenus dans ces deux répertoires seront à traiter pour la conservation ou l'élimination.

### Le format des e-mails (standard MIME et protocole de transfert SMTP)

Le format de base des e-mails est défini dans la RFC 2822<sup>13</sup>. Cette recommandation spécifie le format des en-têtes et du corps des e-mails contenant du texte, ainsi que les règles d'en-têtes générales comme les champs *To:*, *Subject:*, *From:* ou *Date:*. Le standard MIME définit un ensemble d'attributs additionnels d'en-têtes aux e-mails pour le type de contenu du message et de codage<sup>14</sup>. MIME définit également des règles spécifiques pour encoder des caractères non ASCII dans les en-têtes de messages afin par exemple d'autoriser des caractères accentués dans le sujet d'un e-mail. Le contenu du corps de l'e-mail était initialement réglementé par la norme RFC 2821 qui préconisait l'utilisation de lignes de 1000 caractères maximum et de caractères tronqués à 7 bits. Comme les langages basés sur l'alphabet latin incluant des caractéristiques particulières - des diacritiques - n'étant pas supportés par l'ASCII (7 bits), les e-mails respectant cette RFC ne pouvaient pas être correctement représentés.

Le standard MIME est également un composant fondamental des protocoles de communications comme HTTP<sup>15</sup>, qui requièrent l'envoi de données dans le même contexte que l'envoi d'e-mails, même si les données ne sont pas des e-mails. L'intégration ou l'extraction des données au format MIME est généralement effectuée automatiquement par le logiciel client de messagerie ou par le serveur de messagerie électronique quand l'e-

<sup>8</sup> Loi du 11 mars 2003 sur certains aspects juridiques des services de la société de l'information (Moniteur belge du 17 mars 2003).

<sup>9</sup> Article 2 (h) de la directive « vie privée et communications électroniques » du 12 juillet 2002 (Journal officiel n° L 201 du 31 juillet 2002).

<sup>10</sup> Contraction de l'anglais « electronic mail ».

<sup>11</sup> Contraction de « courrier électronique » ; l'abréviation « mél » doit quant à elle être utilisée pour indiquer une adresse e-mail sur une carte de visite, au même titre qu'un numéro de téléphone y est abrégé par « tél ».

<sup>12</sup> Il peut arriver que d'autres répertoires soient créés par l'utilisateur pour classer ses e-mails entrants et sortants. Dans ce cas, la conservation des e-mails s'effectuera (également) au départ de ces répertoires. Il est également utile de conserver le carnet d'adresse de la messagerie électronique dans le but de pouvoir lier les coordonnées complètes des contacts avec les messages.

<sup>13</sup> RFC 2822 (*Request for Comment 2822*) est en quelque sorte un pré-standard. La procédure complète pour la transcription d'une RFC en standard est la suivante : RFC → Internet Draft → Proposed Standard → Draft Standard → Internet Standard.

<sup>14</sup> Le codage étant la façon de traduire en ASCII (7 bits) les données 8 bits du message.

<sup>15</sup> HTTP (*HyperText Transfer Protocol*) signifie littéralement le « protocole de transfert hypertexte ».

mail est envoyé ou reçu. En outre, ce standard est extensible car sa définition inclut une méthode pour enregistrer de nouveaux types de contenus ou d'autres valeurs d'attributs.

## Exemples de formats d'exportation des e-mails

Outre le format intrinsèque des e-mails exposé précédemment, chaque logiciel de messagerie électronique offre la possibilité de sauvegarder les e-mails dans un format d'exportation (soit propriétaire, soit ouvert). La plupart de ces logiciels peuvent en utiliser plusieurs, tandis que certains de ces formats ne sont utilisés que par un seul logiciel de messagerie électronique<sup>16</sup>. Le choix de ces formats d'exportations pourra s'avérer crucial lors de l'archivage définitif des e-mails car ils permettront - ou non - de conserver l'entièreté des informations tant du message lui-même que celles liées au contexte (métadonnées). Le format XML est à ce titre une voie tout à fait adaptée pour la conservation structurée des e-mails ainsi que de leurs données contextuelles. Ce format, outre sa grande souplesse d'utilisation et d'extension, peut également être généré à partir de certains formats de sauvegarde existants comme TXT, EML et RTF notamment. Il n'est par contre pas (encore) utilisé par défaut par les logiciels de messagerie électronique, ce qui rend son usage subordonné à la mise en place de fonctionnalités supplémentaires (au sein des logiciels-clients de messagerie électronique) afin de permettre la sauvegarde des e-mails. Nous reviendrons sur cet aspect dans la partie consacrée à la conservation des e-mails.

## L'utilisation de la signature électronique dans les échanges d'e-mails

Le législateur belge a transposé en 2001 la Directive européenne du 13 décembre 1999 sur un cadre communautaire pour les signatures électroniques. Cette transposition en droit belge s'est effectuée par l'adoption de deux lois : la loi du 20 octobre 2000 introduisant l'utilisation de moyens de télécommunication et de la signature électronique dans la procédure judiciaire et extrajudiciaire<sup>17</sup> qui modifie notamment l'article 1322 du Code civil et la loi du 9 juillet 2001 fixant certaines règles relatives au cadre juridique pour les signatures électroniques et les services de certification<sup>18</sup>. De manière formelle, la signature électronique est « un mécanisme permettant d'authentifier l'auteur d'un document électronique et de garantir son intégrité, par analogie avec la signature manuscrite d'un document papier. Un mécanisme de signature numérique doit présenter les propriétés suivantes : - il doit permettre au lecteur d'un document d'identifier la personne ou l'organisme qui a apposé sa signature ; - il doit garantir que le document n'a pas été altéré entre l'instant où l'auteur l'a signé et le moment où le lecteur le consulte »<sup>19</sup>. Pour cela, elle se distingue de la signature « simple » qui elle n'est constituée que d'une suite de caractères. La signature électronique - rendue possible grâce à la cryptographie asymétrique<sup>20</sup> - est différente de ce type de signature car elle n'est pas visuelle, mais correspond à une suite de nombres. Quand on parle de signature électronique, les conditions suivantes doivent pouvoir être réunies : authenticité (l'identité du signataire doit pouvoir être retrouvée de manière certaine), non falsifiabilité (la signature ne peut pas être falsifiée), non réutilisabilité (la signature n'est pas réutilisable et fait partie du document signé et ne peut être déplacée sur un autre document), inaltérabilité (un document signé est inaltérable car une fois qu'il est signé, il ne peut être modifié), et irrévocabilité (la personne qui a signé ne peut le nier). La signature électronique (avec cryptographie asymétrique) est considérée actuellement comme un des systèmes les plus sûrs et surtout un des plus répandus pour signer dans l'environnement électronique. Cette technique repose sur une relation triangulaire entre le signataire, le destinataire du message et une autorité de certification. Dans la pratique, toute personne désirant signer électroniquement doit demander préalablement à l'autorité de certification<sup>21</sup> de lui délivrer une clé privée<sup>22</sup>, secrète et détenue exclusivement par le signataire. Parallèlement, l'autorité de certification doit créer une clé publique complémentaire à la clé privée. Le signataire peut dès lors, s'il désire envoyer un message signé au

<sup>16</sup> EML, HTML, MSG, RTF, TXT, OFT sont quelques uns des principaux formats d'exportation d'e-mails.

<sup>17</sup> *Moniteur belge* du 22 décembre 2000.

<sup>18</sup> *Moniteur belge* du 29 septembre 2001.

<sup>19</sup> Définition reprise du site <http://fr.wikipedia.org/> (consulté le 22 août 2007) ; cf. également le site internet du SPF Economie, PME, Classes Moyennes et Energie : [http://mineco.fgov.be/information\\_society/e-signatures/home\\_fr.htm](http://mineco.fgov.be/information_society/e-signatures/home_fr.htm) (consulté le 21 septembre 2007).

<sup>20</sup> Appelée également cryptographie à clé publique.

<sup>21</sup> Les prestataires de service de certification sont définis par la loi du 9 juillet 2001 comme étant « toute personne physique ou morale qui délivre et gère des certificats ou fournit d'autres services liés aux signatures électroniques » (*Moniteur belge* du 29 septembre 2001). En Belgique, le système d'accréditation des prestataires de service de certification (« BE.SIGN ») est défini par l'AR du 6 décembre 2002 (*Moniteur belge* du 17 janvier 2003). Une liste de ces prestataires est disponible sur le site internet du SPF Economie, PME, Classes moyennes et Energie ([http://mineco.fgov.be/information\\_society/e-signatures/list\\_e\\_signature\\_fr.pdf](http://mineco.fgov.be/information_society/e-signatures/list_e_signature_fr.pdf)).

<sup>22</sup> Créée à l'aide d'un algorithme.

destinataire, appliquer la clé privée au message. Le message (en règle générale, il s'agit uniquement de l'empreinte du message obtenue par une fonction de « hachage ») sera alors crypté selon l'algorithme de la clé privée. Quand le message est réceptionné par le destinataire, celui-ci tentera de le décrypter à l'aide de la clé publique de l'émetteur théorique du message. Si ces deux clés correspondent, le message peut être décrypté et le destinataire aura alors la certitude que le message a bien été signé par la clé privée de l'émetteur. Par contre, si le décryptage échoue, cela signifie que le message n'a pas été signé avec la clé privée de l'auteur présumé<sup>23</sup>.

La signature électronique est néanmoins encore très peu utilisée en Belgique à ce jour dans le cadre de la correspondance électronique au sein des administrations publiques. Mais il se pourrait que dans un avenir proche, son utilisation s'étende progressivement, notamment pour certains types de correspondance nécessitant une preuve légale. Certains travaux récents démontrent en effet une certaine évolution en la matière. Un exemple significatif est celui lié à l'informatisation de la justice et la mise en place d'une procédure électronique dans le cadre de l'application du nouveau Code judiciaire<sup>24</sup>. Ce projet, initialement baptisé « Phénix »<sup>25</sup>, devrait offrir la possibilité de notifier les actes de procédure par le biais d'une correspondance électronique (e-mails) si ceux-ci disposent d'une signature électronique garante de leur authenticité. Dans ce cadre particulier, l'usage de la signature électronique lors de certains échanges d'e-mails peut apparaître indispensable en vue d'apporter la preuve absolue - s'il en est - que l'envoi est authentique<sup>26</sup>.

## ■ Quand utiliser un e-mail

Dans le cadre de la communication interne et externe d'une institution, l'utilisation des e-mails est souvent liée à la nature même de l'organisation voire à sa culture propre. Il est donc impératif que chaque institution définisse individuellement l'ensemble des processus de travail pour que les utilisateurs puissent choisir en connaissance de cause les moyens de communication qu'ils devront utiliser (courrier classique, téléphone, fax ou e-mail). L'institution doit dès lors définir des règles de base qui détermineront ces choix afin de maximiser les avantages de l'envoi d'e-mails tout en minimisant ses désavantages<sup>27</sup>. Souvent considérés comme une forme de communication temporaire et personnelle, les e-mails sont néanmoins utilisés pour communiquer des décisions importantes voire cruciales pour l'activité professionnelle d'une institution. La raison principale est qu'il est souvent difficile de distinguer - dans le cas de communications internes - une communication temporaire et personnelle d'une communication formelle et professionnelle. Il est dès lors important de rédiger avec toute l'attention requise ce type d'e-mail sinon le risque de non compréhension ou de non prise en compte est réel<sup>28</sup>.

La forme et le contenu des e-mails constituent, dans ce cadre, une incontournable gageure. L'importance de sa forme rédactionnelle ne doit pas être négligée et au besoin, l'institution (globalement et/ou par département) peut joindre des recommandations plus spécifiques sous forme de « guide des bonnes pratiques » pour permettre une utilisation adéquate et normalisée des e-mails. En outre, certains contenus considérés

<sup>23</sup> GUINOTTE L., *La signature électronique après les lois du 20 octobre 2000 et du 9 juillet 2001*, in *Journal des tribunaux*, n° 6063, 14 septembre 2002, pp.553-561, Bruxelles, Larcier, 2002.

<sup>24</sup> Loi du 5 août 2006 (*Moniteur belge* du 7 septembre 2006) modifiant certaines dispositions du Code judiciaire en vue de la procédure par voie électronique.

<sup>25</sup> Loi du 10 août 2005 (*Moniteur belge* du 1<sup>er</sup> septembre 2005) instituant le système d'information « Phénix » ; cf. également HENROTTE J.-F., *Phénix et la procédure électronique*, Commission Université-Palais (vol.85), Bruxelles, Larcier, 2006 ; Le projet avait été provisoirement interrompu depuis 2007 mais a été relancé depuis peu (avril 2008) sous le nom de « projet Cheops ».

<sup>26</sup> Cette preuve légale nécessaire lors de certains échanges électroniques a conduit en Belgique à la création de la co-entreprise Certipost (Joint-venture créée en 2002 entre Belgacom et La Poste, spécialisée dans la communication électronique sécurisée, les applications *e-Government* et d'autres services par Internet). Certipost regroupe en outre les activités E-Trust de Belgacom et les services de Belgium Post Group (BPG). Belgacom a fourni la *E-Trust Certification Authority*, instrument pour la création et la gestion de certificats numériques. BPG a fourni la plate-forme Postbox, aujourd'hui appelé MyCertipost. Grâce à l'un des services de cette entreprise, MyCertipost, qui est un compte en ligne permettant de gérer son administration électroniquement, l'utilisateur a un point d'accès unique et gratuit pour recevoir de manière sécurisée des informations confidentielles (factures, fiches de paie, documents administratifs ou commerciaux et messages de toute nature).

<sup>27</sup> Ces inconvénients se situent d'ailleurs souvent au niveau des destinataires.

<sup>28</sup> *Utilisation efficace du courrier électronique*, COMM collection n°1, Bruxelles, 2004 (Groupe de travail des administrations fédérales) ; ainsi que de ce même groupe de travail : *10 astuces pour une utilisation efficace du courrier électronique*, Bruxelles, 2004 ; cf. aussi les règles de base de la communication par e-mail issues des règles de la Netiquette RFC 1855 (repris du site <http://www.dtcc.edu/cs/rfc1855.html>, consulté le 18 juin 2007).



comme trop sensibles pour être envoyés par e-mails demandent une précaution supplémentaire de la part de chaque institution vis-à-vis de ses employés notamment pour éviter toute transgression. Ces contenus devraient dès lors pouvoir faire l'objet d'une liste détaillée à exclure de l'envoi. L'utilisation privée doit également être permise en vertu de l'article 8 de la Convention européenne des droits de l'homme<sup>29</sup> : « 1. Toute personne a droit au respect de sa vie privée et familiale, de son domicile et de sa correspondance. - 2. Il ne peut y avoir ingérence d'une autorité publique dans l'exercice de ce droit que pour autant que cette ingérence est prévue par la loi et qu'elle constitue une mesure qui, dans une société démocratique, est nécessaire à la sécurité nationale, à la sûreté publique, au bien-être économique du pays, à la défense de l'ordre et à la prévention des infractions pénales, à la protection de la santé ou de la morale, ou à la protection des droits et libertés d'autrui ». Dans le respect de ces libertés, un certain nombre de limites peuvent être indiquées par l'institution afin d'éviter tout abus. Chaque institution peut dès lors mettre en place un système de contrôle du flux des e-mails, mais il est impératif que les personnes soient tenues informées et que ce contrôle se fasse dans le respect de l'article 8 susmentionné.

### La sélection des e-mails à conserver

Les objectifs clés de l'implémentation d'un archivage des e-mails réussis peuvent être regroupés en quatre axes principaux :

- déterminer quand préserver les e-mails ;
- sélectionner les e-mails à archiver ;
- identifier les responsabilités dans l'archivage des e-mails (et de leurs pièces jointes) ;
- mettre en œuvre un système d'archivage de ses e-mails et donc de ses documents numériques en établissant par ailleurs l'endroit où ils seront préservés et la manière dont ils seront traités.

Tout d'abord, il est indispensable que l'institution (le cas échéant le département, la direction ou le service) puisse déterminer sans équivoque quels sont les messages électroniques qui pourront être considérés comme de la correspondance électronique formelle, apparentés à des documents administratifs, et qui, à ce titre, devront être enregistrés et le cas échéant conservés. En d'autres mots, il s'agira de pouvoir distinguer entre les e-mails qui revêtent une valeur administrative (une valeur professionnelle, *business*), archivistique et/ou historique par rapport à ceux qui revêtent une valeur purement informelle, non pertinente ou personnelle. Seuls les e-mails (et leurs pièces jointes) ayant un statut formel ou officiel devront être conservés et dès lors archivés (de préférence en dehors du logiciel de messagerie électronique<sup>30</sup>). Les e-mails personnels, les e-mails à faible valeur informationnelle ne seront par contre pas conservés par l'institution.

- Un e-mail doit être enregistré et, le cas échéant, conservé quand il reflète les activités de l'institution, de l'organisme, formalise les différentes étapes d'une tâche - la mise en œuvre, l'autorisation et l'accomplissement d'une opération liée à cette tâche - dans le cadre des missions liées à l'activité d'une institution, d'un organisme, d'un département, d'un service voire d'une cellule. Il peut s'agir de la correspondance liée à des réunions, des actes, des ordres du jour et des comptes rendus de réunion. En règle générale, les principes de sélection des e-mails à conserver puis à archiver s'applique suivant les mêmes règles de conservation en vigueur pour la correspondance classique<sup>31</sup> (sur format papier).
- Un e-mail ne doit pas être enregistré ni conservé quand il s'agit de messages officieux, de brouillons, d'informations de service (réunion de service, congés, duplicata d'informations, etc.), d'informations de type « liste de diffusion » et de messages résultant de l'usage personnel. De manière générale, l'ensemble des e-mails ne comportant pas de valeur administrative - e-mails à caractère informatif ou à usage interne sans valeur informative additionnelle - ne doivent pas être archivés et pourront, le cas échéant, être éliminés une fois le délai d'utilité administrative écoulé<sup>32</sup>.

---

<sup>29</sup> Cf. la Convention de sauvegarde des Droits de l'Homme et des Libertés fondamentales.

<sup>30</sup> Cf. *infra* la partie relative à la conservation des e-mails notamment concernant le choix entre un classement numérique des dossiers et un système plus élaboré d'archivage électronique.

<sup>31</sup> Cf. les tableaux de tri des archives de chaque administration publique déjà ou prochainement réalisés par les Archives de l'Etat.

<sup>32</sup> Cf. note précédente.

## ■ Les responsabilités dans l'archivage des e-mails

La stratégie d'archivage des e-mails - et par extension l'ensemble des documents électroniques - doit être définie dans le contexte organisationnel de l'institution. Cette politique doit s'articuler autour de la définition des responsabilités par utilisateur (suivant le niveau hiérarchique auquel il appartient) lors de chacune des étapes de la sélection, de l'enregistrement et de la conservation des e-mails. L'institution devra ensuite définir et mettre en œuvre les moyens pour assurer une conservation optimale des e-mails sur la plan pratique et technique.

Les responsabilités dans les processus de sélection, d'enregistrement et de conservation des e-mails doivent idéalement être organisées en deux étapes. D'une part, les responsabilités en matière de sélection des e-mails (dits administratifs<sup>33</sup>) et d'autre part celles concernant l'enregistrement et la conservation des e-mails.

### La sélection des e-mails

- L'expéditeur (ou le destinataire) des messages sera responsable de la gestion de sa messagerie électronique, c'est-à-dire du tri et de la conservation éventuelle des messages. Cette tâche exige des utilisateurs une évaluation de la pertinence et de l'importance que revêt chaque message ainsi que les risques potentiels en cas de non-conservation. Chaque utilisateur est donc responsable du tri des messages à caractère privé et ceux à caractère professionnel (ce qui, en outre, permet de prévenir les problèmes liés au non respect de la législation sur la vie privée). C'est lui qui décidera de conserver ses e-mails professionnels en respectant les règles générales définies par l'institution et les recommandations des Archives de l'Etat.
- Chaque institution doit pouvoir établir et expliquer clairement les règles et les procédures en place pour la sélection des e-mails et le classement (notamment en cas d'envoi multiple ou par exemple quand il s'agit de note de service). Ces règles doivent impérativement être adoptées en conformité avec la législation en vigueur notamment celle liée aux archives publiques. A cet égard, un encadrement et une formation adaptée sont indispensables pour veiller à la bonne application des mesures prises au sein d'une institution. Il peut en effet arriver que l'utilisateur final ne suive pas (ou de manière incorrecte) les procédures de sélection et de conservation, qu'il construise son propre classement en dehors de tout lien avec la structure organisationnelle de l'institution ou encore qu'il procède mal à propos à la destruction de documents d'archives.

### L'enregistrement et la conservation des e-mails

- L'enregistrement devra être réalisé par l'expéditeur (ou le destinataire) des messages suivant des modalités devant être (pré)définies par l'institution. Une partie de cet enregistrement peut néanmoins se faire de manière (semi)automatisée sans que l'intervention de l'utilisateur final ne soit requise notamment pour la liaison des données de transmission des e-mails. Pour ce faire, l'aide d'un spécialiste de la gestion de l'information (records manager)<sup>34</sup> doit permettre la mise sur pied d'un cadre général de classement servant à organiser les différents documents enregistrés puis conservés (non seulement les e-mails mais également tous les autres documents électroniques). Dans ce cadre, concernant les éventuels codes de classement, l'utilisateur final est le mieux placé pour pouvoir déterminer le contexte de formation d'un e-mail entrant ou sortant. Il effectuera cette opération en plaçant les e-mails dans des dossiers spécifiques (de préférence au sein d'un classement numérique en dehors de la messagerie électronique - cf. *infra*) ou pourra, le cas échéant, les transmettre à un *record manager* qui aura pris soin de réaliser le classement de ces dossiers.
- La conservation (à court, moyen et long terme) des e-mails sera idéalement assurée voire coordonnée sur le plan logique et technique par un service spécifique différent de celui auquel appartient l'utilisateur. Il peut par exemple s'agir d'un spécialiste de la gestion de l'information (records manager), qui pourra au besoin se faire épauler pour la partie technique par le service informatique

<sup>33</sup> Ces e-mails administratifs (ceux à conserver) doivent pouvoir être identifiés dans les tableaux de tri des archives des institutions publiques (réalisés ou en cours de réalisation par les Archives de l'Etat).

<sup>34</sup> Les directives suivantes sur l'archivage numérique démontreront si besoin en est combien il est crucial qu'un tel spécialiste soit présent au sein de chaque institution publique pour assurer une gestion globale et homogène de tous les archives numériques (*digital records*) et donc pas uniquement des e-mails.

(ICT). Quel que soit le système mis en place (comme celui d'un archivage électronique des e-mails<sup>35</sup> - cf. chapitre suivant), il devra permettre le choix entre différents modes opératoires :

- soit les utilisateurs pourront classer (après sélection) leurs messages électroniques au sein d'un classement numérique des dossiers (enregistrement puis conservation éventuelle) ;
- soit un système d'archivage électronique (SAE) est prévu (conjointement élaboré par le spécialiste en gestion de l'information et le service informatique) qui comportera un procédé (semi)automatique de classement de tous les messages entrant et sortant. Si un tel système est mis en place, il doit permettre aux utilisateurs de procéder au classement des messages entrant (les messages sortant étant classés de manière quasi automatique) à partir de leur logiciel de messagerie électronique. L'utilisateur doit pouvoir classer chaque message de la boîte de réception au sein du système de messagerie tout en étant capable de voir chaque message avec signalement de pièces jointes (le cas échéant), voir les contenus de pièces jointes à l'aide d'un visualiseur multi-formats, enregistrer le message et ses pièces jointes comme nouveau document dans le système, lier le message et ses pièces jointes à un document déjà archivé dans le système. Ce système d'archivage électronique devrait en outre, et dans la mesure du possible, garantir la capture de l'adresse électronique sous une forme compréhensible par l'homme, associant un nom au message original (par exemple « Marc Durand » plutôt que « mdd@arch.be »)<sup>36</sup>.

## ■ Comment archiver ses e-mails ?

Puisqu'il est tenu de respecter et d'appliquer les règles d'archivage prévues par la législation en vigueur, chaque institution (et donc chaque fonctionnaire) se doit de conserver le texte original de l'e-mail ainsi que tout autre message lié formant un ensemble homogène (contexte de création). Au même titre que les autres documents d'archives, les e-mails doivent être enregistrés puis conservés dans leur contexte archivistique de production et doivent dès lors pouvoir être interprétés avec leurs pièces jointes. L'e-mail doit en outre pouvoir être conservé dans sa forme originale c'est-à-dire sous forme électronique, non seulement l'e-mail avec sa(s) pièce(s) jointe(s) mais aussi le(s) lien(s) avec tout autre document de quelque nature que ce soit qui traite de la même affaire ou concerne le même sujet. On constate d'emblée que la simple impression des e-mails ne constitue pas en soi un archivage adéquat pour ce type de document car elle restreint le nombre d'informations (notamment les métadonnées contextuelles) qui seront conservées. Mais comme au sein de nombreuses institutions il n'existe pas encore de processus adapté d'archivage numérique des e-mails, il est conseillé, à titre transitoire et provisoire, de conserver (après sélection) ses e-mails sur support papier. Dans ces cas précis, il est demandé à l'institution de tout mettre en œuvre pour déployer rapidement un archivage numérique des documents, non seulement pour les e-mails mais également pour tout autre document numérique créé ou reçu dans le cadre de ses activités.

Pour fournir aux utilisateurs la possibilité de capturer une sélection de messages et leurs pièces jointes, il est indispensable que l'institution mette à la disposition de chaque utilisateur des solutions techniques souples et efficaces<sup>37</sup>. Celles-ci devront être élaborées par un spécialiste en gestion de l'information (*records manager*) en collaboration avec le service informatique (ICT). Notons par ailleurs que le fait de placer ses e-mails de manière plus ou moins organisée dans des répertoires (*folders*) au sein du logiciel de messagerie électronique ne constitue pas de l'archivage *stricto sensu*. Outre son contexte de création, il est important de conserver l'ensemble du contenu et de la structure d'un e-mail (données d'en-tête, champ de l'objet du message, pièces jointes) ainsi que la mise en page qui dépend souvent du logiciel de messagerie électronique (client de messagerie) utilisé. Les données de transmission de l'e-mail - à savoir les métadonnées (*metadata*) - doivent pouvoir être également archivées. Ces métadonnées sont considérées comme étant les données pouvant identifier de manière unique un document numérique : un identifiant unique, le nom et l'adresse de

<sup>35</sup> Il s'agira de préférence d'un système d'archivage numérique (SAE) ou en anglais *Electronic Record Management System* (ERMS).

<sup>36</sup> *MoReq Specification, Model Requirements for the management of Electronic records*, Mars 2001 (traduction française novembre 2004) et *MoReq2* (2ème version publiée en 2008); pour les adresses dites de service, il sera utile de les documenter pour qu'elles restent compréhensibles dans la durée (ex : [digit\\_arch@arch.be](mailto:digit_arch@arch.be) devra être renseignée comme l'adresse professionnelle de la Section "Archivage électronique et Numérisation" des Archives de l'État).

<sup>37</sup> L'annexe de la directive sur l'archivage des e-mails présente quelques conseils de base pour développer une politique organisationnelle en vue de gérer des e-mails au sein d'une institution.

l'expéditeur, la date et l'heure de l'envoi, le nom et l'adresse de chaque destinataire (A, Cc, Cci), la date et l'heure de réception ainsi que le nombre de pièces jointes. Le principe de pérennité de la conservation est également essentiel dans l'archivage des e-mails. Pour anticiper d'éventuels problèmes ultérieurs, il est impératif de pouvoir maintenir une indépendance par rapport à des systèmes matériels (*hardware*) et logiciels (*software*) afin d'éviter d'importants écueils notamment celui de l'obsolescence.

## **Les différentes étapes de l'archivage des e-mails**

### ***Sélectionner les e-mails à conserver (cf. supra – Quand archiver ses e-mails ?)***

#### ***Sélectionner les métadonnées d'identification et les métadonnées d'intégrité***

Afin que les e-mails sélectionnés deviennent des documents d'archives complets et authentiques, il est indispensable qu'un certain nombre de métadonnées soit préservées lors de l'enregistrement dans un système d'archivage. Ces métadonnées doivent impérativement fournir des informations sur l'identité du document d'archive, sur le lien qu'il entretient avec son contexte de création ainsi que sur un certain nombre d'éléments techniques. Ces métadonnées peuvent être regroupées en deux catégories : les métadonnées d'identification - celles relatives à leur description et au contexte de leur formation - et les métadonnées dites d'intégrité (comprenant notamment les données techniques) - celles relatives à la leur transmission - qui devront toutes être enregistrées de manière structurée et indissociable du message.

- Les métadonnées d'identification offrent une description générale des e-mails et permettent de les replacer dans leur contexte de production : le nom de l'auteur du processus ou du responsable de l'archive, l'endroit dans le classement numérique où il est conservé (cf. point suivant) ainsi que, le cas échéant, la relation entre l'e-mail et sa(s) pièce(s) jointe(s). On pourra ajouter d'autres métadonnées (dites administratives) comme le délai d'utilité administrative ou le délai de conservation<sup>38</sup> ainsi que la destination finale de l'e-mail<sup>39</sup> (conservation ou élimination).
- Les métadonnées d'intégrité : l'adresse e-mail de l'expéditeur (le cas échéant, nom et adresse du mandataire), la date et heure de l'envoi ou de la réception ainsi que la composition de la correspondance. On peut y ajouter certaines métadonnées dites techniques comme la version du logiciel de messagerie électronique utilisée renseignant ainsi le format dans lequel chaque e-mail sera enregistré (qui sera utile notamment en cas de migration ultérieure).

Il est conseillé d'enregistrer les métadonnées descriptives (de transmission et contextuelles) quand les e-mails sont encore présents dans le logiciel de messagerie électronique. Idéalement, ce processus devrait avoir lieu le plus rapidement possible après l'envoi ou la réception d'un courrier. Une des méthodes à privilégier est celle de l'encapsulation des métadonnées avec le message lui-même de telle façon qu'elles deviennent une composante de l'e-mail lui-même. Cette méthode met par ailleurs en évidence le problème qui est constaté lorsqu'un e-mail est imprimé puis conservé sous forme papier car la plupart des métadonnées ne sont pas préservées avec le message. Cela peut partiellement remettre en cause l'authenticité voire l'intégrité du message lui-même.

### ***Exporter les e-mails (et leurs pièces jointes) du logiciel de messagerie électronique***

La structure de dossiers créée dans le logiciel de messagerie électronique doit être considérée comme temporaire et certainement pas comme un dépôt permanent pour des documents d'archives. Plusieurs raisons peuvent être évoquées. Tout d'abord, cette utilisation pose d'évidents problèmes de durabilité et de pérennité compte tenu du format d'exportation propriétaire de certains logiciels de messagerie électronique

<sup>38</sup> Le délai d'utilité administrative constitue la période durant laquelle l'e-mail présentera une valeur administrative pour le formateur d'archives (l'institution). Le délai de conservation est quant à lui la période pendant laquelle les e-mails seront conservés et au terme de laquelle sera appliquée la décision concernant la destination finale.

<sup>39</sup> La destination finale constitue le sort réservé à l'e-mail dès que le délai de conservation est écoulé. Elle peut soit être l'élimination soit la conservation (durée illimitée).

contraignant l'utilisateur et son institution à utiliser les mêmes logiciels (ou leurs dérivés) à plus ou moins long terme. En outre, les possibilités d'archivage proposées par certains de ces logiciels conduisent souvent à des pertes d'informations consécutives aux compressions des fichiers. Une autre raison est que cette utilisation pose un problème d'accessibilité car chaque utilisateur gérant individuellement les e-mails de sa boîte aux lettres électronique est, en principe, le seul à y avoir accès. Cette utilisation pose également un problème au niveau de l'espace de stockage. Compte tenu du volume d'e-mails échangés, les dossiers créés au sein des logiciels de messagerie électronique peuvent ralentir l'accès au serveur. Les dossiers « *offline* », de par leur grande étendue, sont quant à eux facilement corruptibles et donc sujets à une instabilité et à un manque de fiabilité. Du point de vue archivistique, la solution réside donc dans le choix d'exporter ses e-mails du logiciel de messagerie électronique. Même si le lien avec le serveur mail est rompu (ce lien permet notamment de générer automatiquement les adresses e-mails après l'encodage des noms de personne s'ils sont présents dans le carnet d'adresses) et que les messages archivés sont rendus « statiques » (donc en théorie non réutilisables), cette solution rend possible la liaison des e-mails avec d'autres documents électroniques ayant trait à la même activité ou au même processus de travail. Cela permet notamment d'accéder en une seule requête à l'ensemble de l'information (e-mails et documents numériques de toute type) relative à un sujet ou une affaire traitée. Quel que soit le système d'archivage envisagé, les documents rassemblés dans une structure de dossiers commune en dehors du logiciel de messagerie électronique assurent une gestion centralisée et offrent de nombreux avantages notamment en matière de sécurité (*back-up*), d'accessibilité et de durabilité. Cette solution trace également la voie vers une gestion archivistique contrôlée et centralisée. Par ailleurs, des solutions commerciales d'archivage des e-mails existent. Elles basent leurs applications au sein du logiciel de messagerie électronique en offrant des solutions *user-friendly* et sophistiquées en matière de recherche mais sont payantes et rendent l'utilisateur tributaire dans le choix des logiciels de messagerie électronique. Anticipant en cela sur les évolutions futures conduisant à l'intégration de l'ensemble de l'information d'une institution dans des systèmes de *records management*, un grand nombre d'entre-elles se tournent déjà vers la mise en place de système de classement numérique de leurs dossiers de travail afin d'organiser au mieux la gestion des documents. Deux solutions sont possibles : l'une intermédiaire qui consiste à la mise en place d'un classement numérique par le biais du système d'exploitation et l'autre globale et intégrée qui consistera à la mise en place d'un système d'archivage électronique - un *records management system* - également basé sur un classement numérique.

### **Organiser le classement de ses e-mails (et de ses pièces jointes)**

Les e-mails et leurs pièces jointes seront de préférence conservés séparément ce qui les rendra plus aisément identifiables et réutilisables. De plus, les pièces jointes se présentant sous des formats très différents (texte, photos, audio, vidéo, etc.), leur enregistrement nécessitera un archivage spécifique suivant le format utilisé. Mais cette solution est difficile à mettre en place car elle engendre des procédures fastidieuses qui conduiront inévitablement à un manque de suivi. Il apparaît donc nécessaire de mettre sur pied des solutions sur mesure qui n'alourdiront pas le travail des utilisateurs.

#### **Solution intermédiaire : classement numérique interne au système d'exploitation**

La création et l'organisation de dossiers électroniques suivant un plan de classement propre à l'institution permet de structurer l'ensemble de ses documents numériques en ce compris les e-mails. Le classement des e-mails et de tout autre document numérique sont généralement basés sur les tâches et les activités du formateur d'archives afin de garantir la compréhension et la réutilisation de l'ensemble des documents ainsi conservés. Il est essentiel que l'institution accorde une attention particulière lors de la mise en place du classement numérique de ses dossiers. Dès lors, la structure du classement numérique des dossiers doit être élaborée en étroite collaboration avec tous les groupes concernés au sein de l'institution ainsi qu'avec la personne désignée comme responsable de l'archivage (idéalement un spécialiste en gestion de l'information). Ce classement pourra soit suivre le modèle structurel de l'institution (par direction, service, etc.) soit celui fonctionnel (par nature d'activité). A ce titre, il est d'ailleurs conseillé de se baser sur le classement déjà utilisé pour la gestion des documents papier ainsi que sur les tableaux de tri réalisés ou à réaliser par les Archives de l'Etat pour les institutions publiques belges soumises à la surveillance de leurs archives. La mise en place d'un tel classement doit répondre à quelques principes de base afin qu'il soit compréhensible du plus

grand nombre<sup>40</sup>. Le classement doit être logique et facilement utilisable, basé sur l'organisation des tâches et des activités du formateur d'archives (suivant le modèle structurel ou fonctionnel), construit du général au particulier (d'abord affaires internes puis externes), être de préférence identique au classement des documents sur support papier, établi suivant des codes structurés (basé éventuellement sur celui des documents papier), être limité en terme de niveau de classement (éviter d'aller au delà de 5 niveaux hiérarchiques), utiliser des noms ou des abréviations claires et explicites en se rapportant directement aux tâches et aux activités, éviter de réutiliser les mêmes noms de dossiers (notamment pour des dossiers inférieurs), respecter la norme ISO-9660 qui impose l'utilisation de maximum 31 caractères pour les noms des dossiers, pas d'espace entre les mots (soit les mots seront collés les uns aux autres soit un tiret bas ou *underscore* est utilisé comme séparateur) et l'utilisation exclusive des caractères « A-Z », « 0-9 » et « \_ », s'accorder sur une utilisation uniforme des acronymes, des abréviations ou des contractions (le cas échéant, les renseigner et les documenter) et enfin dénommer les versions des documents (V01, V02, etc.).

### ***Solution globale et intégrée : la mise en place d'un système d'archivage électronique ou de records management (SAE/ERMS)***

La mise en place d'un système d'archivage électronique (SAE) conduit à envisager la conservation des documents d'archives que constituent les e-mails au sein d'un système de *records management*. Dans le SAE, également connu sous l'acronyme ERMS (*Electronic records Management System*), le terme « archivage » doit être replacé dans le cycle de vie du document et non dans le sens réducteur de « stockage de fichiers ». Trois grands standards régissent la mise en place de tels systèmes : la norme internationale ISO 15489 (1&2) ayant comme objectif la mise en place de pratiques et de procédures normalisées pour une gestion des documents d'archives (*records*) afin de leur garantir des valeurs de preuve, le standard américain DoD 5015-2 mis à point par le département de la Défense des États-Unis qui constitue la référence pour la définition des fonctionnalités des applications en matière d'archivage électronique, et la spécification européenne MOREQ, précisant les exigences de l'archivage électronique à des fins de preuve. L'utilisation des normes ISO 15489(1&2) et DoD 5015-2 et des spécifications de MOREQ constitue une base solide pour la mise en œuvre d'un système d'archivage électronique au sein d'une institution. En outre, l'étude de la norme ISO 14721:2002 - plus connue sous l'acronyme OAIS<sup>41</sup> - peut s'avérer être d'un grand secours notamment dans l'établissement d'un cadre descriptif général des fonctions, des responsabilités et de l'organisation d'un système qui voudrait préserver sur le long terme de l'information (plus particulièrement des données numériques) et en garantir l'accès à des utilisateurs identifiés.

La mise en pratique de ces normes doit permettre à chaque institution désirant mettre en place un système d'archivage - notamment de ses e-mails - d'effectuer les choix opportuns. En outre, les spécificités propres aux systèmes d'archivage - qu'ils soient de gestion électronique des *records* (ERMS ou SAE) ou de gestion électronique de documents (EDMS ou GED) doivent être attentivement analysées afin de retenir les fonctionnalités les plus adaptées en fonction de l'utilisation souhaitée. Il est à noter qu'un système d'archivage électronique (SAE) offre la possibilité de lier les *records* à une activité ou les *records* entre eux tandis qu'un système de gestion électronique (GED) de documents doit pour le permettre mettre les e-mails au préalable en relation avec un système de classement de fichiers.

De manière générale, tout système d'archivage électronique doit suivre les principes suivants :

- au niveau de la conservation, les e-mails doivent demeurer intacts et doivent donc être conservés dans des systèmes de documents identifiables afin de préserver leur intégrité et de pouvoir les reproduire et les visualiser dans leur format d'origine ;
- au niveau du classement des *records*, l'ensemble des documents électroniques (en ce compris les e-mails) doivent respecter un classement de fichiers qui permet de maintenir les liens entre tous les documents (comme les messages avec leurs pièces jointes ou avec tout autre document connexe) ;
- savoir quand il est impératif d'archiver un e-mail (ou tout autre document électronique) car il ne faut pas conserver un même *record* sous plusieurs formats ;

<sup>40</sup> Ces éléments de recommandations sont largement inspirées de celles préconisées dans le cadre du projet e-David : Directive e-David (Richtlijn & advies nr. 3).

<sup>41</sup> *Open Archival Information System* est un modèle conceptuel destiné à la gestion, à l'archivage et à la préservation à long terme de documents numériques.

- éviter de conserver (trop longtemps) les e-mails dans des répertoires ou des fichiers privés car cela en restreint l'accès ;
- imprimer les e-mails si il n'y a pas d'autre solution à court terme (pas de système électronique de gestion des records par exemple) et si et seulement si les informations connexes au message (métadonnées) sont conservées elles-aussi (éventuellement en les imprimant).

### **Enregistrer les e-mails (et ses pièces jointes) dans des formats d'archivage pérenne**

Le choix des formats d'exportations pourra s'avérer crucial lors de l'archivage définitif des e-mails car ils permettront - ou non - de conserver l'entièreté des informations tant du message lui-même que celles liées au contexte (métadonnées). La plupart des logiciels de messagerie électronique offrent la possibilité d'utiliser plusieurs formats d'exportation comme .EML, .TXT, .HTML, .MSG, .OFT, ou RTF pour ne reprendre que les principaux (cf. chapitre sur la structure et le format des e-mails). Ce choix n'est cependant pas évident quand on connaît la rapide obsolescence des formats informatiques. Nous recommandons néanmoins d'utiliser prioritairement des formats aussi ouverts que possibles, souples, non-propriétaires (non liés à quelques logiciels de messagerie électronique) et, le cas échéant, permettant de structurer les données enregistrées avec les messages. L'institution se doit dès lors de définir un format d'exportation qui sera idéalement le même format utilisé pour l'archivage. Il faut néanmoins savoir que dans la pratique, certains formats d'archivage pour des documents textuels structurés (comme XML, PDF ou PDF/A<sup>42</sup>, ODT<sup>43</sup> ou TIFF) ne peuvent, à l'heure actuelle, être utilisés avec le logiciel de messagerie électronique notamment pour répondre ou envoyer un message<sup>44</sup>. Les Archives de l'Etat recommandent la prudence dans le choix du format d'enregistrement des e-mails en reconnaissant que le format idéal pour l'archivage définitif n'existe pas encore. Les alternatives, comme les formats TIFF, PDF (voire PDF-A), ODT existent mais, contrairement au format XML<sup>45</sup>, ils ne permettent pas une structuration aussi souple des données et des métadonnées. Le format XML est à ce titre une voie tout à fait adaptée pour la conservation structurée des e-mails ainsi que de leurs données contextuelles. Ce format, outre sa grande souplesse d'utilisation et d'extension, peut également être généré à partir de certains formats de sauvegarde existants comme TXT, EML et RTF notamment. Dès lors, son utilisation peut être envisagée comme solution pour l'archivage des e-mails. Si ce choix est considéré, il s'agira de procéder à une conversion du message avant tout archivage. Cette conversion consiste principalement en l'ajout de balises (*tags*) aux différents champs de données du message ainsi qu'à la structuration des éléments. Le résultat est un fichier qui comprendra une combinaison des métadonnées et du contenu de l'e-mail dans un format XML<sup>46</sup>. Comme il n'existe pas de solution au sein de la plupart des logiciels de messagerie électronique actuels répondant aux exigences archivistiques, une solution *ad hoc* devra être préconisée. Il conviendra dès lors d'utiliser des logiciels de conversion ou d'élaborer des fonctionnalités supplémentaires au sein du logiciel de messagerie électronique<sup>47</sup>.

Pour les formats d'archivage des pièces jointes, il est essentiel de distinguer chaque document numérique et ses propriétés spécifiques. En effet, chaque type de document numérique exigera des modalités d'archivage spécifiques tant du point de vue du format choisi que des métadonnées à conserver. C'est pour cela que les e-mails et leur(s) pièce(s) jointe(s) seront idéalement séparés les uns des autres lorsqu'ils sont conservés en

---

<sup>42</sup> PDF/A : Portable Document Format/Archive - spécification de l'ISO utilisée par les organismes de normalisation du monde entier pour garantir la sécurité et la fiabilité de la diffusion et des échanges de documents électroniques.

<sup>43</sup> ODT : Open Document Text.

<sup>44</sup> Certains systèmes développés récemment permettent toutefois de pouvoir consulter les e-mails archivés y compris les pièces jointes et tout autre document électronique lié (cf. projet DISSCO).

<sup>45</sup> Standard ouvert du consortium WWW.

<sup>46</sup> Ces fichiers seront de préférence sauvegardés sur un serveur à part ; un contrôle doit également être effectué pour s'assurer que l'enregistrement est conforme au schéma XML. Il peut être par ailleurs envisagé d'utiliser un *XSL stylesheet* pour la conversion de XML à HTML (pour l'affichage web notamment) avec contenu de l'e-mail, les métadonnées ainsi que les données sur lay-out de l'e-mail. Les transformations XSLT (Transformations XSL - *eXtensible Stylesheet Language Transformations*) ont pour objet de convertir un fichier XML d'un format de document à un autre. Par exemple pour afficher un document XML sur un navigateur web, en le convertissant en XHTML. Elles permettent ainsi une mise en forme des données plus adaptée à l'impression ou à l'affichage sur un ordinateur, une extraction automatique des éléments utiles d'un document, pour en générer un résumé ou une table des matières et une conversion du contenu d'un document en un format plus aisément manipulable.

<sup>47</sup> Des solutions plus concrètes seront prochainement testées aux Archives de l'Etat et seront proposées aux institutions publiques. En attendant, des solutions - comme celles émanant du projet *e-David* à Anvers (<http://www.expertisecentrumdavid.be/>) ou de la plateforme *Digitale duurzaamheid* des Pays-Bas (<http://www.digitaleduurzaamheid.nl/>) - offrent déjà des réponses adaptées pour l'archivage des e-mails.

dehors du logiciel de messagerie électronique<sup>48</sup> tout en conservant le lien qui les unit. Ils seront donc exporter et enregistrer de préférence dans leurs formats d'origine et dans un second temps seront migrés vers des formats d'archivage adaptés.

Quant aux solutions commerciales présentes sur le marché (quelques-unes d'entre-elles ont été analysées dans le cadre de cette directive), elles n'apportent rien de plus qu'un outil efficace de gestion - parfois de conservation - de ses e-mails et ne sont donc pas adaptées aux objectifs d'un archivage numérique définitif.

Nous présentons dans le tableau ci-dessous quelques exemples de formats d'archivage de documents numériques<sup>49</sup>. Compte tenu des évolutions technologiques, cette liste fera l'objet de mises à jour fréquentes.

Types de document	Formats d'archivage envisageables
Document texte	Structuré : XML, ODT, (OOXML) Non-structuré : PDF/A, TXT
Tableur (graphique)	Réutilisable : XML, ODS, CSV, (OOXML) Non-réutilisable : PDF/A
Base de données	Réutilisable : XML, CSV, (OOXML) Non-réutilisable : PDF/A
Document de présentation	Réutilisable : ODP, (OOXML) Non-réutilisable : PDF/A
Image	- matricielle : TIFF 6.0 (non compressé), JPEG2000, PNG 1.0 - vectorielle : CGM, SVG
Audio	WAV (PCM non compressé)
Vidéo	AAF ou MXF
Dessin (CAO)	DXF

<sup>48</sup> Cf. la prochaine directive sur l'archivage numérique des fichiers bureautiques (en préparation).

<sup>49</sup> Les formats de conservation des documents numériques évoluant rapidement, il est conseillé de se référer aux recommandations des Archives de l'Etat (cf. notamment la directive en préparation sur l'archivage des documents numériques dits « bureautiques » (documents textes, tableurs graphiques, bases de données, documents de présentation).



## ■ Conclusion

Les recommandations présentées dans la première directive sur l'archivage numérique des e-mails ouvrent la voie vers une meilleure gestion (électronique) des e-mails et de leur archivage approprié. Ils nécessiteront des compléments plus pratiques et plus synthétiques dans le but d'offrir aux institutions publiques belges des guides d'usage mieux calibrés en fonction de leurs spécificités.

Dans le cadre de la mise en place d'un archivage numérique des e-mails, chaque institution doit d'ores et déjà définir une politique globale et structurée en vue de conserver - provisoirement ou définitivement - ses e-mails. Elle doit également mettre sur pied un dispositif prévoyant l'entretien régulier des données (supports et formats) en prévision de leur transfert vers les dépôts numériques des Archives de l'Etat.

Chaque institution devra progressivement faire évoluer les pratiques actuelles vers plus de rigueur notamment en préparant des directives internes pour un archivage approprié des documents électroniques. L'archivage électronique ne doit plus être apparenté à une simple opération informatique de sauvegarde (*back-up*). C'est un processus spécifique complexe devant mettre en œuvre une organisation, des méthodes et des outils pour exploiter et conserver des informations - quel que soit le support - sur le moyen et le long terme.

L'archivage des documents électroniques - et donc des e-mails - peut apparaître comme un véritable défi pour chaque institution publique mais la garantie de son succès doit impérativement passer par l'implication et la collaboration de chacun. Le résultat attendu n'est pas seulement de garantir la survie des documents et des informations qui y sont stockées mais également et surtout d'assurer à chaque agent une amélioration de la qualité des processus de travail tout en veillant à simplifier au maximum les procédures mises en place.

Avec le présent outil, nous espérons pouvoir sensibiliser les acteurs des institutions publiques sur l'importance de la gestion et de la conservation des archives électroniques car « *les archives de l'avenir seront électroniques ou ne seront pas...* »<sup>50</sup>

---

<sup>50</sup> Cité le 2 octobre 2007 par Laurent Moutinot, Vice-président du Conseil d'Etat (Suisse).

# ARCHIVAGE DE LA BANQUE D'INFORMATION RADIOPHONIQUE DES DEUX-SÈVRES (BIRDS)

---

Brigitte Pipon

Directrice des Archives départementales des Deux-Sèvres

**Présentation** : La BIRDS, créée par le Conseil général des Deux-Sèvres en mai 2005, recueille les émissions (journaux, magazines,...) des quatre radios associatives du département ainsi que les émissions réalisées par le Conseil général pour permettre un partage et un échange des données. Le serveur, mis à disposition par la communauté de communes de Parthenay, stocke ces données une année durant. Une convention passée pour 10 ans entre le Conseil général et les radios associatives fixe les conditions d'utilisation des informations versées dans la BIRDS. La deuxième partie concerne le don des documents aux Archives départementales.

**Collecte** : L'archivage définitif de ces données est effectué par les Archives départementales qui téléchargent chaque nuit les nouvelles émissions à partir du serveur de Parthenay. À chaque fichier son correspond une fiche descriptive remplie par l'auteur de l'émission (ill. 1). Aucun support matériel n'est utilisé, les différentes opérations s'effectuant par télétransmission. Un bordereau de versement, conforme aux recommandations de la DAF (standard d'échange) est généré chaque jour et un accusé de réception envoyé au service émetteur. La mise en place de ces éléments a été réalisée en collaboration avec la Direction des Archives de France (Département de l'innovation technologique et de la normalisation).

**Traitement** : Les fiches descriptives sont vérifiées et corrigées, si besoin est, par les Archives départementales qui vérifient également que les fichiers son (MP 3) s'ouvrent correctement et que leur transmission a été correcte (ill. 2). Téléchargement et vérification s'opèrent via une application spécifiquement développée en interne par le service informatique du Conseil général (ill. 3). Aucun tri sélectif n'est pour l'instant pratiqué. Le fonds BIRDS est coté en 14 AV. À chaque radio correspond une sous-série, de 14 AV 1 à 14 AV 4, puis les émissions reçoivent un numéro d'ordre (exemple : 14 AV 1 / 1).

**Conservation** : Les données sont stockées sur un serveur situé aux Archives départementales. Une sauvegarde différentielle s'effectue chaque jour sur un deuxième serveur, situé dans des locaux du Conseil général ; s'y ajoute une sauvegarde complète tous les week-ends et tous les mois. En fin d'année, l'ensemble des fichiers fait l'objet d'une copie sur disque dur externe, disque conservé aux Archives départementales.

**Communication** : Les fichiers sont communiqués en salle de lecture via le logiciel Action Archive qui permet à la fois la recherche documentaire par mot clé et l'écoute des émissions. Trois postes informatiques sont équipés de casques audio. Le site Internet du Conseil général ([www.deux-sevres.com](http://www.deux-sevres.com), rubrique Culture et patrimoine) offre l'écoute d'une semaine de journaux et magazines. La reproduction des émissions est autorisée pour le strict usage privé mais soumise à autorisation de la radio pour un usage public. Au 23 mai 2008, 4.192 émissions étaient disponibles (ill. 4).

Cette première expérience de gestion de documents numériques par les Archives départementales a été riche dans la réflexion méthodologique tout en bousculant certains critères archivistiques traditionnels :

- versement quotidien avec génération automatique de bordereau,
- fonds mêlant archives privées et archives publiques,
- intégration de l'instrument de recherche dans le logiciel documentaire sur le même plan que les autres fonds avec consultation directe.

[Consulter la présentation : partie 1 ; partie 2.](#)

## **Archiving sound files from local radios in the Deux-Sèvres département**

The Banque d'information radiophonique des Deux-Sèvres (BIRDS) was set up in May 2005 by the Conseil Général to archive broadcasts (such as news and magazine programmes) made by the four voluntary radio stations in the Deux-Sèvres Département together with programmes made by the Conseil Général, to allow them to be shared and exchanged. The server, which is provided by the community of municipalities of Parthenay, stores the data for 12 months.

A ten-year agreement between the Conseil Général and the voluntary radio stations sets the conditions of use of the information stored in the BIRDS. The agreement also covers donation of documents to the Deux-Sèvres Archives.

### **Collection**

Final archiving of these data is performed by the Deux-Sèvres Archives, which downloads the new broadcasts every night from the Parthenay server. Each sound file has a corresponding description file completed by the creator of the broadcast (Figure 1). No physical media are involved, as all operations are performed online. An acceptance note is generated every day in accordance with the transfer standards recommended by the French Archives authority (DAF) and an acknowledgement is sent to the broadcaster. These procedures were implemented in collaboration with the Technological Innovation and Standards Department of the DAF.

### **Processing**

The description files are checked and corrected, if necessary, by the staff of the Deux-Sèvres Archives who also verify that the MP3 sound files open correctly and have been properly transferred (Figure 2). Downloading and verification are handled by a specially-developed application written by the Conseil Général's IT department (Figure 3). No sorting is applied as yet. The BIRDS collection is classified under 14 AV. Each radio station has a sub-sub-series number, from 14 AV 1 to 14 AV 4, with each broadcast being given an order number (e.g. 14 AV 1/1).

### **Storage**

The data are stored on a server located at the Deux-Sèvres Archives. A differential backup is made daily to a second server, located on the premises of the Conseil Général; a full backup is made every weekend and every month. At the end of each year all files are copied to an external hard disk, kept by the Deux-Sèvres Archives.

### **Consultation**

The files can be consulted in the reading room using Action Archive software which allows the user to search for a document by keyword and listen to the broadcasts. Three workstations are fitted with audio headsets. News programmes and magazines are also available for one week on the Conseil Général's website ([www.deux-sevres.com](http://www.deux-sevres.com), 'Culture et patrimoine' section). Broadcasts may be reproduced for private use only, but authorisation from the corresponding radio station is required for public use of any broadcasts. As of 23 May 2008, 4,192 broadcasts were available (Figure 4).

This first experience of the management of digital documents by the Deux-Sèvres Archives has proved very stimulating in terms of rethinking our methodology, while revolutionising certain traditional criteria concerning archiving:

- daily input with automatic generation of description sheet,
- repository with mixed private and public archives.

[See the presentation: part 1 ; part 2.](#)

## ■ Genèse du projet

Dans le département des Deux-Sèvres, quatre radios locales associatives existent, qui couvrent chacune un territoire défini (fig. 1) :

- D4B, la plus ancienne (1981), basée à Melle, couvre le sud du département (Niortais, Mellois)
- Radio Gâtine à Parthenay couvre le centre du département (Gâtine)
- Radio Val-d'Or à Airvault couvre le nord-est (Thouarsais)
- Collines FM à Cerizay couvre le nord-ouest (Bressuirais, Mauléonais).

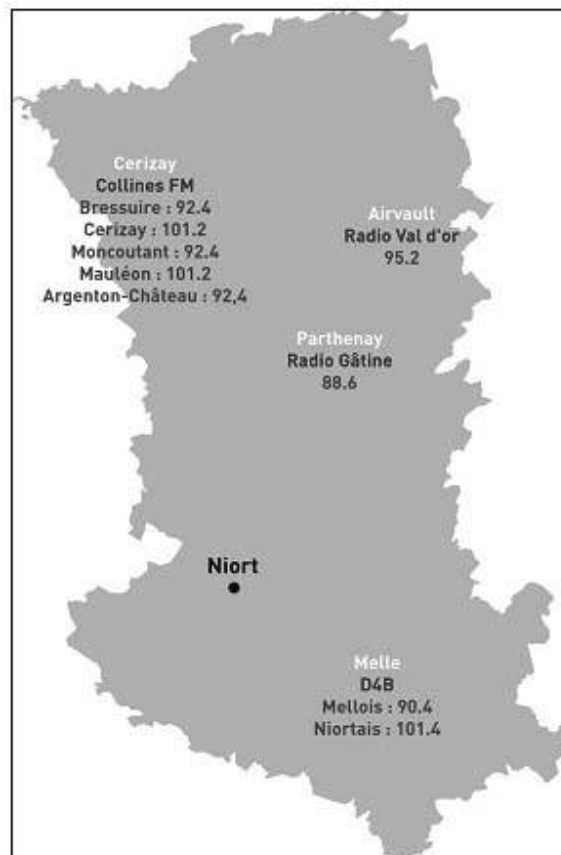


Fig. 1 : Carte des Deux-Sèvres présentant l'implantation des 4 radios locales (DAO : Service Communication du Conseil général des Deux-Sèvres).

Elles produisent journaux d'informations, magazines, reportages,... menant un travail de terrain sur des territoires éloignés les uns des autres. Soucieux de s'appuyer de manière pérenne sur leur capacité de mobilisation au profit de l'animation, du développement et de l'aménagement du territoire départemental, le Conseil général propose aux radios, à la fin de l'année 2004, de mettre à leur disposition un serveur informatique, baptisé BIRDS (banque d'information radiophonique des Deux-Sèvres), pour permettre l'échange et le partage de données.

Cinq enjeux sont posés :

- assurer la pérennité de l'information en la rendant accessible à tous ;
- engendrer des économies en faisant gagner du temps et des frais de déplacement, en multipliant les sources de production d'information, en imaginant des actions communes productrices de recettes ;

- renforcer la cohésion départementale par les échanges d'émissions, de sorte que l'information puisse être mutualisée, conservée et disponible pour l'ensemble des deux-sévriens ;
- promouvoir l'information départementale par la diffusion notamment des émissions institutionnelles et de l'agenda des fêtes et manifestations ;
- organiser la conservation et l'archivage des émissions.

Pour ce qui est des engagements réciproques :

- chaque radio met son agenda en ligne à destination des autres radios, offre à la BIRDS ses journaux quotidiens et un choix de modules ou de magazines et met en œuvre le plan annuel de l'information départementale ;
- le Conseil général accorde une subvention annuelle aux radios, plafonnée à 25 000 €, met en place le serveur BIRDS, produit certaines émissions avec son studio ; enfin il organise le référencement des informations sonores stockées et prend en charge la conservation définitive via les Archives départementales.

La BIRDS est mise en service en mai 2005. Le serveur, situé à Parthenay, grâce à une collaboration avec la Communauté de communes, stocke les données une année durant (puis elles sont effacées). Chaque radio envoie au serveur BIRDS par internet ses journaux quotidiens et un choix de modules ou de magazines. Les autres radios peuvent aller les récupérer sur le serveur et les diffuser. Le site internet du Conseil général les met à disposition du public pendant huit jours ([www.deux-sevres.com](http://www.deux-sevres.com), rubrique Culture et patrimoine).

Si les radios s'engagent à mettre quotidiennement en ligne leurs informations, elles restent libres des contenus et de ce qu'elles acceptent de diffuser comme modules et magazines : l'indépendance de chaque radio est ainsi respectée. Après trois années d'expérimentation, les radios s'accordent sur le bilan positif qui se dégage - coopération des équipes, coordination des sujets, économies de frais- même si, dans la réalité, certaines radios « jouent le jeu » plus que d'autres dans le partage des informations. Le pourcentage de reprise des émissions s'élève à plus de 50 %, une même émission pouvant être reprise par plusieurs radios. La BIRDS joue ainsi un rôle de « service public », à ce jour unique en France semble-t-il.

En 2006, le Conseil général apporte aux radios un service supplémentaire : un studio d'enregistrement à Niort et une équipe chargée de créer des émissions citoyennes dont les sujets, à dimension départementale, complètent l'offre de programmes des quatre radios. En concertation avec celles-ci, le plan d'émissions est élaboré et validé et les émissions réalisées sont versées dans la BIRDS. Quinze jours après, elles basculent sur le site internet du Conseil général où elles restent accessibles pendant un an.

Une convention passée le 7 janvier 2005 entre le Conseil général et les radios s'attache dans sa première partie aux modalités de la cession des droits d'auteurs relatifs aux informations versées dans la BIRDS : les radios s'autorisent mutuellement à reprendre les informations qu'elles ont versées et à les réutiliser, ainsi que le Conseil général, dans « le cadre de leurs activités professionnelles journalistiques », en indiquant, dans tous les cas, « le nom de l'auteur de l'information » (art. 5).

D'emblée, et il s'agit d'un but affiché dès l'origine, il est prévu que les documents sonores soient conservés aux Archives départementales. Le « don aux Archives départementales » fait l'objet de la deuxième partie de la convention. Tous les documents versés dans la BIRDS sont concernés ; les Archives départementales procéderont « à une sélection des documents qui seront conservés [...]. Les autres documents seront détruits et un bordereau des documents éliminés sera établi » (art. 9).

## ■ Mise en œuvre de l'archivage des émissions

L'archivage définitif étant décidé, il fallait le mettre en œuvre. Jusqu'alors, les Archives départementales des Deux-Sèvres n'avaient eu aucune expérience de stockage de données nativement électroniques. Réunions et concertations eurent lieu avec nos différents partenaires : le service informatique du Conseil général - dont dépendent les Archives départementales -, le service informatique de la Communauté de communes de Parthenay et les quatre radios. Parallèlement contact était pris avec le Département de l'innovation

technologique et de la normalisation de la Direction des archives de France<sup>1</sup> afin de mettre en application le « standard d'échange »<sup>2</sup>.

Il fallait par ailleurs développer une application spécifique afin que les Archives départementales puissent procéder à la vérification des données transférées.

La réflexion s'engage en septembre 2005 pour aboutir au mois de mai suivant par les premiers téléchargements.

## Modalités du transfert des données

Chaque jour les Archives départementales téléchargent les nouvelles émissions radio à partir du serveur de Parthenay : l'opération s'effectue automatiquement via internet. À chaque fichier son (format MP3) est attachée une fiche descriptive (format base de données) (fig. 2) mise au point par le service Documentation du Conseil général. Étant donné que cette fiche est renseignée par l'auteur de l'émission et est obligatoire, elle a été conçue d'utilisation aisée : nombre de champs restreint, dont quelques uns indispensables signalés par un astérisque, menus déroulants. Un dernier champ « Commentaire » permet d'apporter d'autres précisions de contenu. Les champs « Personnalité » et « Organisme » concernés par l'émission sont facultatifs.

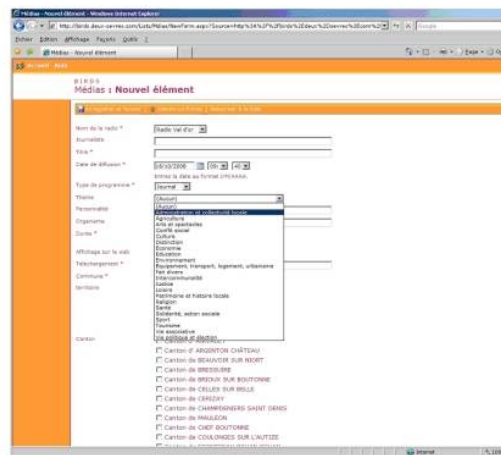


Fig. 2 : Fiche descriptive des émissions radios.

*(Cliquez sur l'image pour la voir en pleine taille)*

Les champs obligatoires sont :

- le nom de la radio
- le titre de l'émission
- la date de diffusion (en réalité il s'agit parfois de la date de création)
- le type de programme (journal, magazine...)
- le thème (avec liste déroulante d'une vingtaine d'occurrences)
- la durée
- le chemin de téléchargement
- la commune (avec liste déroulante des communes du département).

Un bordereau de transfert (fig. 3), au format du standard d'échange, est généré à partir de ces éléments : il contient un aperçu global des transactions, dont les coordonnées des services émetteurs et récepteurs, puis les informations contenues dans les fiches descriptives.

<sup>1</sup> Je remercie tout particulièrement Claire Sibille et Olivier de Solan pour le précieux concours qu'ils m'ont apporté.

<sup>2</sup> Instruction DITN/RES/2006/001 du 8 mars 2006 de la direction des Archives de France sur le « standard d'échange de données pour l'archivage ».

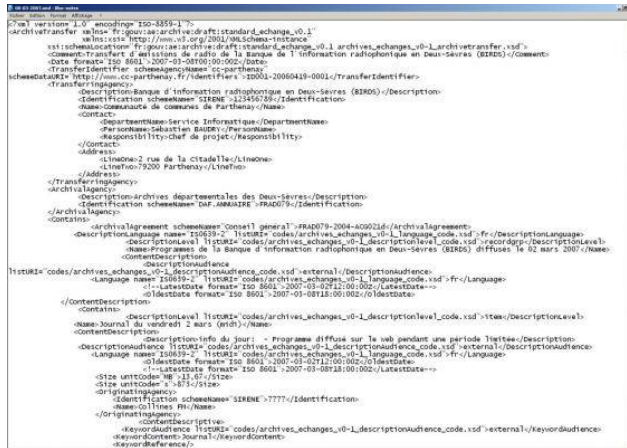


Fig. 3 : Bordereau de transfert, au format du standard d'échange.

*(Cliquez sur l'image pour la voir en pleine taille)*

Il s'agit donc d'une description des émissions à la pièce ; les métadonnées sont alimentées directement à partir de la base de données. Pour l'accusé de réception, a été retenue une formule simplifiée, sous forme d'un fichier Log (fig. 4), indiquant le numéro d'ordre, la date et le nom de l'émission ainsi que la date de téléchargement et son issue (« réception terminée » ou « erreur »). L'accusé de réception, envoyé au service émetteur, est conservé également aux Archives départementales. Le bordereau de transfert est généré après correction des fichiers son et texte ; il est identifié par la date (jj-mm-aaaa) suivie de .xml.

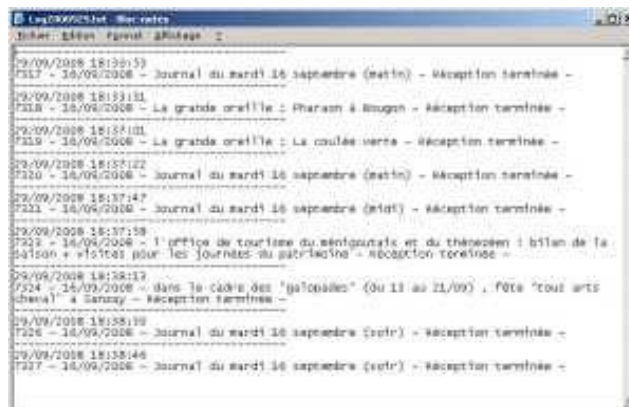


Fig. 4 : Accusé de réception sous forme de fichier log.

*(Cliquez sur l'image pour la voir en pleine taille)*

## Traitement

Les fiches descriptives sont vérifiées et corrigées, si besoin est, par les Archives départementales qui vérifient également que les fichiers son s'ouvrent bien et que leur transmission a été correcte. Dans le cas inverse, un nouvel essai de téléchargement est effectué.

Vérification et téléchargement s'opèrent via une application spécifiquement développée en interne par le service informatique du Conseil général, sur une base Access (fig. 5 et 6). Un manuel Utilisateur a été mis au point, à l'usage de l'agent chargé de la vérification.

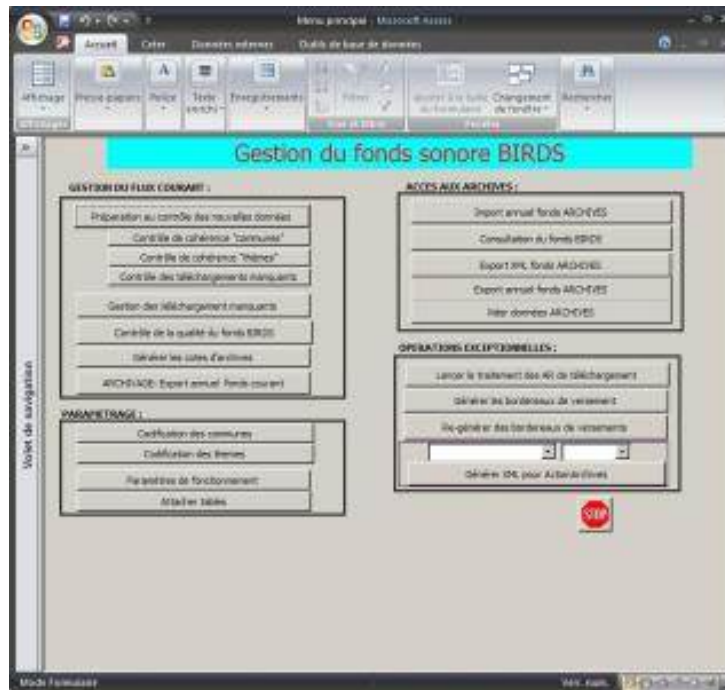


Fig. 5 : Application développée sur Access permettant la gestion du fonds sonore.

*(Cliquez sur l'image pour la voir en pleine taille)*



Fig. 6 : Idem ; écran de contrôle des données.

*(Cliquez sur l'image pour la voir en pleine taille)*

Le fonds BIRDS a été coté « archivistiquement » en 14 AV. À chaque radio correspond une sous-sous-série, de 14 AV 1 à 14 AV 4 (14 AV 5 pour les émissions produites par le Conseil général), puis les émissions reçoivent un numéro d'ordre (exemple : 14 AV 1/1 : première émission de Collines FM) : ces cotes sont générées manuellement via Access en cliquant sur « Générer les cotes d'archives ». Le fichier son aurait dû, selon les indications données par les Archives départementales, recevoir la même cote ; pour des questions techniques, le service informatique a souhaité maintenir le système de nommage adopté par lui dès l'origine : n° incrémenté (qui correspond au numéro informatique de la fiche), suivi du format du fichier. Cette solution n'est donc pas complètement satisfaisante ; l'occasion nous sera peut-être donnée un jour de remédier à ce qui peut paraître comme doublonnage inutile de cote.

Quatre fichiers sont téléchargés en moyenne par jour.

### Conservation

Les données (texte et son) sont stockées sur un serveur NAS situé au Conseil général. Une sauvegarde différentielle s'effectue chaque jour sur un deuxième serveur situé dans un autre site du Conseil général ; s'y



ajoute une sauvegarde complète tous les week-ends et tous les mois. En fin d'année l'ensemble des fichiers fait l'objet d'une copie sur disque dur externe, disque conservé pendant cinq ans aux Archives départementales (puis rotation) ; la première sauvegarde annuelle a été effectuée en 2007.

Le poids moyen des émissions d'une année représente 28 Go (fichiers son et textes, bordereaux de transfert, accusés de réception).

## Communication

Les fichiers sont communiqués en salle de lecture via le logiciel ActionArchive (fig. 7) qui permet à la fois la recherche documentaire - simple ou avancée - et l'écoute des émissions (les fichiers son s'ouvrent grâce à Windows media player). Trois postes informatiques sont équipés de casques audio.

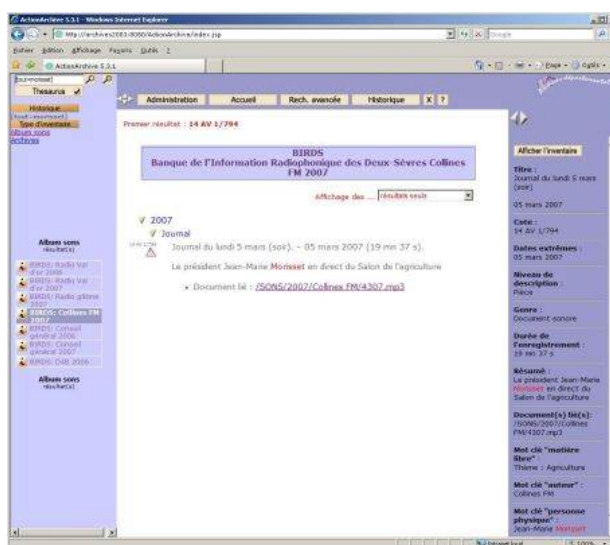


Fig. 7 : ActionArchive, permettant la recherche documentaire et l'écoute des émissions.

*(Cliquez sur l'image pour la voir en pleine taille)*

Le site Internet du Conseil général, comme nous l'avons déjà indiqué, offre l'écoute d'une semaine de journaux et de magazines.

La communicabilité des informations est immédiate (art. 10 de la convention). L'utilisation de celles-ci est autorisée par les radios pour le compte du Département aux fins d'activités culturelles. Pour le grand public, la reproduction est autorisée « pour le strict usage privé du copiste » mais soumise à autorisation de la radio pour un usage public (art. 12).

Pour la dernière année - septembre 2007-août 2008 - 1300 émissions environ ont été téléchargées, représentant 416 heures.

## ■ Quelques réflexions archivistiques

Cette première expérience de gestion de documents nativement électroniques par les Archives départementales des Deux-Sèvres a été riche dans la réflexion méthodologique, tout en bousculant certains critères archivistiques traditionnels.

Le fonds, d'emblée, semble mêler archives privées (émissions des radios associatives) et archives publiques (émissions du Conseil général), ce qui apparaît « anti-archivistique ». Dans le cas des notaires par exemple les minutes - publiques - sont classées à part des dossiers de clients - privés. Il faut, en fait, considérer que la

BIRDS n'est qu'un intermédiaire par où transitent les émissions, et pourrait disparaître sans modifier pour autant les contenus et le statut des informations. La cotation permet de bien distinguer les producteurs.

Le « versement » quotidien, générant un accusé de réception quotidien et un bordereau de transfert en moyenne hebdomadaire (en fonction du rythme de validation aux Archives départementales), sort également de l'archivistique traditionnelle, plutôt habituée à des services administratifs versant en moyenne tous les vingt ans... Le bordereau, très long (l'équivalent de 23 pages imprimées pour 16 émissions) n'est bien sûr pas conservé sous forme papier mais uniquement sous format électronique (XML).

Deux éléments descriptifs ne trouvent pas place, actuellement, dans les champs du standard : le champ « auteur » où devrait figurer l'auteur de l'émission et le champ « statut juridique du document » où serait indiquée la provenance publique ou privée des émissions. Ces points, notés par la direction des Archives de France, devraient être pris en compte lors des prochaines évolutions du standard.

Comme convenu dans la convention, les Archives départementales avaient proposé d'effectuer un tri sélectif des émissions, en ne gardant par exemple qu'un journal par jour et par radio, alors que certaines en produisent jusqu'à trois. Les radios n'ont pas souhaité d'échantillonnage. Les émissions sont donc intégralement conservées pour l'instant. À terme, on peut penser qu'il faudra cependant procéder à un échantillonnage, du simple point de vue de l'intérêt historique, sans même aborder la question du coût du stockage.

Le travail de vérification du téléchargement, s'il n'est pas en soit compliqué, requiert cependant attention, régularité et temps. Le rythme - qui, pour des raisons d'organisation n'est pas journalier mais plutôt hebdomadaire - ne souffre pas d'interruption trop longue. Les fiches descriptives sont passées en revue une à une pour correction : au-delà des fautes d'orthographe - récurrentes chez certains journalistes ! - il faut notamment développer les sigles, trop présents encore malgré des rappels de notre part, ce qui peut obliger à écouter l'émission. En fonction des radios, les fiches sont plus ou moins bien renseignées. En moyenne, vingt minutes sont consacrées à la correction chaque semaine.

La cotation retenue - 14 AV -, générée au fur et à mesure du contrôle, permet un répertoire numérique dont la logique répond à la date de diffusion (pour les radios locales) ou de versement dans la BIRDS (pour le Conseil général). Cependant les problèmes qui peuvent survenir dans le téléchargement impliquent que certaines émissions sont récupérées postérieurement : elles prennent rang alors par date de téléchargement, mais le cas reste minoritaire. Le répertoire apparaît donc à l'écran dans l'ordre numérique des cotes mais il serait logique d'envisager une présentation méthodique par date de diffusion/versement.

Pour l'instant, aucune demande de communication en salle de lecture, ni de reproduction, n'a été formulée. En cas de reproduction et d'exploitation, les tarifs afférents seraient appliqués.

Du point de vue juridique, deux démarches ont été menées :

- Une déclaration à la CNIL a été effectuée - qui n'a d'ailleurs reçu aucune réponse : il s'agissait davantage d'une précaution que d'une obligation ; la base de données contient en effet des noms de personnes mais aucune information relative à la vie privée.
- Le Conseil supérieur de l'audiovisuel (CSA) a indiqué que, la BIRDS n'étant que fournisseur de contenu et non pas éditeur de service de radio, il revenait aux radios intéressées - et non au Conseil général - de faire une demande en règle au CSA pour diffuser les émissions et d'effectuer les démarches nécessaires auprès de la SACEM pour le cas particulier des œuvres musicales programmées dans les émissions.

Pour conclure, cette première expérience revêt un caractère d'une relative simplicité puisque, une fois calée, la chaîne des opérations est en grande partie automatisée et répétitive. Mais, si les opérations se passent globalement sans difficulté, il faut souligner les liens étroits qu'il convient de maintenir en continu avec le service informatique du Conseil général et la vigilance à avoir en matière de stockage et de sauvegarde.

# THE NATIONAAL ARCHIEF OF THE NETHERLANDS AS A TRUSTED DIGITAL REPOSITORY FOR RECORD CREATORS AND ARCHIVAL INSTITUTIONS IN THE NETHERLANDS

---

Jacqueline Slats

Head of Digital Longevity Department, Nationaal Archief - National Archives of the Netherlands

*Dutch government agencies are producing a mass of information and records that are created in various formats. As the national repository of all records of permanent value, the Nationaal Archief of the Netherlands has made procedures for transfer of paper records that are in line with Dutch archival legislation. Implementation of digital information systems within government agencies has prompted the Nationaal Archief to build an infrastructure that can handle digital records with permanent value in a similar manner as the paper equivalents. The information systems that are used within agencies are not necessarily designed to support long-term archival functionalities, as they are primarily focussed on the work processes they facilitate and thus not capable of providing long-term solutions for records earmarked as of permanent value. The digital repository that is built in the Nationaal Archief will be complementary to these records management systems at the agency level, and will take the function of trusted digital repository.*

*However, the digital repository that the Nationaal Archief is currently building will be much more than a simple 'depot' that is capable of storing the digital records of government agencies. The architecture of the digital repository allows for a more integrated approach of ingest, data management, storage, preservation, and access. It is designed as a secured, online service.*

*A complicating factor in the construction of the digital repository is that it is designed to become a vital element in an even broader way of approaching the problem of digital longevity. For government agencies, it will act as a multifunctional repository. It will be capable of ingesting records after the legal period of administrative use (20 years), but it will also enable the transfer of records at an earlier stage. Finally, it will function as a repository for intermediate storage (i.e. storage of records that are projected to be destroyed after some years, but have to be preserved for a period that is too long for secure storage within an agency's records management system).*

*For regional archival institutions lacking the expertise and facilities to develop their own repository infrastructure and functions, the Nationaal Archief digital repository will take the role of a trusted digital repository, enabling ingest, data management, storage, preservation, and providing access to these records.*

*This presentation will discuss how the digital repository is designed, developed and built. It will explain the architecture of the repository, but also the process of design and development as a collaborative effort of the Nationaal Archief, regional archival institutions and government agencies.*

[See the presentation.](#)

## **Les Archives nationales néerlandaises servent de plate-forme d'archivage électronique sécurisée aux producteurs d'archives et aux institutions d'archivage des Pays-Bas**

*Les agences gouvernementales néerlandaises produisent un volume d'informations et d'archives créées sous divers formats. En tant que plate-forme d'archivage nationale de l'ensemble des archives à valeur permanente, les Archives nationales des Pays-Bas ont mis en place des procédures de transfert d'archives papier conformes à la législation néerlandaise relative à l'archivage. La mise en place de systèmes d'informations numériques au sein des agences gouvernementales a incité les Archives nationales à installer une infrastructure capable de traiter de façon identique les archives numériques à valeur permanente et les archives papier. Les systèmes d'information utilisés au sein des agences ne sont pas nécessairement conçus pour supporter des fonctionnalités d'archivage à long terme car ils se concentrent principalement sur les processus métier qu'ils facilitent et ne sont donc pas capables de fournir des solutions de conservation à long terme pour les archives identifiées comme étant de valeur permanente. La plate-forme d'archivage aménagée au sein des Archives nationales viendra compléter ces systèmes de records management au niveau de l'agence et il détiendra le rôle d'archivage numérique accrédité.*

*Cependant, la plate-forme que les Archives nationales aménagent actuellement sera plus qu'un simple « dépôt » capable de stocker les archives numériques des agences gouvernementales. Son architecture permet une approche plus intégrée en termes d'ingestion, de gestion, de stockage, de conservation et d'accessibilité des données. Elle est conçue pour être un service en ligne sécurisé.*

*La difficulté en ce qui concerne la construction de la plate-forme réside dans le fait qu'elle est conçue pour devenir un élément essentiel dans la façon d'aborder le problème de la longévité numérique. Pour les agences gouvernementales, elle agira en tant que dépôt à fonctionnalités multiples. Elle sera capable d'ingérer des archives après la période légale d'utilisation administrative (20 ans), mais permettra également le transfert d'archives avant la fin de la période légale. Enfin, elle fonctionnera en tant que dépôt destiné au stockage intermédiaire (par ex. stockage des archives dont la destruction est prévue au bout de quelques années mais devant être conservées pendant une période trop longue pour permettre un stockage sécurisé au sein d'un système de gestion des archives d'une agence).*

*S'agissant des institutions d'archivage régionales qui manquent d'expertise et d'équipement pour développer leur propre infrastructure et fonctions d'archivage, la plate-forme des Archives nationales jouera le rôle d'infrastructure d'archivage électronique accréditée permettant l'ingestion, la gestion, le stockage, la conservation des données et fournira l'accès à ces archives.*

*Cette présentation analysera la façon dont la plate-forme est conçue, développée et construite. Elle présentera l'architecture de la plate-forme mais également le processus de conception et de développement dans le cadre d'un effort de collaboration des Archives nationales, des institutions régionales d'archivage et des agences gouvernementales.*

[Consulter la présentation.](#)



# **Séance 4**

## **La conservation pérenne des données numériques**

---

---

sous la présidence de Francesco BARBEDO

Instituto dos Arquivos Nacionais, Portugal

# **Session 4**

# **Long-term preservation**

---

---

*chairing Francisco BARBEDO*  
*Instituto dos Arquivos Nacionais, Portugal*

# LA SÉLECTION DES DOCUMENTS DANS LE SYSTÈME NUMÉRIQUE

---

Roberto Guarasci

Anna Rovella

Maria Taverniti

Università della Calabria, Rende (CS), Italia

*Dans la transition du document papier au document numérique, une des erreurs les plus fréquentes, c'est de croire que - comme pour le passé - le changement du support documentaire n'a aucune influence sur les choix méthodologiques qui président aux différentes phases du cycle de vie d'un fond documentaire.*

*En particulier, pour ce qui est du plan de préservation, la numérisation a éliminé, d'un côté, l'occupation des espaces - souvent coûteuse - qui constituait l'aspect fondamental du tri tout en accentuant, de l'autre, celui de la redondance et du bruit informatif.*

*La théorie traditionnelle, qui voyait le tri comme un des moments de la réorganisation des archives papier, se basait sur l'affirmation que le respect du lien entre les documents véhicule un contenu intellectuel, ce qui demande d'abord la reconstruction de la structure globale des archives et, ensuite, l'évaluation des unités méritant d'être conservées.*

*Actuellement par contre, comme les documents numériques n'existent pas en tant qu'entités physiques proprement dites et que rarement l'emmagasinement effectif des signaux électroniques qui forment un document conserve quelque connexion avec le document, soit-il affiché sur l'écran ou imprimé, la relation physique perd toute signification. Et ce phénomène, loin d'être négatif, favorise en réalité les relations logiques entre les documents. Le résultat en est le changements des paramètres et des méthodiques qui ont présidé - jusqu'ici - à la préservation et à la sélection des documents d'archives. En même temps, on met fortement en discussion la nécessité absolue de la réorganisation des fonds comme phase propédeutique à la sélection ex post surtout en considération de la nécessité, de plus en plus évidente, de prédéterminer ex ante le cycle de vie de chacune des typologies documentaires.*

*Dans ce contexte, des méthodologies d'analyse textuelle et d'extraction terminologique qui se basent, entre autres, sur des algorithmes de fréquence et de pertinence statistique des termes appartenant à des archives numériques et numérisés peuvent représenter un suivi d'application extrêmement intéressant, capable de faire face aussi à l'augmentation de la production documentaire qui n'a pas toujours été correctement classée d'un point de vue archivistique.*



## Selecting documents in a digital system

*One of the errors most frequently made in any transition from paper to digital documents is to believe that – as in the past – the change of document format has an effect on the choice of methodologies governing the different phases of an archive's life-cycle.*

*Concerning the conservation plan in particular, although digitisation has removed one of the basic motives for the sorting required to optimise the use of space (often a costly item), it has resulted in an increase in redundancy and superfluous information.*

*The traditional theory that considered sorting as an integral part of the reorganisation of paper archives was based on the conviction that respecting the relationships between documents involved an intellectual content that required, firstly, the reorganisation of the archives' overall structure and, secondly, the evaluation of which elements ought to be kept. Electronic documents, in contrast, do not exist as actual physical entities and it is rare for their storage as electronic signals to have any connection with the documents themselves; whether displayed on a screen or printed out, the physical relationship loses all meaning. Indeed, this phenomenon strengthens the logical relationships between documents. Consequently, we are witnessing a change in the parameters and methods that – until now – were paramount for the conservation and selection of archived documents. At the same time, there no longer seems to be the same absolute necessity to reorganise repositories as a preliminary phase to ex post selection, even considering the increasingly ambiguous need to pre-establish ex ante the life-cycle of each document typology.*

*In this context, methodologies for textual analysis and terminology extraction, also based on algorithms of frequency and statistical relevance of terms applying to digital and digitised archives, can represent the development of extremely useful applications, that can also deal with the increase in the production of documents that have not always been correctly classified as archives.*

Si l'on exclut des raisons purement idéologiques comme celles de l'an II de la Révolution française (1794) qui prévoyaient l'élimination de tous les documents "*portant l'empreinte honteuse de la servitude*", jusqu'au début du XIX<sup>e</sup> siècle "le problème du triage ne se posait pratiquement pas car la mission fondamentale des archives consistait à inventorier et à conserver les documents versés par les administrations" (Marta Maffei, 1997-1998). Ce n'est pas par hasard que dans le «De Archivis» de Baldassarre Bonifacio, publié à Venise en 1632, l'accent est mis sur les "violatores archivii" et sur les mesures pour la sauvegarde des documents, mais jamais sur une éventuelle sélection des mêmes.

Il faut attendre le début du XX<sup>e</sup> siècle pour que naissent, en Europe, des règles et des dispositions sur la réglementation des opérations de sélection et de tri des documents.

Les décennies suivantes font enregistrer une tendance progressive des archives nationales à la centralisation des procédures ou, au moins, du contrôle des procédures de sélection en vue d'une utilité surtout historique du document d'archives.

L'archiviste conservateur des actes se pose le problème de leur sélection presque exclusivement en fonction de l'évidence du matériel historiquement pertinent. Tout cela même si, en France - au contraire de ce qui se passe, par exemple, en Italie - le Décret du 21 juillet 1936 prescrit un rapport stable entre les archives de conservation et les administrations productrices qui sont appelées à élaborer, périodiquement, des "tableaux d'élimination" à déposer aux Archives Nationales au moment du versement.

A la suite des dégâts au patrimoine archivistique provoqués par la première guerre mondiale, l'attitude envers les procédures de sélection est, de toute façon, extrêmement prudente: la tendance est à les accepter comme un mal nécessaire en niant souvent la validité de l'expression "actes ou documents inutiles" contenue dans beaucoup de dispositions prises sur le sujet.

L'augmentation rapide de la quantité de documents produits et dupliqués dans la période qui suit le second conflit mondial change rapidement la perspective du problème en confiant, de fait, aux administrations la charge de la sélection; elle pousse aussi certaines nations à la construction d'archives intermédiaires pour garantir et gérer la collaboration entre les sujets producteurs et les Archives nationales.

Il existe aussi quelques élaborations théoriques qui essaient de répondre à une demande urgente de la société civile: l'augmentation exponentielle de la quantité des documents produits et la nécessité des administrations d'accéder aux informations ne permettent pas la conservation totale des actes. On rappelle ainsi, entre autres:

- la théorie des trois âges de Théodore R. Schellenberg;
- les solutions proposées en Angleterre par la Commission Grigg;
- les tentatives d'application de ces théories en France, poursuivies par (Yves Pérotin, 1966) qui envisage la réduction du processus de sélection à deux moments, et son articulation en trois phases dans l'arc de 100 ans pendant lesquels il faut gérer la "transition fragile de l'Administration à l'histoire";
- l'élaboration théorique de (Giorgio Concetti, 1970) qui qualifie les différentes étapes de la vie des archives comme "nécessités techniques" n'invalidant pas l'unicité de l'ensemble et accepte substantiellement le tri en tant qu'opération de rupture nécessaire d'un équilibre parfait à l'intérieur d'un ensemble homogène gouverné par le lien original entre les actes.

Après avoir précisé que les documents publics possèdent deux types de valeurs: les valeurs primaires qui se réfèrent au sujet producteur et à ses fonctions opérationnelles et les valeurs secondaires qui se manifestent au cours du temps et concernent des sujets et des utilisateurs différents du producteur, Schellenberg «considéra le rôle de l'archiviste prioritaire pour la détermination des valeurs secondaires qui peuvent, à leur tour, être de deux types: valeurs probatoires et valeurs informatives. La valeur probatoire est une valeur dépendant de l'importance du témoignage, c'est-à-dire, de l'organisation et du fonctionnement du bureau producteur des documents. [...]. Par conséquent, face à la constatation de leur valeur d'information ou de recherche, les documents devraient être analysés du point de vue de leur contenu» (Kent M. Haworth, 2000).

L'irruption massive de l'informatique dans la production documentaire ne provoque pas, d'ailleurs, les débâcles apocalyptiques prédites par certains archivistes ni l'avènement galopant d'une société sans papier. Cela signifie que, vers la fin des années 1970, les prophètes des deux coalitions n'avaient pas compris que les êtres humains, les seuls utilisateurs de ces machines, n'évoluaient pas avec la même rapidité que la

technologie. «Quelle que soit la transformation survenue dans les bureaux, les changements dus aux machines se sont produits à une vitesse supérieure à celle des gens qui les utilisent», c'est là presque une confirmation de l'énoncé de (Marshal Mc Luhan , 2004) «toute technologie inventée et extériorisée par l'homme a le pouvoir d'affaiblir la conscience humaine pendant la période de sa première intériorisation» (*ibidem*).

Pendant les *mêmes années*, la législation de la plupart des pays occidentaux ne reconnaît d'ailleurs pas encore la pleine validité juridique au document numérique, ce qui oblige, presque toujours, à rédiger une copie sur support papier ayant valeur légale. Cela donne aussi aux archivistes le droit de penser que les méthodologies de sélection traditionnelles bien consolidées sont complètement adaptables à la nouvelle situation, quitte à les préciser et à les renforcer éventuellement.

Cependant la conscience qu'il existe beaucoup de documentation échappant à l'analyse et à la validation des archivistes commence à se développer si bien que, en 1971 (Hans Booms, 1987), propose un modèle où l'archiviste assume la fonction de responsable de la documentation de la totalité de la vie publique «telle qu'elle se manifeste dans les communautés formées autour d'intérêts communs» en faisant l'hypothèse d'une participation collective à la définition des méthodes de tri.

L'augmentation des quantités produites et la conscience que les spécialistes ne peuvent pas présider à l'opération délicate de sélection *ex post*, unies à la prise de conscience qu'à l'ère du numérique l'intervention de l'archiviste devrait changer son arc temporel d'intervention, poussent certains auteurs à préfigurer des méthodes de tri basées sur des systèmes pré-coordonnés.

En 1967, (Pierre Boisard, 1967) théorise une conservation définitive fondée sur une présélection des fonctions de l'administration productrice et par là donc sur une évaluation de l'importance de ces dernières pour la production d'actes importants pour l'histoire future. Cette évaluation doit être effectuée, de toute façon, avec les administrations et les archives nationales et avant le versement final. C'est une théorie qui n'aura pas de succès au moment de sa formulation, mais elle jettera les bases théoriques pour des applications à venir.

La «Politique de gestion des documents inactifs», adoptée par le Québec en 1991 comme intégration à la loi sur les Archives du 23 décembre 1983, établira - par exemple - que chaque administration doit conserver seulement les documents pour lesquels on a prévu la conservation en fonction du plan de conservation approuvé, basé sur le poids différent des fonctions des administrations productrices de la documentation (Maurielle Doyle, 1992)<sup>1</sup>.

*Quelques années plus tard*, (Michel Duchein, 1972) essaiera de corriger le tir en proposant de nouveau une idée du tri comme processus d'évaluation continue des documents qui peut fonctionner, évidemment, seulement à condition que ce soient les archivistes à gérer ou, de toute façon, à superviser le cycle de vie et le «document management» des administrations. Mais les archivistes français, comme ceux des autres pays, ne géraient plus le cycle de vie du document parce que, à l'aide du développement de l'ère de l'Information Technology, d'autres expertises étaient intervenues tandis que les parcours *formatifs* de maintes Ecoles des Chartes parmi les plus prestigieuses continuaient d'être orientés plutôt vers les études en paléographie que vers celles en «document management» et en archivistique contemporaine.

L'archiviste conservateur des documents, effrayé par l'irruption d'un monde peu connu, a eu - au départ - une fermeture préconçue ou un acquiescement superficiel face au changement rapide de la typologie des documents si bien qu'il a négligé, plus ou moins consciemment, la gestion des documents courants pour se réfugier dans le monde privilégié et connu des documents historiques. Il s'agit d'un phénomène dont l'importance change en fonction des situations nationales ou à cause des différences culturelles des archivistes et de l'attention différente qui lui a été réservée par les organismes nationaux préposés.

Le passage décisif survient au moment où les législations nationales reconnaissent la pleine validité juridique au document numérique qui n'est plus alors la copie de travail d'un original en papier ou la donnée brute d'une élaboration dont la forme juridiquement valable reste toujours celle qui est sur support tangible, mais devient, dorénavant, la seule copie valide existante.

---

<sup>1</sup> «Afin d'identifier le patrimoine archivistique et d'assurer une gestion efficace de la masse documentaire aux stades actif et semi-actif, la Loi sur les Archives prévoit que chaque organisme public élabore un calendrier de conservation de ses documents et le soumette à l'approbation des A.N.Q. Le Règlement sur le calendrier de conservation, le versement, le dépôt et l'élimination des archives publiques, édicté en septembre 1985, précise la façon d'établir et de tenir à jour un calendrier de conservation».

«Les archivistes qui s'occupent des documents informatiques à longueur d'année se rendent rapidement compte que les méthodes mises au point à un moment donné peuvent devoir être partiellement ou complètement révisées deux ou trois ans plus tard. Ces révisions intéresseront non seulement les méthodes d'acquisition, de traitement et de conservation des documents informatiques, mais aussi la fonction d'évaluation et de tri» (Harold Naugler, 1986).

Lorsqu'ils évaluent des documents informatiques, les archivistes doivent déterminer s'il y a lieu de conserver les données élémentaires, les données agrégées sous forme de résumés et de rapports statistiques ou les deux.

«Puisque les documents numériques n'existent pas en tant qu'entités physiques proprement dites et que l'emmagasinement effectif des signaux électroniques constituant le document garde rarement quelque connexion avec un document sur écran ou imprimé, le stockage physique des bits et des parties de documents numériques ne demande ni ne véhicule de contenu intellectuel. Et si le fait de sauvegarder l'ordre original des documents numériques n'a aucune conséquence, il est par contre absolument crucial d'en documenter les liens logiques» (Charles Dollar, 1992b).

L'attention se déplace donc sur l'analyse des contenus et sur les relations logiques. Par conséquent, ce sont ces mêmes paramètres que doit assumer une hypothèse de travail sur le tri et celle-ci, en raison des volumes, ne peut qu'être automatique, comme prévu au point 8.3.7. «Conservation et sort final» de la norme ISO 15489 où l'on affirme «il est recommandé est en outre que, le cas échéant, ces mesures soient activées automatiquement, et que les systèmes intègrent une piste d'audit ou une autre procédure pour assurer la traçabilité complète de ces mesures» (Andreas Kellerhals Maeder - Jean Luc Cohard, 1993)<sup>2</sup>.

Et c'est dans cette direction que l'on poursuit les études sur la possibilité de développer des méthodes de classement automatique dans le cadre des archives numériques en utilisant des techniques d'«information retrieval», de «text mining» et de «social network analysis» (Esteva, 2007).

L'attention est portée, bien évidemment, sur les documents numériques, dépourvus d'un original papier, à propos desquels les «evidential information value» des fichiers master, intégrés par des critères d'évaluation concernant leur capacité de relation et leur exportabilité, testés depuis les années 1990, ont démontré qu'ils sont décidément inapplicables. «Une première réponse a été donnée en concentrant l'attention sur la fonction des documents ou leur compétence productrice plutôt que sur les documents eux-mêmes. Si la compétence est fondamentale pour l'organisation, il est probable que les documents créés en cours de route le soient également» (Charles Dollar, 1992a).

Cela suppose une participation active des archivistes aux projets des systèmes informatifs documentaires pour la définition précise des critères d'identification des fonctions productrices et de l'indication du moment du tri concernant le cycle de vie du document.

Une des erreurs les plus répandues revient à estimer que les logiques et les systèmes de l'univers papier peuvent s'appliquer automatiquement aux différents contextes de gestion, là où la simple augmentation exponentielle des volumes détermine, par contre, des scénarios nouveaux non prévisibles; cependant, il est vrai aussi que la sélection des documents est une opération qui demande l'apport de la sensibilité professionnelle des archivistes et une évaluation que l'on ne peut pas toujours reconduire aux fonctions productrices ou à l'arc chronologique de la vie, surtout dans des contextes nationaux où la différence entre la praxis et la structure administrative de référence est très grande.

Aux solutions basées sur des décisions a priori donnent une réponse optimale la définition des éléments des métadonnées qui décrivent les ressources informatives d'un système et l'indication de l'arc chronologique de vie d'un document ou d'une typologie. Il est vrai que la poussée vers des subset minimaux et partagés, comme ceux du Dublin Core, (ISO 15836: 01/2003), nous permet d'apercevoir une intégration des systèmes de plus en plus grande et la diffusion conséquente de ces mécanismes, mais il est vrai aussi que, dans les différents contextes nationaux, cette situation ne sera sûrement pas généralisée. La production de documents de la part des

---

<sup>2</sup> Dans une hypothèse de projet pour un système d'interrogation full text des archives fédérales de la Confédération Helvétique en 1993 on affirme que: "Pour arriver à une représentation plus précise du contenu des documents nous avons choisi d'élargir l'analyse du contenu par une analyse syntaxique et sémantique. Par conséquent, nous sommes confrontés à plusieurs problèmes: 1.L'administration publique s'occupe de presque tous les aspects de la vie et les documents d'archives reflètent cette diversité. 2.La structure syntaxique d'une requête ne sera que très rarement identique à celle d'une partie d'un document mais sera très probablement une paraphrase plus générale que le sujet précis du document. 3.Même les phrases en apparence simples sont très souvent ambiguës. Une analyse complète d'un texte en langue naturelle ne peut être réalisée avec la technique de l'analyse automatique de la langue, disponibles aujourd'hui".

utilisateurs ordinaires non plus ne pourra respecter ces règles dans l'immédiat et cela non pas ou non seulement à cause de l'absence de règlement, mais parce que le temps nécessaire pour que l'innovation devienne patrimoine cognitif commun est bien différent du temps nécessaire à introduire des modifications dans le système normatif des États. La situation italienne, qui voit un cadre législatif complet et techniquement avancé mais peu appliqué, est symptomatique de la possibilité d'une anomalie de ce genre.

On sait qu'atteindre une date préétablie est une précondition nécessaire mais pas suffisante et que non seulement la détermination à priori de fonctions ou d'éléments de contenu décrits dans les métadonnées n'est pas une condition généralisable, mais aussi qu'elle pourrait ne pas tenir compte des événements situés entre la date de création, qui fixe la date limite et celle de la sélection effective des événements qui pourrait modifier de manière significative la perspective de valeur et de conservation conséquente de l'information. Compte tenu de ces facteurs, la possibilité d'intégrer ces systèmes avec d'autres méthodologies d'analyse et d'évaluation du contenu, automatiques ou semi-automatiques, pourrait s'avérer importante et justifier sa raison d'être.

L'idée de départ pour la formulation d'une hypothèse de travail dans ce domaine trouve son origine dans un projet de collaboration entre le Conseil National des Recherches et le Centre National pour l'informatique de l'Administration Publique dont le but était la réalisation d'un logiciel d'application open source pour la classification semi-automatique des documents.

Cela devait permettre de réduire le cadre discrétionnaire des opérateurs en proposant des classifications possibles pour les documents, basées sur l'évaluation d'un set de formation adéquat, utilisable pour l'entraînement.

Le prototype du logiciel réalisé se base sur la considération qu'il existe - dans toute administration - un set plus ou moins vaste de documents, correctement classifiés et validés par des experts, que l'on peut utiliser pour instruire un système de classification automatique.

Le premier outil construit la connaissance de la classification en indexant chaque fichier déjà classifié appartenant au training set.

L'outil suivant, un query builder, utilise les données indexées par le premier outil. Le query est construit avec des termes tirés du champ objet inséré par l'opérateur. Les éléments du champ objet sont divisés en caractères et mots qui sont considérés comme des opérateurs booléens.

Le query ainsi construit est soumis au *search library*: au résultat est attribué un score (Rank) de similarité, calculé à travers la formule suivante:

$$score(q, d) = coord(q, d) \cdot queryNorm(q) \cdot \sum_{t \in q} tf(t \in d) \cdot idf(t)^2 \cdot t.getBoost() \cdot norm(t, d)$$

Le score mesure la similarité de q, le nouveau objet inséré, avec d, un objet indexé par le search engine. Le résultat est constitué du fichier original et de chaque fichier contenant sa propre classification. Les fichiers sont ordonnés selon la méthode du pivot, basée sur la valeur de la classe utilisée comme clé primaire.

Un score de classification est donné par la moyenne entre le score de similarité le plus haut et la moyenne géométrique de tous les scores de similarité. La classification qui obtient le score le plus haut sera proposée comme premier candidat. L'adoption successive d'un dictionnaire de synonymes du secteur a ultérieurement augmenté la précision de récupération. Cela a permis de réduire de manière significative (90%) les erreurs de classification des opérateurs par rapport à la classification réalisée par des experts. L'expertise acquise nous a permis de formuler une hypothèse d'application de la même approche au processus de sélection des documents.

Pour la réalisation des opérations de tri, il faut pouvoir vérifier que la coexistence d'au moins deux conditions «on-off» soit d'ordre purement chronologique et qu'elle soit causée par la réalisation d'une date limite qui permet, à son tour, d'évaluer la possibilité successive. Cette dernière n'est pas si univoque que la première car, face à la décision éventuelle d'adopter le critère des fonctions préétablies, on ne pourrait qu'ajouter un niveau ultérieur «on-off». Tout cela à cause de l'impossibilité de procéder à la sélection, sauf si l'on est en présence aussi d'une forme d'analyse quelconque du contenu probatoire et/ou informatif des documents candidats. Le problème qui se pose est donc celui de reconduire l'analyse scientifique du contenu, traditionnellement faite par des archivistes, à des éléments univoques et mesurables. En adoptant les prémisses définies plus haut, on peut repartir du constat que toute administration possède un corpus

documentaire correctement classifié par des experts humains. Ce set est utilisé, cette fois-ci, pour instruire le système de classification à reconnaître les documents qui doivent être éliminés à partir d'évidences intrinsèques aux documents eux-mêmes. L'analyse des caractéristiques objectivement relevées dans les documents et leur co-occurrence dans les fichiers à éliminer, nous poussent à prendre en considération la possibilité de suivre une règle uniforme lorsque certaines conditions se vérifient.

L'usage de systèmes de classification probabilistes pourrait représenter une solution possible à notre problème. Après avoir défini le training set, ces méthodes sont capables, en effet, de formaliser les événements (features) distinctifs d'un document à éliminer. Pour ce faire, on utilise un classificateur statistique qui exploite, comme méthodologie probabiliste, le principe de la Maximum entropy (Adwait Ratnaparkhi, 1998 - Chris Manning and Hinrich Schütze, 1999).

Très utilisé dans des domaines comme le TAL, le POS tagging, la segmentation et la classification des textes, le «maximum d'entropie» permet de mesurer la distribution des probabilités qu'un événement a de se vérifier en présence de conditions établies et déduites par l'observation empirique des données. Celles-ci sont représentées, dans chaque problème de classification, comme des ensembles d'événements qui sont, à leur tour, l'expression d'une contrainte du modèle. Un ensemble de toutes les features est souvent sélectionné en employant des métriques, dont «l'information gain» (Solomon Kullback, 1997)<sup>3</sup>. Si l'on part donc de la représentation du document et d'un «training corpus» donné de documents, marqués de leur classe d'appartenance, on peut créer un modèle pour prédire la probabilité qu'un document a d'appartenir à une classe plutôt qu'à une autre. En conclusion, le «maximum d'entropie» fournit une méthodologie efficace pour calculer la distribution des probabilités à partir du corpus de formation. Cela veut dire qu'étant donnée un set d'informations et de caractéristiques des documents à l'objet, on doit être à même de calculer combien de probabilités un document présente d'être on/tri (ou bien pas tri/off) (de l'Orletta et al, 2007).

On a déjà dit que le training set est constitué des listes de tri, sélectionnées à l'intérieur d'une population relative à un arc chronologique représentatif. En ce cas-là, si les conditions on/off décrites plus haut se vérifient, l'algorithme déjà réalisé est à même de proposer une liste, à valeur décroissante, des actes à proposer pour le tri. Si à cette proposition, basée exclusivement sur une fréquence statistique incrémentée grâce à la gestion des synonymies, on associe aussi la présence contemporaine, dans les textes analysés, d'un ou plusieurs termes portés descripteurs d'un lexique du domaine, on aura des features capables de déterminer des modèles de documents à proposer pour le tri. À ce propos, on peut utiliser des techniques employées dans l'analyse des corpus textuels.

La combinaison de log-likelihood<sup>4</sup> avec z-score<sup>5</sup> permet -par exemple- l'identification de termes qui peuvent mieux diversifier, de fait, un corpus de l'autre (Murzaku, 2008).

Les caractéristiques seront constituées, par conséquent, de caractéristiques lexicales temporelles ou structurelles dépendant, par exemple, de l'âge du document ou de la date du dernier accès, etc. On crée ainsi un système d'événements multidimensionnel dont le poids pourrait, de fait, être calculé en employant le modèle du «Maximum d'entropie». En outre, en maximisant l'entropie, la co-occurrence des mêmes événements/features fait «comprendre» au système quels sont les fichiers qui peuvent être soumis au tri. Dans le tableau qui suit est présenté un exemple de données et caractéristiques dont les valeurs peuvent être calculés en utilisant l'algorithme GIS - Génératif Implemented Scaling - implementé à travers des logiciels tels que, par exemple, AMIS (A Maximum Entropy estimator for feature forests, Miyao and Tsujii 2002):

<sup>3</sup> On cite aussi: Document frequency, la Likelihood Ratio, Optimal Orthogonal Centroid, Term Discrimination, et la méthode d'extraction *Probabilistic Latent Semantic Analysis*.

$$^4 - \ln \lambda = 2 \sum_i \text{Observed}_i \ln \left( \frac{\text{Observed}_i}{\text{Expected}_i} \right)$$

$$^5 z = \frac{\text{Observed} - \text{Expected}}{\text{std}}$$

Doc	Terms (caractéristique lexicale)					Classification					Date_imput	Date dernier_accès	Évènements
	t 1	t 2	t 3	t 4	t 5	1	2	3	4	4.1			
D 1	0	2	3	0	3	0	1	0	0	0	15.05.2001	18.07.2001	On
D 2	2	1	0	1	1	1	0	0	0	0	3.3.2000	16.10.2008	Off
D 3	4	5	0	0	2	1	0	0	0	0	16.10.2008	16.10.2008	Off
D 4	1	0	4	0	4	0	1	0	0	0	20.01.2006	20.01.2006	On

Exemple 1 : features des vecteurs

Les deux tableaux ci-dessous représentent respectivement, pour chaque vecteur, la somme des évènements qui le caractérisent et leur normalisation.

Terms (caractéristique lexicale)					Classification					Évènements
t <sub>1</sub>	t <sub>2</sub>	t <sub>3</sub>	t <sub>4</sub>	t <sub>5</sub>	1	2	3	4	4.1	
1	2	7	0	7	0	2	0	0	0	On
6	6	0	1	3	2	0	0	0	0	Off

Exemple 2 : somme des vecteurs des évènements

Terms (caractéristique lexicale)					Classification					Évènements
t <sub>1</sub>	t <sub>2</sub>	t <sub>3</sub>	t <sub>4</sub>	t <sub>5</sub>	1	2	3	4	4.1	
0,14	0,03	1	0	0,7	0	1	0	0	0	On
0,85	0,75	0	1	0,3	1	0	0	0	0	Off

Exemple 3 : normalisation des vecteurs des évènements

Les valeurs des caractéristiques des documents sujets au tri, comme on l'a déjà dit, seront donc pesées et formeront le prototype qui représentera notre modèle de fichier à éliminer. À ce moment-là, chaque document de l'archive A sera représenté par un set de caractéristiques ( $d_i$ ) qui le marquent et qui en déterminent le statut  $D_{on}$  ou bien  $D_{off}$

$$p(D_{on}|d_i)$$

où  $D_{on} \in A$  et la distribution des caractéristiques  $f_j(D_{on}|d_i)$ , de  $d_i$  est tirée du training corpus et indique les valeurs des  $k$  couples des caractéristiques ( $D_{on}|d_i$ ) et ( $D_{off}|d_i$ ), enfin, est une fonction de normalisation des facteurs. Pour chaque paramètre  $\alpha$  les valeurs de la distribution  $\alpha_1, \dots, \alpha_k$  correspondent aux poids des caractéristiques. La probabilité distributionnelle qui respecte les propriétés susmentionnées est celle qui présente le degré d'entropie le plus haut. La formule est:

$$p(D_{on}|d_i) = 1/Z(d_i) \prod_{j=1}^k \alpha_j^{f_j(D_{on}|d_i)}$$

En conclusion, étant donné un modèle de tri établi à priori, la méthode basée sur la maximisation de l'entropie nous permettra de calculer la probabilité qu'un document soit *on* ou *off* en respectant l'ensemble des caractéristiques qui le représentent. Comme résultat, on aura deux vecteurs,  $D_{on}$  et  $D_{off}$ , qui dérivent de la somme des valeurs de chaque caractéristique:

$$D_{on} \cdot v \langle e_1, e_2, e_3, e_4, \dots, e_n \rangle \text{ on};$$

$$D_{off} \cdot v \langle e_1, e_2, e_3, e_4, \dots, e_n \rangle \text{ off};$$

Et, pour finir, nous établirons aussi une fourchette au-delà de laquelle le système ne pourra pas avancer de manière autonome car l'action de tri peut avoir besoin d'évaluations ultérieures. Dans les situations d'incertitude, le système devra donc produire un alert file, en demandant ensuite une autorisation à procéder avec le tri. Le système, conceptuellement et théoriquement défini, est encore sous testing, mais les premiers résultats, sur des échantillons significatifs, donnent un pourcentage d'actes correctement proposés pour le tri tout à fait comparable à celui d'un groupe d'opérateurs experts, utilisé comme test.



## ■ Bibliographie

- BOISARD, P., (1967), «Pour une politique des élimination: réflexions sur la pratique des Archives de la Seine», *La Gazette des Archives*, 1967, 59, pp. 205-237.
- BOOMS, H., (1987), «Society and the formation of a documentary heritage », in *Archiviaria*, 1987, 24, pp. 69-107.
- CENCETTI, G., (1970), *Scritti Archivistici*, Il Centro di Ricerca, Roma, 1970.
- DOLLAR, C. M., (1992a), *Archivistica Informatica*, L'impatto delle tecnologie dell'informazione sui principi e i metodi dell'archivistica, sous la dir. de ODDO BUCCI, Publication de l'Università de Macerata, 1992, p. 62.
- DOLLAR, C. M., (1992b), «La teoria e la prassi archivistica di fronte all'informatica. Alcune considerazioni», in *L'Archivistica alle soglie del 2000*, Acts de la conférence internationale de Macerata, septembre, 1990, sous la dir. De ODDO BUCCI, Publication de l'Università di Macerata, Ancona 1992, p. 306.
- DOYLE, M., (1992), «La Gestion des documents Administratifs et des archives au sein des ministères et organismes gouvernementaux», in *Archives*, 1992, 24, 1-2, p. 117.
- DUCHEIN, M., (1972), «Les procédures de tri dans les archives départementales», in *La Gazette des Archives*, 1972, 77 pp. 75-97.
- ESTEVA, M., (2007), *Bits and Pieces of Text: Appraisal of a Natural Electronic Archive*, présenté au Digital Humanities 2007, < <http://www.digitalhumanities.org/dh2007/abstracts/xhtml.xq?id=136>>
- HAWORTH, K. M., (1992), «La fedeltà ai principi archivistici. Il futuro della selezione documentaria, della gestione dei documenti e dell'inventariazione nel Nord America», in *L'Archivistica alle soglie del 2000*, sous la dir. de ODDO BUCCI, Publié par l'Università di Macerata, Macerata, 1992, pp. 141-142
- ISO 15489-1 (Information et documentation : Records management, 2001, 8.3.7)
- KELLERHALS MAEDER, A., COCHARD, J. L., (1993), «Vers l'Archiviste Artificiel(le)? La contribution de la linguistique informatique à un système archivistique de gestion de l'information», in *Archivi & Computer*, 1993, 3, p.160.
- KULLBACK S., (1997), *Information Theory and Statistics*, (Dover Books on Mathematics), Mineola, Dover, 1997.
- MAFFEI, M., (1998), «Le Moment du tri dans la pratique archivistique française: réflexions théoriques et applications pratiques», in *Archives*, 1997-98, 29, .3/4, p. 4.
- MANNING, C. and SCHÜTZE, H., (1999), *Foundations of Statistical Natural Language Processing*, MIT Press. Cambridge, May 1999
- MC LUHAN, M., (2004), *La Galassia Gutenberg, la nascita dell'uomo tipografico*, Armando Editore, Roma 2004, p. 210.
- MIYAO Y., e TSUJII J., (2002), *Maximum entropy estimation for feature forests*. In Proc. HLT 2002, <<http://www-tsujii.is.s.u.tokyo.ac.jp/~yusuke/amis/manual.html#introduction>>.
- MURZAKU, A. (2008), *Language Unchained - A Quantitative Analysis of Language Change in Albania after the Fall of Communism*, 16th Conference on Balkan and South Slavic Studies, Banff, Calgary, Canada, May 1-4, 2008
- NAUGLER, H.,(1986), *Evaluation et tri des documents informatiques en archivistique: un étude ramp*, accompagnée de principes directeurs, Unesco, Paris 1986, p. 5.
- PEROTIN, Y., (1996), «Administration and the "Three Ages" of Archives», *The American Archivist*, 1966, 29, pp. 363-369.
- RATNAPARKHI A., (1998), *Maximum Entropy Models for Natural Language Ambiguity Resolution*. Ph.D. Dissertation, University of Pennsylvania.

# PRESERVATION PLANNING: WHY & HOW TO IMPLEMENT? A CASE STUDY FROM THE NATIONAAL ARCHIEF OF THE NETHERLANDS

---

Bart Ballaux

Senior Advisor Digital Longevity, Nationaal Archief - National Archives of the Netherlands

*Preservation planning is one of the functions in the OAI reference model that has not yet received the attention of the digital preservation community that it actually deserves. However, if digital repositories should be capable of successfully processing the increasing mass of files that are ingested, a more automated and standardised approach for assessing which processes should be executed and which procedures adhered to, is necessary.*

*At this moment, a couple of archival institutions have already implemented some automated processes in their digital repositories, e.g. identification of ingested file formats, batch migration to identified long-term durable formats, etc. A full and working preservation planning functionality however is still to be developed. There is hardly any controversy about the need and importance of preservation planning, but there is still a huge gap between the abstract OAI model and its preservation planning functionality on the one hand, and an implemented and/or a semi-automated way of handling preservation planning, on the other hand. The Nationaal Archief of the Netherlands aims to use and fully incorporate preservation planning functionalities in its emerging digital repository, and is therefore testing and experimenting with preservation planning tools from the Planets project to come to standardised and automated preservation plans that can be implemented once electronic records are to be ingested in the digital repository.*

*The goal of this presentation is twofold. It will place preservation planning functionality in the wider context of digital preservation, and will indicate which elements preservation planning should consist of to be implemented in a digital repository. This theoretical framework will be used as an entry point for the second, practical part of the presentation, in which an example of preservation planning implementation is shown. It will focus on one process of the preservation planning functions in the OAI reference model.*

*On the basis of an experiment with records, it will be demonstrated:*

- how the Nationaal Archief of the Netherlands has gathered and identified requirements for these records,*
- how these requirements have been clustered and valued, and*
- how these gathering and valuation efforts have resulted in recommendations about a preferred strategy for the records.*

[See the presentation.](#)

## **Planifier la conservation : pourquoi et comment la mettre en oeuvre ?**

### **Etude du cas des Archives nationales des Pays-Bas.**

*La planification de la conservation est l'une des fonctions du modèle de référence OAIS, modèle qui n'a pas encore reçu toute l'attention qu'il mérite de la part de la communauté de la conservation du numérique, alors qu'en réalité il aide cette dernière. Cependant, s'il est vrai que les plates-formes d'archivage électronique devraient réussir à traiter le volume croissant des documents ingérés, il est nécessaire d'adopter une approche plus automatisée et plus normalisée pour analyser les processus à mettre en oeuvre et les procédures afférentes.*

*Actuellement, deux ou trois institutions d'archives ont déjà mis en place dans leurs plates-formes d'archivage électronique des processus automatisés comme, par exemple, l'identification des formats des documents intégrés, la migration des paquets vers des formats identifiés et pérennes, etc. Mais il reste à développer une fonctionnalité de planification complète et opérante de la conservation. Il n'y a pratiquement aucune controverse sur l'importance et la nécessité de planifier la conservation, mais il y a encore un fossé entre le modèle abstrait OAIS et la fonctionnalité de planification de la conservation d'une part, et un mode de gestion exécuté et/ou semi-automatisé pour planifier la conservation d'autre part. Les Archives nationales des Pays-Bas ayant l'intention d'utiliser et d'intégrer entièrement la fonctionnalité de planification de la conservation dans leur nouvelle plate-forme d'archivage électronique, elles sont en train d'expérimenter et de tester les outils de planification de la conservation du projet Planets afin d'obtenir des programmes de conservation normalisés et automatisés qui pourront être exécutés dès que les documents électroniques devront être intégrés dans la plate-forme.*

*L'objectif de cette présentation est double : elle replacera la fonctionnalité de planification de la conservation dans le contexte plus large de la conservation du numérique et identifiera les éléments qui doivent figurer dans la planification de la conservation pour qu'elle puisse être mise en oeuvre dans une plate-forme d'archivage. Ce cadre théorique sera utilisé comme point d'entrée de la seconde partie pratique au cours de laquelle il sera présenté un exemple d'application de la planification de la conservation. Cette partie sera centrée sur l'une des méthodes des fonctions de planification de la conservation du modèle de référence OAIS.*

*A partir d'une expérience réalisée sur des documents il sera démontré :*

- comment les Nationaal Archief des Pays-Bas ont réuni et identifié les spécifications de ces documents,*
- comment ces spécifications ont été regroupées et analysées,*
- comment ces travaux de collecte et d'analyse ont permis de formuler des recommandations sur la stratégie à adopter avec ces documents.*

[Consulter la présentation.](#)

# ARCHIVER ET DIFFUSER L'INFORMATION AU SEIN DU SYSTÈME SPAR DE PÉRENNISATION DE L'INFORMATION NUMÉRIQUE

---

Charlotte Fabre

Consultante EIM, responsable fonctionnelle SPAR, Atos Origin

Thomas Ledoux

Chef de projet SPAR, BnF

Laurent Sollier

Architecte, responsable technique SPAR, Atos Origin

*Pour préserver le patrimoine culturel numérisé, permettre d'accéder à ses archives et de les conserver sur le long terme, la BnF a confié à Atos Origin la conception et la mise en œuvre de son Système de Préservation et d'Archivage Réparti (SPAR). Ce système sera capable de stocker des masses de données, dépassant le million de giga-octets en 2010, en leur adjoignant les métadonnées indispensables à leur pérennisation et en leur appliquant les opérations de migration nécessaires pour assurer de maintien de leur communicabilité. Stocker et garder à jamais cet héritage culturel numérique, grâce aux nouvelles technologies, représente l'enjeu majeur de ce projet innovant de grande envergure.*

*De plus, au-delà de cet aspect de conservation SPAR, se base sur la norme OAIS (Open Archival Information System), doit offrir un accès facile et performant à ces données à ses différents utilisateurs (professionnels, lecteurs, internautes...).*

*Ces deux processus, conservation et accès, sont bien indissociables.*

*Comment alors satisfaire les utilisateurs en leur fournissant des objets numériques rapidement tout en les stockant sur des supports pérennes ?*

*Afin d'optimiser les coûts, et compte tenu de l'autonomie matérielle et de la masse de données, la Bibliothèque nationale de France a choisi d'utiliser une infrastructure de stockage basée, pour le stockage de conservation, sur des lecteurs de bandes. De plus, en cas de perte ou de problème sur certains supports ou sur la partie matérielle, les bandes valides restantes fourniront une information cohérente et autoporteuse.*

*Inversement, pour le stockage d'accès, la BnF a choisi d'utiliser une technologie disque plus performante.*

*Ces choix d'infrastructure associés à la nécessité de copier les différentes données sur plusieurs sites physiques et la recherche de l'indépendance par rapport aux changements technologiques des systèmes de stockage amènent à concevoir un système d'abstraction de stockage. Ce dernier est chargé de gérer les différents paquets d'informations tout au long de leur cycle de vie, en prenant en compte les impératifs de performance, d'accessibilité et de pérennité. C'est pourquoi l'application doit être capable de choisir automatiquement le meilleur support de stockage en fonction de l'action en cours.*

*Or, ces différents choix et exigences constituent une spécificité forte par rapport aux autres projets de même nature menés à travers le monde et impose une réflexion importante du point de vue de l'architecture du système.*

*Après une présentation succincte du projet SPAR et de son contexte, la présentation détaillera les particularités liées au stockage d'un point de vue fonctionnel et technique. Elle traitera aussi de la norme OAIS et de sa correspondance avec le projet.*

[Consulter la présentation : partie 1 ; partie 2 ; partie 3 ; partie 4.](#)

## **Archiving and disseminating information using the SPAR system for the long-term preservation of digital information**

*In order to conserve its digital cultural heritage, improve access to its archives and ensure their long-term preservation, the Bibliothèque nationale de France (BnF - National Library of France) commissioned Atos Origin to design and implement its distributed conservation and archiving system, the Système de Préservation et d'Archivage Réparti (SPAR). The system will be able to store very large quantities of data exceeding a million gigabytes in 2010, while adding the metadata that are indispensable for their long-term exploitation and performing the migration operations necessary to ensure that they remain communicable. The permanent storage and protection of this digital cultural heritage, using new technologies, is the major challenge addressed by this innovative and large-scale project.*

*SPAR, which is based on the Open Archival Information System (OAIS), must guarantee not only preservation but also easy and efficient access to the data for the various types of user (professionals, readers, web surfers etc).*

*These two processes go hand in hand.*

*How can users' needs be satisfied by providing both rapid access to digital objects and media capable of long-term preservation?*

*For greatest cost-effectiveness, bearing in mind the quantity of data and the need for equipment to function independently, the BnF chose to use digital tape drives for preservation storage. In addition, in the event of loss or other problems on any media or hardware, the remaining valid tapes will provide coherent and free-standing information.*

*For accessible storage, on the other hand, the BnF chose to use more efficient disk technology.*

*These infrastructure choices, together with the need to copy data to several locations and the desire to remain independent of any technological changes to the storage systems, led us to design a system with a storage abstraction layer. This is responsible for managing the different information packets throughout their life-cycle, taking account of imperatives concerning performance, accessibility and long-term preservation. For these reasons the application must be capable of choosing the best storage medium for a given action, automatically.*

*These choices and requirements are therefore particularly demanding compared with similar projects being carried out in other parts of the world, and oblige us to consider the appropriate system architecture very carefully.*

*The presentation will first give a brief overview of the SPAR project and its context, before detailing the functional and technical aspects of its specific storage requirements. It will also deal with the OAIS standard and how it applies to the project.*

[See the presentation : part 1 ; part 2 ; part 3 ; part 4.](#)

## ■ Introduction

Pour préserver le patrimoine culturel numérisé, permettre d'accéder à ses archives et de les conserver sur le long terme, la BnF a confié à Atos Origin la conception et la mise en œuvre de son Système de Préservation et d'Archivage Réparti (SPAR). Ce système sera capable de stocker des masses de données, dépassant le million de giga-octets d'ici à 2010, en leur adjoignant les métadonnées indispensables à leur pérennisation et en leur appliquant les opérations de migration nécessaires pour assurer le maintien de leur communicabilité. Stocker et garder à jamais cet héritage culturel numérique, grâce aux nouvelles technologies, représente l'enjeu majeur de ce projet innovant de grande envergure.

Pour ce faire, SPAR s'appuie sur les bonnes pratiques de la communauté des utilisateurs de la conservation du numérique ; plus spécifiquement il se base sur la norme OAIS (Open Archival Information System). De plus, un accès performant à ces données doit être offert à l'ensemble de ses différents utilisateurs (professionnels, lecteurs, internautes...).

Les trois aspects du processus, collecte, conservation et accès, restent indissociables.

## ■ Problématique

Comment alors satisfaire les utilisateurs en leur fournissant des objets numériques rapidement tout en les stockant sur des supports pérennes ?

Afin d'optimiser les coûts et compte-tenu de la masse sans cesse croissante des données, la Bibliothèque nationale de France a choisi d'utiliser, pour le stockage de conservation, une infrastructure basée sur des bandes magnétiques de dernière génération. Chacun de ces supports offre une grande autonomie : ils sont indépendants, ne consomment pas d'énergie par eux-mêmes et peuvent être dupliqués ou même déplacés aisément en cas de besoin.

Inversement, pour le stockage d'accès ou de collecte, afin de garantir un accès performant, la BnF a choisi d'utiliser une technologie disque.

A ces choix d'infrastructure, s'ajoute la nécessité de copier les différentes données sur plusieurs sites géographiques et la volonté de conserver une indépendance forte par rapport aux changements technologiques des systèmes de stockage. Il est donc nécessaire de concevoir un système d'abstraction de stockage chargé de gérer les différents paquets d'information tout au long de leur cycle de vie, en prenant en compte les impératifs de performance, d'accessibilité et de pérennité. Ainsi l'application cliente sera en mesure de choisir automatiquement le meilleur support de stockage en fonction de la finalité de cette action.

Or, ces différents choix et exigences constituent une spécificité forte par rapport aux autres projets de même nature menés à travers le monde et impose une réflexion importante du point de vue de l'architecture du système.

Après une présentation succincte du projet SPAR et de son contexte, la présentation détaillera les particularités liées au stockage d'un point de vue fonctionnel et technique. Elle traitera aussi de la norme OAIS et de sa correspondance avec le projet.

## ■ Le projet SPAR en quelques mots

Le système de préservation et d'archivage réparti a pour but de garantir la pérennité des données sur du très long terme. Les risques majeurs liés à la perte de l'intelligibilité ou de la lisibilité des informations proviennent de l'obsolescence technologique (formats, supports, logiciels). SPAR offre une réponse à ces problématiques par la mise en place d'une traçabilité aussi exhaustive que possible de l'ensemble des éléments mis en cause au cours de la vie des données.

Le planning du projet est articulé en deux grands moments : il s'agit dans un premier temps d'élaborer le noyau SPAR qui devrait être mis en œuvre opérationnel au premier semestre 2009 puis dans les 3 prochaines années d'étendre la couverture du système à l'ensemble des types de données numériques gérées par la BnF.

Il s'agit en effet pour SPAR de gérer les informations en provenance de multiples applications de production de données, de les archiver, tout en les diffusant au travers d'applications multiples (dont la plus emblématique pour la BnF est Gallica.).

La production d'un ensemble de données cohérent et identifié, couplée à la diffusion de ces mêmes données à des utilisateurs avec des règles de gestion spécifiques au sein du système SPAR forme une filière. On peut citer notamment :

- la filière numérisation de conservation,
- la filière production administrative (« Records Management »),
- la filière dépôt légal du Web,
- ...

## ■ SPAR, OAIS et les fonctionnalités liées au stockage

La conception de SPAR est fortement basée sur le modèle OAIS. Cela implique un système modulaire avec une définition des responsabilités claires entre les différents acteurs : Archive, producteurs, utilisateurs, administration...

Cela implique aussi un modèle de données structuré comprenant des métadonnées descriptives, administratives, de pérennisation... ainsi qu'une structuration de l'information en paquets (paquet de versement dit SIP, paquet d'archive dit AIP, paquet de diffusion dit DIP).

Toutefois le système SPAR se distingue de l'OAIS sur plusieurs points et notamment un module de gestion des droits (lié aux droits d'auteurs sur les objets numériques), mais aussi un service d'abstraction de stockage (SAS) permettant de dissocier stockage logique (réalisé par le module Stockage) et stockage physique (réalisé par l'infrastructure).

Cependant, l'abstraction ne peut être « opaque » car l'OAIS mentionne certaines fonctions telles que le remplacement des supports, le contrôle d'erreurs de l'infrastructure, mais aussi les migrations liées aux objets numériques. Or toutes ces fonctions au sein de SPAR sont supportées par le service d'abstraction de stockage ou SAS.

## ■ Fonctionnalités du SAS

Le SAS a plusieurs clients qui vont lui demander de prendre en charge une information numérique (des paquets ou de simples fichiers) pour la stocker sur les éléments de stockage de l'infrastructure ou la récupérer. Aujourd'hui trois modules communiquent avec le SAS :

- le module Versement qui va demander l'écriture (temporaire) d'un paquet qui vient d'être versé.
- le module Stockage qui va demander l'écriture d'un paquet dont le versement a été validé et qui sera le master à conserver sur une durée très longue
- le module Accès qui va demander l'écriture d'objets données transformées qui seront utiles pour construire les paquets de diffusion.

C'est au SAS de donner les instructions d'écriture à l'infrastructure. Pour cela, il s'appuie sur les exigences fournies par le module Client qui impliquent la mise en œuvre de règles de gestion différentes selon le type d'enregistrement :

- Le module Versement aura besoin d'écrire rapidement son paquet, tout en ayant l'assurance que ce dernier soit répliqué sur un site de secours.
- Le module Stockage aura besoin d'écrire son paquet sur des supports pérennes (des bandes vraisemblablement), d'en faire plusieurs copies sur des sites distants, de requérir ou non du chiffrement, de la compression etc.
- Le module Accès aura besoin d'un accès rapide aux informations qu'il aura confiées au SAS (des supports de disques vraisemblablement).

Pour gérer toutes ces spécificités, le SAS propose des capsules de stockage qui pourront répondre aux exigences spécifiées par les clients. Ces capsules de stockage vont gérer les enregistrements du SAS.

C'est au module Client de trouver la capsule de stockage adéquate parmi une liste proposée par le SAS. Pour ce faire, le SAS publie la liste des classes de services associés à chaque capsule. Ces dernières indiquent les caractéristiques saillantes pour permettre ce choix. Il peut s'agir des temps moyens de lecture ou d'écriture, de la capacité à offrir plusieurs copies, de l'utilisation d'un algorithme de chiffrement, ... Au vu de ces caractéristiques, le module Client peut alors sélectionner la capsule

### ***Quelques précisions de vocabulaire :***

#### *Enregistrement :*

Un enregistrement est une information gérée par le SAS. Ce dernier ne sait pas qu'elle peut correspondre à un paquet d'information au sens OAIS. Ce sont les clients du SAS qui ont la vision sur le contenu et qui peuvent ouvrir les paquets d'information et auditer leur contenu par exemple. Pour le SAS, l'enregistrement est figé, non modifiable. Chaque enregistrement appartient à une capsule et une seule.

#### *Copie :*

Selon les besoins, un enregistrement peut avoir plusieurs copies dans le SAS. Toutes les copies sont évidemment identiques. Au sein de son référentiel, le SAS ne connaît qu'un enregistrement qui a plusieurs copies. Chaque copie est écrite sur un élément de stockage.

#### *Capsule de stockage :*

Il s'agit d'une entité permettant de connaître les caractéristiques de stockage. Les capsules sont paramétrées par l'administrateur qui définit, entre autre, combien d'éléments de stockage leur sont rattachés ainsi que le nombre de copie des enregistrements. Tous les enregistrements d'une même capsule ont les mêmes caractéristiques de stockage.

Par ailleurs, le SAS offre plusieurs services :

- création d'un enregistrement,
- lecture sécurisé d'un enregistrement,
- audit d'un enregistrement,
- recherche d'un enregistrement,
- ajout/modification d'attributs d'un enregistrement,
- liste des capsules
- état d'une capsule
- modification d'une capsule par ajout/remplacement d'un élément de stockage.



En particulier, il est important de noter que chaque enregistrement est accompagné d'une clé de contrôle d'intégrité qui permet à chaque lecture de vérifier que la copie fournie est intègre. Dans le cas contraire, le mécanisme d'audit est mis en œuvre.

Ce mécanisme consiste à auditer toutes les copies d'un enregistrement. Il s'agit de vérifier la clé de contrôle d'intégrité de chaque copie et en cas de copie non intègre de planifier la réécriture de cette copie en partant d'une copie intègre.

De même, chaque modification des capsules de stockage (liée à l'évolution de l'infrastructure) doit être accompagnée de la migration de réplication (passage des copies d'un ancien élément de stockage à un nouveau) nécessaire.

## ■ iRODS et implémentation

Dans ses choix techniques de mise en œuvre, Atos Origin a choisi le framework iRODS.

En effet, bien qu'il existe différents produits couvrant cette notion d'abstraction de stockage, seul iRODS répond au plus grand nombre d'exigences et de contraintes formulées par la BnF.

Par exemple : iRODS se positionne comme un « middleware » de stockage et à ce titre, il offre un mécanisme de règles permettant de mettre en œuvre les différents comportements souhaités selon les types d'enregistrements soumis.

De plus, il n'est pas attaché à une technologie matérielle et cet aspect agnostique lui permet d'envisager une adaptation aux futurs standards de stockage. Ainsi, à ce jour, il exploite un mécanisme de système de fichiers proposé par le système d'exploitation hôte, mais demain, il peut s'appuyer sur un système de base de données. Pour ce faire, il est construit sur le principe de « drivers ».

Enfin, iRODS est un produit respectant une licence « Open Source ». De ce fait, il garantit aux utilisateurs une indépendance vis à vis de l'actuel éditeur, il offre le bénéfice des évolutions apportées par la communauté des utilisateurs et enfin s'inscrit dans la pérennité de l'information du fait qu'aucun point technique n'est masqué.

## ■ Scénarios catastrophes évités

Tous ces choix sont essentiels pour mettre en œuvre le système SPAR et ainsi prévenir les différents scénarios catastrophes que tous, nous voulons éviter.

### Exemple 1 :

Nous montrerons ce qu'il adviendra dans SPAR en cas d'incendie du site principal.

En effet, une bibliothèque est grandement vulnérable dans le cas d'un incendie touchant les magasins de documents. Les données numériques n'échappent pas à ce risque et il convient de s'en protéger.

Pour répondre à ce besoin, le SAS, à travers la capsule de stockage sélectionnée par le module client, permet une réplication automatique des enregistrements sur un site distant. Ce site distant bénéficie d'une infrastructure à même de fournir le service souhaité mais en mode dégradé.

Il faut noter que les paquets stockés sont « autoporteurs » d'un point de vue fonctionnel et technique. Ainsi, si cet incendie était le fruit d'une déstabilisation majeure de la société (ex : guerre, catastrophe naturelle, ...) entraînant une perte des données stockées sur les baies de disques, il serait possible de reconstruire l'information à l'aide des métadonnées intégrées aux paquets présents sur le stockage sur bande. Ainsi, il serait possible de reconstruire le magasin sans perte d'informations.

### **Exemple 2 :**

Nous montrerons aussi ce qu'il se passera quand un élément de stockage deviendra obsolète

L'évolution technologique passée prouve l'existence de sauts technologiques récurrents. Aujourd'hui il est possible de dire que tous les 10 ans un palier technologique majeur est franchi. De ce fait, il est certain que les supports physiques retenus sur l'actuelle infrastructure matérielle de SPAR seront obsolète à moyen terme et qu'il faudra envisager leur remplacement. Cette étape ne doit pas avoir d'impact ni sur les données hébergées ni sur l'application matérialisant SPAR.

Ainsi, le SAS offre des services de migration de support permettant de transférer les enregistrements d'un type de support à un autre type de support. Cette migration est accompagnée d'une étape de contrôle permettant de garantir l'intégrité des enregistrements migrés. De plus, cette migration est transparente pour les modules clients, qui continueront à accéder aux enregistrements par le biais du même canal d'échange sans se soucier de la technologie mise en œuvre sur l'infrastructure.

### **Exemple 3 :**

Nous montrerons aussi comment seront gérées les copies identifiées non intègres.

Selon la capsule de stockage sélectionnée, chaque enregistrement présent dans le SAS peut bénéficier de plusieurs copies. Il convient donc de s'assurer que chacune de ses copies soit intègre au regard de l'enregistrement fourni par le client.

Pour ce faire, chaque enregistrement reçu bénéficie d'une clé de contrôle d'intégrité (par exemple au format md5) et chacune des copies attachées se voit attacher sa propre clé d'intégrité. La cohérence de ces différentes clés est effectuée de façon systématique à chaque consultation d'un enregistrement et régulièrement un plan d'audit valide la cohérence de la totalité des copies d'un enregistrement.

Si toutefois l'une des copies venait à ne pas être intègre, le SAS effectue un remplacement de la copie défectueuse à l'aide d'une copie identifiée comme intègre. Cette correction s'effectue à la volée lors de la consultation mais aussi lors de plan d'audits récurrents.

Ainsi, le SAS garantit l'intégrité des enregistrements hébergés.

## **■ Conclusion**

Le SAS est indispensable pour que SPAR puisse assurer ses missions. Cette présentation montre bien que tout ne dépend pas du matériel et que les choix faits à un instant T peuvent et doivent évoluer pour s'abstraire de l'obsolescence technologique.

# LA PLATE-FORME D'ARCHIVAGE ÉLECTRONIQUE DU DÉPARTEMENT DES YVELINES : DES ENJEUX STRATÉGIQUES AUX CHOIX TECHNIQUES

---

Elisabeth Gautier-Desvaux

Directeur des Archives départementales des Yvelines

Gaëlle Mignot

Chargée de mission informatique aux Archives départementales des Yvelines

*Le Département des Yvelines, engagé dans une politique d'innovation technologique multiforme, a développé une plateforme d'archivage électronique destinée à conserver toutes les données numériques natives des services publics situés sur le territoire départemental. Ce projet se situe aujourd'hui à une étape-clé : pendant dix-huit mois, il a en effet représenté un enjeu expérimental ; il constitue désormais une réalité quotidienne.*

*Le projet a débuté en juillet 2005, avec la rédaction d'une fiche projet relative à la mise en œuvre d'une plateforme d'archivage électronique. Sur la base des résultats de l'enquête engagée à l'instigation de la Direction des Archives de France sur ce type d'infrastructure, les Archives départementales et la Direction des systèmes d'information ont ensuite déterminé les spécifications fonctionnelles et techniques de cet outil informatique d'échanges et de conservation.*

*La réflexion initialement limitée à l'archivage des actes soumis au contrôle de légalité, s'est progressivement élargie à d'autres sources de données, sous réserve de leur conformité avec le standard d'échange de données pour l'archivage établi par la Direction générale de la modernisation de l'Etat (DGME) et la Direction des Archives de France (DAF).*

## **A première vue, un projet technique...**

*La démarche adoptée est pragmatique : les premières fonctionnalités développées répondent aux besoins immédiats en matière de transfert et de réception d'archives électroniques, de gestion des données descriptives, de consultation et de communication, de stockage et d'administration. De même, les choix d'architecture technique répondent aux maîtres mots « simplicité » et « automatisation », en prévision des grosses volumétries prévisibles. Les formats sont filtrés à l'entrée mais aucune politique de conversion n'a encore été mise en place. Quant aux coûts, ils restent assez modérés au regard de l'ampleur du projet.*

**...mais dont les enjeux politiques et juridiques sont forts, voir bloquants s'ils ne sont pas pris en compte à temps**

*Si les questions techniques ont focalisé l'attention des acteurs à la genèse du projet, les paramètres politiques et juridiques se sont révélés beaucoup plus délicats à manier. Les archivistes ont notamment dû s'interroger sur le moment où le transfert devrait intervenir (archives intermédiaires / archives définitives), sur la manière de formaliser leurs relations avec les services versants (conventions de transfert) et sur l'ouverture éventuelle de la plateforme, par voie de subsidiarité, aux services d'autres collectivités du territoire départemental (conventions de dépôt).*

## **Des procédures organisationnelles déterminantes à la réussite du projet**

*Outre l'instauration d'une gestion de projet classique, basée sur un comité de pilotage et une équipe projet, les archivistes ont conçu de nouvelles procédures organisationnelles. Celles-ci s'appuient sur des pratiques déjà éprouvées pour les supports traditionnels mais ont nécessité quelques adaptations : articulation Archives départementales / collectivité déposante, contrôle de flux de l'ordre d'une centaine de transferts par semaine, validation définitive.*

*L'impact de cette plateforme d'archivage électronique est encore difficile à mesurer. Ce qui ne change pas, en tout état de cause, c'est le rôle joué par les archivistes en matière de sélection documentaire, d'organisation et de gestion de l'information.*

*La communication à deux voix proposée par les Archives départementales des Yvelines permettra de présenter ce projet sous plusieurs facettes : politiques, juridiques, techniques et organisationnelles.*

[Consulter la présentation.](#)

## ***The Digital repository for the Yvelines Département***

*The Yvelines Département pursues a policy favouring innovative multi-format technologies and has developed a digital repository designed to preserve all digital data issuing from the public services located within the Département. The project is currently at a key stage: after an experimental phase lasting 18 months, it has now become operational.*

*The project began in July 2005, when a project outline was drawn up for the implementation of a digital repository. Based on the results of a survey commissioned by the French Archives authority (DAF) on this type of infrastructure, the Yvelines Archives and the Information Systems Directorate wrote the functional and technical requirements for a computerised system for the transfer and preservation of documents.*

*The analysis began by considering archiving documents subject to legal obligations, before broadening its outlook progressively to take in other sources of data, on condition that these complied with the data exchange standard for archiving laid down by the French Directorate for the Modernising of State Affairs (DGME) and the DAF.*

### ***At first sight, a technical project ...***

*A pragmatic approach was taken: the first functions developed addressed the immediate needs concerning transfer and reception of electronic archives, management of descriptive data, consultation and communication, storage and administration. In the same way, the watchwords for the choice of technical architecture were 'simplicity' and 'automation', in preparation for the large quantities expected. Formats are filtered on input but no conversion policy has yet been implemented. Costs remain moderate considering the scale of the project.*

***... but with significant political and legal challenges, that could block progress if not taken into account at an early stage***

*Whereas technical issues were the focus of attention at the start of the project, political and legal questions turned out to be far more difficult to manage. In particular, the archivists had to consider the best time to make the transfer (from intermediate to final archives), how to formalise relations with the depositing services (via transfer agreements) and whether the platform should be opened to the services of other local authorities in the Département (through deposit agreements) by way of subsidiarity.*

### ***Organising procedures determine project success***

*As well as implementing a traditional project management structure, based on a steering committee and a project team, the archivists instigated new organising procedures. These were based on practices that have proved successful for traditional media, although they required a few modifications to cover the relationships between the Yvelines Archives and depositing bodies, how to monitor the flow of about a hundred transfers per week, and applying final validation.*

*It has so far proved difficult to measure the impact of this digital repository. One thing that has not changed, in any case, is the role played by the archivists in document selection and the organisation and management of information.*

*The joint presentation planned by the Yvelines Département Archives will illustrate several different facets of the project: political, legal, technical and organisational. Elisabeth GAUTIER-DESVAUX, who holds the title of conservateur général du patrimoine (Archives curator) and is Director of the Yvelines Archives, will speak on the strategic challenges and organisational choices, while Gaelle MIGNOT, attaché de conservation du patrimoine (curator assistant) with responsibility for IT, will talk about the technical options and the lessons learned from the first applications.*

[See the presentation.](#)

Engagé dans une politique d'innovation technologique multiforme, le département des Yvelines a conçu une plateforme d'archivage électronique destinée à conserver les données numériques natives des services publics implantés sur son territoire. Ce projet se situe aujourd'hui à une étape-clef : durant 18 mois, il a en effet représenté un enjeu expérimental, appliqué au traitement des actes des collectivités soumis au contrôle de légalité de l'Etat, initialement conservés sur la plateforme de la Caisse des dépôts et consignations (CDC); il constitue désormais une réalité quotidienne, appliquée à des sources diversifiées telles que le sommier foncier de l'Etablissement public d'aménagement de Saint-Quentin-en-Yvelines (EPASQY).

Les prémisses de cet investissement numérique résident sans nul doute, dès le début des années 2000, dans la mise en oeuvre d'un ambitieux programme de numérisation de masse et de mise en ligne des documents les plus précieux ou les plus consultés (près de 3,7 millions de pages numérisées dont 3 millions mises en ligne et 3,5 millions proposées à des annotations coopératives), à l'occasion du transfert de la Direction des Archives départementales dans de nouveaux locaux, en Ville nouvelle de Saint-Quentin.

Mais il convient aussi et surtout de relier cette initiative à la « culture d'innovation » développée conjointement par les services informatiques et d'archives, notamment via le partenariat de recherche appliquée souscrit avec l'Institut national de recherche en informatique et en automatique (INRIA) et l'Institut national des sciences appliquées (INSA) de Rennes pour la reconnaissance automatique de structures et d'écritures manuscrites. Cette culture, qui anime leurs projets de service respectifs, les a ainsi conduits, d'une part à coopérer dès 2005 à l'étude menée par la société PARKER-WILLIBORG pour le compte de la Direction des Archives de France (DAF), sur la mise en oeuvre d'une plateforme d'archivage électronique<sup>1</sup>, d'autre part à s'impliquer dans la démarche d'élaboration du standard d'échange de données pour l'archivage<sup>2</sup> engagée par cette direction du Ministère de la culture et la Direction générale de la modernisation de l'Etat (DGME).

## ■ Des enjeux stratégiques au principe de réalité

### Une impérieuse nécessité technique

Il faut bien reconnaître que le « principe de réalité » s'est avéré moteur, lorsqu'en 2005, la CDC a alerté ses partenaires sur la perspective de saturation de la plateforme d'échange expérimentale FAST, développée depuis début 2004 dans le cadre d'une convention avec le département des Yvelines : dès l'année suivante et sauf solution de transfert, l'ensemble des données électroniques ainsi stockées se trouvait menacé de disparition.

Il convenait donc d'être en mesure de prendre le relais et d'archiver les quelque 3,5 Mo engrangés<sup>3</sup>, que ce soit par le recours à un tiers-archivageur du marché, par voie de mutualisation avec une autre structure publique ou grâce au développement d'une solution intégrée.

A cela s'ajoutait le constat des versements électroniques déjà pris en charge selon des standards plus ou moins atypiques, comme les procédures classées sans suite du Tribunal de Grande instance de Versailles, au titre de l'application de la « Nouvelle chaîne pénale »<sup>4</sup>, ou le sommier de l'EPASQY<sup>5</sup>.

### Une philosophie d'action partagée : le choix d'un pragmatisme raisonné

Très rapidement, et dans le sillage du partenariat soudé entre ces deux services, la Direction des systèmes d'information et celle des Archives ont opté pour la mise au point d'une plateforme d'archivage électronique expérimentale intégrée, conciliant rigueur et pragmatisme. Un consensus se dégagait en effet pour trouver un juste milieu entre le nécessaire investissement conceptuel et la mise au point d'un outil simple, robuste et ergonomique.

---

<sup>1</sup> Etude qui a mis en évidence la pertinence d'un recours au modèle *Open archival information system(OAIS)*.

<sup>2</sup> Standard diffusé en mars 2006.

<sup>3</sup> Portés à 8,5 Mo, fin 2006.

<sup>4</sup> Versement de la NCP intervenu à compter de mai 2004, au format ASCII, pour 292,2 Mo.

<sup>5</sup> Versement intervenu en décembre 2005, au format XML (standard d'échanges de données, pour l'archivage, version soumise à commentaires), pour 22 Mo.

L'investissement conceptuel, nous l'avons en partie engagé en transcrivant l'étude PARKER-WILLIBORG déjà citée. Nous allons le poursuivre également en nous associant à la réflexion initiée par la DAF pour déterminer les contours de la plateforme pilote PILAE. Nous allons enfin approfondir le dialogue entre informaticiens et archivistes, sur le mode « prouver le mouvement en marchant », adapté à une collectivité de la taille des Yvelines dont les moyens permettent de passer assez souplement de l'intention à l'action.

Sans pour autant perdre de vue la nécessité d'une architecture d'ensemble transposable à des occurrences diversifiées, nous avons donc décidé de répondre aux besoins immédiats de transfert et de réception d'archives électroniques, de stockage, de gestion et de consultation des données, tout en nous adaptant aux grosses volumétries prévisibles. C'était d'emblée reconnaître que du chemin resterait encore à parcourir en ce qui concerne la description fine et la communication des sources ainsi constituées, en étroite relation avec les instruments de recherche encodés ouvrant l'accès aux archives papier et les données numériques produites au terme des opérations de numérisation dont le traitement relèvera du progiciel documentaire en cours d'acquisition.

### **Une dimension juridique insoupçonnée, ou du principe de subsidiarité**

On le voit, nécessité faisant loi, les paramètres techniques ont fortement pesé sur la genèse du projet. C'était compter sans une composante juridique importante de l'archivage des actes soumis au contrôle de légalité : le principe de subsidiarité à l'égard des collectivités ne disposant pas des moyens de stockage électronique et ses limites.

Nous substituant en effet aux obligations des communes et groupements de communes auxquelles incombe une obligation d'archivage et de communication des documents qu'elles ont produits, en dehors de tout lien hiérarchique entre le niveau départemental et le niveau communal, il nous fallait prévoir des dispositions contractuelles telles que prévues pour le dépôt volontaire des archives définitives des communes de plus de 2000 habitants.

Cette étape franchie, demeure aujourd'hui la question du périmètre de cette subsidiarité (en termes de seuils démographique et budgétaire) et des dispositions transitoires à envisager à l'égard des collectivités les plus importantes, dans l'instance de leur mise en capacité d'héberger et traiter ces données. Demeure aussi l'appréciation de la charge représentée par les demandes de communication qui devraient incomber aux collectivités déposantes, moyennant une gestion fine des droits d'accès des utilisateurs, sur la base d'un annuaire détaillant leurs profils.

Il faut toutefois reconnaître que la transposition de ce principe à l'échelle nationale impacte plus largement la relation entretenue par toute collectivité avec son territoire de compétence, et rejoint certains questionnements politiques sur les risques de recentralisation implicite induits par les regroupements électroniques sectoriels au sein de ministères tels que l'Intérieur, la Justice ou les Finances.

### **Un repositionnement des étapes de traitement archivistique**

Une autre incidence notable du développement de la plateforme réside dans le positionnement de plus en plus précoce de l'intervention archivistique, à la source même du processus administratif auquel elle est liée. C'est ainsi que les transferts de données électroniques relatives au contrôle des actes interviennent dès leur validation par les services préfectoraux, deux mois tout juste après l'émission du *recepisse* correspondant<sup>6</sup>.

On le verra également ci-après dans le propos de Gaele Mignot sur l'architecture technique et fonctionnelle de la plateforme : des étapes successives de contrôle par les archivistes en charge du traitement des archives électroniques, et de validation portant valeur juridique de prise en charge par le directeur sont prévues. La traçabilité de ce processus se trouve assurée par la fonction "journal" enregistrant l'ensemble des interventions.

Initialement, les transferts depuis le service versant étaient programmés "au fil de l'eau", dans des "paquets d'informations"<sup>7</sup> de volume très variable mais souvent restreint. Les différents échanges intervenus, notamment lors des journées de présentation de la plateforme, nous inciteront très vraisemblablement à

---

<sup>6</sup> Deux mois, soit le délai légal au terme duquel les actes sont réputés valides, sauf observation préfectorale.

<sup>7</sup> Dans le cas de figure exposé, un « paquet d'informations » est toujours accompagné de son « information de description ».

revoir ce principe et à envisager un allotissement destiné à éviter des interventions trop segmentées, privées des indispensables liens contextuels. Ici encore, seul un dialogue étroit et précoce avec les services producteurs permettra de déterminer la périodicité des transferts, de même qu'en interne, une régulation des flux garantira la fluidité de leur traitement, entre les niveaux de contrôle et de validation.

### **Pérenniser et sécuriser la mémoire numérique, tout en assurant la continuité de l'accès**

Dès la conception de la plateforme, et au regard de l'expérience engrangée pour les fichiers résultant de la numérisation des fonds les plus précieux ou les plus consultés, le principe d'une réplique exhaustive et quasi immédiate a été adopté. Deux serveurs, l'un implanté sur le site du Conseil général, l'autre aux Archives départementales, conservent donc des versions redondantes des données transférées, qui permettront de sauvegarder et, si nécessaire, de restaurer une version intègre des documents<sup>8</sup>.

Ainsi s'esquisse progressivement la perspective d'une répartition des vocations respectives des serveurs : un serveur d'écriture et de conservation aux Archives, l'autre de consultation au Conseil général, garantissant par cette proximité avec le parc informatique départemental une maintenance 24H sur 24, sachant que la majorité des incidents actuellement enregistrés sur le site intervient en dehors des heures d'ouverture des services et coïncident avec une fréquentation maximale des internautes.

A terme, on l'a vu, se posera la question d'une éventuelle mutualisation de ces données électroniques natives avec les instruments de recherche encodés et le produit des campagnes de numérisation, dans un souci d'élargissement de la ressource archivistique et du potentiel de recherches. Cette perspective impliquera en outre une extension de la capacité de stockage qui devrait pouvoir être portée à 15 To.

Restera également à résoudre la question de la sécurisation des données référentielles, métadonnées et instruments de recherche électroniques, via le recours circonstancié à un coffre-fort électronique.

## **■ L'architecture technique et fonctionnelle de la plate-forme**

Compte tenu des impératifs calendaires et de l'approche pragmatique retenue, la réflexion a intégré prioritairement les besoins spécifiques de l'archivage des actes des collectivités soumis au contrôle de légalité de l'Etat, puis elle s'est élargie au cas du sommier foncier de l'EPASQY.

---

<sup>8</sup> Il est prévu une réplique asynchrone quotidienne.

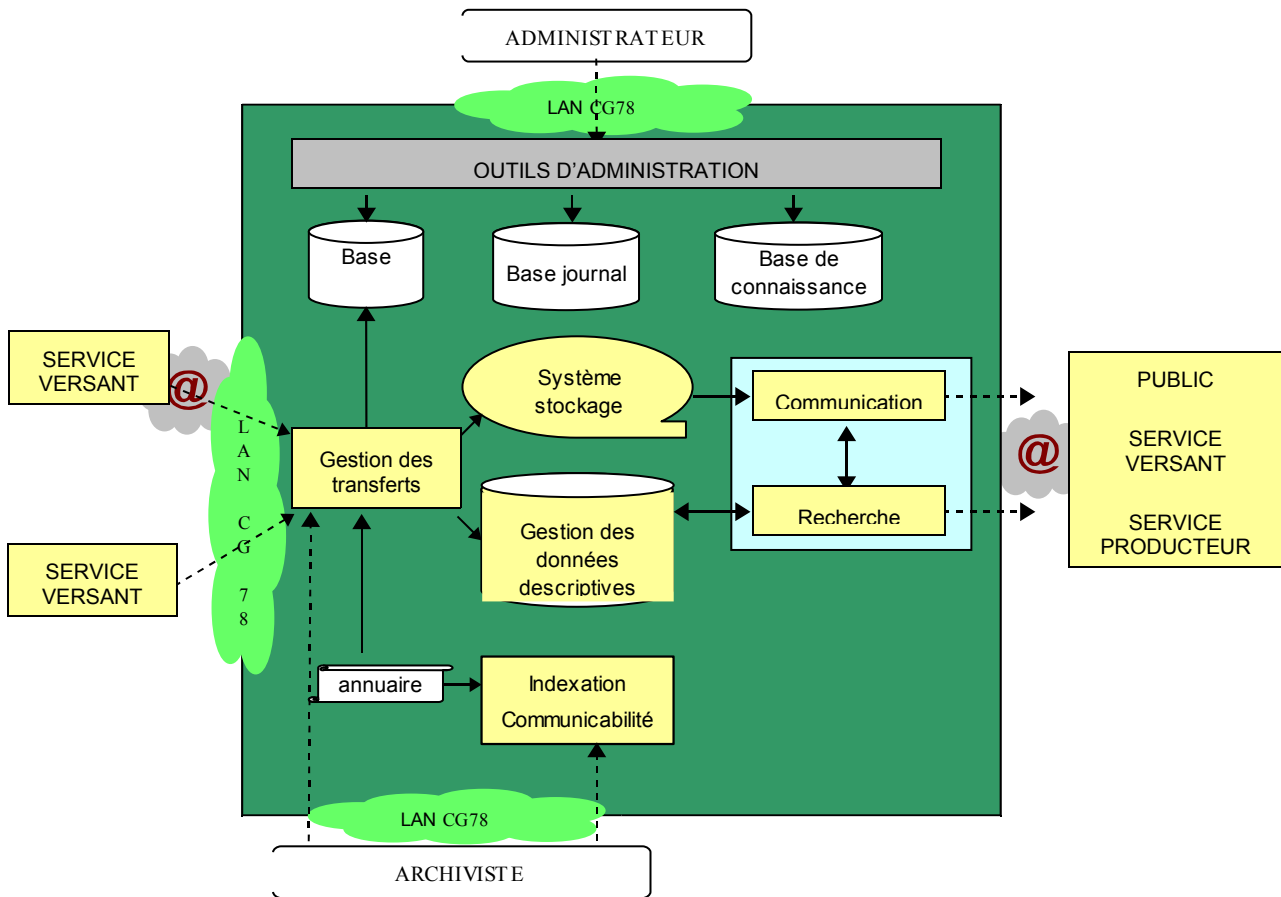


Fig. 1 : Architecture fonctionnelle de la plateforme d'archivage électronique du Conseil général des Yvelines

## Les premières fonctionnalités développées

### Le transfert

Cette fonction gère la transmission, le contrôle, le rejet et le traitement des paquets d'informations en provenance des services versants. Seuls les paquets structurés conformément au standard d'échange de données pour l'archivage peuvent être intégrés dans la plateforme. De même, les formats acceptés en entrée sont définis en amont : à ce jour XML, PNG et PDF. La transmission de ces paquets s'effectue soit automatiquement via le réseau, soit manuellement sur supports amovibles.

Lors de l'arrivée d'un paquet d'informations, une série de contrôles automatiques est réalisée par la plateforme, à savoir :

- La présence de plusieurs balises définies comme indispensables lors de la modélisation (service versant, service d'archives, poids, format, identifiant, etc.),
- L'adéquation entre le nom du destinataire du paquet d'informations et le nom du service d'archives récepteur,
- L'autorisation préalable accordée au service versant, ou au service producteur si distinct, d'envoyer un paquet d'informations vers la plateforme,



- La vérification, en cas de transfert automatique via le réseau, de l'adresse IP<sup>9</sup>,
- La correspondance entre le type MIME<sup>10</sup> et le nom de format déclaré dans le paquet d'informations ; les noms de formats doivent être conformes aux accords passés préalablement au transfert,
- La présence d'une pièce jointe<sup>11</sup>,
- Le calcul de l'empreinte dont le résultat doit être conforme à celui indiqué juste avant l'envoi<sup>12</sup>,
- La vérification que l'espace disque est suffisant sur le système pour stocker le contenu.

Une fois ces contrôles automatiques réalisés, l'archiviste en charge du traitement des archives électroniques effectue, après s'être authentifié, une série de contrôles portant davantage sur le contenu. Il constitue des lots<sup>13</sup> et pratique ensuite un contrôle par échantillonnage. Plusieurs types d'affichages lui sont proposés : vue synthétique d'informations relatives à plusieurs paquets d'informations<sup>14</sup>, détail des informations contenues dans un paquet et visualisation arborescente des sous-niveaux descriptifs<sup>15</sup>. L'archiviste en charge du traitement des archives électroniques formule un avis sur le traitement à réserver au paquet d'informations : refus ou prise en charge.

Le directeur prend connaissance de cet avis, après s'être authentifié. Il dispose des mêmes possibilités d'affichage que ci-dessus et décide du traitement à appliquer au paquet d'informations qui peut être conforme ou différent de l'avis énoncé par l'archiviste en charge du traitement des archives électroniques.

Toutes les opérations intervenues entre la réception d'un paquet d'information et sa sortie définitive du système (restitution à la collectivité ou élimination) sont enregistrées dans la base journal.

Cette fonction gère également le transfert du contenu du paquet d'informations vers le système de stockage et l'extraction de l'information de description en vue de son intégration dans la base dédiée à sa gestion.

### ***Gestion des données descriptives***

Cette fonction traite l'ensemble des métadonnées de description, de conservation et de gestion. Ces métadonnées, notamment celles relatives à la communicabilité et à l'indexation, peuvent être, si besoin, modifiées par l'archiviste en charge du traitement des archives électroniques.

### ***Recherche et communication***

Cette fonction concilie l'ensemble des mécanismes permettant d'accéder et de consulter les informations disponibles, qu'il s'agisse des données descriptives ou du contenu des documents numériques proprement dit.

Le lecteur sélectionne les documents ou données qui l'intéressent à partir des termes descriptifs contenus dans les métadonnées. Il pourra alors avoir communication du contenu, sous réserve que celui-ci soit librement communicable. Actuellement, les droits propres à certaines catégories de public (services producteurs, bénéficiaires de dérogation, etc.) ne sont pas pris en compte.

### ***Stockage***

Cette fonction gère la conservation des métadonnées et des contenus à archiver, l'espace de stockage et les sauvegardes. Les contenus à archiver sont répartis dans un plan de classement méthodique. Tenant compte

---

<sup>9</sup> Internet Protocol, identifiant numérique unique d'un ordinateur sur un réseau.

<sup>10</sup> Multipurpose Internet Mail Extension, permet d'identifier ou de lire le format d'un fichier.

<sup>11</sup> Le standard d'échange de données pour l'archivage peut être utilisé pour des versements papier ou électroniques ; dans ce dernier cas, le contenu à archiver est joint à l'envoi.

<sup>12</sup> Clé de sécurité d'une centaine de bits que l'on constitue à l'aide d'un algorithme pour identifier un document.

<sup>13</sup> Par exemple : délibérations du Conseil général des Yvelines pour le mois de décembre 2008.

<sup>14</sup> Notamment, identifiant, intitulé, description, dates extrêmes, communicabilité, indexation.

<sup>15</sup> Affichage des données descriptives d'un paquet et ouverture des pièces jointes.

de la vulnérabilité des supports, l'architecture technique retenue pour cette première phase du projet repose sur la redondance des informations, copiées sur des supports de type bandes et disques magnétiques dont les durées de vie diffèrent. A ce choix, plusieurs avantages : il s'avère peu onéreux, présente un mode d'administration plutôt simple et permet d'accroître aisément la capacité de stockage.<sup>16</sup>

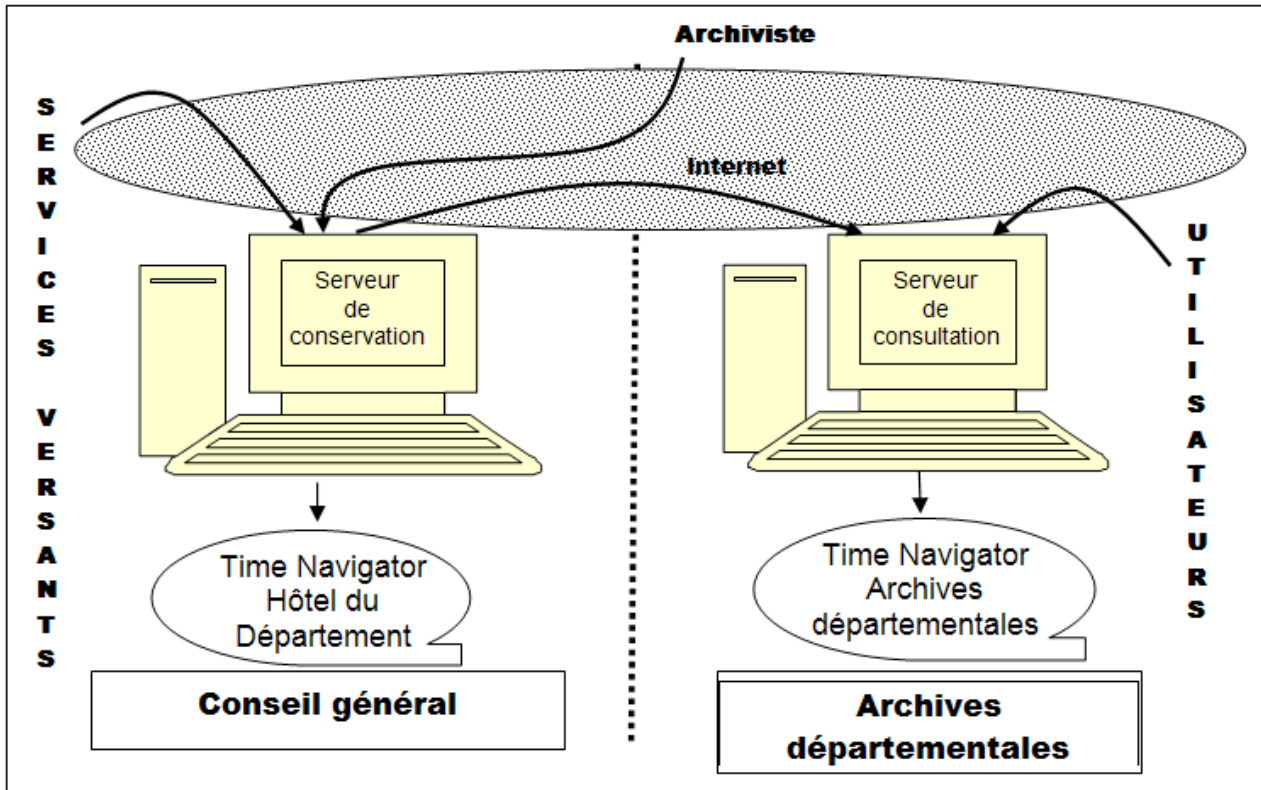


Figure 2 : architecture de stockage de la plate-forme d'archivage électronique du Conseil général des Yvelines<sup>17</sup>

- Les données sont transférées, soit par les services versants via Internet en protocole sécurisé HTTPS avec vérification du certificat « client », soit par les archivistes après chargement des fichiers contenus sur support amovible.
- Chaque soir, le système de fichier du serveur d'écriture et de conservation est synchronisé avec celui du serveur de consultation, les bases MySQL du serveur de conservation sont recopiées en totalité sur le serveur de consultation. Le transfert se fait via Internet via le protocole SSH. L'exécution du script de transfert est supervisée par envoi d'un courriel après chaque exécution.
- Chaque soir, les données contenues dans chacun des serveurs sont sauvegardées (sauvegarde incrémentale). Il n'est pas fait usage d'un catalogue dédié pour l'archivage électronique mais le jeu de cartouche est dédié. Une sauvegarde totale est programmée tous les mois.
- Les archivistes accèdent au serveur de conservation pour le contrôle des transferts (contrôle, indexation, communicabilité) via Internet en protocole sécurisé HTTPS, sans vérification du certificat « client ».
- La consultation des archives s'effectue à partir du serveur de consultation, via Internet, en protocole HTTP.

<sup>16</sup> Deux serveurs Poweredge 1850 équipés de disque RAID 1 de 73 Go ont été mis en place. Chaque serveur est composé d'un système d'exploitation Linux, d'un serveur APACHE (langage PHP), d'un système de base de données MySQL et des binaires Time Navigator.

<sup>17</sup> Comme indiqué ci-dessus, la localisation des serveurs sera inversée à court terme

## **Administration de la plateforme**

Cette fonction permet de suivre les services relatifs à l'exploitation d'ensemble de la plateforme d'archivage électronique. Deux profils d'administrateurs ont été définis : l'administrateur fonctionnel (gestion des droits d'accès des utilisateurs, consultation des journaux, intégration ou modification d'informations dans la base de connaissances, accès aux tables, côté direction des Archives) et l'administrateur technique (architecture de la plateforme et modification de la structure des tables, côté direction des systèmes d'information).

### **Les fonctionnalités programmées à terme**

Les premiers développements permettent de répondre aux besoins immédiats liés à la réception de paquets d'informations. Pour autant, à court ou moyen terme, l'intégration de nouvelles fonctionnalités permettra de mieux répondre à l'ensemble des besoins identifiés.

### **Transfert**

A moyen terme, les services versants pourront effectuer une demande préalable de transfert en ne saisissant que les métadonnées des documents numériques dont l'entrée est programmée. Une fois validées, les métadonnées enrichies des pièces jointes constitueront le paquet d'informations à verser.

Par ailleurs, le traitement des documents ou données électroniques produits ces dernières années et les moyens humains et financiers des institutions concernées limitent parfois la possibilité de générer en amont des enveloppes XML de transfert conformes au standard d'échange. C'est pourquoi, une interface de saisie permettant aux services versants de constituer ces paquets d'informations sera mise à leur disposition. Ceux-ci seront, en outre, formés à son utilisation.

### **Gestion mutualisée des données**

Comme indiqué ci-dessus, il est envisagé de mutualiser à terme la gestion des données descriptives des versements papier et électroniques. Les recherches et demandes de communication seraient, de ce fait, communes à tous les supports.

Dans cette même perspective, la mutualisation des espaces de stockage dédiés à la numérisation et à l'archivage électronique est à poser.

## **Administration de la plateforme**

Toutes les transactions décrites dans le standard d'échange de données pour l'archivage n'ont pas été prises en compte dans les développements. Outre la demande de transfert décrite plus haut, les évolutions du système devront concerner l'éventuelle restitution aux collectivités déposantes, la demande d'élimination et la demande de modification.

Par ailleurs, la prise en compte des droits d'accès spécifiques de certains utilisateurs (notamment, services producteurs, bénéficiaires de dérogations) devra être assurée.

### **Planification de la pérennisation**

Pour répondre au double besoin de contrôle des formats en arrivée et de conversion des formats non pérennes en un format cible d'archivage, des outils spécifiques devront être intégrés dans le système. Les travaux menés actuellement par la DAF sur ce sujet seront transcrits dans le fonctionnement de la plateforme d'archivage.

De même, des fonctionnalités liées à la migration de support devront être étudiées pour parer à l'obsolescence prévisible du système.

## ■ En conclusion, les enseignements au terme de la phase expérimentale

Les conditions de la réussite se dégagent au terme de cette phase expérimentale.

Tout d'abord, **une implication active des responsables** que le caractère technique du chantier ne doit pas déconnecter de ses enjeux stratégiques. Seules des questions claires de leur part appelleront les réponses explicites susceptibles d'assurer un balisage vigilant de la progression, évitant ainsi que par un processus d'enchaînement quasi automatique, les étapes clefs appelant de leur part un choix politique, juridique, archivistique ou organisationnel, ne soient banalisées. A cet égard, on se référera utilement au propos introductif consacré par Milad Doueïhi<sup>18</sup> à son essai paru cette année sous le titre *La grande conversion numérique*.

Le **tandem archiviste-informaticien** constitue un atout déterminant pour la réussite du projet, selon une dialectique reliant transcription des principes fondateurs de l'archivistique, normalisation descriptive et potentialités des systèmes électroniques, le tout en recourant à des formats ouverts, susceptibles de garantir une interopérabilité optimale des données.

Une **pédagogie attentive** doit être développée auprès des futurs utilisateurs, tout au long du chantier, sauf à déconnecter une partie de l'équipe associée et à marginaliser la plateforme d'archivage électronique qui fera pourtant inéluctablement part de leur futur environnement professionnel. Au service de cette pédagogie, les qualités ergonomiques de l'outil et sa simplicité d'utilisation constituent un levier majeur qu'il convient de ne pas sous-estimer.

Moyennant ces précautions, le constat rassurant de la **permanence opérationnelle du référentiel archivistique** peut être formulé. Sous réserve d'une présentation lisible du cahier des charges transmis aux informaticiens, la mise en forme des étapes de traitement des archives électroniques et d'accès à leur teneur apparaît parfaitement accessible à des utilisateurs même non experts, via les écrans pré-formatés. Le cœur de métier de l'archiviste peut ainsi rester voué au recueil et au contrôle de la validité des métadonnées contextuelles de production et de versement, et à la description normalisée des articles pris en charge, conformément aux tableaux de gestion en vigueur. On pourrait même prévoir que la mission de contrôle scientifique et technique ne s'en trouve que renforcée : l'histoire de la phase opérationnelle nous le dira.

---

<sup>18</sup> Milad DOUEIHI, *La grande conversion numérique*, Paris, La librairie du XXI<sup>ème</sup> siècle - Seuil, 2008, 278 p.



# Intervenants

---

---

# ***Speakers***

---

---

## ■ Kuldar Aas

Kuldar Aas holds a masters degree in Information Technologies from the University of Tartu. 2002 he joined the National Archives of Estonia and has since then been actively involved in issues regarding description, transfer and ingest of electronic records and datasets. He has been involved in several national and international research projects including *Protage*, a small targeted research project (STREP) funded by the European Commission in FP7, as well as projects developing national metadata standards.

[kuldar.aas@ra.ee](mailto:kuldar.aas@ra.ee)



## ■ Bart Ballaux

Since December 2006, Bart Ballaux is Senior Advisor Digital Longevity at the Nationaal Archief of the Netherlands. Before joining the Nationaal Archief, Bart worked at the Royal Library of Belgium and the University of Antwerp. He studied History and Archival Science in Belgium and Canada. Since October 2007, he is part-time lecturer on Records Management in the Master programme in Archival Science at the Free University of Brussels.

Bart Ballaux

Senior Advisor Digital Longevity

Nationaal Archief - National Archives of the Netherlands

Postal box 90520

2509 LM The Hague

The Netherlands

tel.: +31 (0)70-331 5487

[bart.ballaux@nationaalarchief.nl](mailto:bart.ballaux@nationaalarchief.nl)

[www.nationaalarchief.nl](http://www.nationaalarchief.nl)

[www.planets-project.eu](http://www.planets-project.eu)

[www.digitaleduurzaamheid.nl](http://www.digitaleduurzaamheid.nl)

## ■ **Martin Berendse**

Martin Berendse (1963) studied Law at the University of Utrecht and joined the office of the University Board between 1984 and 1986. He was consultant for the management- and PR-divisions of the advertising company FHV/BBDO (Amsterdam). He moved in 1990 to the theatre business and managed a theatre festival, a national agency for theatre and dance and the Rotterdam Repertory Theatre Company. He started at the Dutch Ministry of Education, Culture and Science as head of general cultural policy division in 1998 en worked as (dep.) director of the Arts Department/dep. director-general Culture and Media. Since 2008 he is managing director of The National Archives of The Netherlands (Nationaal Archief).

## ■ **María Teresa Bermúdez-Muñoz**

Mastère en Études et gestion de l'information (Université du Costa Rica (UCR), 2005) et diplômée de cette même université en archives administratives (1988), María Teresa Bermúdez-Muñoz y enseigne l'archivistique, depuis 2004, et mène des recherches sur divers sujets, en particulier la gestion du document numérique. Elle a prêté ses services à la Direction des archives nationales du Costa Rica, pendant plusieurs années, où elle a été la responsable de la coordination du Système national des archives entre 1997-2004.

Engagée et passionnée par les nouvelles technologies de l'information et leur impact sur l'archivistique et la gestion publique, María Teresa Bermúdez-Muñoz, a suivi plusieurs stages de formation tant dans son pays (*Norme 15489: Gestion des archives ; Les nouvelles technologies et les archives*) qu'à l'étranger (*Théories et pratiques archivistiques : le document numérique, Direction des archives de France, à Paris ; Deuxième Semaine du document numérique, à Fribourg ; La gestion des archives, La gestion des archives municipales, et La gestion des systèmes nationaux d'archives, à Madrid*). Elle s'est investie dans la coordination, entre 2000-2004, de la Commission de normalisation archivistique, ainsi que du Congrès archivistique national (1998-2005), et a occupé le secrétariat de la Commission nationale de triage entre 1990-2005. Elle a également contribué à la réflexion sur l'archivistique du XXI<sup>e</sup> siècle avec la publication de plusieurs articles et sa participation dans des congrès et conférences internationaux.

María Teresa Bermúdez-Muñoz agit comme consultant auprès d'organismes publics et privés.

## ■ Martine de Boisdeffre

Martine de Boisdeffre, ancienne élève de l'École normale supérieure (Sèvres), est titulaire d'une maîtrise d'histoire et diplômée de l'Institut d'études politiques de Paris.

Elle a été ensuite élève à l'École nationale d'administration (promotion Solidarité 1981-1983), Auditeur (1983) puis Maître des requêtes (1986) au Conseil d'État, Rapporteur près la Commission spéciale de cassation des pensions (1983), Secrétaire général du Comité national d'éthique (1985-1992), Rapporteur général des travaux du Conseil d'État sur les sciences de la vie (1988), Chargée de mission à la Mission interministérielle pour l'Europe centrale et orientale (1990), Conseiller technique au cabinet d'Élisabeth Guigou, ministre délégué aux affaires européennes (1990-1993), Maître de conférences à l'Institut d'études politiques de Paris (1984-1988), Secrétaire général adjoint (1994-1995) puis Secrétaire général (1995-2001) du Conseil d'État, Conseiller d'État (depuis 1998), Présidente du conseil d'administration de l'Institution nationale des Invalides (depuis 2000).

Depuis 2001, elle est directrice des Archives de France. La direction des Archives de France est chargée de proposer au ministre de la Culture et de la Communication les choix stratégiques à opérer en matière d'archives et de les mettre en œuvre. Elle exerce sur l'ensemble du réseau des archives un rôle de conseil, d'incitation, de réglementation, d'évaluation et de contrôle en ce qui concerne la collecte, le tri, le classement, la description, la conservation et la communication tant aux chercheurs qu'à l'ensemble des citoyens, des archives publiques autres que celles des ministères des Affaires étrangères et de la Défense.

## ■ Marie-Anne Chabin

Marie-Anne Chabin a fondé et dirige le cabinet de conseil et d'expertise en archivage *Archive 17* ([www.archive17.fr](http://www.archive17.fr)). Elle a commencé sa carrière dans le secteur public (Direction des Archives de France) puis décidé d'élargir son expérience professionnelle dans une société privée comme consultant en GED, avant de rejoindre l'Institut national de l'audiovisuel (INA). En 2000, elle décide de créer sa propre structure, *Archive 17*, comme expert en records management et archivage électronique, pour assister les entreprises dans la mise en oeuvre de politiques et procédures d'archivage, de référentiels de conservations, de systèmes d'archivage et de formations. En 2007, elle développe le concept de « catégories de conservation » et crée une nouvelle méthode, ARCATÉG™, pour simplifier et piloter l'archivage.

Marie-Anne vient d'être recrutée comme professeur associé au CNAM (Conservatoire national des Arts et Métiers) pour enseigner le records management et l'archivage.

Elle a publié de nombreux articles, ainsi que plusieurs ouvrages dont le dernier est : *Archiver, et après?* (Djakarta, 2007). Son expérience et ses travaux l'ont impliquée dans la traduction en français d'ISO 15489, MoReq, OAIS et MoReq2.

## ■ Adam B. Cowling

Adam Cowling is currently pursuing his Master of Archival Studies and Master of Library and Information Studies at the University of British Columbia (UBC), at Vancouver. As a Graduate Research Assistant, he has been involved in developing an electronic record preservation strategy for the Morris and Helen Belkin Art Gallery as part of Phase 3 of the InterPARES project (International Research on Permanent Authentic Records in Electronic Systems). Adam has also served as Digitization Project Manager of a rare collection of bookplates at UBC's Rare Books and Special Collections. Between terms, he has been primarily based out of Geneva, first honing his arrangement and description skills in the archives of the International Federation of Red Cross and Red Crescent Societies (IFRC), and then implementing an opensource library information system at a small trade-related organization. His interests lie in the management of records of international organizations, and he hopes to pursue further endeavours in this area after completing his studies in the new year.

## ■ Marius Delail

1981 : Ingénieur au Centre National d'Etudes Spatiales, spécialiste robotique spatiale jusqu'en 1998, en particulier : simulation de missions pour le bras robotique de la navette Hermès puis mise en place d'un site d'essais pour la robotique mobile lunaire et martienne. A partir de 1999, Auditeur interne pour la Direction Générale du CNES et chef d'entreprise (Robots mobiles applications) jusqu'en 2004. Depuis : chef du service Gestion de l'Information et de la connaissance au sein de la DSI. Ce service regroupe les activités : Gestion des Connaissances, Veille et Intelligence, Traduction/Interprétariat/Terminologie, Archives, Documentation, Mémoire d'Entreprise.

Dans ce cadre je gère deux projets dans le domaine couvert par DML :

mise en place des chaînes de traitement relatives aux archives "papier" et aux archives "électroniques"

mise en place d'une base de connaissances dédiée au domaine spatial permettant, en particulier, de valoriser les travaux liés à l'archivage

Ces deux projets fusionnent et se concrétisent à travers la mise en place de la Mémoire d'Entreprise du CNES.

Service DSI/SD/GI,  
CNES 18 avenue Edouard Belin  
31401 Toulouse.

[mario.delail@cnes.fr](mailto:mario.delail@cnes.fr)

## ■ Charlotte Fabre

Charlotte Fabre est consultante EIM (*Enterprise Information Management*) chez Atos Origin. Elle s'occupe plus particulièrement des problématiques fonctionnelles de gestion de l'information, de la dématérialisation à la conservation long terme. Sur le projet SPAR de la Bibliothèque nationale de France, elle intervient en tant que responsable fonctionnelle.

Dans ce cadre, elle participe aussi aux réunions du groupe PIN, Pérennisation de l'information numérique.

Auparavant, elle a fait partie de la société Jouve pendant 4 ans où elle a travaillé sur la structuration et la dématérialisation des données patrimoniales en tant que consultant Avant-vente et lors de l'initialisation et la gestion de projets.

Issue d'une formation en sciences humaines (ethnologie mais aussi histoire de l'art), elle a poursuivi son cursus à l'école du Louvre ainsi qu'à l'ENSSIB (Ecole Nationale Supérieure des Sciences de l'Information et des Bibliothèques).



## ■ Elisabeth Gautier-Desvaux

Conservateur général du patrimoine, Elisabeth Gautier-Desvaux dirige actuellement les Archives départementales des Yvelines, après avoir successivement occupé les fonctions de directeur des Archives départementales de l'Orne et du Calvados, d'inspecteur général des Archives de France, et de directeur régional des Affaires culturelles de Basse-Normandie, puis de Bretagne. Ancienne élève de l'Ecole des Chartes où elle avait soutenu une thèse d'archéologie, elle a adapté ses centres d'intérêt, ses interventions et ses publications, aux caractéristiques des services dont elle a assumé la responsabilité, notamment à l'action éducative et culturelle des Archives (après avoir lancé le premier « archivobus » en France) et aux archives sonores et audio-visuelles (auteur d'un manuel sur ce sujet, en collaboration avec Chantal de Tourtier-Bonazzi). Dans un département pionnier en matière de technologies de l'information et de la communication, elle a perçu l'ampleur des perspectives offertes par l'exploitation des systèmes numériques, tant pour sauvegarder un patrimoine prestigieux et en élargir la diffusion, que pour s'adapter aux évolutions les plus contemporaines de la dématérialisation administrative. Elle s'est donc investie, avec ses collaborateurs, dans une quadruple direction : poursuivre un ambitieux programme de numérisation, souder un partenariat de recherche appliquée avec l'INRIA et l'INSA de Rennes sur la reconnaissance automatique de formes et d'écritures, normaliser l'ensemble des instruments de recherche pour les mettre en ligne, concevoir avec la direction départementale des systèmes d'information une plateforme territoriale d'archivage électronique, tout en participant aux travaux du comité de pilotage du pilote d'archivage électronique de la Direction des Archives de France (Pil@e).

Elisabeth Gautier-Desvaux assume la co-direction du Master professionnel « Métiers de la culture-archives » de l'Université de Versailles-Saint-Quentin-en-Yvelines dont elle coordonne le module « Enjeux et nouveaux outils de l'information et de la communication ».

## ■ **Stephen Harries**

Stephen Harries is currently an independent consultant in information and records management, with a particular interest in improving the integration of electronic records with knowledge systems and organisational change. Previously, he led the UK initiative to implement electronic records management across UK government departments and agencies, working from the National Archives (earlier the Public Record Office); this included the development of policies, standards and specification for EDRM in the context of e-government. He has also worked on initiatives to improve the delivery of public service and policy programmes across government by developing capabilities and practices in the public sector. He has a long background in information science and knowledge systems, encompassing records, libraries and business management systems, including a ten-year period of teaching and research in a university. He initially studied history, then information science, and also holds Masters degrees in Information Systems Management, and Public Policy Management.

## ■ Stephen Howard

I graduated in 1991 from the University of Manchester with a degree in Economics & Social Science. I worked in various Records Management posts for Instant Library & TFPL before joining Westminster City Council, London in 1995. As Corporate Records Manager I successfully prepared the council for Data Protection & Freedom of Information compliance, and implemented a hybrid EDRMS.

I worked as Head of Corporate Information & Records at Suffolk County Council between July 2003 and July 2007. I led the Corporate Information & Records Team - responsible for the strategic management and development of information resources. The team had oversight of the Council Web site, electronic & manual records management, information security, data protection and freedom of information. The team communicated advice throughout the organisation via a network of 'Information Agents', and co-ordinated their knowledge management initiatives.

Between July 2007 and March 2008 I held the P4 post of Chief, Records & Archives Unit for the United Nations Mission in Kosovo (UNMIK). My primary responsibility was to ensure the satisfactory capture of records for HQ in New York prior to the transfer of authority to the Kosovo government. I led a team of 50 local records staff who archived key records series to minimise risk to the UN and ensure an adequate historical record of UNMIK's extensive functions of interim administration.

I have held senior positions within the UK Records Management Society Executive and Local Government Group and in 2005 gained an MSc Econ in Records Management via the distance-learning course at Aberystwyth, Wales.

### Previous speaking engagements include:

ARK conference "EDRM Strategies in the Public Sector" June 2007, London - "Information Governance"

RMS Local Government Group meeting June 2005, London - "Email Management"

Cimtech conference "Electronic Document, Content and Records Management in Public Authorities" May 2005, Hatfield - "Reviewing the options for moving from multiple EDM systems to an integrated EDRM solution"

Chair of Capita conference "Electronic Records Management" February 2003, London.

RMS Annual Conference April 2002, Cardiff - "Hybrid Records Management"

### Previous professional activities include:

Vice-Chair of GB Records Management Society (RMS) August 2002-April 2004

GB RMS Local Government Group Chair Feb 2001-Feb 2003

GB RMS member since May 1997

Member of the joint GB RMS/SoA Working Party responsible for "A draft Code of Practice for Records Managers and Archivists on the implementation of the Data Protection Act"

Member of ARMA International July 2000 - July 2005

Co-creator and administrator of the UK and international Records Management professional listserv Jan 2003 - Jun 2007 (<http://www.jiscmail.ac.uk/lists/Records-Management-UK.html>)

## ■ Thomas Ledoux

Thomas Ledoux est le chef du projet de réalisation du système SPAR. Depuis 11 ans au Département des Systèmes d'Information de la Bibliothèque nationale de France, il a participé à toutes les phases de l'élaboration de la bibliothèque numérique du Poste de Lecture Simple à la mise en ligne sous Gallica. Diplômé de l'école nationale supérieure de Techniques Avancées, il est également titulaire d'un "*master's degree*" en informatique de l'Université du Michigan.

## ■ Yves Marleau

Yves Marleau est le président de Cogniva Inc. Diplômé en Communications organisationnelles de l'Université d'Ottawa et en Informatique de l'Université du Québec au Canada, M. Marleau a débuté sa carrière en 1985 comme gestionnaire au sein de la fonction publique canadienne. Il a été responsable de plusieurs projets d'envergure reliés à la gestion de l'information (GI) dont, entre autres, le déploiement du premier système de gestion électronique de documents et de dossiers au Gouvernement fédéral. En 2000, il s'est joint au secteur privé en tant que Vice-président, Gestion du savoir chez FSG Consultants. M. Marleau a été impliqué dans plus d'une soixante de projets de gestion de l'information dans une cinquantaine d'organismes gouvernementaux et paragouvernementaux.

C'est en 2005 que M. Marleau a créé Cogniva afin de mettre sur pied un programme de recherche et développement visant à apporter des solutions aux problèmes de GI et de « Records Management » qui sont communs à tous les projets de gestion électronique de l'information. Après plusieurs années de réflexion et d'expérience auprès de grandes organisations le concept ISIS (Information Semantic Infrastructure Services) a vu le jour. Ce concept propose un nouveau modèle intégré basé sur le principe de classification par facettes permettant de créer une infrastructure sémantique unique qui regroupe et intègre la fonctionnalité de plusieurs outils de GI tels que les plans de classification, les thésaurus, les normes de métadonnées ainsi que les taxinomies. Avec ce nouveau modèle il est possible de mieux décrire à la fois le contenu l'objet informationnel et le contexte dans lequel il a été créé, d'améliorer la précision de la recherche, de gérer l'information à travers son cycle de vie et aussi d'améliorer l'automatisation du processus de classification et d'organisation de l'information, délestant ainsi l'utilisateur d'une tâche qu'il accomplit rarement très bien.

M. Marleau collabore régulièrement avec le milieu universitaire, notamment avec l'Université de Montréal, l'Université d'Ottawa et l'Université de technologie de Troyes en France. Il a également contribué à l'élaboration de normes canadiennes portant sur la GI telles que la Norme de métadonnées de la gestion des documents du gouvernement du Canada et le Système de classification selon les activités opérationnelles structurées (SCAOS).

## ■ Sabine Mas

Sabine Mas est chercheur postdoctoral à l'Université de Technologie de Troyes en France. Diplômée en histoire et en archivistique, elle a travaillé en 1996 au Centre rhénan d'archives et de recherches économiques à Mulhouse. À partir de 1999, elle participe à un projet de recherche sur la notion de dossier subventionné par le Conseil de recherches en sciences humaines du Canada. Cette expérience développe son intérêt pour les problématiques touchant à la notion d'archives dans un contexte électronique et plus particulièrement à leur organisation.

Dès l'automne 2000, elle entreprend une thèse de doctorat à l'École de bibliothéconomie et des sciences de l'information (EBSI) de l'Université de Montréal qui porte sur les schémas de classification hiérarchiques et le repérage des documents administratifs électroniques dans un contexte de gestion décentralisée. Conjointement à sa recherche doctorale, elle participe en 2001 à un travail d'analyse des facettes pour la classification des documents institutionnels au sein du Gouvernement du Québec. Elle demeure également impliquée dans différents projets de recherche relatifs à la typologie des documents et dossiers administratifs, à la classification et aux schémas de classification hiérarchiques ou à facettes de ressources documentaires dans le domaine des archives, des bibliothèques numériques ou des musées canadiens. Toutes ces recherches ont régulièrement donné lieu à plusieurs communications et publications scientifiques et professionnelles dont une qui a reçu le prix du meilleur article de la revue *Archives (Revue de l'Association des archivistes du Québec)* en 2005.

Parallèlement à ses activités de recherche, elle enseigne depuis 2000 comme chargée de cours au certificat en archivistique à l'EBSI. Elle a été membre de la Commission des archives et du Groupe interdisciplinaire de recherche en archivistique à l'Université de Montréal. Elle travaille actuellement au sein du laboratoire Technologie de la coopération pour l'innovation et le changement organisationnel (Tech-CICO) de l'Université de Technologie de Troyes, dans le cadre d'une recherche financée par le Fonds québécois de la recherche sur la société et la culture. Le projet de recherche vise à explorer la faisabilité d'une classification à facettes pour l'organisation et le repérage de documents électroniques au sein des entreprises qui tiendrait compte à la fois des besoins informationnels et documentaires spécifiques des employés dans le cadre de leurs activités quotidiennes et des exigences archivistiques (par exemple, gestion du cycle de vie du document, mise en valeur du patrimoine documentaire, respect du fonds d'archives). Dès septembre 2008, elle retourne au Canada pour occuper la fonction de professeure en archivistique à l'Université de Montréal.

## ■ **Gaelle Mignot**

Après des études d'histoire à l'Université François Rabelais de Tours, Gaelle Mignot obtient le DESS Information Communication Archivistique dispensé par l'Université de Lyon III en 1999. Elle consacre ses deux premières années d'activité professionnelle à organiser la gestion d'archives intermédiaires au sein d'une direction opérationnelle du secteur public puis d'un cabinet d'avocats. C'est en 2001 qu'elle rejoint la Direction des Archives départementales des Yvelines. Tout d'abord responsable des archives contemporaines, elle conjugue le suivi et le traitement des archives papier avec celui des archives électroniques. Dans ce cadre, elle est amenée à participer à la rédaction des spécifications fonctionnelles de la plateforme d'archivage électronique du Département des Yvelines puis est associée, avec les informaticiens de la direction des systèmes d'information, au projet Pil@e (pilote d'archivage électronique) de la Direction des Archives de France. En 2008, elle devient chargée de mission informatique et coordonne à ce titre l'ensemble des projets informatiques de la direction des Archives départementales (plateforme d'archivage électronique, numérisation et travaux de recherche associés, acquisition d'un progiciel de gestion documentaire, conversion rétrospective en XML/EAD des instruments de recherche). Elle intervient régulièrement sur l'archivage électronique en direction de professionnels des archives et d'étudiants en archivistique.

## ■ **Markku Nenonen**

PhD Markku Nenonen is a Deputy Director General in the National Archives of Finland since the beginning of 2007. At the moment Dr Nenonen is responsible for the information services department of the National Archives. Dr Nenonen is also in charge of developing the eServices in the National Archives Services of Finland. At the moment the main project for developing eServices is the concept for eRecords management system. The system covers the whole life cycle of a record from birth to long term preservation. The system has also to be seen from the viewpoint of architecture of public administration and a tool for the better data management in producing public services. Deputy Director Markku Nenonen came in the service of the National Archives from the Ministry of interior. In the ministry his focus was the co-operation between state and municipalities especially on eServices - branch. Mr Nenonen had a role in the first Information Strategy Programme launched by Prime Minister Matti Vanhanen's first government in 2003 - 2007. Mr Nenonen came to the Ministry of interior in March 2002. Before that he had worked almost 15 years in the municipal sector as a secretary of data management.



## ■ Brigitte Pipon

Brigitte Pipon est depuis 2005 directrice des Archives départementales des Deux-Sèvres à Niort, après avoir été de 1992 à 2005 directrice-adjointe des Archives départementales de Maine-et-Loire à Angers. Archiviste-paléographe (École nationale des chartes, promotion 1991), conservateur en chef du patrimoine (Institut national du patrimoine, promotion 1991), elle a publié sa thèse «Le chartrier de l'Abbaye-aux-Bois (1202-1341)», dans *Mémoires et documents de l'École des chartes* (n° 46) en 1996. Elle anime régulièrement des formations et participe à l'élaboration de nombreuses expositions d'ampleur locale et nationale (par exemple: "Des bords de Loire aux rives du Saint-Laurent. Les Angevins et le Canada du XVI<sup>e</sup> au XX<sup>e</sup> siècle" en 2000; «Anjou-Pologne: une histoire contemporaine» en 2004, « Archiv'et vous. De la fabrication à la restauration du papier » en 2007). Elle s'investit depuis 1999 dans la publication de manuels d'archivistique (collection : *Les archives : c'est simple !*); elle a coordonné la parution de *Abrégé d'archivistique* (1<sup>ère</sup> et 2<sup>e</sup> éditions, 2004-2007, 315 p.) et de *Archives privées : un patrimoine méconnu. Petit guide à l'usage des propriétaires* (2005, 48 p.), publiés par l'Association des archivistes français, dont elle est vice-présidente depuis 2007. Elle participe au niveau local aux associations de promotion du patrimoine et à l'animation de sociétés savantes.

Brigitte Pipon,

Directrice des Archives départementales des Deux-Sèvres

26, rue de la Blauderie

BP 505

79022 Niort Cedex

Tél. : 05 49 08 94 90 - Fax : 05 49 06 63 80

[b.pipon@cg79.fr](mailto:b.pipon@cg79.fr)

<http://archives.deux-sevres.com>

## ■ Raivo Ruusalepp

Raivo Ruusalepp of the Estonian Business Archives Consultancy has collaborated with the Digital Curation Centre on audit and certification services in the EU-funded project DigitalPreservationEurope and is technical coordinator of the FP-7 EU-funded digital preservation project Protage. Previously he was involved in producing the JISC funded reports on the Feasibility and Requirements Study on Preservation of EPrints. (2003) and the Assessment of UKDA and TNA Compliance with OAIS and METS Standards. (2005) ), and most recently has been involved in developing the Data Audit Framework (2008) for UK HE/FE institutions.

[raivo@eba.ee](mailto:raivo@eba.ee)

## ■ Dr Elizabeth Shepherd

Dr Elizabeth Shepherd is a Reader in archives and records management at University College London, School of Library, Archive and Information Studies (UCL, SLAIS). Between 1992 and 2002, she was programme director of the MA in Archives and Records Management at UCL and she now teaches on the Masters programmes. She has established a research centre, ICARUS (International Centre for Archives and Records Management Research and User Studies) within SLAIS (<http://www.ucl.ac.uk/slais/research/icarus/>). Elizabeth's research interests include the relationships between records management and information policy compliance and the development of the archive profession in England, which is the subject of her PhD and of a forthcoming book (2009). She serves on the editorial boards of Archival Science and the Records Management Journal, is a member of the Arts and Humanities Research Council's Peer Review College and Higher Education Funding Council for England's Research Assessment Exercise 2008, Panel 37. She has published numerous articles and (with Geoffrey Yeo) the internationally best selling book *Managing Records: a handbook of principles and practice* (Facet Publishing, 2003).

### Contact details

School of Library, Archive and Information Studies

University College London

Gower Street

London WC1E 6BT

[e.shepherd@ucl.ac.uk](mailto:e.shepherd@ucl.ac.uk)

## ■ Laurent Sollier

Laurent SOLLIER est Architecte au sein du centre de compétences Open Source d'Atos Origin. Intégrant depuis huit ans des applications n-tiers au sein des systèmes d'information de grandes entreprises, il a développé son expertise dans les portails d'entreprises et les logiciels libres. Laurent Sollier a conçu de nombreux systèmes pour le compte de multinationales et collectivités territoriales, et participé à leurs mises en œuvre.

## ■ Jacqueline Slats

After her study in Information Communication Technology, Jacqueline worked for 7 years at the computer centre of the Ministry of Transport, Public Works and Water management. In 1994, she joined the Dutch State Archive Service, where she was responsible for different Information Technology projects. Last few years she was the program manager of the Digital Preservation Testbed and the Taskforce Digital Longevity, which were sponsored by the Dutch State Archive Service and the Ministry of the Interior and Kingdom Relations. Now she is Head of Digital Longevity Department of the Nationaal Archief of the Netherlands.

Jacqueline Slats

Head of Digital Longevity Department

Nationaal Archief - National Archives of the Netherlands

Postal box 90520

2509 LM The Hague

The Netherlands

tel.: +31 (0)70-331 5438

[jacqueline.slats@nationaalarchief.nl](mailto:jacqueline.slats@nationaalarchief.nl)

[www.nationaalarchief.nl](http://www.nationaalarchief.nl)

[www.digitaleduurzaamheid.nl](http://www.digitaleduurzaamheid.nl)

## ■ **Jan Dalsten Sørensen**

Jan Dalsten Sørensen holds an MA in History from the University of Aarhus and is a special consultant at the Danish National Archives, Department of Appraisal and Transfer, where he has been employed since 1999. He is the head of the newly established Section for Digital Archiving. He has participated in the theoretical and practical handling of erecords, especially appraisal and transfer of e-records and the development of regulations and requirements. His current projects include a revision of the standards and specifications for the transfer of e-records. He represents the Danish National Archives in the DLM Forum and has written several articles and presentations on the appraisal and transfer of e-records.

## ■ Sébastien Soyez

### *Expérience professionnelle*

*Depuis août 2006* Attaché scientifique sous mandat : Préparation de directives sur l'archivage électronique (e-mails et fichiers bureautiques) pour les administrations publiques fédérales ; Surveillance des archives (*Archives de l'Etat - Bruxelles*)

*D'août 2000 à juillet 2006* Administrateur de programme de recherche : « Programme Société de l'Information » et « Plan de digitalisation du patrimoine culturel et scientifique des Etablissements scientifiques fédéraux et de la Cinémathèque Royale de Belgique » (*Politique scientifique fédérale - Bruxelles*)

*De mars 1999 à août 2000* Attaché scientifique contractuel : Travail historique et archivistique (*records management*) basée sur l'exploitation du fonds d'archives des séquestres de la Première guerre mondiale (*Archives de l'Etat - Bruxelles*)

### *Formation supérieure*

*De 1997 à 2000* Diplôme d'Etudes Spécialisées (Master spécialisé - 3ème cycle) en Gestion de l'Information et de la Documentation (INFODOC) – orientation Documentation d'entreprises (*Université Libre de Bruxelles*)

*De 1992 à 1998* Licence en Histoire (orientation époque contemporaine) et Agrégation en Histoire (*Université Libre de Bruxelles*)

### *Activités scientifiques*

*Université Libre de Bruxelles* Exposé scientifique sur le *records management* appliqué au fonds d'archives des séquestres des Archives de l'Etat dans le cadre du cours d'Archivistique du Professeur Patrick LEFEVRE le 29 mars 2000

*Archives de l'Etat (Bruxelles)* Divers exposés scientifiques entre 1999 et 2000 concernant les archives d'entreprises du Fonds des Séquestres conservées aux Archives de l'Etat Co-formateur du cours sur l'archivage électronique *Mundaneum (Mons)* Membre du Comité scientifique - musée/centre d'archives (depuis octobre 2007)

### *Participations à des colloques / séminaires / groupes de travail*

*DLM-Forum Conference : Barcelone du 6 au 8 mai 2002 et Budapest du 5 au 7 octobre 2005 DLM-Forum Meeting : Nationaal Archief, La Haye, du 18 au 20 octobre 2004*

Workshop « Innovations in Digital Asset Management » (COLLATE) , Fraunhofer IPSI, Darmstadt (Allemagne), du 6 au 8 octobre 2003 (« Advanced concepts, approaches and tools in archiving systems »)

Meeting concerning the Replication of Electronic Documents and Metadata of the International Tracing Service, La Haye, du 7 au 8 mars 2007 (Groupe de travail sur la gestion des archives du Service International de Recherche (SIR) de Bad Arolsen - Allemagne) 8ème Congrès de l'Association des Cercles francophones d'Histoire et d'Archéologie de Belgique, Namur, du 28 au 31 août 2008 (Communication sur « La problématique de l'archivage des documents électroniques : enjeux, perspectives et esquisses de solutions d'une sauvegarde pérenne de l'information »)

## ■ Dr Alice Stevenson

BA, MA, MPhil, PhD

### Contact details:

School of Library, Archive and Information Studies  
University College London  
Gower Street  
London WC1E 6BT  
[alice.stevenson@ucl.ac.uk](mailto:alice.stevenson@ucl.ac.uk)

### ***Teaching and research :***

2008-: Research Associate, The impact of the Freedom of Information Act (2000) on records management in the public sector, School of Library, Archive and Information Studies, UCL

2007-2008 Research Fellow, Collections for People: Access to the stored collections of English Museums as a public resource, Institute of Archaeology, UCL.

2007-08: Tutor in archaeology, University of Cambridge, Department of Continuing Education

2004-2007: Supervisor in archaeology and Egyptology

Includes supervision of masters and undergraduate students.

### ***Other relevant employment :***

2007-2008 Librarian and Archivist, The Egypt Exploration Society

### Education :

2007: PhD in Oriental Studies (Prehistoric Egyptian archaeology), University of Cambridge

2005: MPhil in Oriental Studies (Prehistoric Egyptian archaeology), University of Cambridge

1983/85: MA in Museum Studies, University of Leicester

1982: BA (Hons) Archaeology and Anthropology, Class 1, University of Cambridge



## ■ **Maria Taverniti**

Université de Calabre - Département de Linguistique

Via Pietro Bucci 17 B

87036 Rende (CS)

[maria.taverniti@unical.it](mailto:maria.taverniti@unical.it)

2004: Licence de médiation linguistique, Université de Calabre, mémoire en documentation;

2006: Master d'informatique appliquée aux sciences humaines, mémoire en documentation et informatique,

2006: Bourse d'études régionale en science archivistique;

2006 - 2009: Doctorat d'histoire économique: Démographie, institutions et société dans les pays méditerranéens, Université de Calabre;

2006 - 2007: Enseignement universitaire en documentation et informatique.

2008 : Recherche en extraction terminologique à l'Institut de linguistique informatique du CNR (Conseil National de la Recherche).

## ■ Evelyn Wareham

Evelyn Wareham is Programme Manager for Digital Sustainability at Archives New Zealand. Her team is responsible for developing a national strategy, capability and systems for maintenance and preservation of digital information across New Zealand's government. Until 2008, Evelyn managed the Government Recordkeeping Programme at Archives New Zealand. In this role, she led the development of recordkeeping standards and establishment for audits under New Zealand's Public Records Act 2005. From 2002-2004, she worked in the headquarters of the International Council of Archives (ICA) in Paris, where she co-ordinated ICA's professional programme of projects and conferences. Evelyn has been involved in international standard setting through Standards New Zealand, Standards Australia and ISO. She is Secretary General of the Pacific Regional Branch of the International Council on Archives (PARBICA). Evelyn has spoken at conferences in Australia, Canada, Fiji, Germany, Malaysia, the USA and New Zealand. She is the author of numerous articles on recordkeeping topics.

Ms Evelyn WAREHAM

Programme Manager, Digital Sustainability

Archives New Zealand

PO Box 12050, Wellington 6144, New Zealand

Phone: +64 4 495 6237

Mobile: +64 21 799 349

Fax: +64 4 495 6210

Email: [evelyn.wareham@archives.govt.nz](mailto:evelyn.wareham@archives.govt.nz)

## ■ **Karl Wessbrandt**

Karl Wessbrandt is a Strategic Development Officer at the Swedish Administrative Development Agency. He is currently the project manager for the development of a proposed National Interoperability Framework for the Swedish public sector.

Karl represents Sweden in the European Commission's expert groups of the IDABC Program's European Interoperability Framework, Semantic Interoperability Centre and Common Assessment Method of Standards.

He is the chairman of the CEN/ISSS Workshop on Discovery of and Access to eGovernment Resources (eGov-Share) and a member of the Swedish Standardisation Institute's work groups on Archives/Records Management and on Document Description and Processing Languages. He is a frequent speaker and author of reports and articles on interoperability, standardisation and information management. Karl holds a Bachelors degree in Informatics from the University of Göteborg. In 1997 he received the AIIM Master of Information Technology Award.

[karl@wessbrandt.se](mailto:karl@wessbrandt.se)

## ■ Yun Petersen Yoneyama

M.A., Ph.D

Danish National Archives.

Jun is Special Consultant (information technology) currently assigned as project manager for the Access Project (2006 - today). The Access Project (2005-2008) creates software and methods for public access to archival records ingested as database structured datasets, as well as the documentation for these records.

Jun holds a M.A in Information Science (1993) and a Ph.D in Computer Science (2004). Scientific articles published and/or presented within Organizational Studies and Computer Supported Cooperative Work 1996-2003.

Contact info:

[jpy@ra.sa.dk](mailto:jpy@ra.sa.dk)