

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 2

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 2**

**2009**

# Sommaire/Summary

## Volume 2

### ***Séance 5 - La certification*** \_\_\_\_\_ **8**

sous la présidence de Martin Waldron, Consultant, In-Form Consult, Royaume-Uni

### ***Session 5 - Certification*** \_\_\_\_\_ **9**

chairing Martin Waldron, Consultant, In-Form Consult, United Kingdom

#### *Le COREF : Une organisation pour l'audit des outils et des services dans le domaine de la dématérialisation et de l'archivage électronique* \_\_\_\_\_ **10**

Jean-Louis Pascon, Consultant, Hénon Conseil, Lyon

#### *Data seal of approval - assessment and review of the quality of operations for research data repositories* \_\_\_\_\_ **17**

Dr Henk Harmsen, Deputy Director of Data Archiving and Networked Services (DANS), The Hague, The Netherlands

#### *Accreditation of document management and archiving services in Slovenia. From the legislation to the practice* \_\_\_\_\_ **23**

Tatjana Hatjnik, MA, The Archives of the Republic of Slovenia, Head of the Division for E-Archiving  
Pavel Golob, EUROJUST, Acting Head of the Security, Facility Management, General Services and Events Unit

#### *Two Words, two Challenges : distinguishing audit and certification of digital archives.* \_\_\_\_\_ **34**

Dr. Seamus Ross, Professor of Humanities Informatics and Digital Curation, Université of Glasgow  
Andrew McHugh, University of Glasgow  
Hans Hofman, Consultant, Nationaal Archief of the Netherlands  
Perla Innocenti, Co-Principal Investigator for Requirements Analysis and Identification of User Scenarios in the project Sustaining Heritage Access through Multivalent Archiving (SHAMAN)  
Raivo Ruusalepp, Consultant, Estonian Business Archives Consultancy

## **Séance 6 - Stratégies pour de nouvelles ressources et architectures de l'information**\_\_\_\_\_ **38**

sous la présidence de Franck Brady, superviseur du système de gestion des documents électroniques de la Commission européenne

## **Session 6 - Strategies for records management and archiving services**\_\_\_\_\_ **39**

chairing Franck Brady, supervisor of the electronic records management system of the European Commission

*Connecting the dots. Leveraging standards and technology to achieve automation of records management processes.*\_\_\_\_\_ **40**

Conni Christensen, Founding Partner, Director of Synercon Management Consulting

*Creating an Intermediate Archive for Semi-Active Electronic Records - Concepts and Challenges*\_\_\_\_\_ **54**

Michael Hollman, Director of the modern archives of the Federal Archives (Bundesarchiv, Koblenz)

*Transition to Electronic Records Exchange in Estonia*\_\_\_\_\_ **61**

Livii Karpistsenko, Kadi Riisma, State Chancellery of the Republic of Estonia

*Long-Term Digital Preservation using Agent Technology - the PROTAGE Project*\_\_\_\_\_ **69**

Johan Bengsston, InterNIT  
Raivo Ruusalepp, Estonian Business Archives Consultancy

*Building a digital repository: a practical implementation*\_\_\_\_\_ **76**

Filip Boudrez, Expertisecentrum DAVID vzw, Antwerpen

*Processus et architecture pour la conservation à long terme des données numériques : le système mutualisé du Conseil général de l'Aube*\_\_\_\_\_ **84**

Céline Guyon, Attaché de conservation du patrimoine, Archives départementales de l'Aube

## **Atelier 1 - MoReq2 : présentation et prise en main**\_\_\_\_\_ **90**

## **Workshop 1 - Understanding MoReq2**\_\_\_\_\_ **91**

*MoReq2: What it is, and How to Use it?*\_\_\_\_\_ **92**

Marc Fresko, Serco Consulting

*MoReq2 and The Romanian Framework*\_\_\_\_\_ **94**

Bodgan-Florin Popovici, archivist, National Archives, Romania  
Lucia Stefan, Archivaria Ltd, London, UK

*Records Management Market Study for Germany, Austria and Switzerland*\_\_\_\_\_ **101**

Dr. Ulrich Kampffmeyer, Managing Director of Project Consult (Germania, Hamburg)

*Traduire MoReq2 : le cas français*\_\_\_\_\_ **102**

Marie-Anne Chabin, expert en archivage, Archive 17

**Atelier 2 - Sécurité et protection des données\_\_\_\_\_104**

**Workshop 2 - The protection of personal data and the archival requirements\_\_\_\_\_105**

*Using an approach from ISO 27001 to build the business case for records management\_\_\_\_\_106*  
Richard Jeffrey-Cook, CITP, Managing Director In-Form Consult Ltd

*Sécurité des données : les supports et les formats sont-ils au coeur de la problématique ?\_\_\_\_\_113*  
Jean-Marc Fontaine, Ingénieur - Laboratoire d'Acoustique Musicale (LAM) (Ministère de la Culture)  
Jacques Perdereau, Ingénieur - Laboratoire National d'Essais et de Métrologie (LNE)  
Gérard Weisz, Consultant - Sirius System France

**Atelier 3 - Comment choisir les bons formats ?\_\_\_\_\_126**

**Workshop 3 - File formats for long-term preservation\_\_\_\_\_127**

*Choisir les formats d'archivage adaptés\_\_\_\_\_128*  
Philippe Martin, Administrateur de l'Aproged, responsable du Pôle Normalisation, Directeur Associé au Bureau van Dijk Ingénieurs Conseils

*PDF/A: Standard for the long term archiving\_\_\_\_\_135*  
Dr. Uwe Wächter, Product Manager - SEAL Systems  
Saïd El-Morabiti, Sales Manager - SEAL Systems

**Atelier 4 - Les stratégies de conservation à long terme\_\_\_\_\_142**

**Workshop 4 - Long-term preservation strategies\_\_\_\_\_143**

*Digital preservation policies created by the ATON project\_\_\_\_\_144*  
Osmo Palonen, Project Manager, Mikkeli University of Applied Sciences, Finland

*Conversion of the Danish National Archives' Collection of Digital Records\_\_\_\_\_151*  
Kirsten Villadsen Kristmar, Head of the Department of Appraisal and Transfer

*Dioscuri: emulation in practice\_\_\_\_\_157*  
Remco Verdegem, Senior Advisor Digital Longevity, National Archives of the Netherlands

**Table-ronde - Les Archives sont-elles prêtes pour affronter le monde numérique ? Réflexions sur le métier** \_\_\_\_\_ **168**

sous la présidence d'Evelyn Wareham, Archives New Zealand

**Round-table - Are archivists ready for the digital world? Reflection on the profession and its evolution** \_\_\_\_\_ **169**

chairing Evelyn Wareham, Archives New Zealand

**Introduction** \_\_\_\_\_ **170**

*Breaking the barriers of traditional records management. Records, Records Management, Enterprise Content Management and MoReq2* \_\_\_\_\_ **170**

Dr. Ulrich Kampffmeyer, Managing Director of Project Consult (Germania, Hamburg)

**L'expérience française** \_\_\_\_\_ **180**

*Comment transformer un archiviste en spécialiste de l'archivage électronique* \_\_\_\_\_ **180**

Carole Gragez, conservateur du patrimoine, responsable du Service historique de la Défense antenne marine de Toulon

Edouard Vasseur, conservateur responsable des archives au Ministère de la Culture.

*Le métier d'archiviste à l'épreuve de l'archivage électronique : entre permanence et renouvellement des concepts et des pratiques. L'expérience du Service national des archives du Groupe La Poste* \_\_\_\_\_ **189**

Anne Burnel, directrice du Service national des archives du Groupe La Poste

*La collecte des données numériques produites dans le cadre des missions d'aide sociale des départements* \_\_\_\_\_ **197**

Emilie Goubin, responsable des archives contemporaines, Archives départementales du Finistère

Nathalie Regagnon, chef du service des relations avec les administrations, Archives départementales de la Haute-Garonne

**Séance plénière de clôture** \_\_\_\_\_ **206**

sous la co-présidence de Toivo Jullinen, président du DLM-Forum

et d'Hubert Sztaszewski, directeur au secrétariat général de la Commission européenne

**Plenary closing session** \_\_\_\_\_ **207**

co-chairing Toivo Jullinen, DLM-Forum President

and Hubert Sztaszewski, Director, Secretariat General, European Commission

*Achievements and new directions: conclusions of the DLM-Forum Conference 2008 in Toulouse adopted by the conference participants* \_\_\_\_\_ **208**

*Réalisations et nouvelles directions : conclusions de la Conférence 2008 du DLM-Forum, Toulouse, adoptées par les participants à la conférence* \_\_\_\_\_ **210**

*From projects to realisations, conclusions: the DLM-Forum, MoReq2 and beyond...* \_\_\_\_\_ 212

Dr. Seamus Ross, professor of Humanities Informatics and Digital Curation, University of Glasgow

*Discours de clôture* \_\_\_\_\_ 213

Martine de Boisdeffre, Directrice des Archives de France

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

***Speakers*** \_\_\_\_\_ **216**

# Séance 5

# La certification

---

---

Sous la présidence de Martin WALDRON  
Consultant, In-Form Consult, Royaume-Uni

[Consulter l'introduction à la séance de Martin Waldron](#)



# ***Session 5***

# ***Certification***

---

---

*Chairing Martin WALDRON*  
*Consultant, In-Form Consult, United Kingdom*

[See Martin Waldron's introduction to the session](#)

# LE COREF : UNE ORGANISATION POUR L'AUDIT DES OUTILS ET DES SERVICES DANS LE DOMAINE DE LA DÉMATÉRIALISATION ET DE L'ARCHIVAGE ÉLECTRONIQUE

---

Jean-Louis Pascon

Consultant, Hénon Conseil, Lyon

*Le COREF est né de quelques questions simples : Comment auditer un service d'archivage ou une plateforme de dématérialisation de factures ? Quels référentiels utiliser ? De quels types d'experts a-t-on besoin pour réaliser ces audits ?*

*C'est en 2006 que trois fédérations professionnelles françaises impliquées dans le domaine de la dématérialisation et de l'archivage ont décidé de créer le COREF.*

*Les trois membres fondateurs du COREF sont :*

- 1. L'ADAP (Association pour la Dématérialisation de l'Achat Public) qui regroupe les fournisseurs de plateformes de dématérialisation, notamment pour les appels d'offres ;*
- 2. FedISA (Fédération ILM, Stockage et Archivage - [www.fedisa.eu](http://www.fedisa.eu)) se veut un lieu de dialogues entre des utilisateurs, des intégrateurs, des experts techniques et juridiques et des fournisseurs de technologies, de matériels et de logiciels dans le domaine de l'archivage ;*
- 3. La FNTC (Fédération Nationale des Tiers de Confiance - [www.fntc.org](http://www.fntc.org)) qui regroupe les principaux opérateurs de l'archivage électronique en France ainsi que des ordres professionnels et experts juridiques et techniques.*

*Les principales missions assignées au COREF par ses fondateurs sont :*

- 1. Le suivi du déroulement des audits de la signature du contrat d'audit jusqu'à la remise du rapport final ;*
- 2. La sélection des experts nécessaires à la réalisation des audits, leur qualification et la mise en place de règles de déontologie pour ces experts ;*
- 3. La participation à la conception de référentiels au sein d'organisations professionnelles ou dans le cadre d'organismes publics de normalisation comme l'AFNOR ;*
- 4. La création ou l'adaptation de méthodes d'audit pour ces référentiels.*

*Après 18 mois consacrés son l'organisation et à la recherche d'experts, le COREF a audité trois des plus grandes sociétés françaises d'archivage électronique (ORSID et ASPHERIA, toutes les deux filiales du groupe La Poste et ASTERION filiale du groupe Pitney Bowes). Plusieurs audits sont prévus pour la fin de l'année 2008.*

*La communication visera principalement :*

- A faire un bref historique de cette organisation et à présenter ses membres ;*
- A décrire la structure et à expliciter les raisons qui ont conduit à la création de cette structure ;*
- A préciser le positionnement d'une telle organisation par rapport au cadre juridique de la certification en France et en Europe ;*
- A présenter ses principales procédures, notamment pour l'acceptation d'un référentiel, pour la sélection des experts et pour la mise en oeuvre des audits ;*
- A montrer, sur un exemple concret, en l'occurrence le label Tiers Archiveur (pour les opérateurs de services d'archivage électronique), les conditions pratiques d'un audit (durée, coût, suivi, rapport final, gestion des contestations, etc.).*

*La communication se terminera par une présentation :*

- Du Comité d'Orientation Stratégique (COS) qui est une commission indépendante composée de représentants de l'administration, d'organismes de normalisation, d'associations d'experts et dont l'objectif est de fournir les grandes orientations du COREF pour les années à venir ;*
- Des axes de réflexion pour le futur : mise en conformité de son organisation avec les règles de la norme européenne EN 45012, prise en compte de nouveaux référentiels (coffre-fort numérique et plateforme de facturation), formation et contrôle des experts, ouverture vers de nouveaux membres, la recherche de coopération en Europe, etc.*

[Consulter la présentation.](#)

## **The COREF: dedicated to auditing tools and services in the field of dematerialisation**

*The COREF was created in response to a few simple questions: How can we audit an archive service or a dematerialised invoicing platform? What reference framework could be used? What type of expert is qualified to perform such audits?*

*In 2006, three French professional federations from the field of dematerialisation and archiving decided to set up the COREF.*

*The three founder members of the COREF are:*

- 1. ADAP (Association pour la Dématérialisation de l'Achat Public) whose members are suppliers of dematerialisation platforms, particularly for calls for tender;*
- 2. FedISA (Fédération ILM, Stockage et Archivage - [www.fedisa.eu](http://www.fedisa.eu)), a forum for discussions between users, integrators, technical and legal experts and suppliers of technology, hardware and software in the field of archiving;*
- 3. FNTC (Fédération Nationale des Tiers de Confiance - [www.fntc.org](http://www.fntc.org)) which includes the principal French operators of electronic archiving as well as professional bodies and legal and technical experts.*

*The founders laid down the following primary missions for the COREF:*

- 1. to follow audit procedures from signing the contract through to submission of the final report;*
- 2. to select the appropriate experts for carrying out audits, to qualify them and to establish a code of professional conduct;*
- 3. to participate in specifying reference frames with professional organisations or public standards bodies such as the French AFNOR;*
- 4. to create or adapt audit methods for these reference frames.*

*After 18 months devoted to its organisation and the search for experts, the COREF has now audited three of the largest French companies in electronic archiving (ORSID and ASPHERIA, both subsidiaries of the Groupe La Poste, and ASTERION, a subsidiary of the Pitney Bowes Group). Several more audits are scheduled for the end of 2008.*

*The main aspects of the presentation will be:*

- a brief history of the organisation and an introduction to its members;*
- a description of its structure with an explanation of the reasons that led to its creation;*
- an explanation of the position of such an organisation with respect to French and European legal certification;*
- a presentation of its main procedures, particularly for adopting new reference frames, selecting experts and implementing audits;*
- a demonstration, using a real example, of the Tiers Archiveur certification (for providers of electronic archiving services), and the practical conditions under which an audit is carried out (duration, cost, monitoring, final report, dispute management, etc.).*

*The talk will conclude with presentations of:*

- the Strategic Orientation Committee, an independent commission made up of representatives from government and standardisation bodies and of associations of experts; its purpose is to provide the COREF with its major themes for the years to come;*
- subjects for future consideration: ensuring that its organisation complies with the requirements of the European EN 45012 standard; adopting new reference frames (the 'digital safe' and invoicing platforms); training and monitoring experts; opening the doors to new members; seeking cooperation in Europe, etc.*

[See the presentation.](#)

## ■ Un peu d'histoire

Bien qu'il existe plusieurs méthodes et plusieurs référentiels pour définir, tester ou auditer des systèmes ou des services d'archivage (Moreq 2 et TRAC en sont de bons exemples), il se pose de nombreuses questions lorsqu'il s'agit d'évaluer ou de valider la conformité d'un système d'archivage ou d'un service ou d'un opérateur dans ce domaine, par rapport à ces normes ou à ces référentiels : Comment auditer un service d'archivage ou une plate-forme technique ? Quels référentiels ou quels combinaisons de référentiels doivent être utilisés impérativement pour tel ou tel usage ? Avec méthodologie d'audit ? A quels types d'experts faut-il recourir pour réaliser les audits de ces systèmes ou de ces services ? Existence-il des organismes capables de réaliser de telles opérations de contrôle ?

Entre 2003 et 2005, plusieurs réflexions, menées par des organisations professionnelles du domaine de la dématérialisation, ont montré qu'il n'existait pas, en France, de sociétés ou d'organismes aptes à répondre à ces questions de façon satisfaisante.

Ces réflexions ont conduit en 2004 à un premier organisme d'audit mais il s'est avéré très rapidement que cet organisme n'avait pas un fonctionnement satisfaisant. Il lui fallait un fonctionnement plus rigoureux et plus autonome.

C'est pourquoi, en 2006, trois fédérations professionnelles françaises, toutes les trois fortement impliquées dans le domaine de la dématérialisation et de l'archivage, ont décidé de créer le COREF, une structure spécifique pour réaliser des audits.

Le COREF a été créé par :

1. L'ADAP (Association pour la Dématérialisation de l'Achat Public) qui regroupe les fournisseurs de plate-formes de dématérialisation, notamment pour les appels d'offres de l'Etat et des collectivités locales ;
2. FedISA (Fédération ILM, Stockage et Archivage - [www.fedisa.eu](http://www.fedisa.eu)) est une fédération lieu de dialogues entre des utilisateurs, des intégrateurs de systèmes, des experts techniques, des juristes, des avocats et des fournisseurs de technologies, de matériels et de logiciels dans le domaine de l'archivage et de la gestion du cycle de vie du document ;
3. La FNTC (Fédération Nationale des Tiers de Confiance - [www.fntc.org](http://www.fntc.org)) qui regroupe les principaux opérateurs de l'archivage électronique en France ainsi que des ordres professionnels et des experts juridiques et techniques liés à l'archivage et aux métiers de la confiance.

Le COREF a déposé ses statuts (association conforme à la Loi de 1901) en août 2006.

## ■ Les missions du COREF

Les principales missions qui ont été assignées au COREF par ses trois organisations fondatrices sont les suivantes :

1. Le pilotage d'audits de la signature du contrat initial jusqu'à la remise du rapport final ;
2. La sélection et la gestion d'experts nécessaires à la réalisation de ces audits, leur qualification et la mise en place de règles de déontologie pour ceux-ci ;
3. La participation à la conception de référentiels au sein d'associations professionnelles ou dans le cadre d'organismes publics de normalisation (comme l'AFNOR, l'association française de normalisation) ;
4. La création ou l'adaptation de méthodes d'audit pour ces référentiels.

D'une façon plus globale, le COREF a été chargé par ses membres fondateurs d'être un opérateur d'audits dans le domaine de la confiance numérique.

## ■ Pourquoi créer un organisme spécifique ?

Quatre raisons majeures nous ont conduits à la création du COREF.

En tout premier lieu, il était indispensable d'aller vite. En effet, depuis cinq ou six ans, l'offre en matière d'archivage électronique a explosé. On peut trouver aujourd'hui, aussi bien des sociétés qui permettent l'hébergement de photographies que des banques qui assurent un service pour stocker les relevés de comptes de leurs clients. Bien évidemment ces offres sont radicalement différentes. Elles n'offrent pas les mêmes garanties tant en termes techniques, que juridiques. Il devenait donc urgent de définir ce que doit être une offre "professionnelle" d'archivage numérique. L'un des moyens de définir et de vérifier la réalité de cette offre professionnelle était de mettre en place un organisme d'audits indépendant. Du fait de sa forme associative, la création du COREF s'est faite rapidement. Cela a permis de démarrer les audits pour les opérateurs de tiers archivage en seulement quelques mois.

En second lieu, il convenait de définir des processus d'audit clairs et transparents, dans un domaine encore mouvant. En effet, les normes, les standards et les référentiels sont aujourd'hui nombreux dans le domaine de l'archivage. De plus ces normes, ces standards et ces référentiels évoluent rapidement. Il fallait donc disposer d'un organisme souple, apte à modifier rapidement ses approches et à répondre efficacement aux évolutions dans ce domaine.

Troisièmement, il nous a semblé que face à de nouvelles problématiques, il était préférable de mettre en place une nouvelle structure, étudiée spécifiquement à cet effet. Bien sûr, il existe plusieurs organismes spécialisés dans les audits informatiques mais aucun ne recouvre exactement l'ensemble des compétences qui concourent au fonctionnement des systèmes ou des services d'archivage à savoir : technologie des supports d'archivage, méthodes cryptographiques, gestion de l'intégrité des données, etc.

Enfin, il fallait tenir compte de la jeunesse et de l'étroitesse du marché des tiers archiveurs. En termes de potentiel de sociétés auditables, nous sommes arrivés à un chiffre assez réduit : au maximum d'une dizaine de sociétés réellement susceptibles d'être auditées entre les années 2007 et 2010 (nous arriverons à cinq sociétés fin 2009, ce qui est bien en ligne avec nos prévisions). Quelques calculs nous ont conduits à penser que le recours à des sociétés spécialisées induirait des coûts assez élevés qui auraient évidemment rebuté certains des candidats aux audits.

C'est donc sur la base de tous ces éléments qu'il a été décidé de créer le COREF et la forme associative a été retenue.

## ■ Nos principales procédures

Afin de fonctionner de façon efficace, nous avons établi très rapidement des procédures écrites détaillées car le fonctionnement d'un organisme d'audits est fondamentalement basé sur des procédures.

Nos principales procédures sont au nombre de quatre. Elles concernent le fonctionnement du conseil d'administration, la sélection des référentiels et la définition des méthodes d'audit, la sélection et la formation des experts et, enfin, le contrôle et le suivi des audits proprement dits.

Le fonctionnement du Conseil d'Administration repose à la fois sur les statuts de l'association et sur le règlement intérieur de celle-ci. Si les statuts n'ont pas changé depuis la création de l'association, le règlement intérieur se complète lui plus régulièrement. Il a évolué à la fois pour des raisons de fond (par exemple, définir le Comité d'Orientation Stratégique) et pour des raisons mineures (notamment d'ordre rédactionnel et de précision terminologique afin de réduire les risques de mauvaises interprétations du texte). Les buts de ces statuts et de notre règlement intérieur sont avant tout d'assurer la transparence du fonctionnement du COREF et l'impartialité de nos décisions.

La sélection et le suivi des experts sont un des autres piliers de notre fonctionnement.

La sélection des experts se fait d'un premier temps sur dossier (nous avons mis en place des dossiers d'inscription qui nous permettent de bien connaître le futur expert et son parcours professionnel) et dans un

deuxième temps, si le dossier de candidature est satisfaisant, par un entretien approfondi conclu par une décision du conseil d'administration.

Au début, nous avons retenu en majorité des experts judiciaires, en raison de leur expérience, de leur maîtrise reconnue en matière d'informatique et, bien évidemment, de leur déontologie. Mais pour nos dernières sélections d'experts, nous nous sommes ouverts à d'autres profils. En effet, pour disposer de personnes aptes à réaliser de manière efficace les futurs audits que nous souhaitons proposer, nous avons commencé à rechercher et à inclure dans notre "pool" d'experts des spécialistes dans le domaine fiscal (pour la gestion des factures dématérialisées), du chiffrage, de l'horodatage et de la signature numérique (pour les coffres-forts électroniques et les outils d'archivage à long terme).

Le suivi des experts portent lui sur deux aspects : leur connaissance des référentiels et des méthodes d'audit associées d'une part, et le respect de ces méthodes, d'autre part. Afin d'améliorer ce suivi et la qualité de nos experts, nous allons mettre en place des formations pour eux, dans le courant de l'année 2009.

La sélection des référentiels est la fois simple et complexe. Simple, car le COREF n'ayant pour finalité que de prendre en compte ou de réaliser des audits autour de l'archivage et de la confiance dans le domaine du numérique, il est facile de décider si un référentiel peut être pris en compte. Complexe, car pour mettre au "catalogue" du COREF un référentiel, il faut s'assurer que nos méthodes seront efficaces et adaptées et que nous pourrions disposer d'experts compétents.

Pour le contrôle des audits, nous mis en place un état d'avancement hebdomadaire. Pour chaque type de référentiel, nous avons défini des rapports formels de façon que tous les rapports portent exactement sur les mêmes critères soient produits de la même façon quelque soit l'expert.

Enfin, tous les aspects pratiques sont transcrits dans des notes techniques. Ces notes techniques sont librement accessibles aussi bien pour les audités que pour les experts.

Aujourd'hui, nos audits repose sur la méthode CoBit.

## ■ Le COS

Afin d'aider le COREF à mieux appréhender son avenir, il a été décidé de créer un comité de personnes extérieures : le Comité d'Orientation Stratégique (COS).

Le but de ce comité est triple.

Tout d'abord le COS doit aider le COREF a définir son futur. Il est là notamment pour aider notre organisation à sélectionner les futurs référentiels que nous pourrions prendre en compte.

Le deuxième rôle du COS est d'aider le COREF à améliorer son fonctionnement. Le COS fournit des idées pour compléter et ajuster nos procédures.

Enfin, le COREF est le lieu des réflexions liés à l'éthique et, avant tout à l'éthique, de nos experts. Dans ce cadre, le COREF pourrait être amené à se prononcer sur des problèmes de déontologie.

Le COS est constitué de représentants de l'administration française et de personnalités reconnues soit dans le domaine des audits, soit dans le domaine de la dématérialisation.

Les activités du COS ont réellement démarré cette année. 2009 devrait être l'année de l'essor de ce comité.

## ■ La situation actuelle

Après deux années d'existence, consacrées en priorité à la définition de notre fonctionnement et aux choix des experts, le COREF est entré, depuis 2008, dans une phase de fonctionnement stable.

Nos procédures sont robustes et efficaces. Cela ne veut pas dire qu'il ne faille pas y apporter des améliorations mais cela veut dire que les audits se déroulent sans problèmes importants et de façon conforme à nos méthodes.

Nous avons un premier groupe d'une dizaine d'experts qui ont prouvé leur capacité à réaliser les audits dans les domaines couverts par le COREF.

Le COREF possède maintenant une bonne expérience grâce aux audits réalisés pour trois des plus grandes sociétés françaises d'archivage électronique (ORSID et ASPHERIA, toutes les deux filiales du groupe La Poste et ASTERION filiale du groupe américain Pitney Bowes).

Nous attendons pour le premier trimestre 2009 deux nouveaux candidats pour le label tiers archiveurs électronique (ces deux candidats viennent du monde de l'archivage papier et souhaitent accompagner leurs clients vers le monde du numérique). Ces deux nouveaux candidats auront, pour leur audit, des procédures complètement définies et détaillées et des méthodes d'analyse de leurs activités complètement publiées.

Pour ce qui est des nouveaux labels, nous pensons que nous aurons 2 ou 3 fournisseurs de logiciels de coffre-fort électronique en 2009 et pour les opérateurs de dématérialisation de factures, nous devrions réaliser 3 à 4 audits de plate-forme l'année prochaine.

## ■ Quelques chiffres

L'un des objectifs visés initialement était d'avoir des coûts raisonnables, tant au niveau des audits qu'au niveau du fonctionnement du COREF lui-même.

Les objectifs que s'étaient fixés le COREF sont, en très grande partie, atteints.

Pour ce qui est du fonctionnement, le COREF a un budget annuel inférieur à 3 000 € (HT). Ce budget recouvre principalement la tenue des réunions du conseil d'administration, les affranchissements, les achats de normes ou d'ouvrages techniques, la création et maintenance d'un site web (qui va ouvrir très bientôt), des déplacements, etc.

Ce budget est bien évidemment très inférieur à ce que peut être le budget d'une société spécialisée dans les audits ou la certification. Ce budget réduit est dû fondamentalement au fait que le COREF n'a pas de locaux (il est hébergé par le Conseil Supérieur de l'Ordre des Experts Comptables) et qu'il n'a aucun salarié permanent.

Cependant nous savons que ce budget évoluera à la hausse dans les années à venir en raison de notre développement (nouveaux référentiels traités et nombre d'audités en augmentation). Cela étant dit, un objectif de budget de fonctionnement de l'ordre de 10 000 €/HT par an, vers 2011-2012, semble tout à fait raisonnable.

Pour ce qui est des audits proprement dits et à titre d'exemple, un audit complet pour un nouveau tiers archiveurs représente 7 jours pour l'expert, soit un coût pour l'audit de 7 500 €/HT. La durée d'un audit de renouvellement pour un tiers archiveur déjà labellisé est de 3 jours, soit un coût de 3 500 €/HT.

Le futur label pour les coffres-forts électroniques devrait se situer dans une fourchette allant de 4 000 à 6 000 €/HT.

Ces durées d'audit et ces prix doivent être comparés aux durées et aux prix d'autres audits informatiques équivalents.

## ■ Le futur

Aujourd'hui, trois axes majeurs de développement ont été décidés par les membres fondateurs du COREF.

Bien évidemment, nous souhaitons offrir plus de référentiels à l'audit mais toujours dans notre sphère de compétence, c'est à dire le domaine de la confiance dans le monde du numérique.



Le COREF ne réalise pour l'instant que des audits d'opérateurs d'archivage électronique (les tiers archiveurs). Dans un proche avenir (premier semestre 2009), le COREF va s'ouvrir et piloter des audits qui auront pour but, d'une part, de valider des offres de coffres-forts numériques et, d'autre part, de vérifier la conformité de plates-formes de gestion des factures dématérialisées conformément aux directives européennes.

Dans cette même approche, nous allons bien évidemment étudier la possibilité d'évaluer des offres systèmes d'archivage qui souhaitent être conformes à Moreq 2. Pour ce faire, nous devons étudier en détail ce référentiel mais aussi le guide des tests qui lui est associé afin de définir les durées des audits et les coûts de ceux-ci.

Le second est directement lié à l'Europe. Nous souhaitons rechercher des coopérations et des alliances en Europe. En effet, des administrations et des sociétés souhaitent avoir les mêmes garanties quand elles transfèrent ou quand elles archivent des documents dans les pays de l'Union Européenne. La première étape de cette ouverture sera de repérer et de discuter avec des organismes similaires au COREF en Europe.

Le troisième est d'ordre organisationnel. Nous devons être toujours plus efficace et toujours plus rigoureux. Pour ce faire, nous allons encore améliorer nos procédures internes et celles que nous imposons à nos experts. Nous allons nous mettre en conformité avec les normes en ce domaine et, notamment, à la norme européenne EN 45012 qui définit les exigences générales relatives aux organismes gérant l'évaluation et la certification/enregistrement des systèmes qualité. Au travers de la mise en oeuvre de cette norme, nous souhaitons améliorer nos relations tout à la fois avec nos experts mais aussi avec nos audités, en particulier dans les cas de difficultés lors des audits.

## ■ Conclusions

Le COREF est tout jeune (tout juste un peu plus de 2 ans d'existence) mais il est déjà fort d'une bonne expérience.

Son travail et ses méthodes sont efficaces et, il nous semble, reconnus par de nombreux acteurs du domaine de l'archivage numérique.

Maintenant, il nous reste trois défis pour les deux années à venir : nous ouvrir à de nouveaux référentiels (coffres-forts numériques et plates-formes de factures numériques), être encore plus efficace en améliorant nos procédures internes et, enfin, aller vers des partenariats européens.



# DATA SEAL OF APPROVAL - ASSESSMENT AND REVIEW OF THE QUALITY OF OPERATIONS FOR RESEARCH DATA REPOSITORIES

---

Dr Henk Harmsen

Deputy Director of Data Archiving and Networked Services (DANS), The Hague, The Netherland

*When Data Archive and Networked Services (DANS<sup>1</sup>) was established, the founders assigned it the task of developing a Seal of Approval for data, to ensure that digital research data<sup>2</sup> will still be found, recognized and used in the future. With regard to durable storage and maintaining accessibility, the physical place where data are stored is a lot less important than how they are stored.*

*The quality guidelines formulated in the Data Seal of Approval<sup>3</sup> are of interest to researchers and institutions that establish digital research files, to organizations that archive research files, and to users of research data.*

*Objective of the data seal of approval is to safeguard high-quality and reliable processing of research data for the future without it entailing new thresholds, regulations or high costs.*

*It gives researchers the assurance that their research results will be stored in a reliable manner and can be reused.*

*It provides research sponsors with the guarantee that research results will remain available for reuse.*

*In a reliable manner, it enables researchers to assess research data to be reused.*

*It allows data repositories to archive and distribute research data efficiently.*

*The Data Seal of Approval contains a total of 17 guidelines for the application and verification of quality aspects with regard to creation, storage and (re)use of digital research data in the social sciences and humanities. These guidelines serve as a basis for granting a 'data seal of approval'. The criteria for assigning the seal of approval to data are in accordance with, and fit in with, national and international guidelines for digital data archiving such as Kriterienkatalog vertrauenswürdige digital Langzeitarchive<sup>2</sup> as developed by NESTOR, Digital Repository Audit Method Based on Risk Assessment (DRAMBORA)<sup>3</sup> published by the Digital Curation Centre (DCC) and Digital Preservation Europe (DPE), and Trustworthy Repositories Audit & Certification (TRAC): Criteria and Checklist of the Research Library Group (RLG)<sup>4</sup>. The guidelines of the Data Seal of Approval can be seen as a minimum set distilled from the above proposals.*

*In this paper I will give an outline of how the guidelines have been come about. Furthermore I will give an overview of the implementation traject and finally I will say some more about the actual situation:*

- *what are the reactions in the field?*
- *how many archives have already asked to be 'sealed'?*
- *how does DANS plan the implementation of the certification based on trust.*

[See the presentation.](#)

---

<sup>1</sup> DANS - Data Archiving and Networked Services - is an institute of the Royal Netherlands Academy of Arts and Sciences (KNAW), and is also supported by the Netherlands Organization for Scientific Research (NWO). Since its establishment in 2005, DANS has been taking care of storage and continuous accessibility of research data in the social sciences and humanities.

The final certification will be done on base of self assessment and trust rather than on audits.

<sup>2</sup> Digital research data must meet five quality criteria:

1. The research data can be found on the internet.
2. The research data are accessible, while taking into account ruling legislation with regard to personal information and intellectual property of the data.
3. The research data are available in a usable data format.
4. The research data are reliable.
5. The research data can be referred to.

<sup>3</sup> <http://www.datasealofapproval.org>

## Les "D.A.N.S." et la mise en place du sceau d'approbation des données (Datakeurmerk) aux Pays-Bas

Lorsqu'ils ont créé les D.A.N.S.<sup>4</sup> (Data Archive and Networked Services), les concepteurs de ces services leur ont confié le développement d'un sceau d'approbation pour garantir que les données numériques issues de la recherche<sup>5</sup> pourront continuer à être recherchées, reconnues et utilisées. En ce qui concerne le stockage durable et le maintien de l'accessibilité, l'endroit physique où sont conservées les informations compte beaucoup moins que la façon dont elles sont conservées.

Les lignes directrices de qualité spécifiées par le sceau d'approbation<sup>6</sup> s'adressent aux chercheurs et aux institutions qui produisent des documents de recherche numériques, aux organisations qui archivent les documents de recherche et aux utilisateurs des données issues de la recherche.

Ce sceau d'approbation a pour fonction de garantir à l'avenir la fiabilité et la qualité du traitement des données issues de la recherche sans que cela n'induisse de nouveaux paramètres, de nouveaux règlements ou des coûts élevés.

- Il donne aux chercheurs l'assurance que leurs résultats seront stockés en toute sécurité et pourront être réutilisés.
- Il donne aux commanditaires des programmes de recherche la garantie que les résultats des recherches resteront disponibles pour une utilisation ultérieure.
- Il permet aux chercheurs d'évaluer en toute sécurité les données de recherche qu'ils souhaitent réutiliser.
- Il permet aux dépôts d'archiver et de diffuser efficacement les données de la recherche.

Le sceau d'approbation est attribué sur la base de 17 lignes directrices pour l'application et le contrôle des aspects qualité en termes de création, de stockage et de (ré)utilisation des données numériques issues de la recherche en sciences sociales et humaines. Les critères de décernement de ce sceau d'approbation des données sont adaptés et conformes aux recommandations nationales et internationales en matière d'archivage des données numériques telles que le « Kriterienkatalog vertrauenswürdige digital Langzeitarchive » élaboré par NESTOR, la Méthode d'audit des archives numériques basée sur l'évaluation des risques (DRAMBORA) publiée par le DCC (Digital Curation Centre) et le DPE (Digital Preservation Europe) ou encore le TRAC (Trustworthy Repositories Audit & Certification) avec les Critères et la checklist du RLG (Association des bibliothèques de recherche). Ces lignes directrices peuvent être vues comme un ensemble minimal de recommandations dérivé des propositions ci-dessus.

Je décrirai brièvement dans cet article la façon dont ces recommandations ont vu le jour. Je ferai ensuite une présentation globale du parcours de mise en œuvre et je finirai en évoquant un peu plus précisément la situation actuelle à travers les questions suivantes :

- Quelles sont les réactions sur le terrain ?
- Combien d'archives ont-elles déjà demandé à recevoir le sceau ?
- Comment les services DANS comptent-ils introduire une certification basée sur la confiance ?

[Consulter la présentation.](#)

---

<sup>4</sup> DANS (Data Archiving and Networked Services) est une institution de l'Académie des arts et des sciences des Pays-Bas (KNAW) également soutenue par l'organisation néerlandaise pour la recherche scientifique (NWO). Depuis sa création en 2005, DANS assure le stockage et l'accessibilité permanente aux données de la recherche en sciences sociales et humaines.

La certification finale sera effectuée sur une base de confiance et d'auto-évaluation plutôt que sur des audits.

<sup>5</sup> Les données numériques issues de la recherche doivent répondre aux cinq critères de qualité suivants :

1. Elles doivent être disponibles sur Internet.
2. Elles doivent être accessibles dans le respect des règlements en vigueur relatifs aux données personnelles et à la propriété intellectuelle des informations.
3. Elles doivent être présentées sous un format utilisable.
4. Elles doivent être fiables.
5. Elles doivent pouvoir être citées en référence.

<sup>6</sup> <http://www.datasealofapproval.org>

## ■ Introduction

Data Archiving & Networked Services (DANS) is active in the area of data infrastructure, with two main themes, namely (digital) archiving and making research data available. The field of activity of DANS covers both the social sciences and the humanities. DANS also manages its own data repository of research data.

In 2005, the founders of DANS, the Royal Netherlands Academy of Arts and Sciences (KNAW) and the Netherlands Organization for Scientific Research (NWO), gave DANS the formulation of a data seal of approval as one of its assignments. In February 2008, 17 guidelines were presented under the name data seal of approval, nationally at a KNAW symposium and internationally at the first African Digital Curation Conference. This article will explain more about the backgrounds of the seal of approval: what it is and what it isn't, which international seals of approval exist, how this seal of approval matches them, what its unique selling point is, and what the plans for the future are?

## ■ Data seal of approval: what it is and what it isn't?

The data seal of approval consists of 17 guidelines that may be helpful to an archiving institution striving to become a trusted digital repository (TDR<sup>7</sup>). The guidelines have been formulated in such a way that they are easily understandable and leave sufficient room for a broad interpretation. Standardization was not the objective as the point of departure was that the data seal of approval would remain dynamic during its first years. The seal of approval does not express any views regarding the quality of the data to be archived, but does regarding the provisions an archive has made to guarantee the safety and future usability of the data.

The seal of approval mentions 4 stakeholders: the financial sponsor, the data producer, the data consumer and the data repository, which share an interest and are responsible for a properly functioning data infrastructure. The sponsor is advised to use the guidelines as a condition for financing of research projects. The remaining three stakeholders are addressed in the 17 guidelines. For example, the data producer is expected (three guidelines) to place its data in a TDR and to provide the research data as well as the metadata in the format requested by the data repository. The data consumer must, if it has access to or uses the information in a TDR, respect (inter)national legislation, (scientific) codes of behavior and the applicable licenses (three guidelines). The data repository, in its turn, must ensure that the archive is equipped in such a way that data producer and data consumer are able to meet their obligations. In addition, there are eleven more guidelines for the data repository, regarding organization (mission, dealing with legal regulations, quality management, long-term planning and scenarios), processes (transfer responsibility, data references, integrity and authenticity) and technical infrastructure (OAIS and automated processes). In other words, the data repository is the stakeholder of which most is expected.

## ■ Overview of the data seal of approval guidelines<sup>8</sup>

### The data producer

1. The data producer deposits the research data in a data repository qualified according to these guidelines.
2. The data producer provides the research data in formats recommended by the data repository.

<sup>7</sup> The term Trusted Digital Repository (TDR) occurs in almost all seals of approval. However, it is unclear what a TDR is exactly. At the time of writing, Wikipedia does not yet have a description of the concept. Main point of such a repository is 'trust'. It is the basis of the data seal of approval.

<sup>8</sup> Data seal of approval, chapter 0.3 Guidelines. See: <http://www.datasealofapproval.org> [site visited 15 August 2008]

3. The data producer provides the research data together with the metadata requested by the data depository.

### **The data repository**

1. A data repository has an explicit mission in the area of digital archiving and propagates it.
2. A data repository ensures that legal regulations and contracts are complied with.
3. A data repository applies processes and procedures for guaranteeing quality management for the storage of data.
4. The data repository has a long-term plan regarding measures for durable archiving.
5. Archiving takes place according to predetermined criteria.
6. A data repository assumes responsibility from the data producers for access to and availability of the digital objects.
7. A data repository enables the users to utilize the research data and refer to them.
8. A data repository ensures the integrity of the digital objects and the metadata. The information contained in the digital objects and metadata is complete and does not change relative to the originals.
9. A data repository ensures the authenticity of the digital objects and the metadata. This regards the degree of reliability of the originality and origin. Existing relationships between data sets and explicit links are maintained.
10. The technical infrastructure explicitly supports the tasks and functions as mentioned in the OAIS reference model.
11. The technical infrastructure of a data repository supports guidelines 5 through 12

### **The data consumer:**

1. With regard to accessing information, the data consumer is bound by national legislation. The data repository has a separate access regulation, which includes restrictions imposed by the laws of the country in which the data repository is located. Access regulations should be based as much as possible on relevant international access standards (e.g. Creative Commons). The data consumer must comply with the applicable frameworks and access regulation.
2. The data consumer conforms to and agrees with any codes of conduct that are generally accepted in higher education and scientific research for the exchange of knowledge and information.
3. The data consumer respects the applicable licenses of the data repository with regard to use of the research data.

### **Data seal of approval Assessment**

An assessment document has been formulated for the data repository which, when completed, approved and publicly published, will result in the repository being allowed to use the logo of the data seal of approval. The logo makes the repository recognizable to both data produces and data consumer.

A data repository may be able to delegate some of the guidelines to another archive that bears the logo of the data seal of approval. This way, the concerned repository does not need to execute all the guidelines in order to meet the requirements of the seal of approval.

With regard to auditing the repositories, a minimal system was chosen that is based on trust. The repository publishes its own assessment and then applies for an audit. This audit will be carried out by a member of the

international DSA (data seal of approval) assessment group<sup>9</sup> on the basis of the available assessment document. It determines whether the guidelines have been complied with and whether the logo can be awarded.

## ■ International initiatives

The text accompanying the seventeen guidelines states that these 'are in accordance with, and match national and international guidelines regarding digital data archiving'. In this section, I will explore the mentioned initiatives in slightly more detail.

### *Catalogue of Criteria for Trusted Digital Repositories*<sup>10</sup> - NESSTOR

This catalogue has identified criteria that can help in the evaluation of the reliability of digital archives at both the organizational and the technical level. The criteria were defined in close cooperation with a broad range of data institutions and information producers. One of the objectives is to offer a tool enabling archiving institutions to archive and demonstrate reliability. The catalogue is also an opportunity for arriving at the certification of repositories, with a 'standardized national or international process'. Again, 'reliability' or 'trust' plays a role here. The catalogue can be used for conceiving, working out and eventually implementing a 'trusted digital long-term repository' and for working out (in various stages) of a self-assessment. The criteria catalog employs over fifty criteria organized into fourteen sections that are arranged into three areas of attention namely: Organizational framework, Object management, and Infrastructure and Security.

*Digital Repository Audit Method Based on Risk Assessment (DRAMBORA)*<sup>11</sup> of the Digital Curation Centre (DCC) and DigitalPreservationEurope (DPE)

The DRAMBORA toolkit is available to support internal audits of archiving institutions. To this end, the party responsible for the archive has the challenge of tracking down the weaknesses, while at the same time acknowledging the strength of the archive. DRAMBORA helps track down the many risks any archiving institution runs. This takes place in the form of process description:

- A detailed description of the organization (mission, and activities);
- Formulation of possible risks, organizational as well as technical, that may occur;
- Evaluation of the impact of these risks and making them manageable and controllable.

DRAMBORA gives support by means of templates for the description of risks and codes to assess the severity of the risks. Apart from that, it is an open process which must be shaped by the party responsible for the repository. There is, however, a list of examples of possible risks. The philosophy of the DRAMBORA authors is clear: by monitoring closely what people are doing and how they are doing it, a repository is capable of keeping the risks involved in archiving of data under control.

Further, the Research Library Group (RLG)<sup>12</sup> developed the *Trustworthy Repositories Audit & Certification (TRAC): Criteria and Checklist*. This criteria checklist comprises three sections, arranged into a various aspects, in their turn subdivided into more than eighty criteria. The status of this project is at this moment unclear.

The paper *Foundations of Modern Language Resource Archives* of the Max Planck institution in Nijmegen<sup>13</sup> must not remain unmentioned. The document describes a data seal of approval specifically for language bodies. A language resource archive (LRA) must meet nine principles.

The Research Information Network in the UK<sup>14</sup> developed the *Stewardship of digital research data: a framework of principles and guidelines*. This document is built up of 5 principles, spread across 40 guidelines.

---

<sup>9</sup> The international DSA assessment group will be launched in the end of 2008

<sup>10</sup> See: <http://edoc.hu-berlin.de/docviews/abstract.php?id=27249> [site visited 15 August 2008].

<sup>11</sup> See: <http://www.digitalpreservationeurope.eu/announcements/drambora/> [site visited 15 August 2008].

<sup>12</sup> See: <http://www.crl.edu/content.asp?l1=13&l2=58&l3=162&l4=91> [site visited 15 August 2008].

<sup>13</sup> Peter Wittenburg, Daan Broeder, Wolfgang Klein, Stephen Levinson, of the Max-Planck-Institute for Psycholinguistics in Nijmegen, The Netherlands, and Laurent Romary of the Max Planck Digital Library in Munich, Germany. See: <http://www.lat-mpi.eu/papers/papers-2006/general-archive-paper-v4.pdf> [site visited 15 August 2008].

<sup>14</sup> See: <http://www.rin.ac.uk> [site visited 15 August 2008].

The German Initiative for Network Information (DINI) developed the *Certificate Document and Publication Services of the Deutsche Initiative für Netzwerkinformation*<sup>15</sup>, a certificate mainly intended for institutional repositories with their own Document and Publication Services.

## ■ Synthesis

The guidelines of the data seal of approval can be seen as a basic set of the above proposals. The data seal of approval wants to facilitate ‘awareness’ at the archiving institutions. The objective of the data seal of approval is mainly to try and convince archiving institutions to start paying attention to quality management. It can serve as a first step toward a ‘heavier’ assessment and certification. The authors see the data seal of approval as supporting for example TRACK. But they also expect the DSA Assessment to evolve in the coming years to become a professional digital curation tool.

## ■ Unique selling point

The data seal of approval (DSA) as developed by DANS has a number of unique features: The DSA is oriented toward scientific data, not primarily toward publications. The DSA not only pays attention to the archiving institution, but also to the data producer and the data consumer. This encourages the idea of shared responsibility.

As indicated before, the DSA is not in conflict with for example TRAC, but is rather a step toward it. Where TRAC tends to standardization, the DSA opts for ‘trust’. This way of working does on the other hand match the custom of peer review in the scientific world.

The DSA also focuses on smaller organizations. The DSA is relatively light and therefore easy to implement. Openness, dynamics and speed are possible in the actual implementation.

The DSA is formulated as points of attention, not as solutions. Finally, the DSA offers possibilities for subcontracting archiving and still meet the requirements of the DSA. This will be appreciated by research groups with their own data projects.

## ■ Future

In 2009, DANS will comply with the data seal of approval and its policy is aimed toward being on the way to meeting the TRAC criteria. Furthermore, DANS will use the code for information security<sup>16</sup>.

DANS strives toward internationalization of the data seal of approval. The previously mentioned DSA assessment group will be launched in the end of 2008, and that same year. Four pilots will be planned in The Netherlands as a first step in the area of certification of the DSA. Other pilots in Europe (at this moment there is an main interest in France, Germany and the United Kingdom) are planned in the beginning of 2009. A conference about the data seal of approval will be hosted by the EC in spring 2009.

---

<sup>15</sup> See: <http://edoc.hu-berlin.de/series/dini-schriften/2006-3-en/PDF/3-en.pdf> [site visited 15 August 2008].

<sup>16</sup> CVI - The *Code voor Informatiebeveiliging* is the Dutch version of the British Standards 7799, which was later published as ISO/IEC 17799 as international standard for information security in organizations. It is a general code applicable to all institutions that work with information.



# ACCREDITATION OF DOCUMENT MANAGEMENT AND ARCHIVING SERVICES IN SLOVENIA. FROM THE LEGISLATION TO THE PRACTICE

---

Tatjana Hatjnik

MA, The Archives of the Republic of Slovenia, Head of the Division for E-Archiving

Pavel Golob

EUROJUST, Acting Head of the Security, Facility Management, General Services and Events Unit

*When new archival legislation was adopted in Slovenia in 2006, when the Archives of the Republic of Slovenia (ARS) introduced the accreditation scheme for document management software, hardware and services providers, the general public, public service organisations and business community have expressed high interest.*

*In this paper we are presenting the brief history, current state of the play and future plans in the implementation of the accreditation. The accreditation of the services, which presents the top layer of the accreditation scheme, is discussed in more detail, as the document management and archiving services are very complex and require special attention.*

***Key words :** Protection of Documents and Archives and Archival Institution Act, Archival Act, Documentary material, Electronic archive, Provider of equipments and / or services, Registration, Confirmation, Accreditation, Internal Rules, Unified Technological Requests.*

[See the presentation.](#)

## **Certification de la gestion documentaire et des services d'archivage en Slovénie De la législation à la pratique**

*Lorsque la nouvelle législation sur les archives a été adoptée en 2006 en Slovénie et que les Archives de la république de Slovénie ont présenté leur programme de certification de la gestion documentaire, les fournisseurs de logiciels, de matériel et de services, le grand public, les organismes de service public et la communauté des affaires ont tous exprimé le plus vif intérêt.*

*Dans cet article nous ferons un bref historique, puis nous présenterons la situation actuelle et les plans de mise en oeuvre de la certification. La certification des services qui représente la couche supérieure du programme de certification est plus détaillée, la gestion documentaire et les services d'archivage étant très complexes et requérant une attention particulière.*

*Les Spécifications techniques unifiées introduites en 2006 et qui sont conformes à MoReq (Modèle d'exigences pour l'organisation de l'archivage électronique) constituent la pierre angulaire du programme de certification.*

[Consulter la présentation.](#)



## ■ Introduction

The corner stone of the accreditation scheme are The Unified Technical Requirements (UTR), which have been introduced in 2006 and are in line with the MoReq specifications.

The precondition for accreditation of the services is the registration of the service provider, which is simple administrative procedure and is facilitated by the web based form.

Any entity (e.g. service provider), which intent to capture or store materials in the digital form, shall adopt Internal Rules (IR) in accordance with Protection of Documents and Archives and Archival Institution Act (PDAAIA, Archival Act)<sup>1</sup>, the related implementing regulations, uniform technological requirements and rules of profession (e.g. archival profession, information security).

IR may be submitted for confirmation to the National Archives, which verifies their compliance with the requirements of the Archival Act, the related implementing regulations, UTR and rules of profession.

In 2008, the ARS started to set up the accreditation framework for document management and archiving services. Service provider has to register the services before the issuing of the claim. As this process requires significant technical, knowledge and skills, the ARS invited the information systems auditors to participate in this process. The information systems auditors have been selected within the public tender. At the time being there are approximately 5 audit services provider on the reserve list. Confirmation of the IR is mandatory for all service providers and the Entities under public law. IR stipulates the control activities. The confirmation and accreditation are supported with specific Check List (CL).

As the accreditation is quite challenging process, the ARS has decided to pilot the first accreditation. Pilot will be joint effort of the ARS in collaboration with the Slovenian ISACA Chapter and is foreseen to be concluded in the autumn of 2008.

After the hopefully successful introduction of the service accreditation, the ARS is planning to update the technological requirement and related audit lists and guidelines, with the aim to link them to MoReq2 specification.

## ■ Registration of the service

No special permit is necessary for performing equipment supply activities or performing services in the field of capture and storage of material in the digital form. The services provider shall register his activity with the ARS at least eight days before they start delivery of the service. Registration of the provider into the register of providers is concluded with an administrative decision.

The aim of the registration process is to verify the completeness of the form and to update the registry of the providers of the services. ARS does not check the IR or procedures of the service provider during the registration process. The registration is facilitated with a special form which needs to be filled in by the service provider.

The registration form consists of the following attributes:

- Title of the service
- Description of the service
- Staff details (first and last name, education, competences details)
- Statement of the compliance with the regulations

In the legislation the following types of the services has been foreseen:

---

<sup>1</sup> Protection of Documents and Archives and Archival Institutions Act /PDAAIA/ (Official Gazette of the RS, 30/2006)

- SpS - 1: Capture of documents in digital form,
- SpS - 2: Capture of documents in physical form,
- SpS - 3: Conversion of documents from physical into digital form,
- SpS - 4: Conversion of documents from digital form into a long-term preservation form,
- SpS - 5: Arrangement or selection of documents in digital form,
- SpS - 6: Arrangement or selection of documents in physical form,
- SpS - 7 Destruction of documents in digital form,
- SpS - 8: Destruction of documents in physical form,
- SpS - 9: Provision of secure premises for storage of materials in digital form,
- SpS - 10: Other services interfering with integrity, security or authenticity of documents.

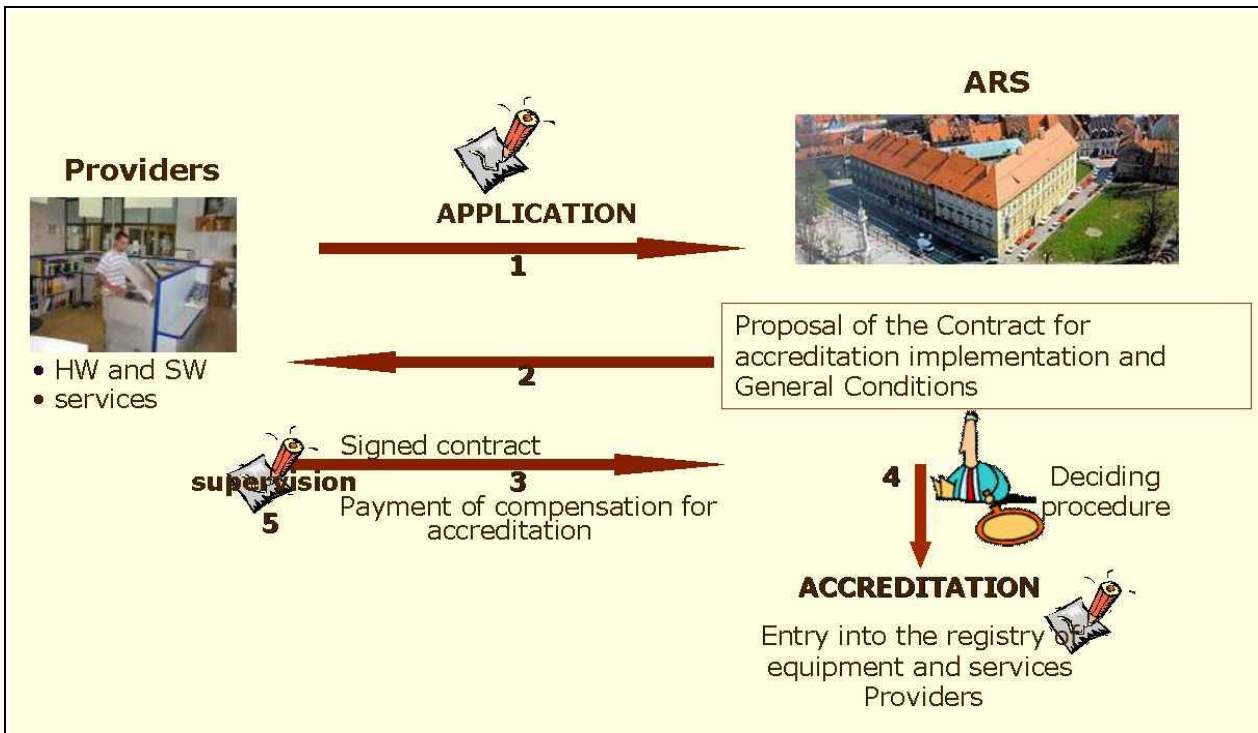
## ■ Confirmation of the IR

Public administration and entities, providing the services to third parties, are obliged to prepare internal rules, which ensure the appropriate level of services in terms of compliance with the relevant legislation, UTR and the best practices. Before the process of accreditation can start, the ARS first confirms the compliance of the entity's IR with the legislative requirements. The confirmation process is based on the CL, which is also published on the ARS Internet pages, and serves as guidelines for IR preparation and implementation.

## ■ Accreditation Process

The services provider may acquire accreditation with the ARS for the equipment or services offered to third parties. In relation to accreditation the ARS shall sign an accreditation contract with the provider defining the relations between the ARS and the provider in compliance with the general accreditation rules as defined by the ARS.

The accreditation process consists of the 5 steps, which are presented on the following workflow diagram:



The steps are as follows:

- Step 1: Application submission
- Step 2: Setting up the contract between the ARS and the provider
- Step 3: closing the contract
- Step 4: Accreditation review and accreditation issuing
- Step 5: Permanent supervision of the service provider

### **Step 1: Application submission**

The accreditation is not mandatory for all service providers. The exception is the providers, who intend to offer the service to the public administration's entities. In this case only the accredited equipment and services for storage documents in a digital form can be deployed.

The provider shall submit a fulfilled accreditation request form, which is published on the ARS web page.

### **Step 2: Setting up the contract between the ARS and the provider**

Before the accreditation process could start the contract between the ARS and the service provider shall be signed. The contract stipulates the terms and conditions of the accreditation, which includes the cost of the accreditation, parties involved, etc. ARS is intitled to involved third parties, e.g. the Information Systems Auditors in the process of the reviewing.

When the contract is closed and agreed the accreditation review could start. When the review is completed with the positive opinion the accreditation license could be issued.

### **Step 3: Payment of compensation for accreditation**

The payment should be put before the accreditation starts.

#### **Step 4: Accreditation review and accreditation issuing**

The most important and time consuming part of the service accreditation is the review of the internal control environment, as it is defined in the IR.

The accreditation review is facilitated with the CL, where are efforts to be taken and tasks to be performed are documented. The CL includes the guidelines for the review of all control activities, which are foreseen to be implemented within the service provider organisation in order to meet all the legal and technical requirements. As each control activity importance is assessed in scale of critical, very important and important the ARS is able to focus to the critical control activities and to efficiently plan the accreditation review.

For the time being the review of the critical control activities is mandatory. The decision on whether to review or not very important and important control activities lies with the review teams and it up to its professional judgment to take that decision.

The outcomes of the review process are:

- The description of the review / test of the control activities
- The assessment on the effectiveness of the control activities
- The findings and recommendations
- Supporting documents (copies of the documentation used, obtained evidences, copies of the audit trails, etc.)
- Overall assessment of the compliance of the control activities with the Service Provider's IR

The overall assessment including the findings and recommendation are communicated with the entity in order to reach the agreement on how to proceed in the process of mitigating the findings and in issuing the accreditation. The ultimate decision making authority is the ARS.

#### **Step 5: Permanent supervision of the service provider**

The ARS shall supervise the accreditation in compliance with general conditions and the contract from the preceding Article. In relation to verification of the activities performed by accredited equipment and services providers, the general conditions and the contract from the preceding Article of this Act shall give the ARS, in relation to the implementation of the existing regulations, UTR and recommendations by the ARS referring to accredited equipment and services, powers equalling the inspection powers on the field of archival institutions and electronic commerce.

In case the accreditation supervision reveals failure to fulfil the regulations in force, UTR and recommendations by the supervisory body for accredited equipment and services, or the provider has terminated the contract with the ARS, the accreditation authority shall delete the provider's services from the register of accredited services.

### **■ Confirmation and Accreditation Check List**

The confirmation and accreditation processes are both facilitated with the Check List which is published on the ARS Internet web pages.

The Archival Act stipulates that every entity, which would like to offer the service of capture and storage of material in the digital form, shall adopt IR. The ARS has the mandate to assess the compliance of the proposed IR against the legislation, national regulations and general requirements, which are stated in the UTR.

Confirmation and accreditation of the IR is mandatory for all public service operators, whilst this is not the case for the other types of organisations. The entities, which would like to offer any kind of the service of capturing and preserving the electronic documents to the third parties shall prepare the IR and confirmed them at ARS.

The CL provides the framework for IR preparation and planning the review of the control activities. The CL may be used to:

- Facilitate the activities of ARS, when they perform:
  - Approval of the IR
  - Review the compliance of the control activities with the IR within the entities, which manage the archive documentation
  - Accreditation of the electronic documents capturing and preserving services
- Support the understanding of the electronic documents capturing and preserving services at this services users in order to enable them set the demands, when acquiring this kind of services
- Registered providers of the archiving services, which would like to accredit this services. The CL could be used as a guideline in setting up the internal control framework for electronic documents capturing and preserving services. It could also be used for self-assessment of the compliance with the relevant legislation and best practices.
- Facilitate the self-assessment of IR within the organisation, which implements the services of capturing and preserving the electronic documents (regardless the nature of the service, e.g. for in-house or service provider)

## ■ Structure of the CL

The CL consists of 5 sections:

- Section A: The methodology of review
- Section B: List of the control activities
- Section C: Final assessment
- Section D: Legal framework
- Section E: Appendixes

### **Section A: The methodology of review**

In the Section A the structure of the documents is described as well the basic guidelines for the CL usage in the process of the approval of IR and in review of the IR implementation.

### **Section B: List of the control activities**

In the Section B the expected control activities are listed in five chapters:

- B.1. General Information on the provider. This chapter is filled in the ARS, when receiving the application for the accreditation
- B.2 Information on the accreditation, hardware and software. Again filled in by ARS, when receiving the application form.
- B.3 Organisational structure and IR. In the chapter all requirements related to the organisational and managerial structure of the service are stipulated. This includes the description of the process IR preparation and Confirmation.
- B.4 Requirements regarding the capture and storage of material in the digital form. In the chapter all requirements, which describes the internal organisational structure, human resources, staff competences, information and communication infrastructure with the emphasis on the information security, are listed.

- B.5 Provisions on capture and storage of material in the digital form. This chapter provides comprehensive descriptions of the requirements with regard to the electronic documents life cycle management.

In total the section consists of 113 requirements. Each of these is presented in the form of the standard table. The structure of the tables, e.g. form is presented in the [Appendix 1](#) of this paper.

### **Section C: Final assessment**

In Section C the final assessment of the Certified Information System Auditor (CISA) is stated. This assessment is considered as an independent opinion on the internal control environment within the capturing and preserving electronic documents service provider, on which the ARS issues a accreditation decision, which could be in favour or against the service deployment.

### **Section D: Legal framework**

In Section D all relevant legislations, including the relevant acts, legislation and other rules issued by the ARS are stated with the relevant references to the Internet pages.

### **Section E: Appendixes**

Section E consists of the appendixes:

- E.1 Mapping of requirements listed in Check List with the standards ISO 27001:2005 and ISO 9001:2000
- E.2. Assessment of the importance in terms of the criticality of the requirement listed in the Check List

## **■ State of play**

As the reviewing of the IR and the control environment at the service provider is quite challenging process, the ARS have decided to pilot the first accreditation. Pilot is joint effort of the ARS in collaboration with the Slovenian ISACA Chapter and is foreseen to be concluded in the autumn of 2008.

In October 2008 ARS has issued the following numbers of the Confirmation certificates:

- 68 registered equipment and services providers
- 6 confirmed internal rules (10 in progress)
- 5 accredited software (5 in progress)
- 15 accredited hardware (58 different type of hardware - system storage, servers, tape library) (3 in progress for 29 type of hardware)

ARS has also started a pilot accreditation with one of the first service providers, which had registered the service.

The aim of the pilot is to test the whole process of accreditation, the documents, which support that process, especially the CL and to assess the timeframe of the process. This outcome is also important for the cost of the accreditation assessment.

The first lessons learned within the pilot accreditation process are as follows:

- When reviewing the IR implementation the special attention should be paid to the way how the IR had been communicated and disseminated throughout the organisation.
- The review should be focused to the key entity's players, who are managing the service, e.g. system administrators, operators, staff responsible for the metadata management and the clients user accounts, etc.

The review team should obtain sufficient evidence that persons in concern are familiar with the IR and that the segregation of duties has been enforced.

- The agreements between the service provider and the outsourced providers should be reviewed carefully and check if the Service Level Agreements have been introduced, monitored on regular bases and appropriate reports have been produced and communicated.

- All critical systems should be regularly monitored and maintained, which shall be checked by reviewing the maintenance plans and logs.

- The compliance with the international standards ISO 27001 and ISO 9001 is important for the IR implementation and accreditation. The control activities foreseen in these two standards are in line with many requirements stated in the UTR and in CL. If the organisation has gained one or both of these standards, the preparation of the IR is simplified, as some of the control activities had been already implemented during the standard implementation. Therefore organisation needs to refer to them in the IR, when appropriate. The compliance with the standards also accelerates the accreditation process, as ARS could rely on the work of other, e.g. the ISO review and ISO certificates.

One of the most challenging parts of the accreditation review is also the fact that in most cases there would not be a lot of audit trails logs recorded yet at the entity, as the service had not been rolled out yet. Therefore the review team should focus to the procedural part of the IR and to the staff's awareness on the implementation of the IR.

After the hopefully successful introduction of the service accreditation, the ARS is planning to update the UTR and CL, with the aim to produce an efficient and effective accreditation framework. The next step would be the compliance with the MoReq2 specification.

## ■ List of abbreviation

1. Check List (CL)
2. Archives of the Republic of Slovenia (ARS)
3. Internal Rules (IR)
4. Unified technological Requirements (UTR)
5. Protection of Documents and Archives and Archival Institution Act (PDAAIA, Archival Act)
6. Certified Information System Auditor (CISA)
7. Systems Audit and Control Association (ISACA)

## ■ References

1. Protection of Documents and Archives and Archival Institutions Act /PDAAIA/ (Official Gazette of the RS, 30/2006)
2. REGULATION On Documents and Archives Protection (Official Gazette of the RS, 86/2006)
3. APPLICATION FOR CONFIRMATION OF EQUIPMENT AND DIGITAL STORAGE SERVICES PROVIDER
4. APPLICATION FOR ACCREDITATION OF EQUIPMENT AND DIGITAL STORAGE SERVICES PROVIDER
5. Regulation on Administrative Authorities within Ministries (Official Gazette of the RS, 58/2003, 45/2004, 86/2004, 138/2004, 52/2005, 82/2005, 17/2006, 76/2006 and 132/2006, 41/2007)
6. Regulation on administrative operations (Official Gazette of the RS, 20/2005)

7. Access to Public Information Act (Official Gazette of the RS, 24/2003)
8. The Rules on the Professional Requirements of Employees of Public Legal Entities and Employees of Services Providers Who Handle Records (Official Gazette of the RS, No. 132/2006).
9. COBIT 4.1 - Framework Control Objectives, Management Guidelines Maturity Model, IT Governance Institute, 2007
10. IT Governance Implementation Guide, 2<sup>nd</sup> Edition, IT Governance Institute, 2007
11. ISO/IEC FDIS 27001, Information technology - Security, ISO/IEC 2005
12. ISO 9001: 2008, Quality management systems - Requirements, ISO, 2008



## ■ Appendix 1: The structure of the tables

Requirement n° :	<i>XX - number</i>		<b>Assesment - Internal Rules :</b> (Filled in by SNA within the process of the IR approval)	
Legal bases	<i>legislation (PDAAIA, regulations, UTR)</i>		Adequate: <input type="checkbox"/>	
Area of implementation	<i>reference to the type of service to be implemented ()</i>		Not adequate: <input type="checkbox"/>	
Reference to Internal Rules	<i>Input from service provider, who needs to state</i> - Document title, version and date - page number of paragraph number		Not applicable: <input type="checkbox"/>	
Control activity importance (Impact with regard to the PDAAIA compliance)	<i>Critical</i> <i>Very Important</i> <i>Important</i>		<b>Finding :</b> (outcome of the accreditation review) Adequate: <input type="checkbox"/>	
Expected frequency of the control activity	<i>Annual</i> <i>Semi annual</i> <i>Quarterly</i> <i>Monthly</i> <i>Weekly</i> <i>Daily</i> <i>Very frequent</i>		Not adequate: <input type="checkbox"/>	
Expected type of the control activity	<i>Manual (no IT support)</i> <i>Automatic (application controls, software input controls, , etc.)</i> <i>Mixed (combination of manual and automatic controls)</i>		<b>Recommendation:</b>	
Relying on work of others	<i>E.g. ISO 9001, ISO 27001, ...</i>			
Review guidance	<b>Requirement</b>	<b>Type of the review activity</b>	<b>Summary of the review</b>	<b>Working documents</b>
	Requirement description, as stated in the legislation bases ( PDAAIA, regulations, UTR). Normally - citation from relevant legal source.	<i>Interview</i> <i>Documentation review</i> <i>Review of the software</i> <i>Posting test transactions</i> <i>Substantial testing</i> <i>Monitoring</i> <i>Evidence review</i>	Filled in within the review by the auditor, following the relevant international standards of auditing.  (test performed, reference to the relevant documentation, etc.)	<i>Filled in within the review.</i>
Compensation controls	<i>Filled in within the review.</i>			
Implemented control activity	<i>Filled in within the review.</i>			

# TWO WORDS, TWO CHALLENGES : DISTINGUISHING AUDIT AND CERTIFICATION OF DIGITAL ARCHIVES.

---

Dr. Seamus Ross

Professor of Humanities Informatics and Digital Curation, Université of Glasgow

Andrew McHugh

University of Glasgow

Hans Hofman

Consultant, Nationaal Archief of the Netherlands

Perla Innocenti

Co-Principal Investigator for Requirements Analysis and Identification of User Scenarios in the project  
Sustaining Heritage Access through Multivalent Archiving (SHAMAN)

Raivo Ruusalepp

Consultant, Estonian Business Archives Consultancy

*The recent trend in digital preservation to rely upon repository infrastructures to provide stewardship services has a number of varied implications, and presents a number of fundamental challenges. Digital repositories are a manifestation of complex organizational, financial, legal, technological, procedural, and political interrelationships. Innate uncertainties characterise each of these, exacerbated by a relative immaturity of understanding prevalent within the digital preservation domain. Criteria have emerged to help determine the services, functions, infrastructures and policies that must be evident within any 'trustworthy repository'. However, a number of critics have suggested that the available guidelines, such as the Trustworthy Repositories Audit and Certification (TRAC) Criteria and Checklist and the nestor Catalogue of Criteria for Trusted Repositories ([1], [2]), lack practical applicability. Confusion continues to surround the evidential burden for formally satisfying individual criteria; gaps persist between the criteria themselves and the ways and extent to which conformance can be adequately, and repeatedly demonstrated. These resources aim to reflect existing best practice in favour of introducing novel theoretical approaches, but practical and implementable meaning is far from implicit. Furthermore, the considerable diversity evident across the repository landscape (in terms of mandates, resource availability, scale, supported content and legal context) limits any attempts to legislate meaningfully on an objective basis. Although apparently intended to provide an intellectual foundation for both undertaking audit and awarding accreditation or certification, some shortcomings are evident.*

*Of critical importance is the synonymy with which the terms 'audit' and 'certification' appear to be perceived within the repository context. The process of certification is well served by documents such as TRAC and nestor. The conferment of a universally acknowledged recognition of success presupposes the availability of an objective benchmarking mechanism. A certification instrument that is overly prescriptive about specific practicalities will only prejudice its breadth of legitimacy. Comparability of results to is an essential quality to achieve an objective view of particular repositories. In contrast, the audit process, although an essential precursor to the award of certification, carries with it a completely different set of requirements. Neither the TRAC or nestor documents provide, in explicit or implicit terms, a sufficiently tangible structure for determining where conformity and success actually exist. Neither offers sufficiently detailed insights into the mechanics of the audit. Which individuals should be involved? What questions should be posed? How should experimental evaluation of systems be conducted? What are the quantitative or qualitative evidence expectations that will adequately demonstrate sufficient check-list compliance?*

*The Digital Repository Audit Method Based on Risk Assessment (DRAMBORA) is a methodology with audit at its very heart. It represents a bottom-up approach that takes risk and risk management as its principle means for determining repository success and charting improvement (see [3]). It encourages repositories to expose their organization, policies and infrastructures to rigorous scrutiny through a series of highly structured exercises, in order to build a comprehensive registry of their most pertinent risks, arranged in a manner that facilitates their effective management. DRAMBORA draws on experiences accumulated throughout six preceding and six evaluative pilot assessments undertaken in an internationally diverse selection of repositories and data centres (also, see [4], [5]). Available online, the DRAMBORA tool presents organizations with insights into risks and responsibilities faced by similarly profiled peers, in order to help ensure the*

*comprehensiveness of their own coverage, and provide a more explicit and tangible sense of how success is realized in various contexts. Emerging repository profiles provide a vehicle within which to present core roles, responsibilities, functions and risks for a variety of repository types. The availability of these instruments facilitates and legitimizes repository assessment and development.*

*This paper will describe the impact of DRAMBORA, as well as some example applications, for its use both in isolation and in combination with established certification metrics. It will address several of the main topics of the DLM Forum Conference, most notably Certification, Long Term Preservation and Strategies for Records Management and Archiving Services.*

[See the presentation.](#)

## Deux termes, deux difficultés : la distinction entre l'audit et la certification des archives électroniques

La tendance récente en matière de conservation du numérique qui consiste à s'appuyer sur des infrastructures de plate-forme pour fournir des services a plusieurs conséquences et présente plusieurs difficultés fondamentales. Les plates-formes d'archivage électronique impliquent des interactions complexes aux niveaux organisationnel, financier, juridique, technologique, procédural et politique. Chaque interaction s'accompagne d'incertitudes inhérentes, exacerbées par une compréhension relativement immature qui prédomine dans le domaine de la conservation du numérique. Des critères sont apparus pour permettre de déterminer les services, fonctions, infrastructures et politiques qui doivent être évidents au sein de toute « plate-forme accréditée ». Plusieurs critiques ont toutefois suggéré que les lignes directrices disponibles, telles que le Trustworthy Repositories Audit and Certification (TRAC) Criteria and Checklist et le NESTOR Catalogue of Criteria for Trusted Repositories, manquent d'applicabilité pratique. La confusion continue d'entourer les preuves à apporter pour répondre officiellement aux critères individuels, des lacunes persistent entre les critères eux-mêmes et les méthodes avec lesquelles la conformité peut être démontrée de façon adéquate et répétée. Ces ressources visent à refléter les meilleures pratiques pour introduire des approches théoriques inédites, mais une signification pratique et réalisable est loin d'être implicite. Par ailleurs, la diversité considérable qui illustre de façon évidente le paysage des entrepôts numériques (en termes de mandats, de disponibilité de ressources, d'échelle, de contenu et de contexte juridique) limite toute tentative de légiférer de manière substantielle sur une base subjective. Même si l'objectif évident est de fournir un fondement intellectuel pour réaliser des audits et attribuer une accréditation ou une certification, il existe incontestablement des lacunes.

La synonymie avec laquelle les termes « audit » et « certification » semble être perçus dans le domaine des plates-formes est d'une importance majeure. Le processus de certification est bien documenté par le TRAC et le catalogue NESTOR entre autres. Une certification universellement reconnue présuppose la disponibilité d'un mécanisme de référencement objectif. Un outil de certification trop contraignant en termes de modalités pratiques spécifiques ne fera qu'entraver la portée de sa légitimité. La comparabilité des résultats est essentielle pour parvenir à une vue objective des plates-formes particulières. Par opposition, le processus d'audit, bien que précurseur essentiel d'attribution de certification, s'accompagne d'un ensemble d'exigences totalement différentes. Ni le TRAC, ni les documents NESTOR ne fournissent, en termes explicites ou implicites, une structure suffisamment tangible permettant de déterminer à quels niveaux se situent réellement la conformité et les bonnes performances. Ils n'offrent pas non plus d'aperçu détaillé du mécanisme d'audit. Quelles personnes devraient participer ? Quelles questions devraient être posées ? De quelle façon l'évaluation expérimentale des systèmes devrait se dérouler ? Quels sont les éléments quantitatifs ou qualitatifs qui prouveront de façon adéquate une conformité à la liste de contrôle ?

DRAMBORA (Digital Repository Audit Method Based on Risk Assessment) est une méthodologie centrée sur l'audit. Elle représente une approche ascendante où le risque et la gestion du risque sont les principaux moyens permettant de déterminer les bonnes performances d'une plate-forme et d'établir une courbe des améliorations. Elle encourage les plates-formes à exposer leur organisation, leur politique et leur infrastructure à un examen rigoureux par le biais d'une série d'exercices très structurés afin d'établir un registre complet des risques les plus pertinents, organisés de façon à permettre leur gestion efficace. L'outil DRAMBORA tire parti des expériences accumulées durant six évaluations précédentes et six évaluations pilotes réalisées auprès de plusieurs systèmes et de centres de données différents sélectionnés à travers le monde. Disponible en ligne, l'outil DRAMBORA présente des organisations et un aperçu des risques et des responsabilités auxquels sont confrontés des chercheurs de profil identique afin de garantir la prise en compte exhaustive de leur propre couverture, et fournit une explication plus explicite et tangible sur les moyens d'atteindre de bonnes performances dans des contextes différents. Des profils de plate-forme émergents fournissent un moyen de présenter des rôles, des responsabilités, des fonctions et des risques essentiels pour plusieurs types de systèmes d'archivage. La disponibilité de ces instruments permet et légitime l'évaluation et le développement des plates-formes d'archivage.

Cet article présente l'impact de DRAMBORA et des exemples d'applications lorsqu'il est utilisé de façon isolée et combinée avec des mesures de certification définies. Il abordera plusieurs des principaux thèmes du DLM-forum et plus particulièrement la certification, la conservation à long terme et les stratégies de records management et les services d'archivage.

**Remerciements :** la recherche et le développement de DRAMBORA ont été possibles grâce au comité conjoint pour les systèmes d'information qui a accordé une subvention au Digital Curation Centre dans le cadre de la phase 1 de son programme et grâce à la Commission européenne dans le cadre du financement du Projet UE FP6 DigitalPreservationEurope (IST-2005-34762).

[Consulter la présentation.](#)



# **Séance 6**

# **Stratégies pour de nouvelles ressources et architectures de l'information**

---

---

Sous la présidence de Franck BRADY

superviseur du système de gestion des documents électroniques de la Commission européenne

# ***Session 6***

## ***Strategies for records management and archiving services***

---

---

*Chairing Franck BRADY*

*supervisor of the electronic records management system of the European Commission*

# CONNECTING THE DOTS. LEVERAGING STANDARDS AND TECHNOLOGY TO ACHIEVE AUTOMATION OF RECORDS MANAGEMENT PROCESSES.

---

Conni Christensen

Founding Partner, Director of Synercon Management Consulting

*In the past decade the records management industry has struggled to come to terms with the unprecedented volumes of electronic data, records and information, in an era when privacy, retention and security requirements have become stringent.*

*Traditional models have seen records management processes undertaken by records management staff, but clearly the manual application of access and security rules and retention policies is not keeping pace with the volumes.*

*On the other hand, transferring the work to end users is not proving successful either. End-users are highly resistant to capturing metadata which does not relate directly to their own business processes.*

*Automation of records management processing appears to be the only viable solution. Only through automation can we take the end-user out of the equation and make the system to do the work.*

*Software developers are responding to the challenge and are building a range of solutions to automate RM processes.*

*Yet despite significant investment, these solutions are yet to deliver the accuracy and consistency required to meet records management compliance objectives.*

*What would it take to make them good enough?*

*In 2002, the International Standard for Records Management (ISO 15489) introduced the DIRKS methodology for the development of records classification schemes and retention schedules. DIRKS (Designing and Implementing Recordkeeping Systems) is based on traditional system design methodologies which have been adapted for the records management environment.*

*The step-by-step analysis in DIRKS identifies and documents an organisation's functions, activities (processes), tasks, documents and records, recordkeeping requirements (risk and value) based on analysis of documentary sources, and stakeholder interests.*

*In many ways the DIRKS outputs can be likened to the DNA of an organization. DNA is the molecule that encodes genetic information in the nucleus of cells and determines the structure, function, and behavior of cells. Similarly DIRKS describes how an organization is structured along functional lines.*

*The DIRKS outputs form the basis of a framework for records management process automation.*

*Every term, task, record in DIRKS has a relationship which can be leveraged to support automation. By building these known relationships into records management solutions, effort by the end-user can be significantly reduced, even eliminated. Furthermore the DIRKS framework can be further extended to incorporate the relationships with other taxonomy facets including roles and subjects / domains.*

*This session presents detailed research undertaken by Synercon on the DIRKS methodology in examining how the DIRKS methodology can significantly assist automation of records management processes including:*

- *Pre-classification of records*
- *Auto-classification of records*
- *System capture (tagging) of metadata*
- *Metadata Filtering*
- *Auto-assignment of lifecycle policies*

*It concludes that current standards - ISO 15489 Standard for Records Management and ISO 23081 Standard for Records Management Metadata - provide suitable reference models for developing the metadata frameworks for automation.*

[See the presentation.](#)



## **Décloisonnement et cohérence. Mobiliser les normes et la technologie pour automatiser les processus de « records management »**

*Ces dix dernières années, le secteur du records management a dû consentir des efforts importants pour pouvoir traiter des volumes sans précédent de données, documents et informations électroniques alors que les exigences de confidentialité, de conservation et de sécurité des données se sont durcies.*

*Dans les modèles traditionnels, les tâches de records management ont été effectuées par le personnel chargé du records management, mais évidemment sans que la mise en œuvre des règles d'accès et de sécurité ainsi que des politiques de conservation en mode manuel ne permette de traiter tout le volume d'informations.*

*Confier cette tâche aux utilisateurs finaux n'est pas non plus une solution satisfaisante, ceux-ci faisant preuve d'une grande résistance par rapport à la capture de métadonnées sans lien direct avec leurs propres processus métier.*

*L'automatisation de la gestion documentaire semble être l'unique solution viable. En effet, seule l'automatisation permet de placer l'utilisateur final en dehors de l'équation et de faire faire le travail par le système.*

*Les développeurs de logiciels s'attaquent actuellement au défi et sont en train d'élaborer une gamme de solutions destinées à automatiser les processus de gestion documentaire.*

*Pourtant, malgré l'importance des investissements, ces solutions ne permettent pas encore d'obtenir la précision et la cohérence permettant de satisfaire aux objectifs de conformité en matière de records management.*

*Que faudrait-il pour les améliorer?*

*En 2002, la Norme internationale de records management ISO 15489 a introduit la méthode DIRKS, dédiée au développement de schémas de classification et de plans de conservation de documents. La méthode DIRKS (Designing and Implementing Recordkeeping Systems) est basée sur les méthodes de conception de systèmes traditionnelles que l'on a adaptées à l'environnement du records management. L'analyse en plusieurs étapes de la méthode DIRKS permet d'identifier et de documenter toutes les fonctions, activités (processus), tâches, documents et dossiers d'une organisation, les exigences (risque et coût) de conservation à partir de l'analyse des sources documentaires, ainsi que les intérêts des différents acteurs.*

*On peut, à bien des égards, comparer les produits DIRKS à l'ADN d'un organisme. L'ADN est la molécule qui codifie les informations génétiques dans le noyau des cellules et qui détermine la structure, la fonction et le comportement de ces cellules. De la même façon, la méthode DIRKS décrit la façon dont une organisation s'articule autour de lignes fonctionnelles. Les produits DIRKS permettent l'élaboration d'un cadre de référence pour l'automatisation des processus de records management.*

*Chacun des termes, chacune des tâches, chacun des documents de DIRKS a un lien qui peut être exploité pour soutenir l'automatisation. Grâce à l'intégration de ces relations connues dans les solutions de records management, la tâche de l'utilisateur final peut être grandement réduite, voire même supprimée. Qui plus est, le cadre de référence de la méthode DIRKS peut être par la suite élargi afin d'inclure les relations avec d'autres éléments de taxinomie, notamment des rôles et des sujets/des domaines.*

*Cette session présente les résultats de l'étude détaillée qu'a menée Synercon sur la méthodologie DIRKS qui a analysé la façon dont la méthodologie DIRKS s'avère une aide précieuse dans l'automatisation des processus de records management dont :*

- *le pré-classement des documents*
- *l'auto-classement des documents*
- *la capture (étiquetage) des métadonnées par le système*
- *le filtrage des métadonnées*
- *l'auto-attribution de principes de cycle de vie.*

*Elle arrive à la conclusion que les normes actuelles (ISO 15489 : norme de records management et ISO 23081 : norme des métadonnées de records management) fournissent des modèles de référence adéquats pour développer les cadres de métadonnées nécessaires à l'automatisation des processus de records management.*

[Consulter la présentation.](#)

## ■ Introduction

Seldom does a day go past without a records management incident in the news. Many have increasingly familiar themes:

- The inability to find records which exist within the system;
- The destruction of records which should have been retained, and the retention of records which should have been destroyed;
- The public disclosure of private records;
- The inability to discover records required in litigation.

Most organizations we know are struggling to come to terms with unprecedented volumes of electronic information - and information privacy, security and records compliance are high risk issues for management and government.

Records managers who have been traditionally responsible for core records management processes (classifying and appraising material) are no longer able to manage such vast amounts of material, but transferring records management tasks to end users has proved difficult. End-users are highly resistant to taking up tasks outside of their domain and which don't directly relate to their own business processes.

Is the automation of records management processes a viable option?

## ■ Automation – a silver bullet for records managers?

Software developers are responding to the challenge and auto-classification and automated disposition management options are emerging in content management systems.

Auto-classification is a process by which document content is classified by analyzing the document text and assigning predefined classes using taxonomies and defined semantic rules. Text analysis identifies concepts using techniques such as word stemming, redirection of stop words, and elimination of prepositions and articles. The developed taxonomies are tested against documents and adjusted to improve consistency and accuracy. Semantic alternatives are built into a thesaurus linking preferred terms with synonyms and non preferred terms.

Automated disposition management is appraisal is a process by which documents are assigned retention rules based on a predetermined classification or on recognition of specific metadata elements.

While many of these products are technically clever, there are still no ground breaking success stories. My theory for this is that developers of auto-classification tools have focused on subject based taxonomies (as used in libraries) to assist with search and retrieval. However records classification schemes and disposition schedules tend to be business activity based, using terminology that describe the context in which records were created.

For example, if the subjects of a document are Racehorses and Equine Influenza (EI), how can we tell how long to keep the document? Clues to retention come from a range of sources including the context in which the document was created e.g. research into the disease, monitoring horse movements, testing for EI or treating horses with EI.

To automate records management processes we need to understand how build and apply taxonomies that are linked to disposition management.

Lets take a closer look at how classification and disposition management models fit together.

## ■ Records Classification Models

Classification is the systematic arrangement of material into groups or categories (taxonomies) according to established criteria.

The criteria used depend on the intended purpose. For example, classification schemes in web sites are designed for the purpose of **arranging** and **finding** web content.

In records and information management (RIM) systems classification schemes are designed to achieve a wide range of purposes including:

- Aggregation of content (grouping records together for business or evidential purposes)
- Arrangement for browsing and navigation
- Access & security management
- Disposition management
- Filtering and segmentation of records
- Managing provenance
- Reporting
- Common criteria used in RIM classification schemes include:
  - Business activity i.e. functions, activities, processes, transactions
  - Subject matter i.e. products, services, disciplines, domains, assets
  - People and organisations
  - Case based aggregates i.e. accidents, matters, claims
  - Record or document types
  - Geographic locations
  - Time frames or sequences

No one single criterion is able to realise all of the collective classification needs of the organisation.

The following matrix illustrates how different classification criteria (or facets) meet different needs. Some meet the aggregation and retrieval needs of business users, others meet the compliance needs of the organisation.

	Function	Activity	Subject	Case	Record type	Business Unit	Roles	Location
Aggregation	✓	✓	✓✓✓	✓✓✓	✓✓	✓	✓	✓✓
Discovery		✓	✓✓✓	✓✓✓	✓✓		✓	
Filtering	✓✓✓	✓	✓	✓	✓	✓✓✓	✓	✓✓
Reporting	✓	✓	✓✓	✓✓	✓	✓✓	✓	✓
Access & Security	✓✓✓	✓✓	✓✓	✓		✓✓	✓	
Disposition	✓✓✓	✓✓✓	✓✓		✓✓✓		✓	

Source: Synercon Management Consulting 2006

The table illustrates that classification schemes may need to include a broad range of criteria to meet the collective needs of the organisation.

## ■ Metadata Models

The classification facets described above are, in fact, a subset of recordkeeping metadata - structured information that describes and/or allows users to find, manage, control, understand or preserve records over time.

As many government jurisdictions have imposed mandatory metadata requirements, the taxonomy developer needs to understand how classification facets within in the taxonomy fit into to the metadata model and what relationships between taxonomy terms and other metadata must be acknowledged i.e. the relationship between classification facets and disposition and/or access rules.

## ■ Disposition Management Models

Disposition management involves a range of processes associated with implementing records retention, destruction or transfer decisions which are documented in disposition authorities and schedules, or other instruments.

The development of disposition schedules involves identifying classes of records with similar recordkeeping requirements. The rationale by which classes are organised into schedules is varied and include function based, subject based, class based and series based arrangements.

Retention periods for classes are derived from an analysis of the business, accountability and community requirements. While the grouping of records into classes may appear subjective, recordkeeping requirements are generally linked to recognisable classification criteria as the following example illustrates.

BIOSECURITY (Function), ENFORCEMENT (Activity)

2.4.1 Records relating to the enforcement of movement controls as a result of an outbreak of disease or infestation of pests at a state or nation wide level. Includes orders for detention notices, orders for destruction, pest control orders, maps of control areas and other supporting documentation.

**Required as State Archives**

2.4.2 Records relating to seizure or impounding of plants or animals, of plant or animal products, fodder or fittings, or of vehicles or equipment in connection with the control of pests and diseases. Includes seizure documentation, inventory and reports.

**Retain minimum of 10 years after last action, then destroy**

2.4.3 Records relating to orders given for controlling the spread of identified diseases, pests, weeds, and non-indigenous animals at a local level.

**Retain minimum of 10 years after last action, then destroy**

Source: Records Disposal Authority (FA258) NSW State Records Web Site  
<http://www.records.nsw.gov.au/recordkeeping/docs/fa0258.pdf>

In the first and third class the guide to retention value is the **subject** of, and the **business activity**. It is the type and scale of the activity which determines retention.

In the second class it is the **activity** only that determines the retention.

## ■ DIRKS Methodology

The above records classes were developed in accordance with the **DIRKS** (also known as **DIRS**) **Methodology**, outlined in the **Standard for Records Management ISO 15489**.

**Classification** is defined in the Standard as the systematic identification and arrangement of business activities and/or records into categories according to logically structured conventions, methods, and procedural rules represented in a classification system.

These “logically structured conventions, methods and procedural rules” are expressed in the DIRKS (Designing and Implementing Recordkeeping Systems) methodology. DIRKS is based on traditional system design methodologies which have been adapted for the records management environment.

DIRKS steps A-C define the methodology for developing records classification and disposition instruments. The analysis identifies and documents an organisation's **functions, activities (processes), tasks, documents and records, recordkeeping requirements** (risk and value) based on analysis of documentary **sources, and stakeholder interests**. A representation of the DIRKS Framework shows how these entities are connected together.

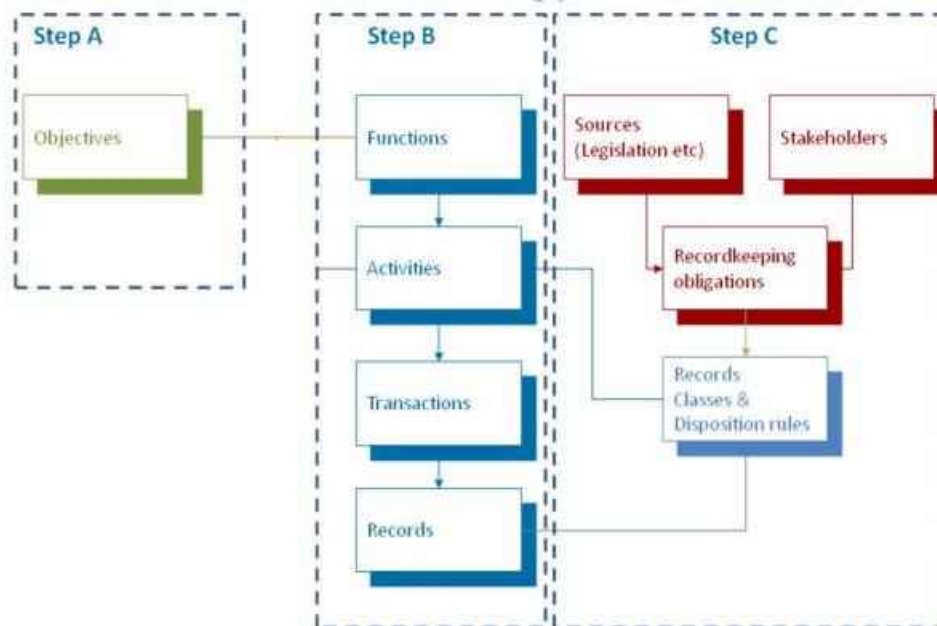


Figure 1 DIRKS A-C Framework

Source: Synercon Management Consulting 2008. Adapted from National Archives of Australia DIRKS Manual

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

In many ways the DIRKS outputs can be likened to the DNA of an organization. DNA is the molecule that encodes genetic information in the nucleus of cells and determines the structure, function, and behavior of cells. Similarly DIRKS describes how an organization is structured from the broadest functions to the minutest detail of business activity.

The DIRKS framework can be extended to incorporate the relationships with other taxonomy facets including roles and subjects/domains as illustrated.

## DIRKS Methodology

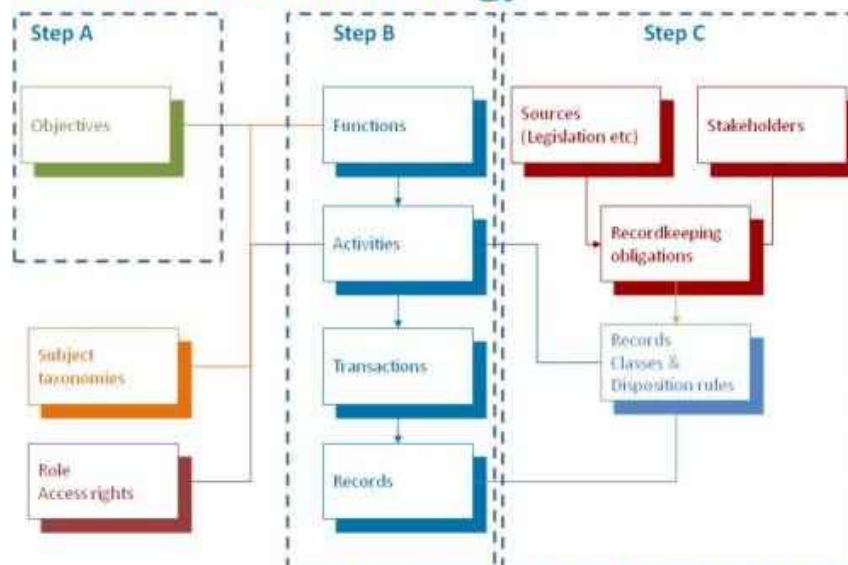


Figure 2 - Extended DIRKS Framework

Source: Synercon Management Consulting 2008. Adapted from National Archives of Australia DIRKS Manual

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

Outputs and instruments derived from of the DIRKS analysis include:

- The Business Activity Classification (BACS) scheme comprising Functions and Activities, and in some jurisdictions, transactions.
- File Plans comprising Functions, Activities, Subjects
- Keyword Thesauri, comprising preferred terms of Functions, Activities, Subjects, and non preferred terms
- The Disposition Schedule, comprising Records Classes linked to the BACS
- Security Classification Hierarchy, comprising Roles linked to the BACS
- Workflows and process maps comprising Transactions and Records linked to the BCS.

## ■ The role of the BACS

From the DIRKS Framework diagram you can see that the **Functions** and **Activities** in the BACS are connected (directly or indirectly) to all other facets within the framework. Thus the BACS provides the axis to which other taxonomies, workflow definitions and retention schedules can be linked together.

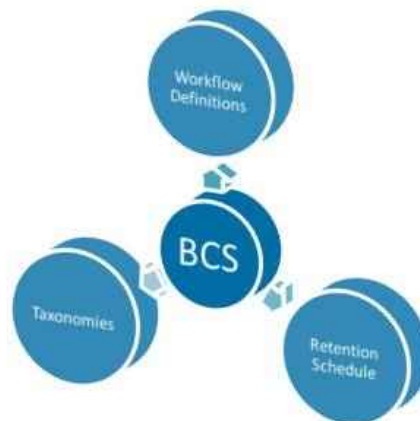


Figure 3 - Role of the BACS

Source: Synercon Management Consulting 2006

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

Given the central role of the BACS, how can we best use it?

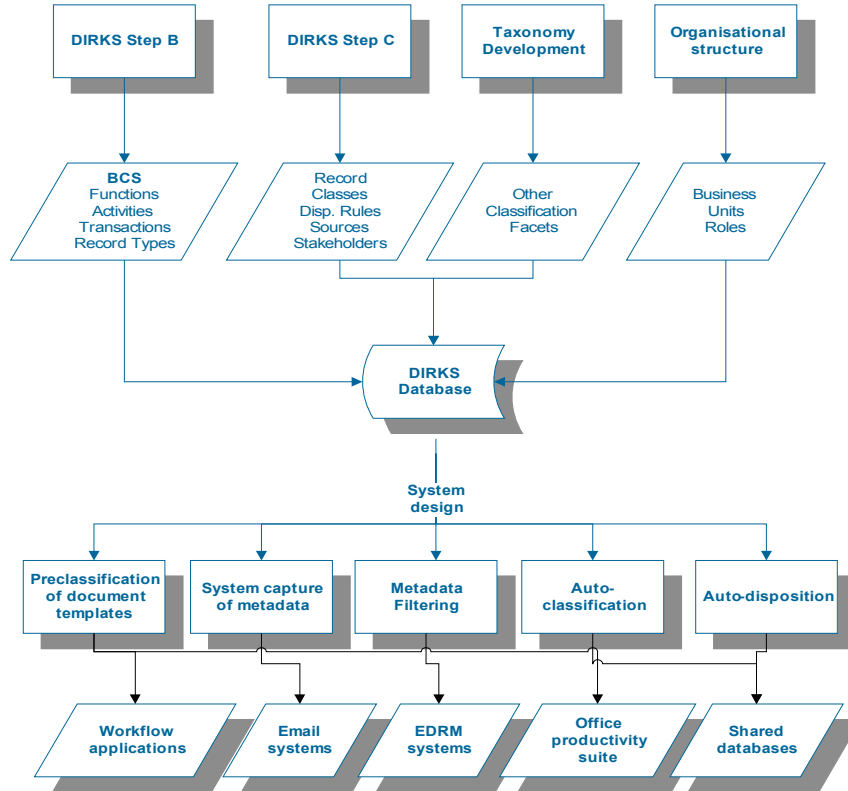
## ■ Opportunities for automation

Effective automation demands consistent standards for description, ideally ones which are universally accepted. DIRKS provides a framework based on industry standards to describe records and records management processes. Technology provides the enabling tools to effect this.

To make effective use of the DIRKS outputs we need to slice and dice the data - to systematically break down the information into smaller parts or views that yield more information about the individual facets and the links between them. In this way we can use the information in a variety of different and useful ways.

Technology, in the form of relational databases, makes this possible.

The following diagram illustrates the development process, gathering data through a range of analysis and development processes to identify and document recordkeeping facets within a relational (DIRKS) database. The database delivers records metadata to a range of target systems which automate recordkeeping processes according to the design opportunities/constraints of these systems.

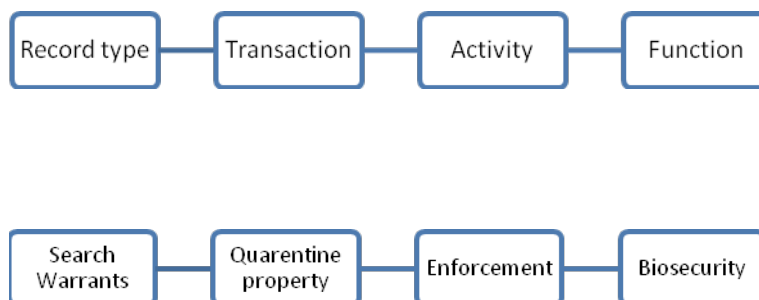


Source: Synercon Management Consulting 2008

This framework delivers a range of different automation options, described here in more detail.

## ■ Pre-classification of records

Records are the inputs and outputs of business activity (workflow), and the DIRKS analysis process identifies where different types of records are located within the framework. Templates for record types can be informed (pre-classified) with the Function/Activity terms.





Pre-classification is achievable with well defined work processes, i.e. acquisition, accreditation, authorization, construction, design, disposal evaluation etc. It is already well established in workflow systems where document templates are linked to business processes. The DIRKS framework makes it possible to extend the same techniques in other RIM systems.

## ■ System capture of metadata (automated meta-tagging)

System capture of metadata makes use of known relationships reducing the effort of data capture for the end user.

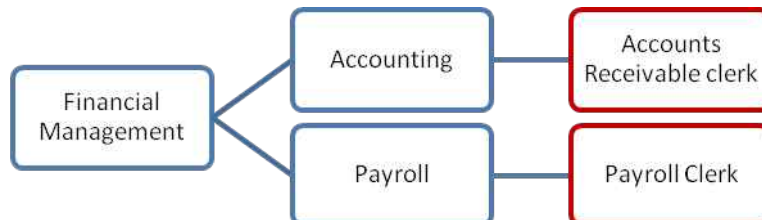
System capture of metadata depends on the type of relationship within the metadata model. Note: Relationship structures will vary considerably according to function.

- One- to-one - enables inheritance in both directions between levels
- One-to-many - enables upward inheritance
- Many-to-one - enables downward inheritance
- Many-to-many - doesn't support inheritance

### Examples

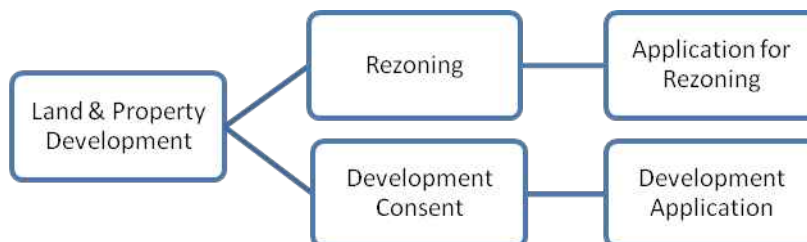
Function can be derived from the role of the user or the business unit to which they belong or conversely role and business unit can be derived from the function, activity, and /or record type.

E.g. a **Payroll Clerk** works within the **Financial Management** Function and within the **Payroll** Activity.



Function and Activity can also be derived from Record Types and conversely Record Types derived from the Function/Activity/Transaction path.

E.g. an **Application for Rezoning** can only be created within the function **Land and Property Development**, activity **Rezoning** and transaction **Application for Rezoning**.



## ■ Metadata Filtering

One of the major issues with current EDRM system design is the issue of presenting too many choices to users i.e. for classification or retention. The presentation of too many choices or irrelevant choices leads to inconsistent classification and poor metadata quality.

Metadata filtering is a technique that can be used when one-to-one inheritance isn't possible but the number of choices is reduced due to known relationships. Browsing is a form of metadata filtering but intelligent design should eliminate the need for the user to have to always navigate from the top of the tree if terms lower down the tree have already been captured. For example:

### Filtering of record classes

The capture of Function and Activity terms as metadata (whether as the file title or the folder path) should narrow the presentation of retention classes to only those related to the Function/Activity pair. Unfortunately this does not happen in most EDRM systems.

### Filtering of functions and activities

Through the relationship of Role and Business Unit to Function, the presentation of the Functions and activities in the Taxonomy can be restricted to those relevant to the role during data capture. In this way the quality of classification can be improved. The full set is made available again when searching.

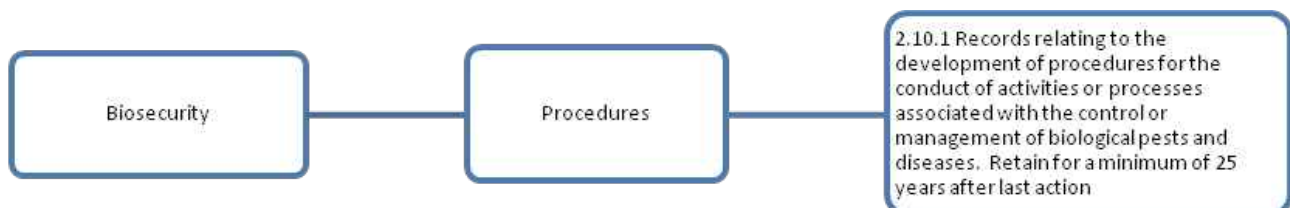
### Filtering of subject taxonomies

Many subject taxonomies are related to a particular discipline and often specific to one or few functions, i.e. a taxonomy of insect pests is specific to Bio-security but to Technology. In the library world this is known as Keyword in Context. The presentation of subjects should be restricted to the relevant functions and activities.

## ■ Auto-assignment of lifecycle policies

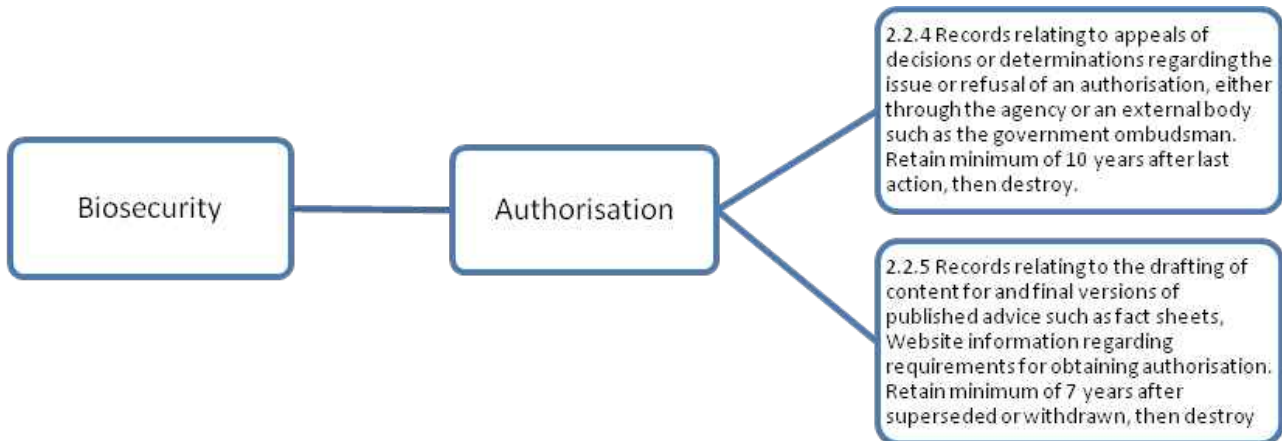
Records classes are grouping of records with the same disposition requirements.

Where there is a one-to-one relationship, the Record class can be derived from the function/activity path e.g:



Source : Records Disposal Authority (FA258) NSW State Records Web Site

However in practice records classes are usually arranged in many-to-one relationships, as illustrated.



Source: Records Disposal Authority (FA258) NSW State Records Web Site.

Records classes are differentiated by a range of record attributes such as:

- Subject (i.e. significance of the event, whether national or state),
- Record Type (i.e. applications, appeals, orders, permits)
- Role of the persons involved in the events (i.e. investigator, ombudsman)
- Status of the documents (draft or final, approved).

Relational database technology now makes it possible to document individual attributes and assign retention values for each component. The DIRKS framework provides the means by retention values can be assessed according to scales of business need, accountability and community value.

## ■ The way forward for automation

At this point in time, no solution exists that is able to automate completely, all aspects of records management

We are however on the threshold of automation.

The DIRKS framework provides us with the tools to create taxonomies and metadata models which will support automation. The identification of recordkeeping facets and the relationships between them can be leveraged to automate a range of records management processes.

Text analysis techniques can be expanded to include recordkeeping facets and using thesaurus relationships link semantic and synonyms to preferred terms in the framework. Concepts derived from text analysis can be mapped against identified terms, record types, and roles to establish the context of records as well as the subject matter.

Through the intelligent design of recordkeeping systems, this information can be utilised in such a way as to establish the precise context of the records within the DIRKS framework. Manual capture of metadata is eliminated and user involvement confined to selecting templates or confirming subject terminology.

“Let the computer do the work”.

## ■ Glossary of terms used

### Activity

Activities are the major tasks performed by an organisation to accomplish each of its functions.

### Appraisal

The process of evaluating business activities to determine which records need to be captured and how long the records need to be kept to meet business needs, the requirements of organisational accountability and community expectations.

Source: *Adapted from Standards Australia AS 4390, Part 1, Clause 4.3.*

### Auto classification

Auto Classification programs attempt to classify content by analysing their text and then assigning them to a pre-defined class.

Source: <http://www.cmscalendar.com/cmspros-glossary.html>

### Classification

The systematic identification and arrangement of business activities and/or records into categories according to logically structured conventions, methods, and procedural rules represented in a classification system.

Source: *ISO 15489 Standard for Records Management*

### Relational Database

A method of structuring data in the form of records so that relations between different entities and attributes can be used for data access and transformation.

Source: <http://www.extension.umn.edu/distribution/naturalresources/components/DD6097hr.html>

### Declaration

The act of assigning to a captured document the status of a 'record'.

Source: <http://www.curaconsortium.co.uk/glossary.html>

### DIRKS

The acronym for 'designing and implementing recordkeeping systems', a methodology for managing records and other business information that is outlined in the Australian Standard for Records Management (Standards Australia AS ISO 15489, Part 1, Section 8.4).

### Disposition

A range of processes associated with implementing records retention, destruction or transfer decisions which are documented in disposition authorities and schedules, or other instruments.

Source: *ISO 15489 Standard for Records Management*

### Disposition Schedule

A comprehensive instruction covering the disposition of records to assure that they are retained for as long as necessary based on their administrative, fiscal, legal and historic value.

### Faceted classification

Allows the assignment of multiple classifications to an object, allowing searching and browsing of related information through several classes. Elements may include subject, geographical, temporal and form of an item.

Source: [www.Wikipedia.com](http://www.Wikipedia.com)

## Function

Functions represent the major responsibilities that are managed by the organisation to fulfil its goals. They are high-level aggregates of the organisation's activities.

Source: *National Archives of Australia*

## Metadata

1. Data describing context, content and structure of records and their management through time.

2. Structured information that describes and/or allows users to find, manage, control, understand or preserve other information over time.

Source: *ISO 15489 Standard for Records Management*

## Records Class

The basic unit of a disposition schedule.

A Records (Disposal) Class is a component of a disposal authority, implemented within an electronic records management system as a set of rules made up of a disposal trigger, a retention period and a disposition action, which may be applied to a record plan entity.

Source: *National Archives of Australia*

## Record series

A group of identical or related records that are normally used and filed as a unit and that permit evaluation as a unit for retention scheduling purposes.

Source: [www.library.utoronto.ca/utarms/info/glossary.html](http://www.library.utoronto.ca/utarms/info/glossary.html)

## Sentencing

The process of identifying the records class a record belongs to and applying the disposition action specified in the relevant disposition authority to the record. Sentencing is the implementation of decisions made during appraisal.

Source: *National Archives of Australia*

## Thesaurus

Classification tool comprising an alphabetical presentation of a controlled list of terms linked together by semantic, hierarchical, associative or equivalence relationships.

Source: *National Archives of Australia*

## Transaction

The smallest unit of business activity. Uses of records are themselves transactions.

Source: *National Archives of Australia*

## Unstructured Information

Unstructured data (or unstructured information) refers to (usually) computerized information that either does not have a data structure or has one that is not easily usable by a computer program. The term distinguishes such information from data stored in fielded form in databases or annotated (semantically tagged) in documents.

Source: [www.wikipedia.com](http://www.wikipedia.com)

# CREATING AN INTERMEDIATE ARCHIVE FOR SEMI-ACTIVE ELECTRONIC RECORDS - CONCEPTS AND CHALLENGES

---

Michael Hollman

Director of the modern archives of the Federal Archives (Bundesarchiv, Koblenz)

*In october 2008 the Bundesarchiv will open its new Digital Archives for records from the German Federal administration. After two years of testing and designing the first "real records" from electronic records management systems will be transferred to the new system. The core element of the system is SAM, which stands for "Standardised archiving module". It is a flexible ingest tool which allows to process data from various ERDMS without a standardised data exchange schema. In a first step the retention period of the records accepted by the federal archives will have to be expired. Agencies will be able to dispose of there records in an secure and efficient way. But this can only be a first step, the next step will be to set up an intermediate digital archive for semi-active records to offer a real service for the federal administration.*

*The benefit for the administration will be that they do not have to set up an infrastructure for longterm preservation on their own and for the Bundesarchiv that it keeps and strengthens its position within the adminstration.*

*New partners and political support are needed to realise this aim. The paper will present the technical concept of the existing system and its necessary extensions (e.g. secure online-access, preservation of digital signatures), the challenges and the political and strategical implications of the plan.*

[See the presentation.](#)

## **Création d'un niveau d'archivage intermédiaire pour les documents électroniques semi-actifs : concepts et défis**

*En octobre 2008, le Bundesarchiv ouvrira son nouveau centre d'archives numériques destiné aux documents de l'Administration fédérale allemande. Après deux années de tests et de développement, les premiers « vrais » documents des systèmes d'archivage électronique vont être transférés sur le nouveau système. L'élément central de ce système est SAM, un module d'archivage normalisé (Standardised archiving module). C'est un outil flexible d'intégration de données qui permet de traiter des informations issues de différents systèmes de gestion et d'archivage électronique sans schéma d'échange normalisé. Dans un premier temps, les Archives fédérales n'accepteront que des documents ayant atteint leur date d'expiration de stockage et les administrations publiques pourront disposer de leurs documents en toute sécurité et efficacité. Mais il ne s'agit que d'une première étape, la seconde consistant à créer un niveau d'archivage électronique intermédiaire pour les documents semi-actifs afin d'offrir un réel service à l'Administration fédérale.*

*L'avantage pour l'administration est qu'elle-même ne sera pas obligée de mettre en place d'infrastructure dédiée à la conservation à long terme et l'avantage pour le Bundesarchiv est qu'il conserve et conforte sa position au sein de l'administration.*

*De nouveaux partenaires et un soutien politique sont indispensables pour mener à bien cet objectif. Cet article présentera le principe technique du système existant et les aménagements nécessaires (tels qu'accès en ligne sécurisé ou conservation des signatures numériques, par exemple) ainsi que les défis et les implications politiques et stratégiques de ce programme.*

[Consulter la présentation.](#)

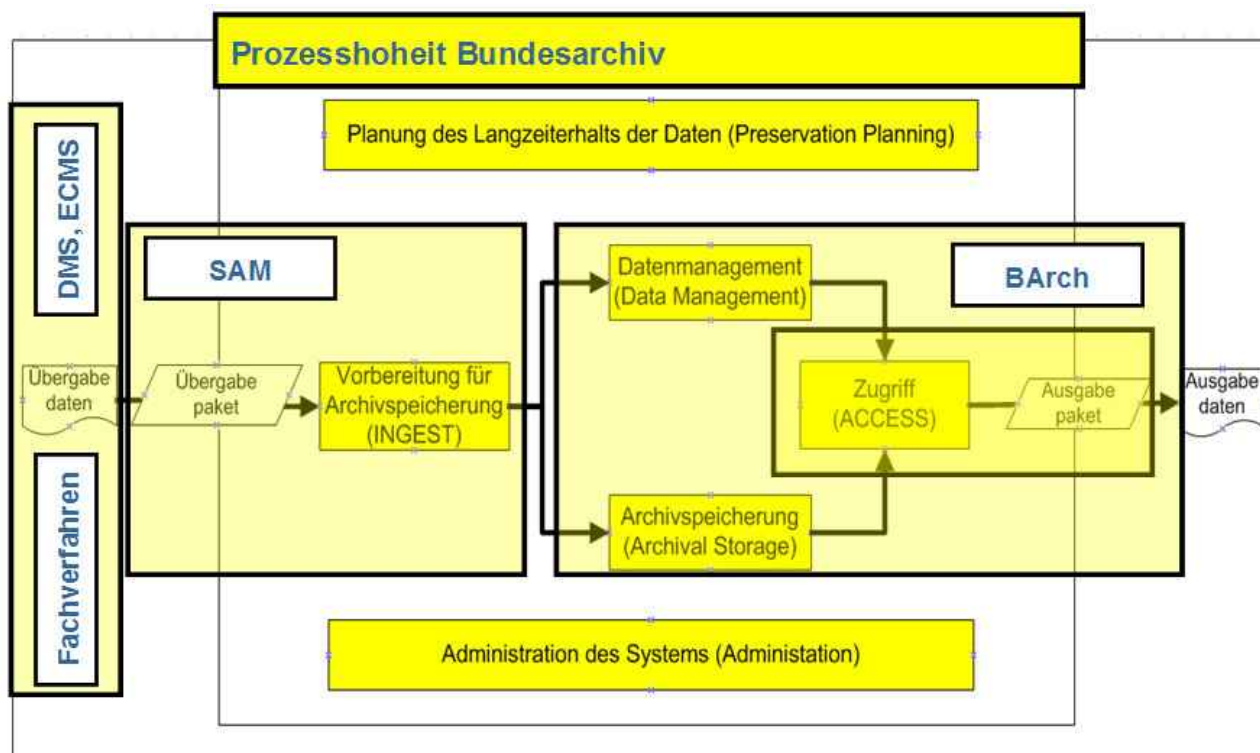


Fig.1: Process modelization of the „Digitales Archiv“ project of the Bundesarchiv<sup>1</sup>

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

## ■ Introduction

It is one of its most important tasks of the Bundesarchiv to save the records of the federal government and its predecessors since 1871 as well as the records of the administration of the German Empire, the German Democratic Republic from 1949 to 1990 and the Federal Republic of Germany since 1949, to appraise their archival value, to describe them and - the final goal of it all - to make it accessible to public use. That is the content of §§ 1 and 2 of the Federal Archives Act from 1988. The Bundesarchiv and especially its Department B - that means Bundesrepublik Deutschland - comply with that legal request in an experienced manner. The Department B, which is responsible for the records of the government and top level administration of the Federal Republic, has developed special methods to ensure the takeover of the records of the named administrations. Especially the so called Zwischenarchiv, the Intermediate Archives, is to mention. Also for the archival appraisal of the records, the databased description and the utilisation we have approved and partly formalised processes to deal with the annual inflow of round about 5 to 6 kilometer of records.

The main goal is to make archival material - historical sources - accessible for the purposes of the historical science, of the needs of private persons, that means everybody and not only German citizens, and of the needs of the state in the area of legislation, jurisdiction and administration. The Bundesarchiv as long term memory of the state.

Records in the sense of the Federal Archives Act are not only paperbased records but also maps, moved and still pictures and - of course - electronic, that means genuin digital records. They become important more and more; the often used keyword is e-government. So in a not to far futur the Bundesarchiv will have to deal with huge amounts of digital record submitted to the archives.

In the course of the German reunification the Bundesarchiv had to acquire several huge databases with personnel data collected by agencies of the former German Democratic Republik, for example the so called Social Working Funds with detailed data of round about 17 million working people in the GDR or the "Auslandskader"-database with data of people who were allowed to travel abroad. These GDR-databases are

<sup>1</sup> These schemes were taken from „Entwicklungsstudie elektronische Zwischenarchive“ published in 2008 on behalf of the Bundesarchiv in partnership with the society INFORA GmbH.



very important for the calculation of pension claims. Although the Bundesarchiv kept these data without their former technical environment and often without any sufficient documentation the Bundesarchiv succeeded in saving the data and making the files readable.

The Bundesarchiv also keeps digital photographs, movies or sound files since several years. Partly special platforms have been developed to make these sources available via internet. Especially I advise to you the Digital Still Pictures Archives under [www.bild.bundesarchiv.de](http://www.bild.bundesarchiv.de).

Until now the Bundesarchiv - according to the traditional role of archives in general - was primarily reactive which means that usually the Bundesarchiv considered how to deal with digital records not until they had been delivered to the archives. Now this reactive attitude must be given up.

## ■ The challenge

In a not to far future a greater couple of ministries and agencies will move their internal communication completely from paper based records to electronic records. They will start with selected departments, but the goal is to extend the e-government concept as far as possible. Middle-term ERMS (“Vorgangsbearbeitungssysteme”) and document management systems will be established wherever possible. In fact paper will not vanish completely, but a paper print will just be a form of appearance of an electronic document which will be the legally binding version.

It corresponds to the German administrative tradition that every single ministry and most of the agencies are allowed to have their own processes, practices and systems. The Bundesarchiv doesn't have the authority to give guidelines which have to be followed. So for the moment the evident characteristic of German e-government is to describe by the items heterogeneity and variety of systems and formats. Under these conditions it is particularly difficult to comply with the legal charge of archiving the digital records of the federal government.

Because of these reasons the Bundesarchiv concentrated in the first step on digital records which emerged from ERMS and document management systems. After several years of planning and after the successful realisation of a pilotsystem in autumn 2006 the Bundesarchiv realised its digital Archives in cooperation with Hewlett Packard and SER started it on last October 22.

## ■ The Digital Archives – Concept, launching and future developments

The concept of the digital archives is strictly orientated to the Open Archival Information Standard (OAIS - ISO 14721:2003). The problems of heterogeneity, complexity and perishability are solved by using SAM, the so called StandardArchivierungsModul, Standard Archiving Module. SAM differentiates between metadata and primary data. During the process of offering and transfer SAM will map the XML-metadata of the offering agency onto the standardised metadata-schema of the Bundesarchiv, the XBARCH. The primary data will be harmonised in structure and format, various file formats will be converted automatically to a narrow set of accepted formats. The whole process will be controlled by using agency profiles containing the XML-schema mapping and process information.

In the course of the acquisition of the meta and primary data - in the terms of OAIS: the Submission Information Package SIP - the file formats will be checked and if necessary converted into formats which are applicable for longterm preservation, that means especially into PDF/A. This process step will run automatically using a conversion platform which is integrated in SAM. These conversion processes will be logged.

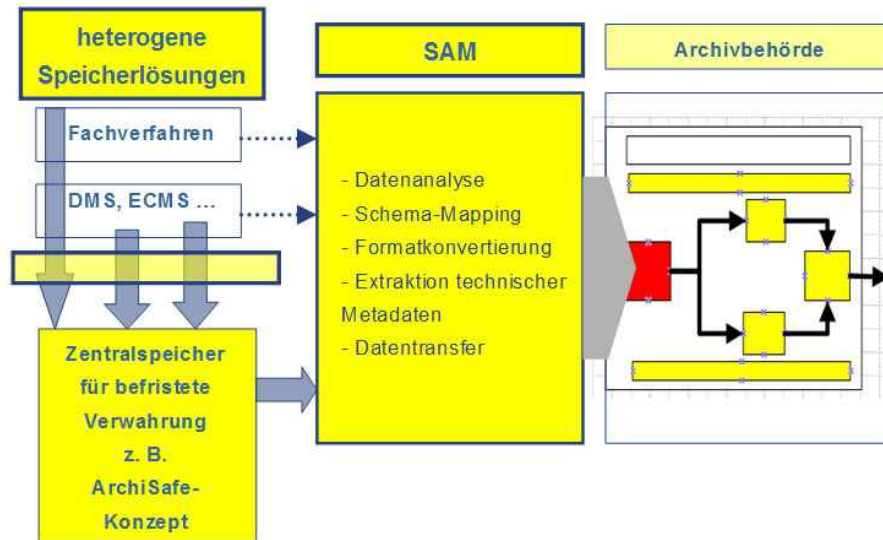


Fig. 2 : the SAM Model

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

A hierarchical storage system designed by Hewlett Packard will store and save both meta and primary data as Archival Information Package (AIP). The HSM will also administrate the technical metadata. Metadata which are relevant for the archival description will be transferred to the archival management system of the Bundesarchiv called Basys via a XML-interface. Via Basys the description data can be researched, the AIP can be uploaded and read by a special viewer.

After this first stage of extension the next steps will be the following: At first the digital archives will be enabled to handle the submission of complex databases and information system, for example geodata. Besides this it is intended to integrate special modules to handle documents with electronic signatures.

## ■ The creation of digital intermediate archives

The most important project to extend the digital archives will be the so called Digitales Zwischenarchiv, the digital intermediate archives. Since more than 40 years the Bundesarchiv provides several intermediate archives in St. Augustin near Bonn, Hoppegarten east of Berlin and Freiburg im Breisgau. The federal ministries can deliver all the records they do not need any longer into an intermediate archives. There the records will be kept for a space of time on which the ministry and the Bundesarchiv has agreed before. During that period of safekeeping the records are under the ministry's power of control, but the Bundesarchiv is allowed to appraise their archival value so the records can be destroyed at the end of the safekeeping period. Only the records of archival value will be transferred to the archives in the narrower sense of the word.

Several federal ministries have already asked if the Bundesarchiv will transfer this well proven and established system into the digital area. The advantage for the administration would be that the ministries and agencies would be released from the obligation to maintain older electronic records and keep them accessible over a longer period. This would be the care of the Bundesarchiv. The ministries would not be obliged to establish their own solutions for long term preservation of older electronic records.

The advantages of the Bundesarchiv would be important too. The electronic records would reach the archives at the most earliest time so the Bundesarchiv would have the technical custody. Problems of file conversion could be provided widely which would be unavoidable if the records would slumber in the ministries storage without any technical care. First of all the intermediate archives system guarantees widely that all records will be submitted to the archives.

By order of the Bundesarchiv the federal office on IT in cooperation with the INFORA company has presented a feasibility study in the summer of 2008<sup>2</sup>. Thereafter the document lifecycle of an electronic records would be the following:

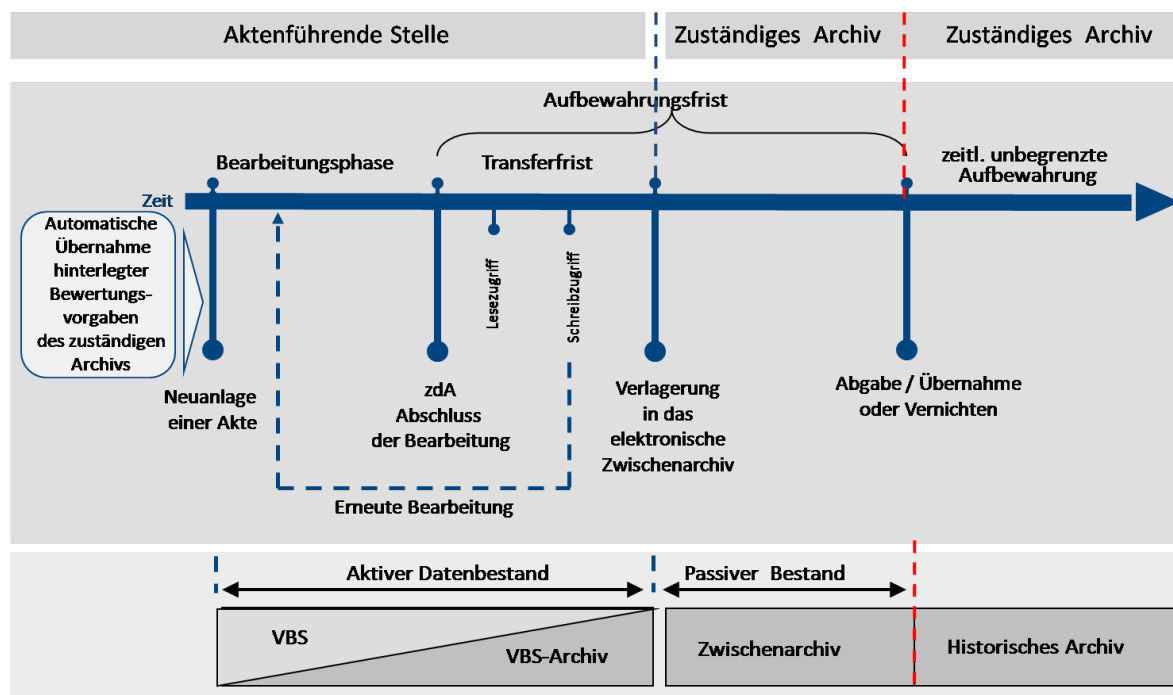


Fig. 3 : Lifecycle of the digital records

During the active period the electronic record is part of the process management system or the document management system of a ministry or an office. When the records will be concluded the second period begins, the storage term. During the first part of that period - the transfer term - the record will stay in the ministries storage so it can be accessed at all times. Under certain conditions electronic records can be taken back into the active records management system of the ministry. At the end of the transfer term the record will be transferred to the storage system of the intermediate archive; from now it will be part of the passive records which only can be read but not called back in the active system. This will be the moment, when the files will be converted into file archiving formats. In other words: SAM will start its work at this moment. During the intermediate term the archivist of the Bundesarchiv will decide on the archival value of the records. At the end of the storage term all records without archival value will be erased, the others will be transferred in the area of open-end-archiving. Basys, the archives management system of the Bundesarchiv, will log the decision of appraisal.

The realisation of the intermediate archives depends on the solution of various questions. The need of the constant availability of access to the electronic records demands a high-capacity system environment which can only be assured in the context of the secure governmental networks IVBB (Information Network Berlin-Bonn - the network of the ministries) or IVBV (Information Network of the Federal Government). For the solution of this problem there exists already a feasibility study of the federal office on IT again in cooperation with the INFORA company<sup>3</sup>.

Another problem which must be solved is the maintenance of the evidential value of electronic records which have been transferred in the intermediate archives. Digital signatures must be handled carefully. The transformation model TransiDoc<sup>4</sup> of the Fraunhofer Institut für Secure Informationstechnology (SIT) and the ArchiSafe-concept<sup>5</sup> of the Physikalisch-technische Bundesanstalt (PTB) open to us ways to solve the problem.

<sup>2</sup> Entwicklungsstudie elektronische Zwischenarchive. Version 1.1. August 2008.

<sup>3</sup> Konzeptionierung und Ausgestaltung sicherer Kommunikationsverbindungen. Version 1.0. May 2008.

<sup>4</sup> <http://www.transidoc.de>

<sup>5</sup> <http://www.archisafe.de>

The establishment of a digital intermediate archives as shared service centre will be connected with serious investments by the federal government. And the Bundesarchiv will need additional financial and personell recourses. On the other side every ministry and every agency in the federal government would be able to save a lot of money and personell capacity. In contrast to the traditional intermediate archives the Bundesarchiv would be able to offer this service to most of the governmental agencies. Without a digital intermediate archives a lot of institution will have to set up an infrastructure for longterm preservation on their own. The costs will exceed the investments needed by the Bundesarchiv by far.

# TRANSITION TO ELECTRONIC RECORDS EXCHANGE IN ESTONIA

---

Livii Karpistsenko

Kadi Riisma

State Chancellery of the Republic of Estonia

*The paper describes the project the State Chancellery of the Republic of Estonia launched in 2005 for connecting records management systems so as to guarantee interoperability and the gradual transition to paperless document exchange in the public sector. Electronic records management systems (ERMS) running on different software will now be able to exchange documents (records) in the secure X-Road data exchange environment through a specially created document exchange centre (DEC).*

*Letters were chosen as the first type of document to be tested in the automatic forwarding process. The scope of paperless document exchange is to be gradually widened to also include other types of documents, such as legislation, EU documents, invoices etc.*

*The state information systems' data transport layer X-Road is a technical and organizational environment that enables secure data transfer between digital state databases and enables secure data transfer between individuals and state institutions. It also coordinates the access of individuals to information being processed in state databases.*

*The DEC is an information system providing a common central document exchange service for various ERMSs as well as other information systems dealing with documents. The objectives of the DEC are the following: interfacing dispersed information systems through the secure data exchange layer X-Road; short-term preservation of records and, in the future, also provision of services that will support the proceeding of documents; long-term preservation and backup of records. Using the DEC as a central asynchronous buffering component on the distributed X-Road infrastructure is justified by the need to simplify the joining of different ERMSs with the state document exchange infrastructure as well as to ensure the reliability of records exchange.*

*ERMS of public sector institutions have an interface with the DEC. They periodically send records to other systems and receive them. Records with metadata are exchanged in XML format.*

*In 2006 a project was launched to develop the legislative drafting process as a whole. Harmonised XML forms are being developed for the creation, processing and automatic transmission of legal acts between institutions through the X-Road and the DEC, and for their publication in the State Gazette.*

[See the presentation.](#)

## **L'évolution vers l'échange d'archives électroniques en Estonie**

*Cet article présente le projet de la Chancellerie d'État de la République d'Estonie lancé en 2005 visant à mettre en relation les systèmes de records management afin de garantir leur interopérabilité et la migration progressive vers l'échange dématérialisé de documents dans le secteur public. Les systèmes de records management appliqués par le biais de différents logiciels sont aujourd'hui capables d'échanger des documents (records) dans l'environnement sécurisé d'échange de données X-Road par l'intermédiaire d'un centre d'échange de données.*

*Les lettres ont été le premier type de document testé dans le processus de transfert automatique. Le champ d'application de l'échange dématérialisé de documents devrait s'élargir progressivement à d'autres types de documents, tels que des documents juridiques, des documents de l'UE, des factures, etc.*

*La couche X-Road d'échange de données des systèmes d'information de l'État est un environnement technique et organisationnel permettant un transfert de données sécurisé entre les bases de données numériques publiques et entre les individus et les institutions publiques. Elle permet également de coordonner l'accès du public aux informations traitées dans les bases de données de l'État.*

*Le centre d'échange de données est un système d'information offrant un service central d'échange de documents destiné à plusieurs systèmes de gestion des ressources électroniques ainsi qu'à d'autres systèmes d'information traitant des documents. Les objectifs du centre d'échange de données sont les suivants : servir d'interface aux systèmes d'information épars ; la conservation à court terme des archives et, dans le futur, la fourniture de services qui assureront le suivi des documents ; la conservation à long terme et la sauvegarde des archives. L'utilisation du centre d'échange de documents en tant qu'élément tampon central asynchrone sur l'infrastructure X-Road répartie se justifie par la nécessité de simplifier le rattachement de différents systèmes de records management à l'infrastructure publique d'échange de documents et de garantir la fiabilité de l'échange d'archives.*

*Les systèmes de records management des institutions du secteur public disposent d'une interface avec le centre d'échange de données. Ils échangent régulièrement des archives avec d'autres systèmes. Les archives contenant des métadonnées sont échangées au format XML.*

*En 2006, un projet a été lancé pour développer le processus d'élaboration de la législation. Des formes XML harmonisées sont en cours de développement pour la création, le traitement et la transmission automatique d'actes juridiques entre les institutions par le biais du X-Road et du centre d'échange de données, et pour leur publication sur le Journal officiel.*

[Consulter la présentation.](#)

## ■ 1. Introduction

In the interoperability of different information systems three components can be distinguished:

- Organizational interoperability is related to the activities of the organizations and inter-organizational agreements
- Semantic interoperability is the ability of different organizations to uniformly understand the meaning of exchanged information (data).
- Technical line interoperability, in which one can distinguish:
  - Infrastructure's interoperability or the ability of the hardware acquired by different organizations to work in mutual association, and
  - Software interoperability or the ability of software used in different organizations to exchange data.

The general context of the semantic interoperability of software systems is to some extent similar to communication between people: individuals exchanging information with each other may prefer to use different languages, their cultural context and knowledge as a rule differ, as do their goals and expectations. In order to ensure that the meaning of a message is conveyed, they choose, however, a language which is mutually understood.

In order for the records (documents) passing between state institutions to be mutually comprehensible, their conformity to common requirements must be ensured. The requirements of paper-based records management are set out as format requirements for documents (designating the visual appearance of documents) and records management regulations, which establish the procedural requirements for the creation and management of documents throughout their life cycle. In the transition to electronic records management there are further technological requirements, and format requirements also play a different role, as in the technical sense they primarily ensure the compatibility of data formats (not a standardized visual appearance). This is an inevitable prerequisite for technological systems to work together, whatever they may be. The creation of software which incorporates the entire life cycle of a document is a complicated task, and even if such software is available it must still be able to communicate with other information systems. In consequence, such software only solves the problem of data format compatibility in part. One of the most suitable data formats in digital records management is XML.

## ■ 2. Background

The possibilities for the application of interoperability can be examined at the following levels:

**1) Record level** - where an important role is played by a document structure that is unique in syntax and a description of content that can be uniformly understood semantically. The creation of standard documents, which so to say describe themselves and their content both for people and for information systems, allow these records to be automatically forwarded from one records management system (RMS) to another together with the necessary metadata. The creation of an inter-institutional interoperable system that would include the entire public sector assumes the control of the XML schemes of records, e.g. with the help of a register of the respective schemes.

**2) The standardization of records and records management metadata** is necessary primarily to make the records management of institutions interoperable at the system level, but also so that citizens could find the information resources of institutions without being able to thoroughly orientate in the division of inter-institutional assignments and functions. The creation of structured standard documents, describing their content, for both people and information systems allow these documents to be automatically forwarded from one (records management) system to another with the necessary metadata. When using standard records management metadata, the records management systems of institutions can forward and capture records



automatically, thereby making the work of the institution more efficient. The standardization of records management metadata also simplifies the compatibility of records management systems and archive management systems, which in summary simplify the work of both the institutions archives as well as the National Archives. According to the law, the records registers of institutions are published on the web, although to date the quality of the information available thereby varies a great deal, and search possibilities are relatively limited. Well-structured records management metadata and the records registers based thereon improve the access of citizens and institutions to public sector informational resources.

**3) Interoperability of information systems.** Because of dispersed development, different records management software is used in the Estonian public sector institutions, which makes the development of inter-institutional electronic documents (records) exchange more difficult. Therefore, the increase of public sector effectiveness is one of the most important assignments of the development of the interoperability of records management systems. The development of inter-institutional electronic communications includes the interface of different information systems and databases and the rational organization of e-mail management. The objective is to increase the relative importance of electronic records exchange. The strategy for Estonian information society and the Government Development Program plan to guarantee paperless records management in the public sector by 2011.

## ■ **3. The transition to electronic records exchange in the Estonian public sector**

In 2005 the State Chancellery of the Republic of Estonia began connecting records management systems so as to guarantee interoperability and the gradual transfer to paperless records exchange in the public sector. Records management systems of governmental institutions running on different software are now be able to exchange records in the secure X-Road data exchange environment through a specially created document exchange centre.

In launching the document exchange centre it was decided that as a simple type of document, letters would be the first to be tested in the automatic forwarding process. The scope of paperless records exchange is to be gradually widened to also include other types of documents, such as legislation, invoices and so on.

### **3.1. The communication layer X-Road**

The state information systems' data transport layer X-Road is a technical and organizational environment that enables secure data transfer between digital state databases and enables secure data transfer between individuals and state institutions. It also coordinates the access of individuals to information being processed in state databases [1].

X-Road is one of the six core state information systems that provide support to all state information systems. It is a standard data communication layer between databases and information systems that allows information systems with differing underlying platforms to transfer data. Platform independence is achieved by using the SOAP and XML-RPC protocols. XML-RPC support will be discontinued at the end of 2008, after which SOAP will remain the only supported protocol.



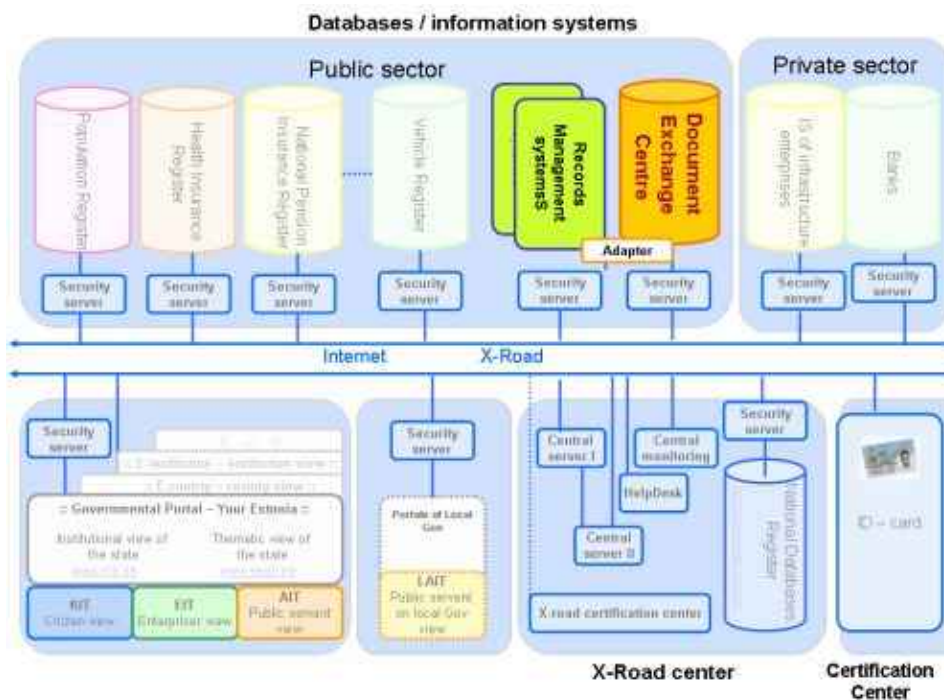


Figure 1. X-Road.

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

The X-Road architecture consists of X-Road servers, X-Road server software and databases/information systems that have joined to the X-Road. The infrastructure consists of security servers, central servers, certificate servers and monitoring servers. The X-Road security server runs on standard software which is installed on a server at each institution that joins the X-Road. All data travelling over the X-Road is transported via the security servers. All security servers are certified with the X-Road certification centre using a hardware module. Information regarding the certified security servers is distributed amongst the security servers via the central server. Security servers encrypt/decrypt data, generate usage logs, control usage rights to services and prevent unauthorized access. The use of security servers gives institutions an assurance that their data is travelling securely over the internet.

### 3.2. Document exchange centre

The document exchange centre (DEC) is an information system providing a common central document exchange service for various records managements systems as well as other information systems dealing with documents [2]. The objectives of the DEC are the following: interfacing dispersed information systems through the secure data exchange layer X-Road; short-term preservation of documents and, in the future, also provision of services that will support the proceeding of documents. Using the DEC as a central asynchronous buffering component on the distributed X-Road infrastructure is justified by the need to simplify the joining of records management systems with the state document exchange infrastructure as well as to ensure the reliability of document exchange.

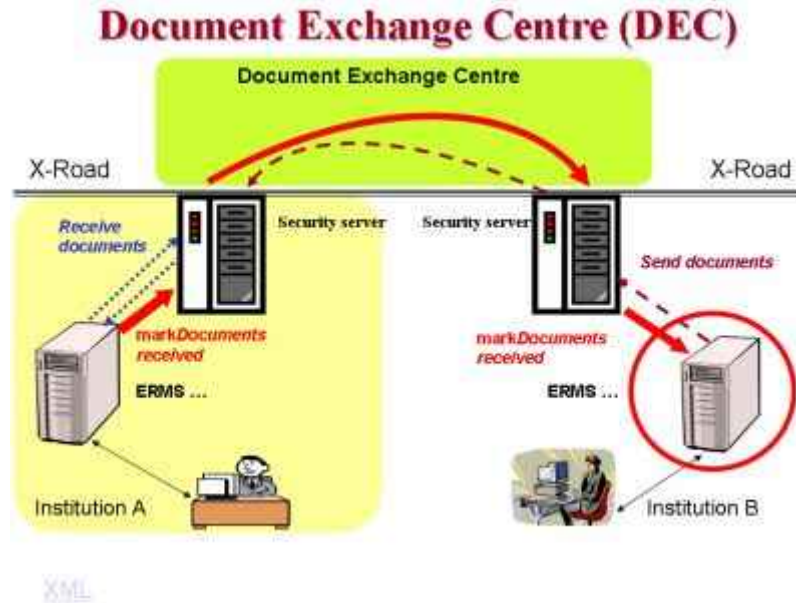


Figure 2. Document Exchange Centre

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

The DEC is, then, an infrastructure for the transmission of records (i.e. a mediation layer for document exchange services of information systems) relying on the X-Road as a transport-level infrastructure. Records (documents) are messages with described semantics and structure. These can be letters, draft legislation, electronic forms, documents related to public procurement procedures etc. Systems exchanging documents can be records management systems or information systems specific to an organization's main activities.

Records management systems of public sector have an interface with the central document exchange point. They periodically send for other systems documents and receive them. No need for traditional post. No need for scanners. Documents with metadata are exchanged in XML format.

### **3.2.1 The records logistics services:**

- uploading
- records delivery
- downloading
- erasing
- forwarding to the on-line system (not through the X-Road, but by e-mail to records management systems not interfaced with the DEC).

Since one of the most important and, at the same time, one of the most complicated questions in the modernization of records is the utilization of digital signatures; then the record's XML-structure is based on the DigiDoc standard for giving digital signatures with the Estonian ID-card.

### **3.2.2 Three basic scenarios for the use of the DEC:**

- A records management system exists in the institution. The records management system is programmatically interfaced with the DEC; the transmission of records to other records management systems and the necessary translation into and from the DEC format occurs through the DEC. Thereby, the DEC operates primarily as a transport environment and a so-called universal intermediate language between the different records management systems formats.

- The institution does not own a records management system. The DEC sends records to the institution and receives them by e-mail. The DEC assumes a (temporary) role as an intermediate archive, guaranteeing the preservation of the records for a certain period.
- A citizen, who is not an official, can send records to the DEC by e-mail to be forwarded to the institution. The institution can answer the citizen either directly by e-mail or by e-mail through the DEC. In the future, the citizen may also get access to the DEC through the web, whereby the same DEC would become the basic system for exchanging state records with citizens.

## ■ 4. The stages of transition to electronic records exchange

1) **The standardization of elements and layouts of records** - the objective was to compile a list of mandatory elements and the respective XML schema for the document type 'letter'. Based on the list of elements with descriptions, the Estonian national standard EVS 882-1:2006 "Information and documentation. Elements of records and requirements for record's layout. Part 1: Letter" was adopted.

2) **The standardization of records management metadata** - the metadata directory was worked out, based on which the institutions could unify the composition of their metadata components, so that metadata would support the exchange of documents between electronic information systems and would allow for detailed searches simultaneously in several information systems. The following characteristics of metadata were taken into consideration:

- Recycling - metadata created for records management must be recyclable in other fields and must conform to the metadata standards in other fields of activity (e.g. archive management).
- Multi-levelness - based on international standards for archive descriptions, multi-level descriptions must be employed in records management. Metadata must be presented in at least the following levels: series, folder, records, and file.
- Modularity - metadata must be presented in groups, by records management events and their groups, which can be employed by organizations either one at a time or in combination with other metadata schemes.
- When employing metadata, a certain amount of metadata components is mandatory for all organizations and additional metadata must be recommended, but each organization should be able to decide whether this is practical based on their records management needs.

1. **In the first stage (pilot project)**, during 2005 a working application was worked out for organizing the electronic records exchange in four ministries for the exchange of digital records by interfacing different electronic records management systems, and introduced in three ministries using identical records management software, and one ministry using different software. The development of the necessary data exchange interface for interfacing the records management systems was based on the specifications of the documents exchange centre and X-road. In the data exchange interface, the XML format was used to display the records and the SOAP protocol for forwarding them. The main emphasis in the pilot project was the creation of preconditions for directed semantic interoperability.

2. **In the second stage of the electronic records exchange project**, the remaining seven ministries, 15 counties and 30 governmental agencies were involved. The emphasis was on the more extensive realization of software interoperability, using a centralized document exchange centre as a buffer for the exchange of desynchronised records between records management systems. The project for the development of electronic records exchange managed by the State Chancellery was closely connected to the project for the creation of a documents exchange centre being executed by the Estonian Informatics Centre of the Ministry of Economic Affairs and Communications, which was focused on the development of the technical interoperability of records management systems. The aim was to establish automatic exchange of XML-based metadata.

3. **The next stage is the implementation of XML-based records** (letters, legal acts, EU documents, invoices etc). It includes the following tasks:
- **The detailed charting of the life cycle of a record** and an optimized procedural scheme and compilation of basic requirements for stages of document forwarding.
  - **The standardization of different types of records** - presents a list of elements and the respective XML schema, includes a list of the global elements, element groups, complex types and attributes, and descriptions of the elements (both technical and personal descriptions).
  - **The creation of a tool which will make the compilation and editing of an XML-based documents** possible, as well as automatic forwarding of them between the records management and other information systems involved in the creative process.
  - **The implementation of a compilation and processing procedure for XML-based records**, the launch of an automatic forwarding system for records created on the basis of harmonized forms, and the integration of the information systems involved in the process of creating.

**The transition to document exchange through the document exchange centre** and the implementation of harmonized XLM forms for documents will make the legislative drafting process and other processes more transparent. The entire life cycle of a record will be reflected in its metadata, enabling us to:

- monitor and control the progress;
- obtain access to the document at every stage of the process;
- guarantee the authenticity, reliability, usability and integrity of the record at all times.

A record can be automatically transmitted through the records exchange centre from one records management system to another without being converted into paper format at any stage. The need for repeated entry of digitally created records disappears, which means that:

- errors produced through the repeated entry are avoided;
- officials are able to work more quickly and easily and with greater efficiency.

In the course of the project conditions will be created for the long-term preservation of digital records. In addition, the transition to an XML-based creative process and automatic forwarding will provide both ordinary citizens and the creators of the records with the opportunity to better monitor the progress.

## ■ References

[1] General principles of operation and administration of the X-Road are defined in the government statute „*Implementation of Information systems data transport layer*” RTI, 23.12.2003, 83, 568. English documentation for the X-Road is available on the Estonian Informatics Centre home page (<http://www.ria.ee/27309>).

[2] Detailed information about document exchange centre is available on the Estonian Informatics Centre home page (<http://www.ria.ee/index.php?lang=en>).

# LONG-TERM DIGITAL PRESERVATION USING AGENT TECHNOLOGY – THE PROTAGE PROJECT

---

Johan Bengsston

InterNIT

Raivo Ruusalepp

Estonian Business Archives Consultancy

*This paper presents a summary of the aims and achievements of the PROTAGE project to date.*

*The PROTAGE project (2007-2009) is funded under the EU FP7 ICT programme under the ‘radically new approaches to digital preservation’ theme. The mission of the project is to make long-term digital preservation easier for all computer users and to help them to preserve their own content. It will also contribute towards reducing the cost of preservation and increasing the capacity of memory institutions to preserve digital information. The PROTAGE project approaches long-term digital preservation in a flexible and distributed way, combining intelligent automated methods - based on software agent technology - with human intervention.*

*The rapidly growing volume of digital information is causing data transfer from active IT systems to digital repositories at an increasing pace. This makes it necessary to find new levels of automation and self-reliance in digital archiving and preservation solutions. The increasing diversity in size and complexity of new digital resources imply that the repository systems must become highly automated and adaptable to various types of input, storage and access.*

*The level of automation in current digital preservation solutions is low, and involves many manual stages. The scalability of existing preservation solutions has been poorly demonstrated, and solutions often have not been properly tested against diverse digital resources or in heterogeneous environments.*

*Based on the latest digital preservation research, the PROTAGE project will build and validate flexible and extensible software agents for digital preservation and access that can be integrated in existing and new preservation systems. Targeted end users of the project are curators of digital repositories and creators of digital content in a variety of domains (including government agencies and publishers), but it will also target individuals managing their own digital collections.*

*The first prototype set of PROTAGE tools will concentrate on the scenarios for appraisal decision making based on current regulatory requirements, similar past appraisal decisions and information use patterns; and also transfer of electronic records from agencies to archives and ingest of digital objects into repositories. The paper will show how a collection of agents collaborate to ensure that relevant digital objects are selected, analysed, and transferred for retention and preservation, and that suitable archival file formats are used throughout the life of digital objects.*

*The project has focused on describing application scenarios, from which the functional requirements and a technical specification have been developed that show which agents and other major system components are needed, and what is their intended use. An End-User-System Agent is envisaged for the final iteration in the project, that will be installed in computers where information is created or received - for example personal computers. Such agents can be left running in the background to observe what files are stored and accessed at different times, and they collaborate with similar agents via the Internet, to collect knowledge on what file formats are in common use for long-term retention, what legal requirements for archiving exist, and then recommend best courses of action to ensure the controlled retention of the stored files. When a situation is detected where digital information might be lost or rendered inaccessible in the future, the End-User-System Agent informs the user or takes automated action to safeguard the file.*

*The PROTAGE project opens up a novel approach to digital preservation. The resulting preservation tools have the potential to enable integrated automated digital preservation operations in digital archives and libraries of different scales throughout Europe. It could also assist individuals in preserving their personal digital collections.*

[See the presentation.](#)



## **La conservation à long terme des données numériques en utilisant l'outil technologique : le projet PROTAGE**

*Cet article présente un résumé des objectifs et des réalisations du projet PROTAGE à ce jour.*

*Le projet PROTAGE (2007-2009) est financé par le programme UE FP7 ICT dans le cadre du thème des « approches radicalement nouvelles en matière de conservation du numérique ». Le projet a pour mission de simplifier la conservation du numérique à long terme pour tous les utilisateurs de l'informatique et de leur permettre de conserver leur propre contenu. Il contribuera également à réduire les coûts de la conservation et à augmenter la capacité des institutions garantes de la mémoire à conserver les informations numériques. Le projet PROTAGE aborde la conservation du numérique à long terme de façon souple et répartie, en combinant les méthodes informatiques intelligentes - basées sur des logiciels - et l'intervention humaine.*

*En raison de l'augmentation rapide du volume d'informations numériques, le transfert de données des systèmes informatiques actifs vers les plates-formes d'archivage électronique est de plus en plus rapide. Il est de ce fait nécessaire de trouver de nouveaux niveaux d'automatisation et une capacité technologique autonome en matière d'archivage numérique et de solutions de conservation. En raison de la diversité croissante en termes de taille et de complexité des nouvelles ressources numériques, les systèmes de plates-formes doivent devenir très automatisés et adaptables à plusieurs types de données d'entrée, de stockage et d'accès.*

*Le niveau d'automatisation dans les solutions actuelles de conservation du numérique est faible et implique bon nombre d'étapes manuelles. L'évolutivité des solutions de conservation actuelles s'est révélée peu concluante et souvent, les solutions n'ont pas été correctement testées avec des ressources numériques variées ou dans des environnements hétérogènes.*

*Basé sur les dernières études sur la conservation numérique, le projet PROTAGE vise à concevoir et valider des logiciels flexibles et extensibles destinés à la conservation numérique et à l'accès pouvant être intégrés dans les systèmes de conservation actuels et futurs. Le projet cible des utilisateurs finaux tels que les administrateurs de plates-formes numériques et les créateurs de contenu numérique dans divers domaines (y compris des agences gouvernementales et des éditeurs) mais également des individus qui gèrent leurs propres collections numériques.*

*La première série prototype d'outils PROTAGE concernera les scénarios de prise de décision en matière d'évaluation basés sur les exigences réglementaires en vigueur, les décisions d'évaluation identiques prises par le passé et les modèles d'utilisation de l'information, ainsi que le transfert d'archives électroniques des agences vers les archives et l'intégration d'objets numériques dans les plates-formes d'archivage. L'article montrera la façon dont une sélection d'agents collaborent afin de garantir que des objets numériques pertinents soient sélectionnés, analysés et transférés pour être retenus et conservés, et que des formats adéquats de fichiers d'archivage soient utilisés tout au long de la vie des objets numériques.*

*Le projet s'est notamment attaché à décrire des scénarios d'application à partir desquels des exigences fonctionnelles et une spécification technique ont été développées pour montrer quels agents et autres éléments importants du système sont nécessaires et quelle est leur utilisation finale. Un agent utilisateur final système est prévu pour la phase finale itérative du projet. Il sera installé sur des ordinateurs où les informations sont créées ou reçues, sur des ordinateurs personnels par exemple. On peut laisser ces agents fonctionner dans l'environnement pour observer quels sont les fichiers qui sont stockés et consultés à des moments différents. Ils collaborent avec des agents similaires via internet afin de recueillir des informations sur les formats de fichiers qui sont utilisés couramment pour la conservation à long terme, les exigences réglementaires en vigueur en matière d'archivage, et enfin de recommander les meilleures actions garantissant la conservation contrôlée des fichiers stockés. Lorsque l'on détecte que des informations numériques pourraient être perdues ou rendues inaccessibles dans le futur, l'Agent Utilisateur du Système Final informe l'utilisateur ou déclenche une action automatisée afin de sauvegarder le fichier.*

*Le projet PROTAGE ouvre une nouvelle approche en matière de conservation du numérique. Les outils de conservation qui en découlent ont la capacité d'intégrer des opérations d'archivage électronique intégrées dans les Archives et Bibliothèques numériques d'importance variable à travers l'Europe. Il peut également aider les individus à conserver leur propre documents numériques.*

[Consulter la présentation.](#)

## ■ The PROTAGE Project

The PROTAGE project (2007-2009) is co-funded by the EU FP7 ICT digital libraries programme under the 'radically new approaches to digital preservation' theme. The aim of the project is to make digital preservation easier for all computer users and to provide tools that help them to preserve their own content. These tools should also contribute towards reducing the overall cost of digital preservation and increase the capacity of memory institutions to manage the preserve of digital information. The PROTAGE approach to digital preservation is flexible and distributed, combining intelligent automated methods - based on software agent technology - with human intervention.

The volume of digital information is growing rapidly everywhere and is causing an increasing pace of data transfer from active IT systems to digital preservation repositories. This makes it necessary to find new levels of automation and self-reliance in digital archiving and preservation solutions. The increasing diversity in size and complexity of new digital resources implies that the preservation repository systems must become highly automated and adaptable to various types of input, storage and access.

The level of automation in current digital preservation solutions is low - many manual stages are still present in the workflow of preservation. The scalability of existing preservation solutions has been poorly demonstrated, and solutions have often not been properly tested against diverse digital resources or in heterogeneous environments.

Based on the latest digital preservation research, the PROTAGE project is building and validating flexible and extensible software agents for digital preservation. These are software tools that can be integrated in existing and new preservation systems. Targeted end users of the project are curators of digital repositories and creators of digital content in a variety of domains, including government agencies and publishers. Individuals managing their own digital collections are also considered as a potential target user group.

The development of intelligent software agent tools for digital preservation will go through two iterations. By December 2008, the PROTAGE project is going through the technical tests of its first prototype of tools. These are, in the first instance, designed to support the appraisal of electronic records in government agencies, and to help preparing electronic records for transfer to archives.

This paper will show how a collection of agents collaborate to ensure that relevant digital objects are selected, analysed, and transferred for retention and preservation, and that suitable archival file formats are used throughout the life of digital objects.

## ■ The Design of Tools

The project has focused on describing application scenarios, from which functional requirements and a technical specification have been developed that describe the necessary agents and other system components, and their intended use.

The aim of the PROTAGE project is to develop an innovative product with no obvious precedent. This process requires an understanding of the users and their capabilities, their current tasks and goals, the use context of the product, and the performance constraints on the products. The first step in the project, therefore, was an analysis of stakeholders needs using document-based methods (state of the art review, analysis of peer projects), focus groups and expert evaluations.

The state-of-the-art review identified several areas where agent ecosystems can contribute to reduced costs, a higher degree of interoperability and co-ordination, and increased capacity of memory institutions to preserve digital information. These areas are submission and ingest of digital material, monitoring preservation and transfer between repositories. The agent technology solutions have to be highly scalable and adaptable to various types of digital objects, as well as economically sound and easy to use and apply. Descriptions of models identified as suitable for implementation with agent technology revealed that they propose solutions

for two different environments, the producer and the archive and as groups of resource allocation, collaborative and/or negotiating agents.

### **Analysis of Stakeholder Needs**

A series of focus groups and expert evaluations were organised that revealed issues relevant for records managers and archivists in government agencies and national archives. The most important issues where agencies need tools to facilitate their work are: managing the classification scheme and retention schedules to meet the changing needs of the organisation; sharing knowledge on existing solutions between agencies when compiling classification schemes, deciding on records' metadata and using automated tools for metadata creation and extraction; appraisal of records; preparing records for their transfer to public archives and handling the transfer from agency to archive. For archives, the focus groups revealed that most important issues are: management of appraisal decisions (e.g., finding background information, past history and context necessary for making appraisal decisions), routine technical transactions (e.g., technical analysis of transferred document files, their structure and characteristics, extracting metadata from them), automatic documentation (creation of process level metadata) of all activities that are carried out in the pre-ingest, transfer and ingest phases in a digital archive.

The stakeholder needs identified in this phase of the project work were converted into a series of application scenarios and use cases.

### **Application Scenarios**

A total of thirteen application scenarios based on the conclusions from the stakeholder needs analysis were developed. The scenarios visualise solutions that would help automating and solving the issues raised by stakeholders in three major areas of digital preservation - pre-ingest, transfer and ingest. During the workshops it became obvious that active records management is also a crucial issue - how to create and manage digital records so that the later archiving process would be easy and cost-effective. To address these issues a fourth topic was added - monitoring of creation and management of digital records.

The PROTAGE multi-agent system will be developed as a prototype system in several iterations. The first prototype contains functionality for pre-ingest, specifically checking submission requirements as specified by an archive, creating overview of records that need to be transferred to an archive and managing the creation of a submission to the archive that meets the stated requirements.

## **■ Using Intelligent Software Agents in Support of Digital Preservation**

Agent technology holds promise for digital preservation, using models for an automatic and self-organizing approach. The PROTAGE project proposes a fresh approach to overcome the fragmented expertise in digital preservation. Software agents will work together with individuals and expert users (curators, preservation specialists, public and private agencies) to provide solutions for appraisal, transfer, ingest and monitoring digital objects. A multi-agent system is composed of a group of possibly heterogeneous and autonomous software agents that share a common goal and work cooperatively to achieve it. A multi-agent system can display an intelligent behaviour. For example, using techniques from artificial intelligence, software agents can maintain a representation of what is happening and offer spontaneous help (pro-active behaviour). To do so, they can use past information. In addition, since agents are developed independently, a multi-agent system can be easily modified by simple addition or removal of agents like in a team where members enter and quit the group at any time. An agent may have mechanisms for learning the user's preferences and the context of the work allowing it to recommend an action.

- Software Agents in PROTAGE must be able to provide the following functions:
- Automatically perform a digital preservation task, monitoring the evolution of the digital object and usage to ensure availability and future access to their content.



- Support and aid users in preserving, retrieving and sharing digital objects.
- Support the collaborative work by managing the communication among the different software components, such as the web services and user interfaces.

To achieve these functions, the PROTAGE multi-agent system is composed of both web services and software agents. Agents are aimed at coordinating all preservation process from pre-ingest, ingest and transfer until the phases for monitoring a digital object. The mission of the agents is to automatically locate, select and employ web services to obtain information, to make decisions and to perform the preservation tasks. Web services are a set of services that provide particular types of services for the agent ecosystem and user interfaces, such as access and retrieval of information from external knowledge bases, virus checks and metadata extraction.

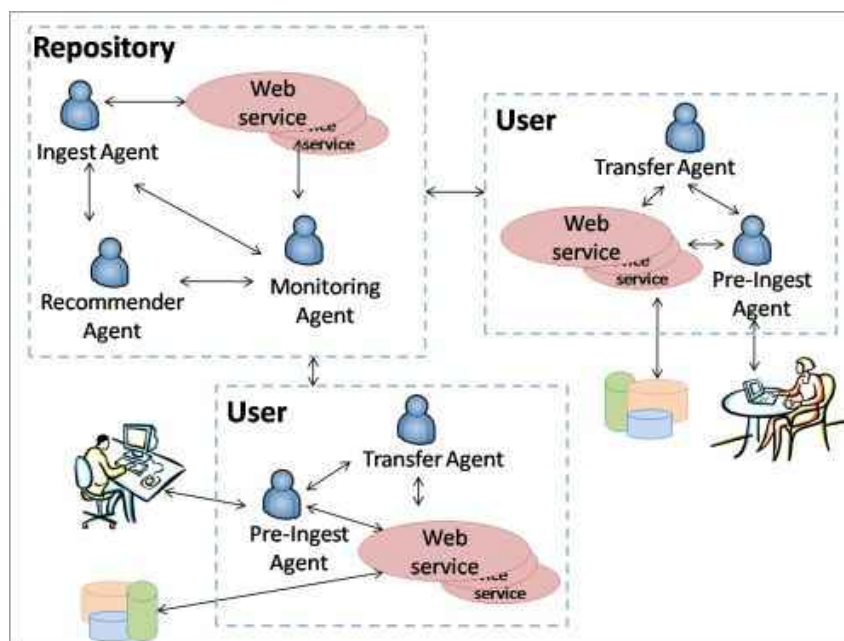


Figure 1. Components of the PROTAGE System: agents and web services.

All the software agents involved in PROTAGE have the next common tasks:

- 1) Interact with other agent requesting or answering information.
- 2) Invoke Web Services to perform a service and retrieval information.
- 3) Make decision and perform specific tasks based on the information received from others agents and web services.

As shown in Figure 1, the PROTAGE system architecture groups agents into Pre-Ingest Agents, Ingest Agents, Transfer Agents, Monitoring Agents and Recommender Agents. These have the following tasks:

- Pre-Ingest Agents are in charge of helping the user to take a decision whether records should be transferred to the repository for permanent preservation or be destroyed or left in the agency until a next decision is made. If the records have to be transferred, the agents help users in checking and preparing the package to be transferred according to the requirements of the given archive or repository.
- Transfer Agents are in charge of transfer of records from users to the repository. One of the most important tasks of Transfer Agents is to test that the created transfer packages and metadata comply with the repository requirements and checking whether the transfer process has damaged the transferred data.

- Ingest Agents are active in the repository side and check the received packages for their completeness and correctness according to the requirements, and data loss that may have happened during the transfer.
- Monitoring Agents are monitoring the evolution of object and usage to ensure availability and future access of their content. For example Monitoring agents send reminders about older records that need to be destroyed, transferred to the repository or retained for later appraisal decisions.
- Recommender Agents: suggest to the users possible actions to perform to help them in decision making process. For example Recommender Agent can suggest to the archivist to accept an appraisal request based on similar past decisions.

Each agent (Pre-Ingest Agent, Ingest Agent, Monitoring Agent, Transfer Agent and Recommender Agent) will be created as a JADE agent. JADE (Java Agent Development Environment) is a robust and efficient environment for distributed agent systems compliant with the FIPA (Foundation for Intelligent Physical Agents) specifications. Figure 2 shows the interaction between two agents Pre-Ingest Agent and Ingest Agent using the interface of JADE platform. JADE includes in its library all the components necessary to develop the agents and to provide a graphical and user-friendly interface to control and monitor the agents in the PROTAGE system.

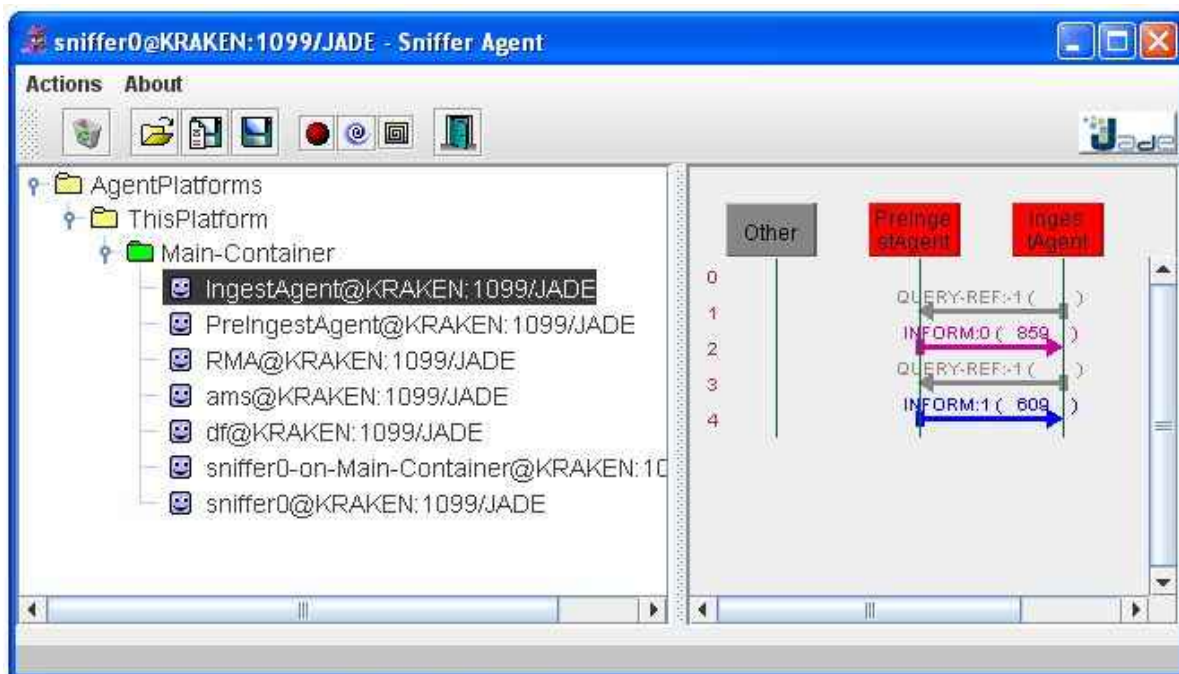


Figure 2. Testing communication between the Pre-Ingest Agent and the Ingest Agent using JADE interface.

## Proposed System Requirements

General requirements for the PROTAGE intelligent software agent tools have been defined. The PROTAGE system should have an online Graphical User Interface (GUI) for users that allows users to inspect, manipulate and administer the system.

The system should trace and analyse its own operation and then fix errors or send a log to the system administrator. The system should have wide support for international character encodings, ontologies and localisation. It should also provide an “escaping functionality” in unexpected circumstances prompting the end-user for further instructions.

The specific functional requirements of individual principal components include Monitoring, Pre-Ingest, Transfer, and Ingest agents, as well as the GUI.

The technical requirements that the system must fulfil are extensibility (the ability to extend features and capabilities without the need to change or extend the core of the system), interoperability (the ability to integrate and be integrated with external systems, legacy systems, other data models and services), scalability, the ability to support intense data and process demands), user perspective (the ability to satisfy its users and provide them the best possible services), system security (the ability to monitor and verify every activity of as user regarding access, execution and view of authorized actions and data), system privacy (the ability to consider data protection and user privacy rules), system trust (informal seminars, handbook and manual to educate the target users and make them understand the benefits of the system), and system usability (the system must be easy to use).

An End-User-System Agent is envisaged for the final iteration in the project that can be installed in computers where information is created or received - for example in personal computers. Such agents can be left running in the background to observe what files are stored and accessed at different times, and they collaborate with similar agents via the Internet, to collect knowledge on what file formats are in common use for long-term retention, what legal requirements for archiving exist, and then recommend best courses of action to ensure the controlled retention of the stored files. When a situation is detected where digital information might be lost or rendered inaccessible in the future, the End-User-System Agent informs the user or takes automated action to safeguard the file.

## ■ Conclusion

The PROTAGE project opens up a novel approach to digital preservation. The resulting preservation tools have the potential to enable integrated automated digital preservation operations in digital archives and libraries of different scales throughout Europe. It will also assist individuals in preserving their personal digital collections.

Software agents possess characteristics (e.g. ability to make decisions based on knowledge of the past) that make them suitable for digital preservation processes requirements. A multi-agent system is composed of a group of heterogeneous and autonomous agents that share a common goal and work cooperatively to achieve it. A multi-agent system can easily be modified and updated by simple addition or removal of individual agents like in a team where members enter and quit the group. This makes the PROTAGE tools scalable and adaptable to many different contexts where digital preservation know-how is necessary. For further up-to-date information, please go to: [www.protage.eu](http://www.protage.eu).

# BUILDING A DIGITAL REPOSITORY: A PRACTICAL IMPLEMENTATION

---

Filip Boudrez

Expertisecentrum DAVID vzw, Antwerpen

*The City Archives of Antwerp is currently building a digital repository for the preservation of digitised and digital-born records. The project has limited resources when it comes to funding and staff (1 FTE), but has the benefit of building on eDAVID-research. The project started by the end of 2007 and should be finalised by the end of 2009.*

*Besides the general requirements for each digital repository, one of the primary design goals is the complete integration of the digital repository within the archival service and the overall IT-infrastructure of the city administration. The digital repository will not be exclusively used by the archival service for the long-term preservation of electronic records. Also process specific information systems of the city administration which create or contain electronic records, will make use of the infrastructure of the digital repository. For this reason, the repository is positioned as a general service provider within the technology infrastructure of the city. As a consequence of this positioning, not only materials with a permanent archival value are stored within the digital repository, but also records with a limited life span. The archival service is responsible for the information architecture and the management of the digital repository.*

*For the City Archives itself, it's very important that the digital repository is an integrated part of its procedures and workflow for the ingest, the management and the dissemination of records. An opportunity for this was the re-designing of the recordkeeping system of the City Archives. For its recordkeeping system the City Archives is not using commercial software or an open source alternative, but instead they are creating their own software in house. The integration of the digital repository within the recordkeeping system is achieved by the use of a common metadatamodel for archival descriptions and by holding on the concept of inventorynumbers. The metadatamodel for archival descriptions of paper and electronic records is based on the three ICA-standards (ISAD(G), ISAAR(cpf) and ISDF) and is built on the basic assumption that records are in the first place intellectual entities. This insight is derived from the world of electronic records, where their physical and intellectual management should be separated as much as possible. As a consequence, some elements of the ISAD(G)-metadataset are not implemented at this description level or are re-purposed for other description needs. Also the ISAD(G)-concept of reference information as unique identification of the unit of description is not applied as prescribed by the standard. Instead, inventorynumbers are still in use as unique identifiers of the lower aggregation levels (file, item), a concept inherited from the paper archives. The new recordkeeping system provides the functionality to link one or more inventorynumbers to ISAD(G)-files. Which representations or manifestations of the records are present within the archives is registered at the level of inventorynumbers. For electronic records, the reference to their storage location within the digital repository is here registered.*

*Besides the descriptive metadata, specific technical and administrative metadata for the electronic records are registered during the ingest and management processes in a fully automated way. Only a minimum set of metadata, required for an efficient management and the quality controls of the digital inventorynumbers, is stored centrally in the database of the recordkeeping system software. The largest quantity of technical metadata is stored in the AIP index-files. These metadata apply to the digital object level and are kept as XML-files within the repository itself. This offers not only a large flexibility, but has also the benefit that the central metadatamodel can be kept relatively simple and easy manageable. The relationships between electronic records and its representations on the one hand, and the representation and their digital objects on the other hand, are also documented in those AIP index-files. These AIP index-files are the starting point for the consultation of electronic records.*

[See the presentation.](#)

## **La mise en place d'une plate-forme d'archivage numérique : une réalisation pratique**

Les Archives de la ville d'Anvers travaillent actuellement à la mise en place d'une plate-forme d'archivage électronique pour la conservation des documents numérisés et des documents numériques natifs. Ce projet a des ressources limitées en termes de budget et de personnel (1 ETP), mais il a l'avantage de s'appuyer sur le programme eDAVID. Commencé fin 2007, il devrait être terminé fin 2009.

Hormis les exigences générales applicables à chaque plate-forme d'archivage électronique, l'un des premiers objectifs de cette étude est l'intégration complète de la plate-forme au sein du service des Archives et de l'ensemble des infrastructures informatiques de l'administration municipale. La plate-forme ne sera pas exclusivement réservée à l'usage du service des archives pour la conservation à long terme des documents électroniques. Les systèmes d'information municipaux destinés à des processus spécifiques qui produisent ou contiennent des documents électroniques utiliseront également les infrastructures de la plate-forme d'archivage électronique. C'est pour cette raison que cette plate-forme se définit comme un fournisseur général de services intégré au sein des infrastructures technologiques municipales. Elle devra donc stocker des documents présentant une valeur archivistique permanente, mais également des documents à la durée de vie limitée. C'est le service des archives qui a la responsabilité de l'architecture informatique et de la gestion de la plate-forme.

Il est très important pour les archives municipales que la plate-forme fasse partie intégrante de leurs procédures et de leur flux d'intégration, de gestion et de diffusion des documents. La refonte du système de conservation des Archives municipales en a été l'occasion. Pour leur système de conservation des documents, les Archives municipales n'utilisent ni logiciel commercial ni solution open source, mais développent leur propre logiciel maison. L'intégration de la plate-forme d'archivage électronique dans le système de conservation des documents est rendue possible par l'utilisation d'un métamodèle commun dédié aux descriptions archivistiques conjuguée au maintien du principe des numéros d'inventaire. Le métamodèle utilisé pour les descriptions archivistiques de documents papier et électroniques fonctionne avec les trois normes ICA : ISAD(G), ISAAR(CPF) et ISDF. Il part du principe fondamental selon lequel les documents sont avant tout des entités intellectuelles. Cette idée provient du domaine des documents électroniques où la gestion physique et intellectuelle des documents doit être, autant que possible, traitée séparément. Par conséquent, certains des éléments de l'ensemble de métadonnées à la norme ISAD(G) ne sont pas exécutés à ce niveau de description ou bien sont adaptés à d'autres besoins de description. Le principe de la norme ISAD(G) selon lequel les informations de référence sont l'unique identifiant de l'unité de description n'est pas non plus appliqué comme le veut la norme. Ici, les numéros d'inventaire continuent à être les seuls identificateurs des niveaux d'agrégation inférieurs (document, article), concept hérité des archives papier. Le nouveau système de conservation des documents a la capacité d'associer un ou plusieurs numéros d'inventaire aux documents ISAD(G). Quelle que soit la représentation ou la forme des documents présents dans les archives, ceux-ci sont enregistrés au niveau des numéros d'inventaire. C'est à ce niveau qu'est référencé le lieu de stockage des documents électroniques dans la plate-forme d'archivage électronique.

En plus des métadonnées descriptives, des métadonnées techniques et administratives spécifiques pour les documents électroniques sont enregistrées de façon totalement automatique lors des processus d'intégration et de gestion. Seul un ensemble minimal de métadonnées nécessaire à une gestion efficace et aux contrôles qualité des numéros d'inventaire numériques est centralisé dans la base de données du logiciel du système de conservation. La plupart des métadonnées techniques sont conservées dans des fichiers AIP. Ces métadonnées concernent le niveau de l'objet numérique ; elles sont stockées au format XML dans la plate-forme elle-même. Cette solution donne non seulement une grande flexibilité, mais elle permet aussi d'avoir un métamodèle central relativement simple et facilement gérable. Les rapports entre les documents électroniques et leurs représentations d'une part et les représentations et leurs objets numériques d'autre part sont également documentés dans ces fichiers AIP qui constituent le point de départ pour la consultation des documents électroniques.

[Consulter la présentation.](#)



## ■ Introduction

Antwerp City Archives began the implementation of their digital archive repository in Autumn of 2007. Such a project however, involves much more than simply installing the required hardware and software. It is not only a matter of many organisational issues; the construction of a digital repository also raises fundamental questions about records management and recordkeeping.

## ■ 1. The digital repository

Digital archiving has, for a long time, been a policy priority for the Antwerp City Archives. A proactive policy has already been in place for many years, focussing on city agencies and services that create records. During this time, the transfer of digital records to the archival service began and also many archival documents were digitised internally. Thus the volume of born-digital and digitised archives managed by the City Archives has gradually been on the increase.

The Antwerp City Archives anticipates that the digital repository will address procedures and infrastructure for the ingest, the management, and the dissemination of the digital archives and collections with medium to long term retention periods. The digital resources archived in the digital repository must be authentic and durable, regardless of whether they are born digital or digitised.

Construction of the Antwerp City Archives digital repository has, to a large extent, been based on the research and recommendations from eDAVID.<sup>1</sup> Development and programming has, for the majority, been carried out in-house. Only the dissemination has been contracted out to an external developer, because this process is carried out separately through the website of the Antwerp City Archives.

## ■ 2. An integrated recordkeeping system

An important and central aspect of the implementation is the positioning of the digital repository within the city archives. It is widely accepted that there are two main options for implementing this. One option is for the digital repository to be constructed alongside the recordkeeping system for analogue records and archives. This option results in only minimal integration between paper and digital records. The only integration envisaged in such instances is the use of metadata, to make the collections easier to search by researchers and civil servants working in the city administration. An alternative option is complete integration of paper and digital recordkeeping. In this case, the same procedures are used for treating and managing both paper and digital archives.

The City Archives has opted to implement option two for their digital repository and is therefore seeking maximum integration between paper and digital archives. There are many reasons for this. The first is so that the digital repository of the Antwerp City Archives may not be an isolated system. The second is that the City Archives wishes to implement a single system that will cater for all archives and collections, regardless of the medium or form of the records. The advantage of this is that the archival processes for intake, management and giving access are consistent for all types of records and only need to be automated once. The same software system can be used for managing all archives so that, in as far as is possible, information is shared and only needs to be registered once. The principle of authority records is thus applied in a de facto manner and the need for developing similar functionality (for example, the ability to construct detailed retrieval aids) across different systems is avoided. Thirdly, integrated management of the archives is completely in line with the records creation practices that exist in a hybrid records creating environment. Finally, integration also contributes to embedding the digital repository within the institution.

---

<sup>1</sup> See: <http://www.edavid.be/eng/index.php>.

The Antwerp City Archives is therefore constructing a single integrated recordkeeping system in which the management of paper and electronic records is fully part of. The integration takes place across numerous levels, including the software for processing and managing the records, the information architecture and the metadata model, and finally also the procedures.

### ■ 3. Recordkeeping software: MACZ

Significant opportunities existed at the start of the implementation project that impacted on the choice of software for archives and records management. With the move to the Sint-Felixpakhuis and the denBell project it was recognised that records management within the Antwerp City Archives was in need of revision. The new location and the application of box placement according to their size meant that a thorough and automated repository system was needed. Prior to the move, work also commenced on structuring and automating physical management practices in the archives.<sup>2</sup>

The denBell project started very soon after the move. This will soon result in administrative departments transferring large volumes of records to the archival service at pace. This must be done in an efficient way so they are quickly processed and accessible after their transfer. To facilitate this and ensure it is most effectively achieved, the way in which records are inventoried has been completely revised and the three international standards for archival description, ISAD(G), ISAAR(cpf) and ISDF, have been implemented. The way in which transfer lists were composed and processed has also been assessed and entirely adapted.

The new recordkeeping system has been named MACZ. MACZ will be completely managed and developed internally. As a result of opting for an integrated system, it was clear from the outset that the digital repository would be involved in the development of MACZ. The archives will not therefore be managed by multiple and different information systems.

### ■ 4. Information architecture and the metadata model

In revising the information architecture and data model for recordkeeping, paper-based records management was not necessarily assumed to deliver the leading principles. This wouldn't be strange as paper records management was a very familiar process. Instead, everything was thoroughly evaluated and optimum functional records management approaches and solutions were sought. Eventually it became apparent that the basic principles of digital archiving would provide the starting point for the new integrated information architecture and the new metadata model.

Two central principles for the new recordkeeping system in the Antwerp City Archives are, on the one hand, the separation of physical records management aspects from intellectual ones and, on the other hand, approaching archival records as abstract entities with one or more representations (paper, digital, microfilm/fiche etc). Separating physical and intellectual records management aspects has however a few consequences. The two aspects are managed and described separately but must remain clearly identifiable. Moreover, they must continue to be clearly linked to one another. In practice, this is achieved by linking the inventory numbers on the ISAD(G) records with descriptions of the records.

Another consequence was that ISAD(G) could not be implemented on its own and that the ISAD(G) implementation at the Antwerp City Archives deviated from the standard on numerous points. Diverse ISAD(G) description fields are related, for example, to representations of the archival documents (e.g. the element 'Extent and medium of the unit of description') and would be better described on the level of the inventory number. Such a ISAD(G) fields were not consequently implemented as proscribed by the standard but were given new functions.

---

<sup>2</sup> See the annual report of the Antwerp City Archives (2006 - 07) for more information. Annual reports are available at [www.felixarchief.be](http://www.felixarchief.be). More information about the denBell project is also available from this site, though in Dutch only.

This approach of identifying archival records using inventory numbers was already in place. This manner of working will be extended from paper record keeping to the digital archives. An inventory number is assigned to archival files and archival items as logical entities. For digital archiving a small amendment will be made: a single inventory number can refer to an entire series of digital archives to facilitate the complete retrieval and consultation of the full series. Request and retrieval of a digital series therefore has no logistical restriction in the way that paper archives do. In fact, this means that for the identification of a digital archive series an inventory number will be used in the meaning of ISAD(G)'s 'reference code' element

Different representations of the same archival record are not allocated individual inventory numbers. On the level of the inventory number itself is registered which representations of the same documents is in the holding of the City Archives; for example, archival records in paper form have the inventory number linked to the physical storage box number (one or more); digital records will have a reference to their location within in the digital repository; and records on microfilm will be denoted by the microfilm number (one or more).

The further allocation of inventory numbers and their extension to digital records had, for the city archives, the significant advantage that the archives staff already was already familiar with this method and did not require any additional training. The manner in which digital records would be prepared for ingest into the repository was communicated over internal email. The new recordkeeping software supports the above in different ways. Furthermore, a built-in functionality enables members of staff to examine for themselves the suitability of the proposed digital inventory numbers.

## ■ 5. New record keeping procedures and development

An important consequence of pursuing a completely integrated records and archives management system is that the procedures for paper and digital records management must be fully aligned. The practical actions for processing paper and digital records are naturally different, but will be embedded in the same basic procedures and follow a common workflow.

The original intention here was to take existing procedures for paper records keeping as the starting point. During this exercise it became apparent that many procedures for paper records keeping were not consolidated and that analogue archives were sometimes processed by archives staff in very different ways. The implementation of a structured and extensively automated recordkeeping system meant that it was consequently necessary to first ensure that consensus was reached between staff and that procedures were agreed across the board. This took a lot of meetings and time, and for this reason, the implementation of the digitale repository slowed down.

In parallel with planning the new recordkeeping and procedures, it was also necessary to programme and test the required software modules for the digital repository. The components of the different OAIS-processes - namely ingest, management, and dissemination - have been clearly defined and implemented. The functional model defined in the OAIS standard results in a basic workflow but this does not have to be strictly followed whilst developing and implementing the different software modules. Instead, the essential components of each step of the process are developed simultaneously. For example, as soon as the first digital inventory numbers are ingested in the digital repository, the essential management quality controls must already be operational and the digital records must be retrievable and accessible. The ingest process does not therefore have to be fully implemented before development and implementation of the management modules is started.

This approach means that after the groundwork has been done on essential modules, work can begin on further refinement and optimisation. As a result, archives staff can already begin learning and gaining experience with the system. One advantage of this internal development is the ease and speed with which feedback can be given. The same approach will also be (more or less) followed to develop the dissemination functionalities retrieval and giving access. Until now, this is the only part of the digital repository that has not been developed by the City Archives itself but has been contracted out to an external partner as front-end access to digital or digitised archives is integrated into the website of the City Archives. Development of the front-end portal is happening in two phases. During the first phase, the basic modules and interfaces will be developed. These will only be available to archives staff, who can retrieve and consult the digital archives. During the second phase, the remaining functionality will be implemented (registration of consultations,



security etc) and will be further optimised according to feedback. The request process will be refined according to feedback from users and will be made more intuitive for the end user.

Gaining experience is valuable for both archival staff and external developers. A digital repository differs in numerous respects from other types of computing systems. Electronic records can have very complex structures (for example, a website), or can be extremely bulky (for example a digitised register). Such factors must be taken into account when the technical issues relating to query and consultation functionality are being worked out.

## ■ 6. Information infrastructure

Alongside development of the first modules, the storage infrastructure was installed and configured. Digital objects will be securely and safely stored in a SAN (Storage Area Network). The SAN is divided in two logical units (LUN's). From a technical point of view there's no real 'air-gap' between the two LUN's, but each LUN has its own security and management policy. The first LUN is used as digital studio and is the storage location for transferred records during their quarantine and processing (migration, transformation, encapsulation) phase. All staff members have read and write privileges in the studio area. The second LUN is the actual digital repository. It holds the public and non-public records. The distinction between public and non-public records is registered in the metadata of an inventory number. The metadata set of an inventory number contains the date field 'Public from' so no specific action is needed when the status of a record changes from non-public to public. Write rights are restricted to the manager of the digital repository and the ingest software modules. A part of this LUN functions as 'reading room'. The permanent DIP's (low resolution or compressed representations) are stored here.

The available storage capacity of the digital repository will be systematically expanded on a step-by-step basis. This step-by-step approach will safely support tests of future development activities to increase storage capacity. Additional and particular points of interest included definition of the security protocols, organisation of the back-up process, and the integration of the storage infrastructure. The storage environment must be secure not only against regular users but also applications such as those used within the front-end access portal. The digital repository differs from regular digital information systems, and this even extends to the backup systems. Due to the high volume of backup data, a standard backup regime is insufficient. The creation of a full backup is a time consuming activity and cannot be carried out weekly or even twice a month. Nevertheless, the creation of good backups remains an essential part of a disaster recovery strategy.

Concern over this last point and a supporting risk analysis resulted in significant attention being paid to the arrangement of the storage infrastructure, which was not selected for technical or efficiency reasons but for providing support to meet minimum archival requirements/logic. Should data be lost at the database level or subsequent references become incorrect, the essential linkage and management data can quickly be generated anew or corrected.

An essential element in the information infrastructure is the applied storage method for the AIP's within the digital repository. The chosen storage method is also based on the conducted risk analysis. Within the digital repository the information objects of the different representations of one record, their technical metadata (identification of the file format, fixity information) and the essential contextual metadata of the record are encapsulated in one XML container file. By doing so, the links and the relationships between the several components are stored in a sustainable way and the digital repository contains its own metadata. Physical encapsulation is applied as much as possible. In the cases in which physical encapsulation is not feasible (e.g. snapshots of websites, very large databases), the XML AIP's contain the metadata, document the relationships between the information objects, and refer by relative URI's to the digital objects<sup>3</sup>.

The XML Schema for the XML AIP's was delivered by eDAVID and further customised during the implementation. Two new subclasses were added to the XML Schema: the subclass 'index' and the subclass 'recordGroup'. The addition of the subclass 'index' was appropriate as the City Archives of Antwerp sticks to the principle of working with inventory numbers. An 'index' AIP contains an overview of all records which are considered to be part of one inventory number. The index AIP is the starting point when consulting the contents

---

<sup>3</sup> The rationale behind this storage method is elaborated in F. BOUDREZ, *Digital containers for shipment into the future*, Antwerp, 2005. The XML Schemas are available on: <http://www.edavid.be/xmlschemas>.

of an inventory number through the website and is the basis for the permanent and systematic quality controls by the DigDep Monitor. This is a management tools which checks at random digital inventory numbers. These checks include: presence and bit integrity of all digital objects, presence of metadata, correctness of the management metadata, etc. The subclass 'recordGroup' is reserved for particular aggregation levels like series or case files. These AIP's contain mainly the records management metadata of the aggregation level and URI's to the records which are part of it.

## ■ 7. Transferring records into the digital repository

As soon as the essential modules for ingest and management were operational in the storage infrastructure, transfer of items into the digital repository could commence. By 2007, the Antwerp City Archives already had collected 1TB of digital and digitised records. These digital records had until then mainly been kept on CD, with some exceptions (f.i. preserved websites) stored on file servers. Transferring these legacy files was thus one of the first tasks.

As a result of the sheer amount of work and the additional activities this task entailed, it was something of a sub-project in its own right. Just copying the contents of the CDs had already taken a long time. Moreover, the contents of the CDs could not simply be transferred direct into the digital repository. In most cases, one or more archival processes had to be undertaken upon the digital records: assigning inventory numbers, abolishing inventory numbers not accepted by the new system and therefore neither by the digital repository, adapting folder names, checking integrity, registering metadata that had been distributed up until then, and so forth. For the most part, these archival processes generally required additional and meaningful checks or analysis so dealing with them took a considerable number of months.

As soon as the first digital inventory numbers were ingested in the digital repository, essential management tasks had to be carried out. This assumed in the first instance that responsibility for management of the digital repository had already been clearly determined. It was expected that the digital repository manager would be thoroughly acquainted with the metadata and database models. This was already the case for the Antwerp City Archives, for not all of the required user interfaces and associated modules for managing the system were ready and a certain amount of work still took place behind the scenes.

A particular activity in this part of the implementation process involved the AVA image database. AVA contains descriptions of digital photographs and makes them available online. Master copies are now stored in the digital repository, whilst low resolution copies are uploaded to AVA for access. The digitised photos are therefore managed in both systems, which means AVA must be integrated with MACZ and the digital repository. However, as AVA was developed externally it has not been possible to integrate all of the functionality in with the recordkeeping system. Eventually it became clear that the most important integration work would be realised within the MACZ environment. The AVA database itself needed only one minimum adaptation. Once this adaptation was complete, work could commence to transfer the AVA master files into the digital repository.

Implementation of the digital repository has led to questions being asked about the future of the AVA database and its content. The digital repository enables direct access to digital photos via the City Archives' website without needing to access AVA itself. Numerous photos and pieces of iconographical work that had been previously stored in AVA are no longer stored there as these contents have been transferred to the digital repository. The AVA records will be deleted once the descriptions have also been extracted.

## ■ 8. Conclusion

From the implementation path followed by the Antwerp City Archives, it is clear that the construction of a digital repository is initially more concerned with structured and efficient record keeping in general than with digital archiving in particular. Consequently, the implementation of a digital repository should ideally begin with a records management and recordkeeping 'vision' in which the best parts of paper and digital records

management are brought together. Once the general procedures have been established for ingest, management, and giving access, implementation of the digital repository should proceed relatively smoothly.

Implementation at the Antwerp City Archives has taken place in a step-by-step manner. Because of this it has been possible to learn about and systematically align records management practices, and thus also improve the digital repository. Implementation is therefore also an iterative process, in which consolidation and improvement continuously alternate with each other. The digital repository remains thus a work-in-progress.

# PROCESSUS ET ARCHITECTURE POUR LA CONSERVATION À LONG TERME DES DONNÉES NUMÉRIQUES : LE SYSTÈME MUTUALISÉ DU CONSEIL GÉNÉRAL DE L'AUBE

---

Céline Guyon

Attaché de conservation du patrimoine, Archives départementales de l'Aube

**Contexte/origine :** Le développement de l'e-administration en France se traduit notamment par la mise en oeuvre de télé procédures (Actes, Hélios, marchés publics) et donc la dématérialisation des données circulant entre administrations (collectivités locales et services déconcentrés de l'Etat principalement) sous la forme de flux. D'autre part, avec le développement de la signature électronique, la valeur probante réside dans ces données nativement numériques, sous réserve qu'elles soient conservées de façon à assurer leur intégrité.

**Etat des lieux dans le département de l'Aube :** le Département est tiers de télétransmission pour le compte des communes et groupements de communes qui le souhaitent depuis avril 2007. Il met à leur disposition, via un Extranet, son infrastructure technique et logicielle et à moyen terme, son infrastructure d'outils de recherche, de consultation et de conservation à long terme des données numériques. D'autre part, à compter de septembre 2008, les actes (délibérations et arrêtés) seront signés électroniquement par le président du Conseil général.

**Enjeux et stratégies du projet de système mutualisé de conservation à long terme des données numériques :** D'un point de vue politique, le projet s'inscrit dans une démarche claire et cohérente de structuration du territoire à l'échelle géographique du département. D'un point de vue organisationnel et administratif, le projet s'inscrit dans une démarche participative associant les connaissances et compétences de la direction des technologies de l'informatique et de la communication et des Archives départementales du Conseil général de l'Aube. D'un point de vue archivistique, le projet prend en compte la complétude de la chaîne archivistique : collecte et sélection des données, gestion du cycle de vie des données, conservation à long terme des données, gestion des droits d'accès (communicabilité), recherche et restitution des données en fonction des profils utilisateurs (producteur/citoyen/chercheur). D'un point de vue technologique/technique, le projet respecte les normes et standards. Ce projet s'inscrit dans un souci fort de coopération et de partage de l'information numérique : mutualisation des connaissances, des infrastructures technique et applicative.

**Les fonctionnalités du système :** Présentation des différentes fonctionnalités du système et de leur compatibilité avec les normes et standards existants (OAIS, standard d'échange pour l'archivage des données électroniques, DTD EAD) : capture et ingestion des données, gestion des données descriptives, stockage du contenu de données, recherche, consultation et restitution des données.

**Les modalités de mise en oeuvre :** Les partenaires : services du Conseil général de l'Aube, communes et groupements de communes, services déconcentrés de l'Etat (préfecture, sous-préfectures, trésorerie générale). L'Extranet du Conseil général : un outil au service des collectivités locales et groupements de communes. L'architecture technique et applicative. Les modalités de réutilisation des outils de gestion des archives papier (progiciel Thot).

**Les conséquences en termes archivistiques :** Les conséquences sur le positionnement de l'archiviste dans la collectivité, les conséquences sur la constitution et la gestion des fonds sur support papier et électronique.

[Consulter la présentation.](#)

## **Procedure and Architecture for the long-term preservation of digital data: the collaborative system of the Conseil général of the Aube département**

**Origins:** the development of e-administration in France can be seen most clearly in the use of online procedures (such applications as Actes and Hélios, and bidding for public contracts) resulting in the dematerialisation of data circulating between administrative bodies (mainly local authorities and decentralised state services) in the form of electronic flows. In addition, with the widening acceptance of electronic signatures, probative value resides in data that are 'natively digital', on condition that they are preserved in a way that guarantees their integrity.

**Current situation in the Aube Département:** the Département has been relaying information on behalf of any communes or groups of communes who choose to use the service since April 2007. They can access its technical and software infrastructure via an Extranet and will soon be able to benefit from its search engines, consultation tools and the long-term preservation of digital data. In addition, as from September 2008, all official documents (whether proceedings or by-laws) will be signed electronically by the President of the Conseil Général.

**Consequences and strategies of the project for a collaborative system for the long-term preservation of digital data:** from a political point of view, the project is an integral part of a clear and consistent strategy to modernise the infrastructure of the entire Département. From an organisational and administrative point of view, it follows a participative approach associating the knowledge and expertise of the ICT Directorate with the Aube Département Archives. From an archival point of view, the project takes into account every aspect of the archiving chain: collection and selection of data, management of the life-cycle of data, long-term preservation of data, management of access rights (communicability), data search and recovery adapted to different user profiles (producer/citizen/researcher). From a technical or technological point of view, the project complies with current standards. Overall, the project expresses a strong desire for cooperation and sharing of digital information: pooling both knowledge and technical and software infrastructures.

**System features:** we shall present the various system features and their compatibility with existing standards (the OAIS model or the EAD DTD): capture and ingestion of data, management of descriptive data, storage of data content, searching, consultation and output of data.

**Implementation:** the partners: the services of the Aube Conseil Général, communes and groups of communes, decentralised state services (Prefecture, Sub-Prefectures, local tax offices). The Conseil Général's Extranet: a tool for local authorities and groups of communes. The technical and software architecture. Ways of reusing tools designed for managing paper archives (Thot software).

**Consequences for archiving activities:** the consequences for the archivist's role in society. The consequences on the creation and management of repositories for paper-based and electronic media.

[See the presentation.](#)

## ■ Enjeux et stratégies

La réflexion menée actuellement au Conseil général de l'Aube, autour du déploiement d'un système d'archivage électronique repose sur une volonté claire et forte de mutualisation : mutualisation des connaissances et compétences, mutualisation des outils matériels et logiciels.

### Enjeux politiques

D'un point de vue politique, le projet s'inscrit dans une démarche claire et cohérente de structuration du territoire à l'échelle géographique du département. Le Conseil général met en effet à disposition des autres collectivités du département (communes, groupements de communes, etc.) via un Extranet, une plate-forme mutualisée de services. Les services en ligne offerts sont multiples et couvrent de nombreux domaines. D'abord de types informatifs (annuaires, modèles d'actes réglementaires, fiches de veille juridique, bourse d'emploi), ils ont évolué vers de véritables services d'e-administration et d'aide à la gestion avec la consultation du cadastre numérisé et la possibilité d'éditer des fiches de propriété, la gestion, via une application mutualisée, du fichier électoral ou du catalogue de la bibliothèque, des outils d'aide à la rédaction et à la mise en ligne des dossiers de consultation des entreprises dans le cadre de la dématérialisation des procédures de passation des marchés publics et plus récemment, une plate-forme mutualisée pour les projets de dématérialisation des échanges actuels et à venir.

De plus, le Conseil général a souhaité se positionner en tant que tiers de télétransmission, au travers du dispositif « démat-aube » qui a été homologué et est également « Autorité d'enregistrement », c'est-à-dire, qu'il dote gratuitement chaque collectivité qui souhaite dématérialiser ses échanges avec les services extérieurs de l'Etat, d'un certificat électronique de classe 3+ qui permet d'authentifier la collectivité et sécurise les transmissions électroniques. Avec ce même certificat, les collectivités peuvent également signer électroniquement leurs actes grâce à un parapheur électronique.

Sur les 433 communes du département, 408 ont adhéré au portail ainsi que 5 établissements de coopération intercommunale ; chaque jour, en moyenne 100 collectivités différentes se connectent à la plate-forme mutualisée de services.

Dans ce contexte, notre réflexion sur le déploiement d'un système d'archivage électronique s'inscrit également dans une volonté forte d'offrir aux communes et groupements de communes, via l'Extranet et la plate-forme mutualisée de services, les moyens de conserver à long terme et de consulter les données électroniques issues des flux de dématérialisation des échanges.

### Stratégies organisationnelles et enjeux archivistiques

D'un point de vue organisationnel et administratif, notre projet s'inscrit dans une démarche participative de type conduite de projet, associant les connaissances et compétences de la direction des technologies de l'informatique et de la communication et des Archives départementales et surtout un renforcement des collaborations entre informaticiens et archivistes.

D'un point de vue archivistique, le projet prend en compte la complétude de la chaîne archivistique et s'inscrit dans une démarche de record management : collecte et sélection des données, gestion du cycle de vie des données avec une intervention de l'archiviste très en amont des procédures, conservation à long terme des données, gestion des droits d'accès (communicabilité), recherche et restitution des données en fonction des profils utilisateurs (producteur/citoyen/chercheur.)

D'un point de vue technologique/technique, le projet s'appuie sur les normes et standards en vigueur (la norme OAIS, le standard d'échange de données pour l'archivage, la DTD EAD) et vise, là encore, sous couvert de mutualisation, à réutiliser, dans la mesure du possible, les outils logiciels déjà déployés au sein de la collectivité, notamment le progiciel Thot utilisé par les Archives départementales pour la gestion des archives papier.



## ■ **Détail fonctionnel du système d'archivage électronique et conséquences du point de vue de la pratique archivistique**

Les fonctionnalités du système d'archivage électronique s'appuient sur les grands domaines fonctionnels décrits par la norme OAIS à savoir les entités « entrée », « stockage », « gestion des données descriptives », « administration », « accès. » Je vous propose de détailler chacune de ces entités sous l'angle fonctionnel en insistant sur les effets induits, par la production documentaire électronique, sur la pratique archivistique.

### **Capture et ingestion des données**

L'entité « entrée » reçoit, contrôle et valide les objets à archiver et vise à transformer les paquets d'informations versés en un ou plusieurs paquets d'informations archivés. Les objets à archiver proviennent de la plate-forme de télétransmission et ont été formatés selon les spécifications du standard d'échange de données pour l'archivage. Les paquets d'informations archivés sont constitués d'une enveloppe XML comprenant le fichier pdf. initial (qui correspond à l'acte envoyé au contrôle de légalité), un fichier avec une extension en lp7. lorsque l'acte est signé électroniquement et les métadonnées nécessaires à leur description et à leur gestion dans le temps.

Les versements ou transferts sont automatisés et s'effectuent en ligne : Chaque soir, l'ensemble des données échangées au cours de la journée, c'est-à-dire l'ensemble des actes transmis au contrôle de légalité au cours de la journée, est transféré sur la plate-forme d'archivage. Les transferts sont préparés par le tiers de télétransmission : dans ce contexte, on peut s'interroger sur la pertinence, dans l'environnement électronique, de la notion de « service versant » puisque ce sont des contingences d'ordre purement techniques qui influent le choix du service versant, sans lien aucun avec l'histoire administrative du service producteur. D'autre part, la notion de « versement » est-elle encore pertinente comme outil de gestion et de cotation des flux dématérialisés ?

Dans notre cas de figure, le Conseil général est tiers de télétransmission ce qui offre l'avantage de réduire le nombre de partenaires et de les limiter à la sphère du domaine public et dans le champ de la collectivité. Et donc de maîtriser l'ensemble du processus de versement. C'est également un gain d'efficacité dans le processus de mise en œuvre du système d'archivage électronique puisque l'implémentation du standard d'échange de données pour l'archivage sur la plate-forme de télétransmission est en cours de réalisation, en interne, c'est-à-dire par un informaticien du Conseil général<sup>1</sup>.

Dans l'environnement électronique, à la différence de l'environnement papier, les objets transférés sur la plate-forme d'archivage sont des archives vivantes et le moment de leur transfert ou versement est quasi concomitant du moment de leur création. Le cycle de vie de ces données électroniques doit donc obligatoirement être déterminé dès la conception du processus de dématérialisation, ce qui induit un positionnement indispensable de l'archiviste en amont des procédures afin de collecter les métadonnées indispensables à la préservation et à la gestion des données électroniques dans le temps.

D'autre part, les concepts de durée d'utilisation courante (DUC), durée d'utilité administrative (DUA), d'âge intermédiaire ou de documents semi-actifs, ne déterminent plus le versement aux Archives départementales ; de même que la notion de place qui, dans l'environnement papier, conditionne souvent le transfert des documents des bureaux vers un espace de stockage intermédiaire ou définitif. Il n'existe d'ailleurs pas d'infrastructure de stockage intermédiaire puisque les données électroniques sont immédiatement versées aux Archives départementales. Dans l'ère électronique, peut-on encore parler de « préarchivage » ?

Dans la pratique, les objets à archiver sont versés sur la plate-forme d'archivage deux mois après leur transfert au contrôle de légalité ; le délai de deux mois correspond au délai d'instruction maximale des services de la préfecture, dans le cadre du contrôle de légalité, c'est-à-dire à la fin de la durée d'utilité juridique.

---

<sup>1</sup> L'implémentation du standard d'échange de données pour l'archivage s'appuie sur le mode d'emploi ou profil pour les actes administratifs soumis au contrôle de légalité et le manuel technique réalisés par les Archives départementales des Yvelines et publié dans les ateliers de la modernisation (initiative IF08, action Adèle 103 : archivage numérique).

## **Stockage**

L'entité « stockage » assure la conservation physique des paquets d'informations archivés et doit garantir la pérennité et l'intégrité de l'ensemble des données conservées. C'est la direction des technologies de l'informatique et de la communication du Conseil général qui est opérateur d'archivage pour le compte des services producteurs et des Archives départementales et qui développe et assure la maintenance des serveurs de stockage.

Le principe de la redondance a été retenu avec deux serveurs de stockage, l'un primaire et l'autre esclave ; le serveur esclave étant répliqué la nuit. Les serveurs de stockage se trouvent sur deux sites géographiques distincts. La capacité de stockage est évolutive, avec à priori, une capacité d'un demi tera octets pour démarrer.

## **Gestion des données descriptives**

L'entité « gestion des données descriptives » assure la gestion des informations descriptives (informations de pérennisation et informations de description) relatives aux contenus d'informations conservés par l'entité « stockage. » Dans le système d'archivage électronique que nous imaginons modulaire et mutualisé, cette fonction est assurée par le progiciel de gestion des archives papier utilisé par les Archives départementales, le progiciel Thot.

Dans la pratique, un fichier d'export conforme aux spécificités du progiciel Thot est généré depuis la plate-forme d'archivage par l'opérateur d'archivage (c'est-à-dire la direction des technologies de l'informatique et de la communication) pour le compte des Archives départementales. A l'heure actuelle, le standard d'échange de données pour l'archivage n'étant pas implémenté dans le progiciel Thot, le fichier d'import est au format txt. et s'appuie, pour la structure, sur un fichier de concordance entre les balises du standard d'échange de données pour l'archivage et les champs et spécificités de notre base de données documentaire.

Le choix de ne pas développer ou acquérir un outil logiciel spécifique pour la gestion des données descriptives relatives aux contenus d'informations électroniques s'explique par le souci de ne pas différencier la gestion des données descriptives se rapportant à un contenu d'informations sous forme papier de celles se rapportant à un contenu d'informations électroniques car il s'agit en réalité des mêmes documents et donc des mêmes métadonnées. D'autre part, notre souci est de ne pas démultiplier les outils logiciels et d'assurer, pour le chercheur, la continuité des sources historiques, quel que soit le support des documents. Nous nous interrogeons d'ailleurs sur les modalités de transfert vers la plate-forme d'archivage des délibérations anciennes, non signées électroniquement, mais dont nous disposons d'une copie numérique. De plus, il s'agit de thésauriser les savoir-faire développés pour la gestion des flux papier et de les adapter, le cas échéant, aux spécificités de l'environnement électronique.

D'un point de vue archivistique, à chaque objet à archiver, c'est-à-dire chaque acte soumis au contrôle de légalité, sont associées des métadonnées de description alors que dans l'environnement papier, les mêmes actes transmis au contrôle de légalité ne sont pas décrits au niveau de la pièce mais plus sûrement au niveau du groupe d'articles. Dans l'environnement électronique, le niveau de description est donc beaucoup plus fin, ce qui démultiplie les potentialités de recherche.

Cependant, d'un point de vue documentaire, le risque est celui d'un déséquilibre entre la description des documents sur support papier et la description de ces mêmes documents dans l'environnement électronique ; déséquilibre dans le degré de finesse de la description et donc dans les réponses offertes à la communauté des utilisateurs, pour une même requête.

## **Recherche, consultation et restitution des données**

L'entité « accès » regroupe tous les services qui sont en interface directe avec les utilisateurs et vise à offrir à l'utilisateur la possibilité de retrouver une information dans un catalogue puis de consulter le contenu d'informations associé.

Deux catégories principales d'utilisateurs ont été identifiées : les services producteurs et les chercheurs (terme pris dans son acception la plus large.) Chacune de ces catégories d'utilisateurs présente des caractéristiques et spécificités fortes, d'autant plus que leurs motivations en terme de recherche et d'accès à



l'information sont différentes. C'est pourquoi, pour mieux répondre aux exigences des utilisateurs, nous avons opté pour deux processus distincts. Ces deux processus sont concomitants mais seul l'un - celui dédié à la communauté des chercheurs - présente un caractère pérenne.

Les services producteurs qui disposent déjà d'une suite logicielle dédiée pour l'accomplissement de leurs missions (par exemple le progiciel AirDelib utilisé par les services du Conseil général pour la rédaction des projets de délibérations), utiliseront, comme ils le font déjà, ces mêmes outils, pour retrouver une information, et dans notre cas de figure, un acte transmis au contrôle de légalité.

Lorsque ce sont les outils bureautiques classiques qui sont utilisés (notamment un traitement texte), une offre de type Gestion Electronique de Documents sera proposée aux services producteurs. Elle s'appuiera sur une recherche de type *full text* dans les contenus d'informations, c'est-à-dire dans les fichiers au format pdf. des actes soumis au contrôle de légalité et sur quelques métadonnées ciblées (ex. nature de l'acte, date, objet.) Pour les services du Conseil général, cette offre sera déployée sur le portail Intranet et pour les autres collectivités du département, sur le portail Extranet.

La question qui se pose est de fixer les bornes chronologiques de ce service. Est-il souhaitable de permettre aux services producteurs de rechercher un acte, quelle que soit sa date ou ne faut-il au contraire calquer l'effectivité cette offre sur le cycle de vie des documents et s'appuyer sur la notion de durée d'utilité administrative pour procéder à l'épuration régulière du système de GED ?

Quant aux chercheurs et autres utilisateurs des archives, l'offre de service qui leur est proposée est la même que celle déjà mise en œuvre pour les archives sur support papier : un accès au catalogue via un portail Intranet en salle de lecture des Archives départementales et un accès web via le site Internet des Archives départementales. En salle de lecture, la solution développée s'appuie sur le progiciel Thot et sur le web, sur une base de données de type MySQL.

Les premiers tests sont prévus en janvier avant un déploiement en interne, vers d'autres services producteurs et, en externe, vers les autres collectivités du département. Notre réflexion a porté initialement, pour une question d'actualité, sur les actes soumis au contrôle de légalité mais le système d'archivage électronique imaginé doit pouvoir accueillir tous types de productions électroniques, quel que soit leur contexte de production (autres processus de dématérialisation, bases de données, images numériques, documents bureautiques, etc.) et qu'elles soient ou non nativement électroniques.

## ■ Conclusion

Au cours de notre réflexion, les questions d'ordre technologique n'ont pas soulevé de grandes inquiétudes, les informaticiens apparaissant presque plus préoccupés par la conception et la mise en œuvre du processus d'archivage, domaine de prédilection des archivistes. Cette expérience a prouvé l'importance du binôme informaticien-archiviste et a renforcé le positionnement de l'archiviste au sein de la collectivité. D'ici quelques mois, un comité de pilotage sur l'archivage électronique sera constitué, co-animé par la direction des technologies de l'informatique et de la communication et les Archives départementales. L'archiviste sera donc nécessairement associé à la conception de nouveaux processus de dématérialisation ou au déploiement de nouvelles applications informatiques. Dans le même temps, l'archiviste devra savoir modifier et adapter ses méthodes de travail et outils conceptuels à l'environnement électronique.

D'autre part, il est une certitude dont nous avons l'intuition et qui s'est confirmée, à savoir, que sans mutualisation (des connaissances, des compétences et des moyens), un projet de préservation à long terme des données numériques, ne peut pas aboutir. Mutualisation et émulation au niveau local mais aussi au niveau national avec la constitution de groupes de travail comme cela fut le cas pour la collecte des données électroniques produites dans le cadre de l'aide sociale des Départements et la diffusion des conclusions de ces groupes de travail à l'ensemble du réseau des Archives.

# **Atelier 1**

## **MoReq2 : présentation et prise en main**

---

---

# ***Workshop 1***

# ***Understanding MoReq2***

---

---

# MoReq2: WHAT IT IS, AND HOW TO USE IT?

---

Marc Fresko  
Serco Consulting

*MoReq2 - Model Requirements version 2 - is the new European de facto standard for Electronic Records Management Systems. Like MoReq which preceded it a few years ago, MoReq2 was initiated by the DLM-Forum, which is also responsible for its ongoing governance.*

*A good understanding of MoReq2 - what is in it, what isn't, how to use it, how not to use it - will be essential for everyone concerned with the good management of electronic records and other electronic information. This interactive workshop, presented by the lead author of MoReq2, will provide the understanding you need.*

*MoReq2 is already very influential in the world of electronic records management, and it will grow to be even more important. It is unusual in several respects, including:*

- *It is for all sectors, not just for government;*
- *It is international - it is used across Europe, and also in Asia, the Americas, and elsewhere;*
- *It includes functionality far beyond "pure" records management;*
- *It was developed using a huge international consultation exercise;*
- *It is backed by the DLM Forum.*

*This interactive workshop will provide a comprehensive introduction to MoReq2, including:*

- *How it was developed;*
- *Where it applies;*
- *The essential record-keeping concepts;*
- *What is new since MoReq was first published;*
- *Its structure, what it contains;*
- *How the "optional modules" fit in;*
- *How the software testing regime should work, and what it means;*
- *How to use MoReq2 for procurement*
- *Translations;*
- *International localisation and "chapter zero";*
- *Suggesting changes and ongoing governance.*

**See the presentation:**

[part 1](#)

[part 2](#)

[part 3](#)

[part 4](#)

## **Qu'est ce que MoReq2 et comment l'employer : atelier interactif**

*MoReq2 - Model Requirements (Exigences-types) version 2 - est de facto la nouvelle norme européenne pour les Systèmes d'Archivage Electroniques (SAE). Comme pour MoReq, qui l'a précédée voici quelques années, c'est le DLM-Forum qui en est l'initiateur, et qui est aussi désormais responsable de sa gouvernance.*

*Une bonne compréhension de MoReq2 (ce qu'il est et ce qu'il n'est pas, dans quels cas l'utiliser ou ne pas l'utiliser) sera essentiel pour quiconque s'enquiert d'une bonne gestion des documents électroniques et de toute autre information sous forme numérique. Cet atelier interactif, présenté par l'auteur principal de MoReq2, pourvoira aux explications dont on aura besoin.*

*MoReq2 est déjà très influent dans le monde de la gestion des documents électroniques, et il est appelé à devenir encore plus important. Il est insolite à plus d'un titre, notamment parce que :*

- *il s'adresse à tous les secteurs, pas seulement le gouvernemental*
- *il est international : il est utilisé à travers l'Europe, mais aussi en Asie, dans les Amériques, et encore ailleurs*

- *il comprend des fonctionnalités allant bien au-delà du « Records management » pur*

- *il a été développé par un large comité international de consultants*

- *il est soutenu par le DLM-Forum*

*Cet atelier interactif proposera une introduction pédagogique à MoReq2, en particulier sur :*

- *son développement*

- *les champs où il s'applique*

- *les concepts essentiels pour la préservation des documents*

- *les nouveautés par rapport à la première publication de MoReq*

- *sa structure et son contenu*

- *l'intégration des « modules optionnels »*

- *le fonctionnement probable du régime de test des logiciels et sa signification*

- *l'utilisation de MoReq2 pour les marchés*

- *les traductions*

- *son intégration internationale et le « chapitre 0 »*

- *les changements suggérés et la gouvernance à venir de la norme*

**Consulter la présentation :**

[partie 1](#)

[partie 2](#)

[partie 3](#)

[partie 4](#)

# MoREQ2 AND THE ROMANIAN FRAMEWORK

---

Bodgan-Florin Popovici

archivist, National Archives, Romania

Lucia Stefan

Archivaria Ltd, London, UK

Romania is located in the South-East of Europe. It was the result of the unification of three principalities, each having different administrative and archival approaches and traditions (French for the Danubian principalities, and Austrian for Transylvania, then Russian under Communist regime). Romania is member of EU since 2007.

Romania's current archival framework, based on different historical settings is confronted with the challenge of managing and preserving digital records.

#### **Strengths at archives level:**

- Unified archival practices at national level. Similar archival understanding, practices and training for central and local (county) archives
- Practice of registration of documents. This makes easier to understand and follow the practices of e-records capture.
- Standardised classification and audit practices.
- The presence and influence of the National Archives to the creators of documents/records.

#### **At IT level:**

- Highly skilled IT staff in programming and software development (Romania is an IT outsource)
- Existing legislation regarding electronic signatures: Electronic Signature Act (455/2001)
- e-Government policy already being implemented

#### **Problems:**

- Political influences and interests, doubled by lack of government support of the National Archives weaken the institution authority
- Important emerging private sector brings under trained employees that deal with "papers" and therefore many times the archival legal duties are accomplished "pro forma" with no understanding of rationale of records management practices.
- Lack of strategic vision of the management of National Archives that neglected the issue of e-records and its importance for public bodies. The outcome is quasi-ignorance of electronic records management by archival specialists and public bodies.
- National Archives not positioned to take strategic decision in matters of electronic archival matters
- Management and archival of electronic records decided by IT specialists and politicians, not archivist (ex: MP Pambuccian author of Electronic Archiving Law)
- Badly conceived legislation of electronic archival as a result of the above situation
- Lack of training for archivists. Faculty of Archives doesn't teach ERM and practically no training exists.
- Bad translations of international standards like ISO 15489 as a result of the IT dominance. Some example (record as „recording”, metadata as “electronic card index” etc.) that leads to misunderstandings (DM = ERM), etc
- Software imports that doesn't comply with archival provisions
- Limited implementation of digital public services

Under these circumstances, what would be the challenges of a Romanian version for Moreq2 ?

- **To adopt or to adapt Moreq2?** Each country is different and have different level of knowledge of ERM. Software development might take these differences into account, to best suit their products on market. **Particular Romanian problems:** possible incompatibility between Moreq2 and current legislation (LAE); Chapter zero may not exist, etc. End result: Moreq2 may be ignored or disregarded.
- **Moreq2 and the national RM practice in Romania.** Semantic problems due to lack of understanding electronic records management theory: Romanian archival practice don't use the notion "record"

and other terms currently applied in electronic archival. A lot of terminology needs to be borrowed from other languages and then agreed at national level. End result: Moreq2 may be mistranslated so the translation needs to be checked and agreed. DLM forum needs to have a say in this issue of the correctness of Moreq2 translation

- Who would be the official translator for Moreq2?
- Who would certify the software, if the National Archives has a little experience for e-records, and the IT sector - no experience of ERM practices?

[See the presentation: part 1; part 2.](#)

## MoReq 2 et le cadre roumain

La Roumanie est située dans le sud-est de l'Europe. Elle a été le résultat de l'unification de trois principautés, chacune ayant des approches et traditions différentes en matière d'administration et d'archivage (françaises pour les principautés du Danube, autrichiennes pour la Transylvanie, puis russe sous le régime communiste). La Roumanie est membre de l'Union européenne depuis 2007.

Le cadre d'archivage actuel de la Roumanie, basé sur différents contextes historiques est confronté à la difficulté de maîtriser et conserver des archives électroniques.

**Forces au niveau des archives :**

- Pratiques d'archivage unifiées au niveau national. Compréhension, pratiques et formation identiques pour les archives centrales et locales (comté).
- Pratique d'enregistrement des documents. Cela permet de comprendre et de suivre plus facilement les pratiques de capture d'archives électroniques.
- Classification normalisée et pratiques d'audit.
- La présence et l'influence des Archives nationales sur les créateurs de documents / archives.

**Au niveau informatique :**

- Personnel informatique hautement qualifié en matière de programmation et de développement de logiciels (la Roumanie est un sous-traitant informatique)
- Législation en vigueur concernant les signatures électroniques : loi sur la signature électronique (455/2001)
- Politique des services publics électroniques déjà mise en œuvre

**Faiblesses :**

- Les influences et les intérêts politiques, associés au manque de soutien des Archives nationales de la part du gouvernement affaiblissent l'autorité de l'institution.
- Le secteur privé important émergeant emploie du personnel insuffisamment formé qui traite des documents papier et les tâches d'archivage sont donc souvent réalisées « pro forma » sans comprendre les pratiques de records management.
- Manque de vision stratégique de la part des responsables des Archives nationales qui ont négligé la question des archives électroniques et son importance pour les instances publiques. En conséquence, les spécialistes de l'archivage et les instances publiques ignorent quasiment le records management.
- Les Archives nationales ne sont pas en mesure de prendre des décisions stratégiques en ce qui concerne les questions d'archivage électronique.
- Les questions de gestion et d'archivage de documents électroniques sont décidées par des informaticiens et par des politiques et non par des archivistes (ex : M. Pambuccian, membre du Parlement, auteur de la loi sur l'archivage électronique).
- En raison de la situation décrite ci-dessus, la réglementation en matière d'archivage électronique est conçue de manière inadéquate.
- Manque de formations disponibles pour les archivistes. La faculté des archives n'enseigne pas le records management et il n'existe quasiment aucune formation.
- En raison de la prédominance du domaine informatique, il existe des problèmes de traduction des normes internationales telles que ISO 15489. Par exemple, « record » et « recording », « metadata » et « electronic card index », etc.) qui conduisent à des malentendus (DM = ERM), etc.
- Des importations de logiciels qui ne sont pas conformes aux dispositions en matière d'archivage.
- Une mise en œuvre limitée de services publics numériques

Dans ces circonstances, quelles difficultés découleraient de la mise en place d'une version roumaine de Moreq2 ?



- *Adopter ou adapter Moreq2 ? Chaque pays est différent et dispose de différents degrés de connaissance en matière de records management. Le développement de logiciels devrait prendre en considération ces différences afin d'ajuster au mieux leurs produits au marché. Problèmes spécifiques à la Roumanie : incompatibilité possible entre Moreq2 et la législation en vigueur (LAE) ; il se peut que le chapitre zéro n'existe pas, etc. Résultat : Moreq2 peut être ignoré ou négligé.*
- *Moreq2 et les pratiques nationales en matière d'archivage électronique en Roumanie. Des problèmes de sémantique découlant du manque de compréhension de la théorie de la maîtrise de l'archivage électronique : les pratiques d'archivage roumaines n'utilisent pas la notion d'« archivage » et autres termes appliqués actuellement dans le domaine de l'archivage électronique. Une bonne partie de la terminologie doit être empruntée à d'autres langues puis approuvée au niveau national. Résultat : il se peut que la traduction de Moreq2 soit incorrecte ; elle doit de ce fait être vérifiée et validée. Le Forum DLM doit se prononcer sur la question de la validité de la traduction de Moreq2*
- *Qui serait le traducteur officiel de Moreq ?*
- *Qui assurerait la certification du logiciel, si les Archives nationales ont peu d'expérience en matière d'archivage électronique et le secteur informatique ignore les pratiques en matière de maîtrise de records management ?*

[Consulter la présentation : partie 1 ; partie 2.](#)

## ■ 1. Historical recall

Romania is located in the South-East of Europe and is the result of the unification of three principalities. Two of them (Moldavia and Wallachia), also known as the Danubian principalities, were in 19<sup>th</sup> century disputed between Russia and Ottoman Empire, and both this two powers influenced the local administrations. For instance, the constitutional settlements in 1831/1832 that founded the State Archives, as a **repository for inactive administration's records** and that introduced the **registration system**, were influenced by Russia<sup>1</sup>. On the other hand, Transylvania, formerly autonomous province of the Hungarian Kingdom, at the beginning of 18<sup>th</sup> century became a part of the Austrian Empire and was profoundly influenced by Viennese administration. One can note the introducing of **registration system** in 1772 for all public offices that enforced the discipline in keeping records<sup>2</sup>.

After the Communists took the power, The State Archives was transferred under the authority of Ministry of Interior<sup>3</sup> in 1951, and several Soviet practices in records and archives management were introduced<sup>4</sup>. During the following decades, the responsibilities of State Archives increased; consequently The Romanian Archives Act of 1971 put the most important records management decisions under the State Archives control. All records had to be registered and organisations were required to design a file plan and implement a retention schedule, to be submitted for approval to the State Archives<sup>5</sup>.

In 1996, a new Archival Act came into force<sup>6</sup>. Regardless of social and economic developments, this Act reiterates the main provisions of the 1971 Decree; not surprisingly the regulations are outdated and The National Archives is not able to cope with the market economy and the decentralized public organisations. Until 2007 The National Archives had a monopole over the records and archives management practices, and the professional instruction. This legislative framework blocked the development of the electronic records management practices and of alternative approaches. As a result, the Information Technology took over the management and archival of digital documents, and all standards, processes and procedures being decided not by the National Archives but Ministry of Communication and Information Technology.

<sup>1</sup> See Istorical Arhivelor Statului [The History of State Archives] (1831-1956) in *Arhivele Statului. 125 de ani de activitate*, București: 1957, p. 23.

<sup>2</sup> *Ibid.*, p. 23-24.

<sup>3</sup> Decree 17/1951 in *Official Bulletin* no. 14/1.02.1951.

<sup>4</sup> Monica Vlaicu, *Arhivele Statului Sibiu [Sibiu County Division of State Archives]* in *120 de ani de arhivă publică în Transilvania [120 Years of Public Archives in Transylvania]*, Sibiu: 1996, pp. 39-40.

<sup>5</sup> Decree 472/1971 in *Official Bulletin* no. 164/30.12.1971.

<sup>6</sup> Act 16/1996 in *Official Bulletin* no 71/9.04.1996.

The personnel of The Romanian National Archives is highly qualified in managing historical archives but lacks training in the management of electronic records and digital archives. Romania is confronted with the challenge of managing and preserving digital records with staff more familiar with old Slavonic and Greek than databases and Web environment.

## ■ 2. Strong points

The picture is not that bleak, however. Due to the 50 years of communist regime and its centralized practices, working with standardized procedures in records management and/or archives management is the accepted way of working. Since 1954, regularly updated rules of archival practices<sup>7</sup> and since 1957, of records management practices<sup>8</sup> have been issued. Currently, The Romanian National Archives has 42 county divisions, with all archivists working according to the same regulations and in the same manner. The organizations that create archives have a unified style of managing records, by using the same forms and same procedures all over the country. Standardized practice in managing of records and archives seems, therefore, natural or everyone working with records. Some of these good practices support the management of digital records:

- *Registration of records* is a standard approach for each organisation. There are very rare cases where this practice is not implemented. In fact, it became natural to ask for a “registration number” when a citizen submits a paper to an authority. Migrating to digital records, any Romanian professional would ask for a registration system that would allow to uniquely identifying a certain record. That makes easier to understand the concept of e-records capture and the assigning of an identification number (ID).
- *Standardised classification and audit practices* are implemented in all Romanian organisations. In 1961, The State Archives legislated the disposition procedures<sup>9</sup> and since then, destruction of records is, at least in theory, subject to approval of the State Archives. Since 1971 to date, every organization has a common style of managing records: same forms, same procedures. In latest years, the training of records managers or archivists has been decentralized, but only from control point of view, because the “archival doctrine” is that of the State Archives.
- *The presence and influence of the National Archives to the creators of documents/records*. Since 1957<sup>10</sup>, State Archives had been the role of auditing the records creators on their records centres. In this respect, the National Archives will continue to control the creation and archiving of digital records.

In the field of Information Technology (ICT), Romania had an early start, fractured by the political management mistakes. Different authors attest the beginning of Romanian informatics in the ‘1950s<sup>11</sup>. In 1970, there were published manuals of Informatics for students. Unfortunately, during the last decade of communist rule, the regime put an end to imports of technology, and the development of IT research and industry was stopped. After the fall of Communism, IT Romania turned into a major IT outsourcing<sup>12</sup> country, with highly skilled IT staff in programming and software development working for companies, nationally and abroad. This IT culture is a good starting point for the development and implementation of digital records management systems, providing that IT staff adequate training in archival and records management.

---

<sup>7</sup> See *Buletin de informare* [Documentation Bulletin for State Archives] from 1955.

<sup>8</sup> *Instructiuni generale pentru organizarea si functionarea arhivelor organelor si institutiilor de stat, ale organizatiilor economice socialiste si ale organelor obşteşti* [General Instructions For The Organization And The Activity Of The Records Centers Of The State Organs And Institutions, Of The Socialist Economic Organisations And Collective Organs], Bucuresti: 1959.

<sup>9</sup> *Instrucţiuni pentru expertizarea materialelor documentare* [Instructions for appraising documentary materials], Bucuresti: 1962.

<sup>10</sup> Decree 353/1957 in *Official Bulletin* no. 19/26.07.1957.

<sup>11</sup> History of computer developments in Romania in *Annals of the History of Computing*, IEEE, Volume 21, Issue 3, Jul-Sep 1999, pp. 58-60.

<sup>12</sup> [http://news.cnet.com/2100-1011\\_3-5074725.html](http://news.cnet.com/2100-1011_3-5074725.html); [http://www.setimes.com/cocoon/setimes/xhtml/en\\_GB/features/setimes/roundup/2007/08/06/roundup-st-03](http://www.setimes.com/cocoon/setimes/xhtml/en_GB/features/setimes/roundup/2007/08/06/roundup-st-03); [http://www.businessweek.com/magazine/content/06\\_34/b3998456.htm](http://www.businessweek.com/magazine/content/06_34/b3998456.htm)

### ■ 3. Half step forward

Before joining the European Union, Romania adopted several acts on digital information. By example, there is an Electronic Signature Act (455/2001) and an e-Government policy that has been already implemented<sup>13</sup>. Although digital public services have been developed, the area covered is still limited: in August, 2008, only 5 000 firms and 50 000 individual users have had a digital signature<sup>14</sup>. Also, the impact of e-Government policy is minimal. There is not a coordinated policy among the government bodies for using digital technologies and the e-Government portal is mainly a place for the e-forms published on the web by various public authorities. The only digital transaction accepted is, for instance, the online<sup>15</sup> payment of driving fines. The development of digital governance in Romania is not uniform, going from a complete digital workflow (Register of firms/Registrul comertului)<sup>16</sup>, to digital workflow of files (Ministry of Justice) or simple portal with information (The National Archives).

The development of e-government was influenced by two factors: the IT lobby and the public administration. The professionals or RM and AM were completely ignored by both. Consequently, the management and archiving of digital records was decided by IT specialists and politicians, not archivists or records managers. The most spectacular example is offered by the Law on Digital Archives Law (LAE) a badly conceived act on electronic archives. Suffice to say that the Act doesn't mention the transfer of digital records to the National Archives, by the public or private sector. The problem of authenticity was resolved by requiring digital signatures, with no thoughts on whether these signatures would be read in fifty years or more from now. Maintenance and long term preservation of electronic records are simply not mentioned in the above named Act, which reflects the myopia and short termism typical of the IT industry.

But that was not all. It was followed by the mistranslation of ISO 15489. The lack of knowledge on archival matters and specific terminology resulted in inaccuracies: by example Document Management became equal to Records Management. In a market with no standards or guidelines, a lot of document management applications created or imported have been advertised<sup>17</sup> as records management systems, although they don't comply with archival provisions. Moreover, in many cases, the rules for filing or appraising were set by IT specialists, completely ignoring the records management professionals and their advice.

The National Archives was not particularly interested in the digital issue until recently and as result they were not positioned to take strategic decisions in matters of electronic archives. The Romanian legislation enforces the transfer of records to National Archives is produces over 30 years since the creation of records; in case of technical archives the time is even greater, of 50 years. No digital record has been transferred so far to the National Archives. There are very only a limited number of cases when records have been created in only in a digital environment, and even less selected for permanent preservation.

Archives staff lacks adequate training. The 'digital' training for archivists targeted only databases and their use. The Faculty of Archival Science offered its first course in "Archives and digital heritage" in 2004, taught by an archivist<sup>18</sup>. The profession of records manager doesn't officially exist; there are no higher education courses on electronic records or training in the management of digital archives.

The public administration and the emerging private sector uses poorly trained employees to deal with "papers" and often the legal duties regarding archival are accomplished "pro forma", with no understanding of rationale of records management practices. The end result is quasi-ignorance of electronic records management by public bodies and commercial companies.

Under these circumstances, would be Moreq2 adopted? And what is needed to make room for the provisions of Moreq2?

<sup>13</sup> [www.e-guvernare.ro](http://www.e-guvernare.ro)

<sup>14</sup> [http://www.realitatea.net/50-000-romani-au-o-semnatura-electronica\\_324131.html](http://www.realitatea.net/50-000-romani-au-o-semnatura-electronica_324131.html) (Sept. 20, 08)

<sup>15</sup> [http://www.ghiseul.ro/eKontentGhiseulFiles%5Cacasa%5Cintrebari\\_frecvente%5Cintrebari\\_frecvente%5CCe-pot-plati-prin-intermediul-acestui-portal.html](http://www.ghiseul.ro/eKontentGhiseulFiles%5Cacasa%5Cintrebari_frecvente%5Cintrebari_frecvente%5CCe-pot-plati-prin-intermediul-acestui-portal.html)

<sup>16</sup> <http://e-forms.onrc.ro/>

<sup>17</sup> See, for instance, [http://www.softnet.ro/library/files/papers/Introducere\\_in\\_DRT.pdf](http://www.softnet.ro/library/files/papers/Introducere_in_DRT.pdf).

<sup>18</sup> He was replaced in the next year by an IT specialist!

## ■ 4. Adoption of MoReq2

Each country is different in terms of understanding the electronic records management discipline and practice. ERM system suppliers might take these differences into account, developing their products to suit the local markets.

The main challenge for a country is the cultural one and this is reflected in the archival tradition, archives legislation, the archival and records management practice, the semantic meaning of terminology and the understanding of records management as an academic discipline.

In the particular case of Romania a successful implementation of Moreq2 would require a *cultural shift* from the old cultural model of archiving to a new a new model of managing records and archives. This paradigm shift entails:

a) A *strategic role* for the National Archives of Romania as a leader of information and electronic records management. The Romanian Archives should provide leadership, guidance and support in all areas of information management to the wider Archive community and promote and develop standards and best practice in all aspects of archival and records management.

b) The implementation of the *lifecycle concept* for managing digital information and electronic archives, leaving behind the notion that management of records starts when records are no longer active. In other words, The National Archives needs to accept a different theoretical model for managing active and semi-active archives and be more involved in setting standards and good practices for this type of archives. This is not only a challenge for Romania but also for other countries or even software suppliers, as demonstrated by Microsoft SharePoint where the records repository is separated from the document management system.

This process also entails replacing the flawed Law of Electronic Archives and other inadequate acts with *new legislation*. In the current form, the Romanian LEA is at odds with Moreq2 so the development of a chapter zero for Romania is out of question. Moreq2 is unlikely to be adopted before a legislation change.

c) Starting the *collaboration between the IT industry and the National Archives*, to ensure the long term readability of digital records across government and public bodies, by developing shared services, trusted guidance and innovative, technical solutions.

d) The *recognition of electronic records management* as a distinct domain within the archival science, which uses Information Technology but is not ruled or controlled by IT. Currently, electronic records management is within the IT remit in Romania. Unfortunately, the Romanian IT specialists have no understanding of archival concepts and this is obvious in the ill conceived Law of Electronic Archives mentioned above. Taking electronic records management out of the control of the Romanian Ministry of Communication and Information technology and putting it under the control of the National Archives is the way forward. This move will also let to the recognition of the records management profession, a profession that still not exist in Romania.

e) Introduction of a *new terminology* to express the new electronic records management concepts of lifecycle and its stages and processes, the record as opposed to document, the ERM, EDRM and ECM systems and so on. Romanian Language, like any Romance language of the family (French, Italian, Spanish, Portuguese, etc) is woefully inadequate to articulate the difference between record and document. Some of the translators of the Moreq1 have replaced the word "record" with the word document but this approach is a wrong in the opinions of the authors because it leads to all sorts of errors and misunderstandings. We came also across of a Romanian translation of ISO 15489 where record was translated as "recording" and metadata as "electronic card index", to mention only a few translation errors.

- When it came to new terms, Romanian language has always adapted terms from French or Italian. Unfortunately, Romance languages are lost for words when it comes to convey the semantic meaning of the word "record" in their language.
- A distinct possibility to translate correctly Moreq2 would be to borrow the term "record" directly from English for the documents that are records according to ISO 15489. It is a highly controversial solution, of which the authors have debated the merits for almost a year. All other ways to express the semantics of the word record in Romanian language have failed and it remains the only valid option for the

authors. Whether the Romanian archivists and emerging records managers will accept the new term is still to be seen.

- Related to the issue above is the official translation of Moreq2 in all European languages. Translation is extremely important. In recent times an EU directive has to be withdrawn because the mistranslation of a single word. An accurate translation of Moreq2 in each of the EU languages is critical for its adoption. The questions that DLM has to answer are: Who will do it? And who will check and validate the translation? It is a challenge that awaits DLM Forum's decisions.

f) Implementation of *a training regime* to help Romanian archivists understand the electronic records management domain, its theory, concepts and practices, and give them the knowledge of Moreq2 and skills to evaluate ERM systems. The training development should go in parallel with the development of records management as an academic discipline, distinct but closely related to archival science.

## ■ 5. Conclusions

For the reasons exposed above, Moreq2 is unlikely to be adopted in Romania, at least in the near future. But a translation in Romanian language will raise awareness of the electronic records management discipline and the ERM systems. That is what we try to achieve by translating Moreq2 in Romanian language and time permitting, creating the explanatory documentation to understand the standards.

There is no doubt that once Moreq2 is translated into Romanian, more and more archivists will understand the rules and the challenges of the digital world. This would encourage The National Archives of Romania to take the lead in the management and preservation of information.

The IT industry in Romania will have to take a close look at the Moreq2 specifications and its testing regime, to remain competitive. The Romanian software suppliers entering the world of document and records management would not be able to ignore Moreq2 without losing their ability to export in Europe and market share in Romania.

Taking account that Romanians have shown during their history an uncanny ability to reinvent themselves, adopt new models and modernise quickly, the authors are confident that Moreq2 will have a big impact on the development of electronic records management in Romania.

# RECORDS MANAGEMENT MARKET STUDY FOR GERMANY, AUSTRIA AND SWITZERLAND

---

Dr. Ulrich Kampffmeyer

Managing Director of Project Consult (Germania, Hamburg)

*The presentation is based on several market studies from roadshows (2007, five cities in Germany; 2008, Berlin, Frankfurt, Zurich, Vienna) but focuses on the results of two internet-based researches from May 2008 with different questionnaires for end-users and vendors. The presentation covers:*

- *problems of the RM terminology in German language*
- *adoption of RM in Germany, Austria and Switzerland*
- *different views of endusers and vendors*
- *consequences for market education and MoReq2 dissemination*

[See the presentation.](#)

## **Étude de marché sur la maîtrise de l'archivage réalisée pour le compte de l'Allemagne, l'Autriche et la Suisse**

*Cette présentation est basée sur plusieurs études de marché réalisées à partir de tournées de présentation (2007 : cinq villes en Allemagne ; 2008 : Berlin, Francfort, Zurich, Vienne) et concerne en particulier les résultats de deux études de recherche réalisées sur internet depuis mai 2008 par le biais de plusieurs questionnaires destinés à des utilisateurs finaux et des fournisseurs.*

*La présentation concerne :*

- *les problèmes de terminologie allemande dans le domaine de la maîtrise de l'archivage*
- *l'adoption de la maîtrise de l'archivage en Allemagne, en Autriche et en Suisse*
- *les différents avis des utilisateurs et des fournisseurs*
- *les conséquences sur l'éducation du marché et la diffusion du MoReq2*

[Consulter la présentation.](#)

# TRADUIRE MOREQ2 : LE CAS FRANÇAIS

---

Marie-Anne Chabin

expert en archivage, Archive 17

[Consulter la présentation.](#)





# **Atelier 2**

# **Sécurité et protection des données**

---

---

# ***Workshop 2***

## ***The protection of personal data and the archival requirements***

---

---

# USING AN APPROACH FROM ISO 27001 TO BUILD THE BUSINESS CASE FOR RECORDS MANAGEMENT

---

Richard Jeffrey-Cook

CITP, Managing Director In-Form Consult Ltd

*Information and records are valuable resources within organisations that need to be protected. Many organisations focus on the protection of their electronic information assets against the risks of loss, misuse, disclosure or corruption. This process is commonly referred to as information security management.*

*Information security management enables the sharing of information in a manner that ensures the appropriate protection of that information. Risk assessment and management, continuity planning and disaster recovery programmes should all form a part of any information security management system. The aim is to protect information from a wide range of threats in order to:*

- *minimise the impact of a security breach;*
- *safeguard the accuracy and completeness of information;*
- *ensure that information is accessible only to those authorised to have access;*
- *ensure that authorised users have access to information as, and when, required;*
- *maximise return on investments and business opportunities.*

*Good information security is essential for implementing successful records management. In this paper, Richard Jeffrey-Cook looks at ISO 27001, the international standard for information security management systems. He will demonstrate how the controls that are applied for information security can be adapted to information management.*

## **ISO 27001**

*ISO 27001 (the current version is formally known as BS ISO/IEC 27001:2005) is the international standard for establishing, implementing, operating, monitoring, reviewing, maintaining and improving an Information Security Management System (ISMS).*

*The decision to adopt an ISMS is a strategic decision for the organisation. The design and implementation of the ISMS is influenced by the needs and objectives of the organisation, the processes employed and the size and structure of the organisation. The standard provides a framework for determining what measures (known as "controls") are appropriate. Guidance on which controls are appropriate is included in ISO 17799.*

*The history of the standard is a little confusing. Originally the standard was developed as BS 7799 in 1995 and just included the controls. A second part, formalising the process for creating an ISMS was added and known as BS 7799 (Part 2). The first part was then adopted as an ISO standard becoming ISO 17799. Part 2 was then adopted as ISO Standard 27001 in 2005.*

*ISO 27001 requires a risk management approach to be adopted to determine the priorities for planning information security improvements and deciding what level of resources should be deployed. This session demonstrates how the same approach can be used to justify an information management strategy and to help build the business case for records management improvements.*

[See the presentation.](#)

## **Utilisation de la norme ISO 27001 comme aide à l'élaboration d'un argumentaire pour améliorer la gestion documentaire**

*Informations et documents sont des ressources précieuses dans les organisations et ils doivent être protégés. Nombre d'organisations accordent une place importante à la protection de leurs fonds d'informations électroniques contre les risques de perte, de mauvaise utilisation, de divulgation ou de corruption, processus communément appelé gestion de la sécurité de l'information.*

*La gestion de la sécurité de l'information permet de partager des données de manière à garantir une protection adéquate de celles-ci. Tout système de gestion de la sécurité de l'information devrait comprendre des programmes d'évaluation et de gestion des risques, des programmes de planification de la continuité des opérations et de reprise après sinistre, l'objectif étant de protéger les informations contre un grand nombre de risques afin de :*

- *minimiser l'impact d'une atteinte à la sécurité ;*
- *protéger l'exactitude et l'intégrité des informations ;*
- *garantir que les informations ne sont accessibles qu'aux personnes habilitées à y avoir accès ;*
- *garantir que les utilisateurs habilités ont accès aux informations selon leurs besoins ;*
- *maximiser le retour sur investissements et les opportunités commerciales.*

*Une bonne sécurité d'information est indispensable à la réussite de la gestion documentaire. Dans cet article, Richard Jeffrey-Cook fait une brève présentation de la norme internationale ISO 27001 sur les systèmes de gestion de la sécurité de l'information. Il montre comment les contrôles qui sont appliqués à la sécurité de l'information peuvent être adaptés à la gestion des informations.*

### **La norme ISO 27001**

*ISO 27001 (la dénomination officielle de la version actuelle est BS ISO/IEC 27001:2005) est la norme internationale dédiée à l'établissement, la mise en oeuvre, l'exploitation, la surveillance, le contrôle, le maintien et l'amélioration des systèmes de gestion de la sécurité de l'information (ISMS).*

*L'adoption d'un ISMS est une décision stratégique pour une organisation. La conception et la mise en oeuvre d'un ISMS sont dictées par les besoins et les objectifs de l'organisme, par les processus utilisés ainsi que par la taille et la structure de l'organisme. La norme fournit un cadre qui permet de déterminer les mesures (appelées ici "contrôles") les mieux adaptées. Les bonnes pratiques auxquelles sont soumis les contrôles sont incluses dans la norme ISO 17799.*

*L'histoire de cette norme n'est pas très claire. Elle a d'abord été élaborée en 1995 sous la référence BS 7799 et ne traitait à l'époque que des contrôles. Une seconde partie formalisant le processus de création des ISMS a ensuite été rajoutée sous la référence BS 7799 (2ème partie). La première partie a alors été adoptée comme norme ISO et est devenue ISO 17799. La 2ème partie de la norme a ensuite été adoptée en 2005 sous la référence ISO 27001.*

*La norme ISO 27001 exige une approche «gestion des risques» pour définir les priorités de planification des aménagements de la sécurité de l'information et décider du niveau de ressources à mettre en place. Cette session montrera comment une seule et même approche peut servir à justifier une stratégie de gestion de l'information et contribuer à l'élaboration de l'argumentaire pour améliorer la gestion documentaire.*

[Consulter la présentation.](#)

Records managers and archivists can often find it difficult to justify the resources they need to undertake their corporate role. By understanding and borrowing techniques identified in ISO 27001, the information security management standard, it is possible to start quantifying benefits associated with operational and reputational risk and develop a framework of information and records management controls justified on reducing the risk. This is also a very relevant and topical issue with the current turmoil in the financial markets which identifies poor accountability and risk as a vital factor in the resulting crisis.

## ■ What is ISO27001

Any organisation requires information in order to operate. This information may be held as physical documents or on computer systems. It can be held in a wide range of formats from documents and e-mails through to audio or video. Information security is the term applied to protecting information and information systems from unauthorised access, use, disclosure, disruption, modification, or destruction.

The aim of an information security system is to protect the confidentiality, integrity and availability of the organisation's information.

- Confidentiality - is about ensuring that only authorised people have access to the information.
- Integrity - is about ensuring that the information is not tampered with or modified in any unauthorised fashion.
- Availability - is about ensuring that the information is available to the people who require it when it is needed.

ISO 27001 (the current version is formally known as ISO/IEC 27001:2005<sup>1</sup>) is the international standard for establishing, implementing, operating, monitoring, reviewing, maintaining and improving an Information Security Management System (ISMS).

## ■ The Benefits of Records Management

The benefits of records management can be broadly summarised into three categories: compliance with legislation and regulation; achieving good governance, the ability to demonstrate good decisions have been taken based upon good information and improving efficiency.

The problem when producing a business case is that the tangible (or cashable) benefits achieved from introducing a record management system are often quite small. The intangible (or non-cashable) benefits are greater and the strategic benefits are usually the most significant. The result in the United Kingdom has been for many organisations to be reluctant to invest in records management systems because they do not see a direct return from the investment.

Compliance with legislation and regulation should be a strong driver for good records management. The challenge for any business case is that the benefit cannot be measured directly. A failure to comply may have no adverse affect on an organisation for many years. The lack of compliance is a risk and management will often ignore risks when taking day to day decisions.

## ■ What is Risk?

Risk may be divided into two main categories, operational risk and reputational risk. Basel II defined operational risk<sup>2</sup> as “the risk of losses resulting from inadequate or failed internal processes, people and

---

<sup>1</sup> The standard may be obtained from [http://www.iso.org/iso/catalogue\\_detail?csnumber=42103](http://www.iso.org/iso/catalogue_detail?csnumber=42103).

<sup>2</sup> Basel Committee on Banking Supervision, 2001b. Working Paper on the Regulatory Treatment of Operational Risk.

systems, or external events”. Operational risk is in the category of “soft risks”. The rationale for referring to them as ‘soft’ is simply because it’s these risks that attack the soft targets of an organisation i.e. when it’s ‘off-guard’. Examples show that the resulting loss can be catastrophic even including the total loss of the business costing many millions. Operational Risk is all pervasive and present in every aspect of the business in the systems, procedures and personnel employed to administer processes. If a business process is inadequately designed, regardless of its purpose then the risk of loss is an operational risk.

Primary elements of operational risk include:

- Assets - risk of damage, misappropriation or theft of an organisation’s physical, logical and intangible assets.
- Systems - risks of inadequately maintained or designed systems.
- Process - risk caused by inadequate or failed internal processes.
- People - risks arising from an inappropriate level of staff, inadequately skilled or poorly managed staff.
- Business - risks arising from competitor activity, supplier unreliability or customer new demands.
- Environment - risk caused by demographic, macro economic, natural environment, technological or cultural change.
- Political - risks caused by actions of government, public interest groups or extremists.
- Regulatory - risk caused by non-compliance with financial or other regulators.

Reputational risk arises from any activity undertaken by a body that could have an adverse impact on the reputation and standing of that organisation. An adverse reputation directly impacts the way in which it is regarded by customers, suppliers, employees, investors, interest groups, regulators and government. An organisation that loses a factory to a fire can get it replaced through insurance. An organisation that loses its reputation is likely to go out of business.

Operational and reputational risks are linked in that when an operation fails it leads to loss of reputation. For example if a retailer repeatedly fails to honour warranty claims, customers will soon regard them as untrustworthy. The fundamental difference between operational and reputational risk is that an organisation owns its processes and as therefore owns its operational risk. It is the public that owns the reputation of the organisation. This difference, unless fully understood, is where the real danger lies because the organisation can only ever strive to manage their reputation in the world as a whole through its actions.

## ■ Records Management and Risk

Integral to the design of all business process must be the internal controls designed to manage the operational risks arising from that business activity. The management of operational risk is concerned with the adequacy, efficiency and integrity of the controls and other risk mitigating tactics established to facilitate the business processes. Its fundamental objective is to protect the assets, earnings and reputation of the organisation from the adverse consequences of process control failure.

Records managers and archivists have an important role to play because they are the guardians of corporate memory. To be effective in reducing operational risks, they need to be involved in the entire information lifecycle and establish record management procedures from the capture and creation of information to its secure disposal.



## ■ Learning from ISO27001

ISO 27001 identifies the elements required for an information security management system. An ISMS policy determines the information that is being protected and what is an acceptable level of risk. Risk assessment is about determining the actual risk. The standard does not mandate a particular approach to risk assessment.

A typical approach will look at the risks associated with Confidentiality, Integrity and Availability. The risks are determined by a consideration of the impacts that would occur if some threat exploits a weakness in the defences to compromise the security of an asset, and how likely is the impact to occur. The risks are then evaluated; basically priority is attached to the more serious risks and the more likely chances that they will occur.

The risk assessment approach is extremely useful to quantify the benefits from improving information management controls. The assessment may indicate that the absence of records management controls are likely to lead (for example) to a 30% chance in the next year of a compliance failure costing an expected £ 20,000 to correct. At the simplest level, this might be used to justify expenditure on mitigation. It can also be used to identify where the greatest risks to the organisation exist and thus where the priority for improvements should be focused.

After completing the risk assessment, the standard requires an approach to risk treatment. This can be by:

- taking steps to mitigate the impact of the risk;
- taking steps to reduce the chance of the risk occurring;
- transferring the risk to someone else (for example paying them to manage a service or insurance) or
- accepting the risk (because mitigation is too expensive).

The ISO 27001 standard requires a risk treatment plan is produced. This is a project plan for introducing any new controls that might be needed.

ISO 27001 is accompanied by ISO 27002. This presents a list of candidate control objectives and controls. The list is not exhaustive and additional control objectives and controls may be added. Not all of those listed may be relevant to every ISMS. The standard requires an organisation to identify all the security controls they have chosen and justify why they are appropriate, and show why those controls that have not been chosen are not relevant. This must be related back to the risk assessment, but the selection of controls can also be related to statements in the ISMS policy.

## ■ Information Management Controls

ISO 27002 identifies over 140 possible information security controls. Many of these controls will either relate directly to records management or will have a similar control but applying to records management instead of information security.

An example of a control that relates directly to records management is A7.2.1 Classification guidelines: "Information shall be classified in terms of its value, legal requirements, sensitivity and criticality to the organization".

An example of an information security control that will have an obvious parallel in records management is A.5.1.1 Information security policy document: "An information security policy document shall be approved by management, and published and communicated to all employees and relevant external parties". For records management this could be re-written as "A records management policy document shall be approved by management, and published and communicated to all employees and relevant external parties".

It can be argued that ISO 15481, the international record keeping standard, would be improved by the inclusion of a set of possible information management controls. This would provide greater guidance to records managers and archivists and would allow greater scope for certification of record keeping systems.

The following list may be considered as a set of controls for starters:

- Information Management Policy Document. Does the organisation have an Information Management Policy? If so, how aware are staff and to what extent do they follow it?
- Inventory of information assets. The organisation should maintain an inventory of information assets.
- Access Control Policy. Does the organisation have a policy of who should have access to what information?
- Classification Guidelines. Does the organisation have a consistent and published approach for protective marking?
- Information management education & training. Has the organisation taken steps to ensure all employees are familiar with their responsibilities?
- Documented procedures. Does the organisation have documented procedures for handling enquiries so that information is not lost?
- Security of Electronic Mail. Does the organisation have an appropriate e-mail policy and measures to ensure information confidentiality is maintained?
- Mobile Computing. Does the organisation have policies to support mobile computing?
- Operational Information Change Control. Does the organisation have proper change control procedures as new information systems are established?
- User Id and Authentication. Procedures for identification and logging-on, and passwords.
- Information Back-up. Ensuring information can be recovered if corrupted or lost.
- Business Continuity Management Process. Can the organisation maintain compliance in the event of a business continuity incident?
- Including Information Management in Job Responsibilities. The responsibilities of employees and third parties with regard to DPA, Companies Act and other legislation should be clearly stated.
- Confidentiality Agreements. An approach to confidentiality agreements, their authorisation and enforcement should be documented.
- Identification of Applicable Legislation. Does the organisation have a process to remain abreast of relevant legislation?
- Specialist advice. There must be appropriate procedures for obtaining specialist advice, for example in addressing DPA enquiries.
- Collection of Evidence. Measures to collect evidence for legislative requirements.
- System Audit Controls. Measures to audit information systems and procedures to ensure staff are following practice.

The controls listed above are not comprehensive and not all of the controls are necessarily appropriate to every organisation. ISO 27001 requires a Statement of Applicability to be created by the organisation to document which controls are applicable to the organisation and the reasons why a control is not applicable. The standard encourages an organisation to identify other controls and the controls themselves must be reviewed on a regular basis.

## ■ The Link between Information Management and Information Security

As should be obvious from the list above, there is a close connection between information management and information security. One relies upon the other. Good information security is impossible without good

information management - you cannot protect your information assets if you do not know what assets you possess! Similarly, good information management is impossible without good information security - nobody will use an information store if they do not trust it to store their information safely so that they can retrieve it later.

There is also quite clearly a link through organisational culture between information security and information management. Organisations with little regard for information management will suffer from poor information security. The recent spate of data security breaches<sup>3</sup> in the United Kingdom derives in part from the lack of interest directors and senior managers take in information management.

## ■ Conclusions

Archivists and Records Managers must be aware of the importance of good information security. Understanding the techniques and standards of Information Security Managers should help them to identify the information management controls which may be missing or need improvement and provide a framework for demonstrating the tangible benefits which will be obtained. Archivists and Records Managers can only do a good job and contribute to their organisation if they can demonstrate the benefits of what they are doing.

---

<sup>3</sup> The loss of two data discs containing personal details of 20 million benefit claimants by Her Majesty's Revenue and Customs has triggered a great deal of media interest and resulted in many further examples of organisations losing personal data.

# SÉCURITÉ DES DONNÉES : LES SUPPORTS ET LES FORMATS SONT-ILS AU COEUR DE LA PROBLÉMATIQUE ?

---

Jean-Marc Fontaine

Ingénieur – Laboratoire d’Acoustique Musicale (LAM) (Ministère de la Culture)

Jacques Perdereau

Ingénieur – Laboratoire National d’Essais et de Métrologie (LNE)

Gérard Weisz

Consultant – Sirius System France

*Les risques de perte d’informations inhérents à la conservation des objets numériques sur le long terme constituent une des préoccupations récurrentes des responsables de systèmes d’archivage électronique (SAE) dès l’initiation du projet et tout au long de la vie de celui-ci.*

*Cette problématique se matérialise notamment au travers de la question des formats qu’il convient de privilégier et des types ou de la nature (CD-R, DVD-R, magnétique, etc.) des supports informatiques utilisés pour le stockage des informations mais également dans les procédures qui devront être mises en place pour surveiller le système, prévenir les causes et corriger les effets pouvant porter atteinte à la pérennité des archives.*

*Devant l’importance des enjeux et l’ampleur de la tâche à accomplir pour assurer la pérennité des informations archivées des efforts significatifs ont conduit en France à la constitution du Groupe d’intérêt scientifique (GIS-DON) pour analyser plus particulièrement la problématique des supports optiques : qualité des supports proposés par l’industrie, stratégie de sélection, compréhension des phénomènes de dégradation.*

*Le GIS-DON associe les compétences des universités Pierre-et-Marie Curie / IJLRA / LAM (Paris) et Blaise Pascal / LPMM (Clermont-Fd), du Laboratoire Nationale d’Essais (LNE) et du Laboratoire d’électronique de technologie de l’Information (CEA-LETI, Grenoble).*

*Cette communication se propose de faire le point, au plan normatif, sur les nouveautés introduites par la révision 2008 de la norme NFZ42-013 dans le domaine de l’archivage et, au plan de la recherche de faire une présentation des données recueillies au cours de plusieurs années d’investigation sur la qualité de collections patrimoniales constituées de disques WORM, sur le comportement de disques tests placés dans des conditions de simulation d’agression ou de vieillissement accéléré ainsi que les premiers résultats d’une récente étude d’approche statistique du vieillissement.*

[Consulter la présentation : partie 1 ; partie 2 ; partie 3 ; partie 4 ; partie 5.](#)

## **Are media and formats the crucial issue?**

*The danger of losing information, inherent to the long-term conservation of digital objects, is one of the most frequent concerns of those responsible for electronic archive systems, from a project's kick-off and throughout its lifetime.*

*This issue arises particularly through the question of the formats to be chosen and the type or nature (CD-R, DVD-R, tape, etc.) of electronic media used for storing information, as well as the procedures to be implemented for overseeing the system, and detecting and correcting problems that could have an adverse effect on permanent conservation.*

*In France, the significance and scale of the task to be accomplished to ensure permanent data conservation have led to the setting up of a scientific interest group (GIS-DON) to study the problems raised by optical media: the quality of media available from industry, selection strategies, understanding how degradation can occur.*

*The GIS-DON brings together qualified personnel from the following Universities and Laboratories: Pierre-et-Marie Curie / IJLRA / LAM (Paris) and Blaise Pascal / LPMM (Clermont-Ferrand), the Laboratoire Nationale d'Essais (LNE) and the Laboratoire d'électronique de technologie de l'Information (CEA-LETI, Grenoble).*

*The presentation will examine the current situation concerning standards (and the new features introduced by the 2008 revision of the NF Z 42-013 standard in the field of archiving) and research (with a discussion of the results of several years of investigating the quality of heritage collections held on WORM disks, the behaviour of disks placed in conditions simulating aggression or accelerated aging and the initial results of a recent study using a statistical approach to aging.)*

[See the presentation : part 1 ; part 2 ; part 3 ; part 4 ; part 5.](#)

## ■ Introduction

Une étude récente d'une université américaine donne une idée des volumes de données gérées (créées, communiquées, stockées) de nos jours. Il en ressort les chiffres suivants :

- Volume de données générées par chaque être humain : 1 To en 2002.
- Contre 250 Mo en 1999.
- En 2000, entre 610 et 1.100 milliards de mails (soit 11.000 à 20.000 To) étaient échangés entre 500 millions de boîtes e-mails. Mais 90% d'entre eux ont heureusement été supprimés très vite.
- Nombre de sites Web recensés : 27 millions en janvier 2001 contre 18.000 en janvier 1995.

L'accroissement du volume des données a entraîné de nouvelles nécessités en matière de sauvegarde et de conservation des informations sous forme numérique.

Les spécialistes s'interrogent sur l'efficacité des moyens utilisés pour conserver ces fonds numériques qui représentent un réel héritage culturel et économique, principalement en raison de l'absence de garantie suffisante malgré les investissements importants consentis pour assurer leur pérennité.

Les utilisateurs sont conscients des risques de dégradation ou de perte des documents numériques mais ne les évaluent qu'à travers l'apparition d'accidents, peu aisément quantifiables et encore moins prévisibles.

Les nouvelles technologies de l'information mettent en avant les performances instantanées plutôt que la qualité et la sûreté de fonctionnement sur la durée, pour les utilisateurs comme pour les fournisseurs.

Cette communication a pour but de faire le point sur les récentes études qui ont permis d'avancer dans la connaissance des performances intrinsèques des différents supports de stockage du point de vue notamment de la capacité à accéder à l'information enregistrée et de fournir quelques pistes de réflexion afin de minimiser les risques de perte sur la durée.

## ■ Les données du problème

La conservation des informations numériques et leur accessibilité dans le temps représentent une difficulté pernicieuse dans la mesure où la dimension technologique, inhérente à la nature même de l'information numérique, oblige à construire de nouvelles méthodologies qu'aucun recul n'est en mesure de valider efficacement.

A priori les risques techniques auxquels est confrontée la conservation des objets numériques sont de trois types :

- La dégradation des supports,
- L'obsolescence des formats d'encodage et, concurremment, la disponibilité des outils permettant d'interpréter ces formats.
- L'obsolescence des environnements matériels et logiciels dans lesquels l'information a été créée ou stockée,

L'absence de prise en compte d'un de ces types de risques dans une stratégie de conservation de l'information numérique peut conduire à la perte de celle-ci soit par la stricte impossibilité d'accéder aux données enregistrées pour les exploiter, soit d'y accéder avec une qualité dégradée, soit d'y accéder mais sans disposer des outils permettant de restituer un contenu informationnel intelligible.

## ■ Analyse

Au regard des risques évoqués ci-dessus, les premiers retours d'expérience tendent à démontrer que les solutions opérationnelles sont de natures techniques, organisationnelles et procédurales et que la pérennité de l'information numérique se pose dès sa création.

En effet, quelles que soient les caractéristiques des systèmes d'archivage électronique, la qualité de la conservation numérique et la capacité à exploiter l'information sur toute la durée prévue s'appuie sur les constituants de l'information elle-même (données et métadonnées), la chaîne de traitement et son environnement de fonctionnement depuis la création jusqu'à l'enregistrement sur les supports et enfin sur la surveillance et le pilotage de l'espace de stockage des données proprement dit.

Dans le cadre de cette communication nous avons concentré nos réflexions principalement sur l'analyse des différents types de supports de stockage au regard des données objectives recueillies sur leur capacité à restituer l'information enregistrée en fonction de différentes périodes de rétention et sur la problématique du choix du format d'encodage en tenant compte de la nature de l'information numérique conservée (document numérique, image fixe, son, vidéo).

Notre analyse tiendra compte également des spécifications de la norme AFNOR NFZ42-013 en tant que référentiel permettant de guider les responsables de projets d'archivage électronique dans les choix techniques et organisationnels qu'ils seront amenés à faire.

### Les supports

L'état du marché en matière de supports de stockage d'informations numériques commercialement disponibles peut se répartir selon la typologie et les technologies suivantes :

	Type de support		
Nature du support	WORM physique	WORM logique	Réinscriptible
Amovible	Optique	Magnétique Magnéto-optique	Optique Magnétique Semi-conducteur
Fixe		Magnétique	Magnétique

Dans cette typologie, il est entendu les caractéristiques suivantes :

- Pour les **supports amovibles**, le support physique contenant les informations enregistrées est extractible du lecteur-enregistreur,
- Pour les **supports fixes**, le support physique contenant les informations écrites est entièrement solidaire de son système d'enregistrement-lecture et ne peut en être extrait,
- Pour le **WORM physique**, l'information est écrite une fois par un processus de modification physique et irréversible du support et, à l'issue de cette opération, n'est plus modifiable ou effaçable,
- Pour le **WORM logique**, le support est réinscriptible, mais un dispositif logiciel et/ou matériel interdit la modification ou l'effacement d'une information.
- Pour les **supports réinscriptibles**, les informations peuvent être enregistrées, modifiées ou supprimées sans restrictions.



A l'analyse de cette typologie il en ressort principalement trois technologies d'enregistrement que nous allons analyser ci-après du point de vue des supports de stockage, de leur qualité et fiabilité au regard des phénomènes de vieillissement et des incidents de fonctionnement.

### **Enregistrement ferromagnétique sur disques durs**

Les risques auxquels sont soumis les disques durs sont issus de la combinaison de défaillances électromécaniques et électroniques que l'on peut analyser comme suit :

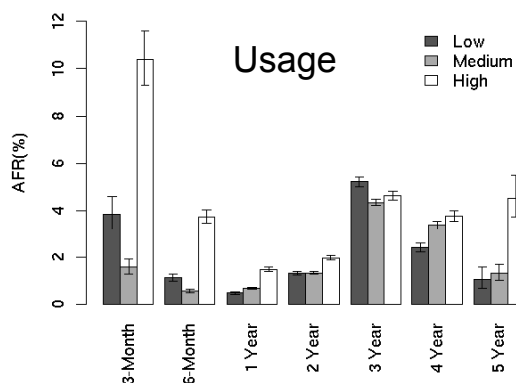
Plusieurs causes de pannes mécaniques dont :

- Têtes de lecture : cassure, contamination, résonance du bras, contact disque (crash),
- Média : rayure, usure, défaut localisé, déformation (voile),
- Moteur, roulements : problème de rotation, d'asservissement, de fatigue en cas de surexploitation (fonctionnement continu),
- Mode d'exploitation : intensité de chargement / déchargement données,
- Manipulation : Chocs (chute), vibrations, variation brutale de température, humidité, projection d'eau, connectique,...
- Pannes électroniques
- Les chocs électriques produits par des surtensions du secteur, des coupures brutales, l'électricité statique sont la cause principale de pannes électroniques de disques durs : les circuits intégrés du disque dur peuvent être endommagés lors d'un pic de tension.
- Problème de connexion (driver et bus)

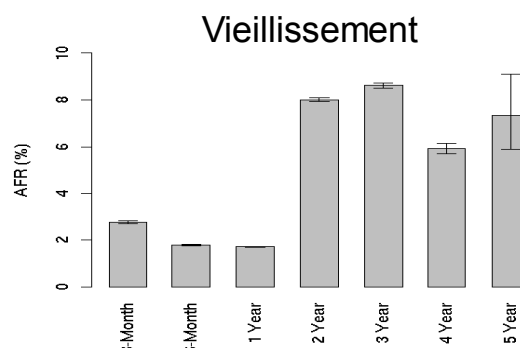
Une étude menée par Google sur une population de 100 000 HDD SATA et PATA (80-400 GB, 5400-7200 rpm) produits après 2001 et maintenu en en fonctionnement permanent a permis de déterminer des tendances malgré la difficulté de dissocier fiabilité et vieillissement lorsqu'on considère des résultats portant sur la combinaison de plusieurs types de fabrication, et des lots différents.

L'utilitaire de contrôle des fonctionnalités SMART fournit un faible score prévisionnel (de l'ordre de 30%) , résultat qui implique que certains critères doivent être reconsidérés car ils sont en contradiction avec l'observation (température trop faible ou trop élevée préjudiciable,) ou se montrent peu pertinents pour décrire une évolution. D'autres critères établis sur les paramètres de réallocations, de détection d'erreurs, etc... confirment leur bonne corrélation avec la probabilité de panne. Enfin, il conviendrait d'apporter des informations complémentaires comme le comportement vibratoire par exemple.

En conclusion de cette étude le rapprochement usage et vieillissement donne une première indication :



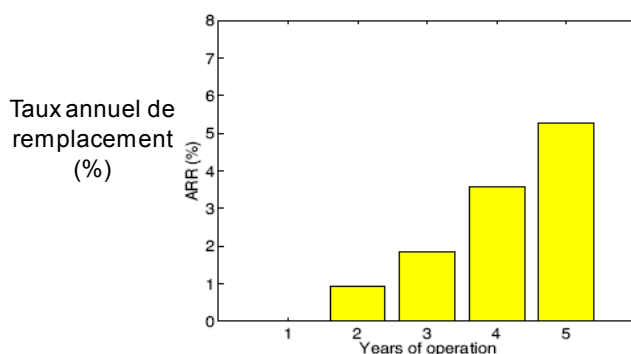
**L'utilisation augmente systématiquement les risques de panne  
L'utilisation intensive s'accompagne d'un effet de *mortalité infantile* important**



**Un effet de *mortalité infantile* est relevé  
L'effet dû au vieillissement est avéré, mais avec des disparités selon fabrication**

Une autre étude menée par Carnegie Mellon University portant également sur un nombre important de disques durs (environ 10000) abouti aux principales conclusions suivantes :

- Les taux fournis par les fabricants en temps moyen avant panne (MTTF: Mean Time to Failure variant de 1 à 1,5 millions d'heures) sous-estiment fortement la réalité en matière d'avaries, surtout au-delà de 5 années d'utilisation
- Un léger "effet de baignoire" (augmentation des avaries en début de vie et à partir d'un certain seuil de vieillissement) doit être pris en compte dans le MTTF
- L'augmentation des pannes commence à ce manifester dès la 2<sup>ème</sup> année d'utilisation.
- Le disque dur constitue la première origine des pannes des systèmes informatiques, mais de nombreux aspects concernant les avaries constatées restent non expliqués : complexité du dispositif et manque d'information constructeur.
- Peu de différence observées entre les types interface étudiés : SCSI, FC et SATA.
- Les conditions d'utilisation priment en regard des composants,
- Des disparités entre fabrications existent, mais l'effet de lots (parfois défectueux) doit être relevé.
- L'analyse du taux de remplacement des disques durs en fonction de leur ancienneté dépasse 5% après 5 ans de fonctionnement.



En résumé les récentes études montrent que les disques durs sont loin d'être aussi fiables que ce qui est annoncé par les fabricants. De plus l'augmentation très importante de la capacité doit être accompagnée par une augmentation tout aussi importante de la fiabilité, sinon le risque de perte de données va s'accroître de façon inquiétante. Si le fait d'incorporer les disques durs dans des systèmes RAID améliorent la fiabilité de l'ensemble, il ne faut pas croire que c'est la solution universelle car des pertes des données ont été constatées aussi dans ces systèmes.

### ***Enregistrement ferromagnétique sur bande***

Les systèmes utilisant les bandes magnétiques comme supports de stockage se caractérisent par des mécanismes complexes (cassette et enregistreur / lecteur) pour assurer, notamment :

- Un positionnement têtes / bande extrêmement rigoureux
- Des tensions optimales de la bande quels que soient les mouvements
- Une parfaite maîtrise des vitesses et des accélérations de la bande
- Ces systèmes exercent des autocontrôles :
  - Par les propres systèmes d'asservissement. Du fait de la disparité des systèmes la normalisation des systèmes et procédures de contrôle n'est pas possible.
  - Par les codes d'erreurs

Des taux d'erreurs importants sont constatés lorsque des encrassements de têtes, des incidents concernant la bande (déformation) ou bien le mécanisme de défilement se produisent.

Aucune étude indépendante des constructeurs n'étant aujourd'hui disponible, il est difficile de fournir une analyse objective de la situation pour ce type de support.

### **Enregistrement optique**

Notre analyse porte sur les média de type CD-R et DVD±R actuellement couramment utilisés, notamment dans les systèmes d'archivage électronique.

Ces supports se caractérisent par des capacités unitaires limitées mais présentent un intérêt du fait notamment des spécificités suivantes :

- Généralisation des Formats
- Produits grand public : longue vie industrielle
- Large champ d'applications
- Des applications exigeantes
- Stockage sur étagères répondant à certaines formes d'utilisation
- Coûts réduits & mise en œuvre aisée

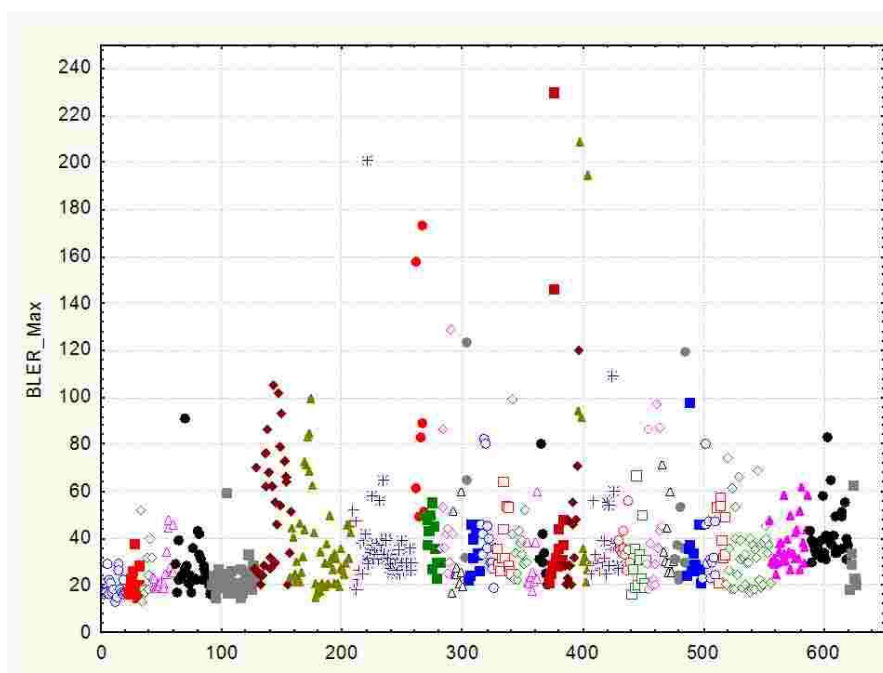
La fragilité des supports optiques est due principalement à leur structure complexe faite d'un assemblage de couches :

- Substrat en polycarbonate
- Surface réfléchissante et semi-réfléchissante dans le cas des disques double couche (DVD)
- Couche active dye ou PC (changement de phase)
- Assemblage avec des colles (pour les DVD)

Le support optique peut être objectivement décrit sur la base des taux d'erreurs obtenus lors de la lecture des données en se basant principalement sur deux indicateurs : le taux de BLER et le Jitter.

Il faut cependant noter que le lecteur de CD utilisé pour évaluer ces taux influence lui-même les résultats.

A titre d'exemple une étude menée sur un ensemble de 626 CD-R répartis en 38 séries conduit, sur la base du taux de BLER, à la dispersion suivante :



*(Cliquez sur l'image pour la voir en pleine taille)*

Des études du LAM et du LNE ont, par ailleurs, montré qu'il existait une grande disparité de comportement entre les différentes marques de disques optiques. Ainsi sur un lot de 60 marques de CD-R gravés entre 2000 et 2002, 15% des CD-R étaient partiellement ou totalement illisibles après 3 ans d'archivage (dont certains disques après moins d'un an d'archivage). A contrario dans l'étude d'un échantillon 1500 CD-R appartenant à une collection de 160 000 disques, les disques les plus âgés (10 ans de conservation) présentaient un parfait état de conservation.

La fragilité relative des supports optiques implique qu'une attention particulière soit portée à leurs conditions de stockage notamment en ce qui concerne la température, le taux d'humidité, la poussière, l'exposition à la lumière, les conditions de manipulation ou la présence de fumée de cigarette,...

La conclusion est que les disques optiques sont de bons média d'archivage mais qu'il faut faire une sélection parmi les marques et comme pour tout média d'archivage assurer un suivi de l'état de conservation des données.

### ***Enregistrement sur semi-conducteurs***

Cette technologie connue depuis de nombreuses années peut être maintenant prise en compte dans le domaine de l'archivage si l'on considère le rapport coût/volumétrie des supports actuellement disponibles.

A priori ce type de support présente de nombreux avantages par l'absence de dispositif électromécanique au niveau lecture/écriture et de faibles contraintes en matière d'hygrométrie ou de température de fonctionnement. Pour l'heure, on dispose de peu d'informations quant à la sensibilité de tels supports aux possibles facteurs d'agression et au vieillissement longue durée.

Les limites connues concernent le nombre de cycles d'écriture auquel une cellule de mémoire flash peut être soumise. En revanche, de telles limites n'existent pas en lecture.

### **Les formats**

Le choix des formats d'encodage des informations archivées est une composante essentielle de la pérennité de leur conservation. En effet, l'exploitation de tout objet numérique est dépendante de la disponibilité d'un logiciel permettant d'interpréter son code pour le rendre intelligible par un humain.

Or l'utilisation de tout logiciel est liée à l'environnement matériel et logiciel pour lequel son fonctionnement a été prévu. En fonction de l'évolution des composants de cet environnement, les éditeurs doivent procéder à des développements ou des adaptations s'ils souhaitent continuer de vendre leurs produits et, en conséquence, les utilisateurs doivent acquérir ces nouvelles versions.

Format et environnement technique sont donc étroitement liés dans la mesure où les éditeurs de logiciels peuvent décider, notamment pour des raisons stratégiques ou économiques, de ne pas assurer la prise en compte de certains formats dans l'évolution de leurs offres. Cette situation peut conduire, dans certains cas, à la perte de documents archivés par l'impossibilité d'exploiter leur encodage.

Dans toute la mesure du possible il convient de choisir des formats d'encodage reposant sur des spécifications publiques accessibles et si possible normalisées. Le choix d'un format de conservation se fera suivant le type de document concerné (texte, image, audio, vidéo) et les caractéristiques qu'on souhaite ou non conserver après conversion dans le format choisi. Il s'agit notamment de savoir comment l'apparence visuelle du document (présentation) doit être conservée, s'il existe des liens vers des documents externes ; si les formules mathématiques ou les macros internes aux documents doivent être maintenues, par exemple.

Le choix du format de conservation et les modalités de conversion doivent en tout état de cause éviter la perte involontaire d'informations significatives.

Les conversions de formats peuvent intervenir à deux stades du processus d'archivage : soit lors de l'entrée dans le système d'archivage, soit une fois les documents archivés au moment où la conversion aura été planifiée ou bien si le format de codage du document numérique archivé devient obsolète et risque d'être un obstacle à sa communication.

Il ne faut pas perdre de vue que le périmètre d'action concernant les formats des documents archivés varie suivant les conditions du contrat passé entre le propriétaire des objets et le service d'archives et la politique d'archivage qui est suivie :

- En entrée du système : contrôle ou non des formats au moment de la prise en charge pour archivage (sur la base d'une table des formats d'entrée reconnus par le système d'archivage) ; conversions ou non de formats à l'entrée sur la base du résultat de ces contrôles ou sur la base des conditions énoncées dans le contrat, en se référant à une table des formats cibles d'archivage. Le contrôle des formats doit se faire à l'aide d'un outil qui permet d'identifier précisément ce format, de le caractériser et de le valider.
- Après entrée dans le système d'archivage : alertes ou non du propriétaire si le format d'encodage des documents devient obsolète ; conversions ou non par le système d'archivage en cas d'obsolescence.

### ***La compression***

La compression permet de réduire les volumes de stockage des données et de réduire les débits de communication de celles-ci. Et donc les coûts.

On distinguera tout d'abord la compression sans perte généralement utilisée dans les systèmes de stockage de masse. Strictement sans perte, le facteur de réduction est généralement de l'ordre de 2. Ces économies sont obtenues grâce à une organisation judicieuse des données binaires de description de l'information (codage entropique)

Souvent, les spécifications d'un produit avancent une capacité double compte tenu d'une telle compression... pour éviter toute ambiguïté, il convient de considérer la capacité native du support, et elle seule.

La compression avec perte concerne les documents visuels ou sonores, elle prend en compte les redondances de l'information en relation avec la tolérance de l'utilisateur vis-à-vis de la qualité de restitution de celle-ci. Les enjeux techniques et économiques sont considérables... Nous ne parlerons que de compression avec perte dans ce qui suit.

Il est important de souligner que la qualité d'un document numérique (transfert d'un original analogique ou création numérique) est déterminée - et ceci de manière définitive - par le choix des paramètres de codage numérique et par celui des traitements de compression mis en œuvre. Nombre de documents n'existent que

sous forme compressée, par exemple les productions audiovisuelles orientées Internet dont la qualité restera limitée et portera les stigmates de leur époque.

Grâce aux moyens de stockage et de communication actuellement disponibles, archiver du texte, des images fixes ou du son ne fait pas nécessairement appel aux procédures de compression, elles sont même fortement déconseillées lorsque l'on dispose, naturellement, de la maîtrise de création du document numérique.

Il n'en va pas de même pour les enregistrements vidéo pour lesquels l'importance des volumétries et des débits nécessite le recours aux opérations de compression. Gain de place, débit, maintien de la qualité de l'information et réduction des coûts constituent les éléments de base des compromis à trouver.

## **Formats audio**

### **Conversion numérique du signal analogique.**

La résolution est définie par la fréquence d'échantillonnage et le nombre de bits affecté à chaque échantillon.

Le débit résultant des principales combinaisons du processus de conversion varie de 1kb/s (32kHz sur 16 bits) à 9,2 kHz (192 kHz sur 24 bits) pour 2 canaux. Le standard correspondant à un débit de 2,3 kb/s (48 kHz sur 24 bits) est largement accepté dans les processus de numérisation des enregistrements analogiques.

Remarque : ces débits ne tiennent pas compte des données supplémentaires (parité pour correction d'erreurs, structure du flux binaire, données de service,...)

Les spécifications normatives se sont adaptées aux développements technologiques : par exemple augmentation de la résolution, augmentation du nombre de canaux de diffusion. Elles s'inscrivent dans le développement des processeurs et d'algorithmes devenus particulièrement performants.

### **Compression audio**

Comme indiqué ci-dessus, les enregistrements audio numériques que l'on se propose d'archiver ne devraient pas faire l'objet de compression. Toutefois, l'usage intensif des enregistrements audio créés, transmis, stockés et diffusés dans des conditions économiques (Internet, baladeur,...) a changé, de manière radicale, la nature de nombreuses collections existantes et en cours de constitution.

Aussi convient-il d'évoquer les codages de compression les plus répandus.

La norme MPEG-1 représente la première génération de codage (1993), elle permet d'effectuer un nombre important de réduction de débit (de 912 à 32 kb/s) sur un ou 2 canaux.

Elle comporte 3 modes correspondant à des taux de réduction croissants :

- Mode 1 : débit de 448 kb/s à 32 kb/s (Fréq. échantil. 48kHz - 44,1kHz - 32 kHz). Un niveau de qualité acceptable est atteint pour une réduction de 1/4.
- Mode 2 (MUSICAM) : encodage différent conduisant à des taux de réduction plus élevés que mode 1, débit de 384 kb/s à 64 kb/s 2 piste stéréo. Un niveau de qualité acceptable est atteint pour une réduction de 1/6 à 1/8.
- Mode 3 (MP3). Un codage et un algorithme supplémentaire permettent des taux de réduction de 320kb/s à 8kb/s pour des fréquences d'échantillonnage de 24 kHz à 16 kHz. Un niveau de qualité acceptable est atteint pour une réduction de 1/10 à 1/12.

La norme MPEG-2 (ISO/IEC 13818-7) ne remplace pas MPEG-1 mais la complète

- Fréquences d'échantillonnage plus basses : 96 kHz à 8 kHz
- Traitement du signal sonore plus efficace

- Peut s'appliquer à un nombre important de canaux : jusqu'à 48
- Comporte 3 niveaux : Main, Low Complexity (LC) et Scalable Sampling Rate (SSR)
- Pour un niveau d'écoute acceptable : réduction de moitié du mode MUSICAM : 1/12 à 1/16

Outre les conditions de codage numérique et de compression, la notion de format des documents sonores s'inscrit dans le cadre de spécifications de formats d'échange (conteneurs) de fichiers audio dans un environnement informatique donné : WAVE (Windows), AIFF (Apple),... L'union européenne de radiodiffusion (UER) a élaborée (1996) une norme d'échanges supportée par le format WAVE intégrant des informations temporelles (Time-Code) et descriptives des contenus (métadonnées) : BWF (Broadcast Wave Format).

### **Formats vidéo**

Toutes les étapes, depuis la saisie du flux d'images jusqu'à l'écran de l'utilisateur final font l'objet de nombreux traitements qui prennent en compte les contraintes d'acquisition, d'enregistrement, de transport, d'usage, de stockage des informations. Deux objectifs sont visés : obtenir un débit le plus bas possible, obtenir une image de la meilleure qualité possible, ceci dans des conditions économiques acceptables. Des objectifs contradictoires qui exigent des consensus qui se traduisent par une multiplicité de formats sans cesse en évolution...

Très schématiquement on distingue 3 étapes de traitement : le codage des composantes primaires, la numérisation, la compression.

Le codage des signaux émis par les capteurs d'analyse des images source distingue les applications professionnelle : extraction des composantes primaires des applications institutionnelles et grand public : assemblage des composantes en un seul signal dit composite sous la forme des formats concurrents : PAL, SECAM et NTSC. On notera la stratégie de répartition géographique de ces standards dans les années 1960, une pratique qui sera reprise par le codage selon les zones internationales des DVD et des Blu-Ray édités.

Les signaux analogiques font ensuite l'objet d'une numérisation. Dans le cas de la conversion des signaux composantes, la sélection et la répartition des échantillons de luminance et de chrominance (différence couleur) sur une ligne (une ligne sur deux peut-être ignorée) définit une structure de référentielle, par exemple 4:2:2 (Bétacam numérique,...), 4:1:1 (DV et DVCPRO,...). Dans le cas des signaux composites, l'échantillonnage porte directement sur le signal porteur de luminance et de chrominance. Lors de la quantification, la longueur du mot binaire décrit l'échantillon sur 8 ou 10 bits généralement).

Les quantités d'informations recueillies et les contraintes liées à la transmission des programmes rendent la compression incontournable. Les techniques de compression reposent sur les redondances d'une image (degré d'uniformité), sur les redondances d'une image à l'autre (degré de changement), sur la qualité souhaitée (tolérance aux imperfections), et enfin sur les possibilités d'économie relatives à la description numérique de l'information.

Les normes relatives au codage, à la conversion, à la compression et à la gestion des contenus constituent des boîtes à outils correspondant aux applications envisagées, par exemple les documents destinés à la sauvegarde de l'information, à l'exploitation dans un contexte donné, au visionnage.

La norme MPEG (Motion Picture Expert Group) :

- MPEG1 : réduction des débits : 1,5 Mbits/s
- MPEG2 : réduction appliquée aux signaux composantes atteint 10 Mbits/s
- MPEG3 vise des débits de 18 à 25 Mbits/s (HD)
- Utilisant des traitements du signal particulièrement élaborés (orientation objet), la norme MPEG-4 (ISO/IEC 14496) a vocation de prendre en charge tout type de programmes multimédia, depuis les plus bas débits (64 kb/s) aux plus élevés (1,2 Gb/s). Ceci pour toutes les applications possibles : création, imagerie scientifique, synthèse sonore, jeux, TV numérique (TNT,...) avec degré d'interactivité élevé.

Parmi les nombreux formats d'enregistrement vidéo numériques normalisés par la SMPTE (Society of Motion Picture and Television Engineers) 2 systèmes ont joué un rôle déterminant :



- La cassette Bétacam numérique (Sony) produite à partir de 1993 constitue une référence en matière de techniques vidéo numériques professionnelles.
- La cassette DV, lancée en 1996 par un consortium de fabricants, destinée au grand public a connu un développement exceptionnel. De format physique réduit (bande 1/4 pouce, l'enregistrement bénéficie d'un système de correction d'erreurs performant, le taux de réduction autorise un débit de 25 Mb/s.

Ce format a été amélioré par la suite (1998) pour atteindre une qualité professionnelle, ceci essentiellement grâce à l'augmentation de vitesse de défilement (DVCPRO par Panasonic, DVCAM par Sony).

Nous retrouvons ces 2 fabricants quelques années plus tard proposer les formats "Full-HD" AVC-HD avec un débit vidéo de 18 Mb/s utilisant pour support le disque Blu-ray, le disque dur ou la carte mémoire.

En matière de compression d'images fixes (en tons continus pour les distinguer des images graphiques), la norme JPEG (Joint Photographic Expert Group) - ISO 10918 définit, depuis son lancement en 1986 des taux variables de compression jugés satisfaisants jusqu'à un facteur 10 (formats source 4:2:2). Elle constitue également un format d'échange de fichiers dont l'usage s'est généralisé.

JPEG 2000 (ISO/IEC 15444-1 et 15444-2) est une norme de réduction de débit à taux variable avec ou sans perte pour les images fixes. Elle met en œuvre de nouveaux traitements qui fournissent une qualité d'image élevée y compris aux faibles débits, utilisant des algorithmes complexes qui prennent en compte des "zones d'intérêt" de l'image par exemple. Orientée vers les applications professionnelles elle prend en compte (11 parties) la gestion des données descriptives des traitements et des contenus, la transmission sécurisée, l'interactivité, les compatibilités d'échanges de fichiers,...

En fait, toutes les normes sont établies seulement du point de vue du décodeur. Les spécifications ne concernent pas l'encodage, elles permettent au fichier compressé d'être décodé. Aussi faut-il s'attendre à des différences de qualité parfois significatives selon le système de décodage utilisé.

## ■ Conclusion

Ainsi que nous l'avons évoqués, dans le cadre de leurs activités, les organismes et les entreprises utilisent de plus en plus des objets numériques, produits par ces organisations ou reçus en provenance d'autres organismes ou d'autres entreprises. Tout ou partie de ces documents peuvent ou doivent être conservés dans des systèmes informatiques pour satisfaire à des obligations légales, réglementaires ou patrimoniales.

Ces systèmes ont pour objectifs de répondre aux attentes des organisations afin d'optimiser l'archivage de leurs documents numériques, de garantir leur pérennité, leur intégrité et de faciliter les opérations de recherche, d'exploitation et de communication des informations.

Un des aspects structurant de la révolution numérique à laquelle nous participons chaque jour est la séparation technique de l'information et de son support. Cette séparation constitue une rupture culturelle majeure dans l'histoire de l'humanité sur ces trois derniers millénaires.

Parallèlement à cette séparation, la notion de complémentarité entre le contenant et le contenu de l'objet numérique s'impose désormais pour la spécification complète de celui-ci.

Dans la mesure où la manipulation de ces différents concepts et leur mise en œuvre dans les systèmes d'informations représentent un enjeu fondamental pour les responsables de projet d'archivage, le recours aux normes du domaine constitue une aide indispensable dans la structuration des projets et un moyen efficace pour maîtriser les offres du marché.

Il s'agit notamment de la norme NFZ42-013 (révision 2008 à paraître prochainement) pour les aspects technologiques des SAE et de la spécification MoReq2 pour les aspects organisationnels.

En complément des aspects techniques, organisationnels et méthodologiques, la maîtrise des coûts de conception et d'exploitation des systèmes d'archivage électronique représente une composante essentielle dans la préservation des patrimoines numériques.

En effet, les choix, notamment en matière de supports et de formats, peuvent induire des charges importantes, s'ils ne sont pas judicieux, lors des opérations de migration ou de conversion dont on sait qu'elles devront être reconduites périodiquement, ce qui implique une continuité des financements.

Force est de constater que les connaissances scientifiques en matière de pérennité de la préservation des informations ne progressent pas de façon cohérente avec la dynamique du domaine car le calendrier des industriels ne correspond pas à celui des utilisateurs ou des chercheurs. Des antagonismes d'échelle de temps de production (fondée sur le court terme) et d'utilisation des documents numérisés (lorsque le long terme est visé), devront être surmontés-

Il est nécessaire, si nous voulons transmettre nos connaissances aux générations qui nous succéderont, de définir un cadre opérationnel permettant de préserver le contenu informationnel de tous les objets numériques qui font et qui ont fait notre quotidien.

# **Atelier 3**

## **Comment choisir les bons formats ?**

---

---

# ***Workshop 3***

## ***File formats for***

### ***long-term preservation***

---

---

# CHOISIR LES FORMATS D'ARCHIVAGE ADAPTÉS

---

Philippe Martin

Administrateur de l'Aproged, responsable du Pôle Normalisation,  
Directeur Associé au Bureau van Dijk Ingénieurs Conseils

*Le choix des formats d'archivage des fichiers est un élément essentiel d'une politique d'archivage pertinente.*

*Deux facteurs principaux gouvernent ce choix : la nature des documents et des informations archivés et le processus de collecte des archives ; la pérennité des formats et leur interopérabilité..*

*Les sources les plus courantes d'archives numériques sont les applications bureautiques, les flux provenant des applications de gestion, les applications produisant des images y compris par la voie de la numérisation. Des sources moins connues, mais produisant des volumes importants sont également à prendre en compte le cas échéant : les systèmes d'information géographique, les systèmes de CAO et la gestion des plans. Le développement de nouveaux canaux de communication ajoute des besoins d'archivage nouveaux : courrier électronique, message interactif, messages vocaux, séquences vidéo, sites web, etc. L'organisation de la collecte des archives doit intégrer la migration des formats, immédiate ou programmée. La numérisation des documents sur support physique est encore souvent un point de passage obligé.*

*La pérennité des formats implique une indépendance par rapport aux formats numériques natifs utilisés par les éditeurs de logiciels. La stratégie actuelle des acteurs majeurs consistant à publier les caractéristiques de leurs formats sous de normes internationales est une démarche intéressante, même si elle conserve quelques limites. La recherche d'une pérennité à long terme peut introduire des contraintes importantes dans la préparation des archives numériques : neutralisation des liens hypertextes intégrés dans les documents, insertion des polices de caractères utilisées dans le document, traitement spécifique des documents signés numériquement, etc.*

*La conférence présentera un panorama des situations les plus fréquentes et des solutions applicables afin de parvenir à produire des archives garantissant à la fois la conservation et l'accessibilité à long terme au contenu informationnel.*

[Consulter la présentation.](#)

## ■ Introduction

Le choix des formats d'archivage numérique est un élément technique clé d'une politique d'archivage. Ce choix doit garantir la pérennité de l'accessibilité aux contenus archivés en minimisant les coûts des opérations d'archivage et de maintenance des archives sur le long terme.

Au cours des dernières années, une meilleure analyse des contraintes et l'arrivée de nouvelles options permettent aujourd'hui de proposer des orientations claires sur les choix possibles et les précautions à mettre en œuvre pour démarrer et réussir un projet d'archivage électronique à long terme.

Cette conférence présente successivement : les postulats de base pour le choix des formats d'archivage, les critères de choix, les options possibles. Un zoom est proposé sur les différents formats Adobe PDF qui sont déjà normalisés ou en voie de l'être. Quelques repères méthodologiques sont préposés à la fin de la présentation.

## ■ Postulats de base

L'objectif d'une solution d'archivage électronique est de garantir la conservation et l'accessibilité des contenus archivés sur le long terme. Cette première assertion contient plusieurs idées qui doivent être explicitées.

L'accessibilité aux contenus implique d'offrir plusieurs services.

Tout d'abord un moyen de recherche qui reposera essentiellement sur les métadonnées associés à chaque document archivé. Sur le plan technique, ces métadonnées pourront être contenues dans chaque document archivé et/ou dans un référentiel externe (base de données, instrument de recherche). En complément, la recherche pourra porter sur le contenu textuel des documents archivés. Ce qui suppose qu'il soit possible d'indexer ce contenu, donc que le document soit codé en mode caractères ou que sa version image soit convertie en mode caractères au moyen d'un logiciel d'OCR (Reconnaissance Optique de Caractères).

Le document retrouvé doit pouvoir être lu sur écran ou imprimé pour qu'un humain puisse en exploiter le contenu. Cela suppose donc que le format d'archivage puisse être interprété par un système informatique pour restituer une version lisible.

Face à l'évolution constante des systèmes informatiques, la garantie d'une accessibilité à long terme implique d'utiliser des formats d'archivage dont toutes les spécifications sont publiées afin d'être capable soit de créer des outils d'affichage fonctionnant dans des environnements informatiques futurs inconnus aujourd'hui, soit de faire migrer les contenus archivés vers des formats nouveaux compatibles avec les environnements futurs. La création de normes internationales rendant publiques les caractéristiques des formats numériques est un facteur essentiel pour garantir le maintien de l'accessibilité à long terme.

L'utilisation de formats d'archivage normalisés permet de s'affranchir des caractéristiques propres aux systèmes de création des documents archivés. Il convient cependant de vérifier que les formats d'archivage choisis n'entraînent pas une dégradation de la qualité de la restitution ou des fonctionnalités offertes.

Le choix de formats d'archivage numérique pérennes renvoie au second plan le problème du choix du support de conservation des données archivées. Il est aujourd'hui admis que la conservation à long terme des données peut être réalisée sur des supports informatiques réinscriptibles à condition de mettre en œuvre des procédures techniques garantissant la sécurité des données : enregistrement des données sur plusieurs supports et sur plusieurs systèmes répartis sur des sites distants, contrôle des accès aux systèmes d'information.

Enfin, tout ce qui précède repose sur l'hypothèse implicite que les données sont conservées par des systèmes permettant la recherche et la consultation des documents sur un poste de travail informatique conventionnel actuel ou futur. Les solutions consistant à enregistrer les documents sous forme d'images sur des supports micrographiques ne sont pas prises en compte dans la présente conférence, quelles que soient leurs avantages particuliers.

## ■ Critères clés pour le choix

Plusieurs points fondamentaux doivent être étudiés lors de la conception d'une solution d'archivage et, en particulier, pour le choix de formats d'archivage.

La nature de l'information à archiver (le contenu) est le premier point à analyser. Il peut s'agir : de texte, de graphiques statiques (illustrations, dessins industriels) ou de graphiques animés en 2D ou en 3D, d'images (photos à tons continus), de son (musique, discours, messages vocaux), de vidéo (analogique ou numérique) et de contenus mixtes. Pour chaque type d'information, un ou plusieurs formats numériques sont utilisables pour l'archivage.

Ces données peuvent être provenir de différentes sources qui déterminent le format des données disponibles pour l'archivage.

Actuellement, les outils bureautiques sont une source majeure de données susceptibles d'être archivées. La multiplicité des logiciels et des conditions d'utilisation créent une situation complexe dans la perspective d'un projet d'archivage organisé à partir des données produites.

L'archivage numérique des documents disponibles uniquement sur support physique (papier, film, etc.) implique la numérisation des documents avec un numériseur adapté aux caractéristiques physiques du support disponible. Des traitements complémentaires des images obtenues peuvent être nécessaires pour en réduire le volume (compression) et en extraire le contenu textuel (OCR / LAD).

Certains outils spécialisés produisent des données qui vont appeler des solutions spécifiques et des formats d'archivage adaptés : messagerie électronique, logiciels de CAO et DAO, appareils de photo numérique, appareils de prise de sons et caméra vidéo.

Enfin l'archivage des bases de données pose un problème majeur.

Pour ce dernier cas, les bases de données, mais également pour les autres, il faudra choisir entre conserver le contenu (un flux de données représentant toute l'information utile) ou conserver le document présentant l'information sous une forme plus habituelle (mise en page, logo, etc.). Par exemple, faut-il archiver le flux de données permettant d'imprimer un ensemble de factures ou les factures elles mêmes avec toutes données redondantes (mentions légales, fonds de page, etc.). Le volume archivé peut varier d'un facteur 10 selon l'option choisie.

La portée des services à rendre en matière d'accessibilité est également un point structurant. S'agit-il simplement de permettre la lecture des documents à l'écran ou en sortie d'imprimante ou faut-il répondre à d'autres besoins tels que la recherche sur le contenu textuel. Dans le premier cas, une simple numérisation en mode image sera satisfaisante, dans le second cas, il faudra faire un traitement OCR, applicable aux documents imprimés, mais pas aux documents manuscrits.

La gestion des métadonnées dans le processus d'archivage est une question clé. Certains formats permettent l'intégration des métadonnées directement dans le fichier des documents. C'est le cas des formats reposant sur le codage en XML. Si cette intégration n'est pas possible, l'accès aux documents passera par la consultation d'une base de données externe.

Le tableau ci-dessous résume les combinaisons possibles.

Contenu Source	Texte	Graphique	Image	Son	Vidéo	Mixte
Bureau- tique	Documents bureautiques			Insertion de contenus audiovisuels dans les documents bureautiques		
Numéri- sation	Numérisation en mode image de tout support physique			Conversion de sources analogiques		
Outils spécialisés	Messagerie électronique	Plans CAO 2D et 3D	Appareils photos	Appareils de prise de sons et caméras vidéo		
Base de données	Fichiers de texte	Données CAO	Contenus binaires			



## ■ Formats candidats pour les documents bureautiques

Plusieurs formats d'archivage sont utilisables pour des documents produits avec des outils bureautiques. Leurs caractéristiques répondent plus ou moins bien aux attentes définies dans la section précédente.

PDF/A est basé sur le format Adobe PDF 1.4. Les spécifications de ce format ont été publiées sous forme de norme ISO validée en Septembre 2005. Les caractéristiques fonctionnelles et techniques de ce format en font une solution sans équivalent pour le moment. Ce format est présenté en détail ci-dessous.

XPS est le format concurrent développé par Microsoft (XML Paper Specifications). Ce n'est pas encore une norme ISO et sa portabilité est encore limitée à l'environnement Windows.

TIFF G4 est le format image dérivé du format des télécopies. La compression G4 est réversible sans perte de données. Sans être une norme ISO, il constitue un standard de fait largement utilisé pour la numérisation des documents. Le contenu textuel des images retraitées par un logiciel OCR peuvent être indexées. D'autres formats plus modernes, tels que JBIG2, apportent des taux de compression plus importants.

JPEG est un format de numérisation et de compression conçu pour les photos et largement utilisé depuis l'essor de la photo numérique. Le format est normalisé (ISO 10918). Des artefacts de numérisation peuvent apparaître lorsque ce format est utilisé pour des graphiques au trait et pour du texte. JPEG 2000 (ISO 15444) apporte une compression plus importante.

.DOC est le format natif du logiciel de traitement de texte Word développé par Microsoft depuis de nombreuses années. Le logiciel est très répandu et bien connu, ce qui ne signifie pas que les documents produits sont homogènes en terme de codage. Les versions diffèrent au fil du temps et des plates-formes de création. De ce fait, ce format est diamétralement opposé aux attentes exprimées pour définir un format d'archivage pour le long terme. En revanche, de nombreuses solutions existent pour convertir un fichier natif Word vers des formats plus adaptés au problème de l'archivage : PDF/A, XPS, voire XML.

XML (Extended Markup Language) est dérivé de la norme SGML ISO 8879 :1986. Cette norme définit une grammaire permettant de construire un système de balisage logique. En pratique, XML peut servir de format d'échange de données structurées entre deux systèmes. XML peut également être utilisé comme format d'archivage de contenus structurés quelconques. Cependant, pour conserver la présentation physique d'un document (polices de caractères, mise en page, etc.) il est indispensable d'appliquer une feuille de styles aux données XML. En d'autres termes, considéré de façon isolée, le format XML est adapté pour l'archivage d'un contenu structuré, mais pas pour la restitution d'un document mis en page.

Le tableau ci-dessous permet de comparer les propriétés des formats qui viennent d'être présentés.

	PDF/A	XPS	TIFF G4	JPEG	.DOC	XML
Norme ISO	✓	X	X Standard de fait	✓	X	✓
Fidélité restitution	✓	✓	✓	✓	X	X Feuille de styles
Polices caractères	✓	✓	✓ Pixels	✓ Pixels	X	X Feuille de styles
Texte indexable	✓	✓	X Avec OCR	X	✓	✓
Données structurées	✓	✓	X	X	✓	✓
Multi plate-forme	✓	X	✓	✓	✓ Disponibilité polices	✓
Afficheur gratuit	✓	✓ Windows uniquement	✓	✓	X	✓

## ■ Formats utilisables pour l'audiovisuel

Les informations fournies pour ce domaine sont volontairement limitées car ces questions sont largement développées dans l'atelier 2 consacré aux supports d'archivage.

## ■ Autres formats utilisables

### Texte seul

Ce format est utilisable pour archiver des contenus textuels sans aucune mise en forme. Sont exclus : les changements de polices de caractères, les mises en valeur (gras, souligné, italique), les effets de mise en page (alinéa, retrait, pied de page), les tableaux, etc.

Trois normes ISO définissent des alphabets de codage : ISO 646 pour l'Ascii 7 bits, ISO 8859 pour l'Ascii 8 bits (apports des lettres accentuées et des signes diacritiques), ISO 10646 pour l'Unicode (Ascii 16 à 32 bits).

### Formats spéciaux

Signalons, à titre d'exemples, quelques formats spéciaux

EDIFACT définit un format pour l'échange de données constituant des factures électroniques (ISO 9735)

La TEI (Text Encoding Initiative) est un projet international visant à mettre au point des directives pour l'élaboration et l'échange de documents électroniques à des fins de recherche érudite, et pour répondre aux besoins les plus variés des industries de la langue en général. Initialement basé sur SGML, le modèle a évolué vers XML. Ce n'est pas une norme ISO.

ALTO est un format conçu pour la numérisation des documents anciens en vue de leur diffusion sur Internet. Ce format très avancé combine une description de la structure physique de chaque page sous forme de données XML, une image de la page complète et de ces composants en PDF transparent et le contenu textuel des composants. Cette solution permet l'affichage de l'image de la page, l'indexation du contenu textuel, et la sélection directe des contenus dans l'image affichée.

## ■ Et les métadonnées ?

Les métadonnées contiennent des informations sur les documents : identification, origines du contenu, descripteurs pour la recherche, etc.

Certains formats permettent d'enregistrer certaines métadonnées directement dans le fichier du document numérique. C'est le cas des formats PDF/A et XPS, mais aussi du format TIFF. Les photos provenant d'un appareil de photos numérique contiennent des métadonnées IPTC (conditions de prise de vue, date, géo positionnement). Cette intégration facilite la gestion des métadonnées.

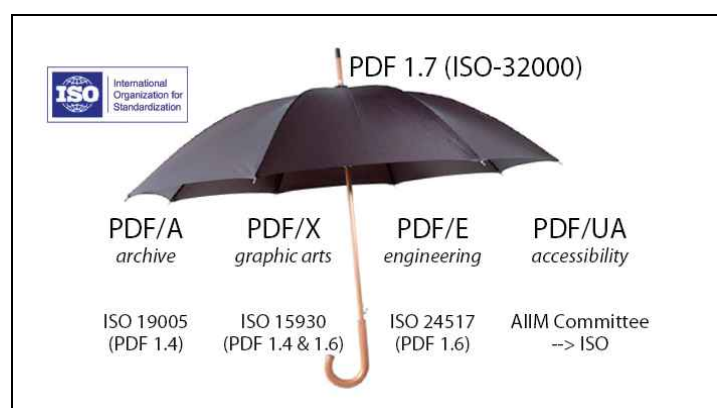
Lorsqu'elles ne sont pas intégrées dans les documents, les métadonnées doivent être gérées dans une base de données de type catalogue. Les fichiers des documents archivés sont accessibles via un identifiant retrouvé dans la base de données.

Dans les deux cas, les métadonnées doivent être soigneusement conservées et archivées avec autant de précautions que les documents eux mêmes.

## ■ Zoom sur la normalisation des formats PDF

La société Adobe, éditrice du format de description de page PostScript, a développé le format PDF (Portable Document Format) à partir de 1993. Le but était de fournir un format de description de page garantissant la restitution fidèle de la mise en page originale des documents créés avec divers logiciels sur des plates-formes différentes. La mise à disposition gratuite d'un afficheur - Acrobat Reader - a été un élément essentiel de cette stratégie éditoriale.

Pour renforcer la position de ses solutions, la société Adobe a entrepris de publier les spécifications de ses formats sous forme de normes ISO à partir de 2003. Ce mouvement est maintenant très avancé comme le montre le schéma ci-dessous.



### PDF/A ISO 19005

Basé sur PDF 1.4., ce format est conçu pour l'archivage des documents bureautiques. Le format peut être produit à partir de sources diverses (bureautique, numérisation). Il supporte plusieurs méthodes de compression normalisées et intègre des métadonnées dont la signature.

Pour assurer une réelle pérennité à long terme, la norme exclut tous les éléments susceptibles de rendre le fichier incomplet ou inaccessible. Sont donc exclus : les scripts, les liens hypertextes, les contenus multimédia, le chiffrement et les codes de sécurité. L'algorithme JPEG 2000 est également exclu. En outre, pour garantir une réelle portabilité, les polices de caractères doivent être intégrées dans le fichier, ce qui augmente considérablement le volume.

La norme actuelle définit deux niveaux de conformité.

- PDF/A-1a implique le respect de la norme complète, y compris de la structure logique et des métadonnées
- PDF/A-1b est une forme allégée qui garantit la préservation de la lisibilité et la restitution à l'affichage et à l'impression.

Cette norme fait l'objet d'importants travaux dans les groupes de travail de l'ISO. Des évolutions significatives sont attendues dans les prochaines années.

En 2009-2010, PDF/A-2 devrait apporter la compatibilité avec les formats PDF 1.5 et 1.7 et autoriser la compression JPEG 2000.

Par la suite (2012 ?), PDF/A-3 devrait intégrer la signature électronique et la gestion des ressources externes autorisant la mutualisation des polices de caractères.

La conception d'un plan de migration vers PDF/A est un projet informatique à part entière. Il convient de choisir les outils de conversion et de validation adaptés en fonction des sources disponibles et des volumes à traiter.

*Bibliographie : le livre intitulé « PDF/A : L'essentiel » co publié en octobre 2008 par l'Aproged et le PDF/A Competence Center présente de façon détaillée tous les aspects de ce format.*

## **Les autres normes PDF**

### **PDF/X**

Le format PDF/X fait est publié dans la norme ISO 15930. Il concerne les échanges de fichiers d'impression dans l'industrie des arts graphiques. Ce standard couvre les aspects suivants : gestion des images hautes résolutions transmises avec les mises en page, gestion des profils colorimétriques et de la transparence, gestion des polices de caractères comme des ressources externes. Cette dernière fonctionnalité devrait faire partie de la norme PDF/A-3.

### **PDF/E**

Ce format concerne les échanges de plans numériques. Il a été publié dans la norme ISO 24517-1 validé en mars 2008. Ce format intègre des fonctionnalités spécifiques des logiciels de CAO : gestion des couches multiples (layers), gestion des versions, gestion des éléments en 3D. Il est important de souligner que cette nouvelle norme vise uniquement l'échange de plans numériques et non pas l'échange de données CAO.

### **ISO 32000**

La norme ISO 32000 résulte de la publication des spécifications techniques du format PDF 1.7. Ce format offre de nombreuses possibilités, en particulier sur les composants qu'il peut gérer : animations 2D et 3D, multimédia. Dans son état actuel, la norme définit plutôt un conteneur pouvant intégrer différents types d'objets dont les spécifications ne sont pas encore toutes normalisées.

## **■ Repères méthodologiques**

Pour terminer cet exposé technique, il convient de souligner que le choix des formats doit s'insérer dans une réflexion plus large de conception d'un projet d'archivage électronique qui devra répondre aux exigences d'une politique d'archivage définie au niveau de l'entreprise ou de l'institution.

Cette réflexion sur les formats doit passer par les étapes suivantes :

- Recenser et caractériser les contenus à archiver : typologie des documents et des flux à conserver, sources des données et mode d'obtention de ces données ;
- Définir les fonctionnalités attendues en termes de services à rendre : simple consultation sur écran avec quelles fonctions, impression, indexation des contenus textuels, gestion des métadonnées intégrées dans les documents, extraction de contenu pour récupération, possibilité de charger les données archivées pour créer de nouveaux documents ;
- Choisir et spécifier les formats retenus pour l'archivage : versions à utiliser, paramétrage des options, contrôles à prévoir ;
- Rechercher et qualifier les outils de conversion et de vérification parmi les produits offerts sur le marché ;
- Organiser la migration : choix des acteurs, planification des opérations, pilotage du chantier.

# PDF/A: STANDARD FOR THE LONG TERM ARCHIVING

---

Dr. Uwe Wächter

Product Manager – SEAL Systems

Saïd El-Morabiti

Sales Manager – SEAL Systems

*Analog archiving (paper documents, microfiches...) was considered as the unique way for secure and save long term storage of documents. Due to the possibilities offered by document management systems, the capacity of networks, the growing amount of documents, the world is more and more changing and documents are archived in a digital form.*

*But this leads to an important question : How to guarantee the correct reproduction of the digital documents over the time? TIFF/G4 was an answer to this question for a long time but it is not efficient enough for text+ research, integration of metadata, colour management and last but not least, it is not an official ISO standard. Since October 2005, the ISO norm ISO 19005+1 : PDF/A is answering all open questions of TIFF as a long term archiving format for digital documents.*

*The offered presentation will introduce PDF/A to the attendees. What is PDF/A, what are the differences to PDF, why should PDF/A be used and how could archives be migrated to PDF/A.*

*We will introduce PDF/A as the ISO standard 19005+1, for long term archiving.*

*Which PDF content is allowed in PDF/A? What are the restrictions? What are the mandatory elements.? When is a PDF file PDF/A compatible? What are the main reasons, why PDF files are not PDF/A. What are the differences between PDF/A 1+A and PDF/A 1+B?*

*We will speak about the arguments: Why to use PDF/A for long term archiving.*

*What are the advantages of PDF/A against other formats as TIFF, JPG. Why should organisations not store documents in their native format (e.g. WORD, EXCEL, ...)*

*We speak about how to create to PDF/A. How can an organisation guarantee that all archived PDF files are PDF/A compatible? How can existing archives be migrated to PDF/A? The attendees will hear about best practices of organisations who already implemented PDF/A as the company standard for archiving.*

[See the presentation.](#)

## **PDF/A : la norme pour la conservation à long terme**

*L'archivage analogique (documents papier, microfiches, etc.) était considéré comme le seul moyen de conserver des documents de façon sûre et pérenne. En raison des possibilités offertes par les systèmes de gestion des documents, de la capacité des réseaux, du volume croissant des documents, le monde évolue et les documents sont archivés sous un format numérique.*

*Dans ce contexte, se pose la question majeure suivante: comment garantir la reproduction correcte des documents numériques à long terme ? Le format TIFF/G4 a longtemps été la réponse à cette question mais ce format n'est pas adapté aux recherches de documents contenant plus que du texte, ni à l'intégration de métadonnées, ni à la gestion des couleurs et enfin, dernier point mais non le moindre, ce n'est pas une norme ISO officielle. Depuis octobre 2005, la norme ISO 19005+1, PDF/A répond à toutes les questions ouvertes posées par le format TIFF en tant que format d'archivage à long terme pour les documents numériques.*

*Cette présentation exposera aux participants le format PDF/A. Qu'est-ce que le format PDF/A ? Quelles sont les différences par rapport au format PDF ? Pour quelles raisons le format PDF/A devrait être utilisé et de quelle façon les archives pourraient migrer vers ce format ?*

*Nous présenterons le format PDF/A en tant que norme ISO 19005+1 pour l'archivage à long terme.*

*Quel contenu PDF est autorisé dans le format PDF/A ? Quelles sont les restrictions ? Quels éléments sont obligatoires ? Dans quelles conditions un fichier PDF et un fichier PDF/A sont-ils compatibles ? Pour quelles raisons principales un fichier PDF n'est-il pas un fichier PDF/A ? Quelles sont les différences entre le format PDF/A 1+A et le format PDF/A 1+B ?*

*Nous argumenterons les raisons d'utiliser PDF/A pour l'archivage à long terme.*

*Quels sont les avantages du format PDF/A par rapport aux autres formats tels que TIFF et JPG ? Pour quelles raisons les entreprises ne devraient pas conserver les documents sous leur format initial (par ex. Word, Excel, etc.) ?*

*Nous expliquerons comment créer un fichier PDF/A. De quelle façon une entreprise peut-elle assurer la compatibilité de tous les fichiers PDF archivés avec le format PDF/A ? De quelle façon les archives existantes peuvent-elles migrer vers le format PDF/A ? Nous exposerons aux participants les meilleures pratiques des entreprises qui utilisent déjà le format PDF/A comme format d'archivage standard de l'entreprise.*

[Consulter la présentation.](#)

## ■ PDF/A:

PDF/A is an ISO norm (19005) which defines a subset of the PDF reference for long term archiving of digital documents.

The norm was published at the end of 2005, after three years of analyses and discussions in the ISO working group (171).

The norm defines the requirements for creation of PDF files fulfilling the needs of long term archiving. It specifies in detail, what is the allowed content, the restrictions and the additional information. These specifications guarantee the faithful document legibility and reproduction independently from their production systems and software.

PDF/A-1 is based on the PDF specification version 1.4. In order to guarantee the portability over the time, certain functions included in PDF 1.4 are not allowed in the norm: Transparency, embedded audio, some compressions like LZW or JPG2000, etc. On the other side in PDF/A certain functions are mandatory: embedded fonts needed for text display, XMP metadata, precise colour management...

## ■ Two levels for the norm 19005-1: 1a and 1b

The PDF/A norm describes two level :

- PDF/A-1a : The strict level is 1a, the structure and semantic has to be clearly defined. Each character has to have its unique encoding. Tags are permitted to define the document content and the logical structure. This level is interesting for displaying PDF files on small digital equipment (PDA; smart phone...), the files are defined as “accessible”. This level is also interesting for future migration, conversion and extraction of some objects.
- PDF/A-1b : This level is sufficient for the long term faithful legibility (representation and reproduction). The logical structure is not mandatory, the most important issue is to guarantee the faithful displaying and reproduction of the document.

It is of course better to produce the 1a level if it is possible. However, this level is often more difficult to obtain, the document structure has to be provided when the document is created (See scanned document).

### Summary:

The production of the PDF/A-1b level is dedicated to scanned documents and documents produced through a conversion of other file formats with a suitable structure. The PDF/A-1a level is possible for documents generated by applications able to define a structure.

## ■ PDF/A main advantages:

Different file formats are used to store or archive documents: native formats, JPEG, TIFF, PDF, XPS... The analyse of the fundamental file format requirements for long term archiving shows that PDF/A is a good answer. In addition, PDF/A brings specific advantages compared to others file formats:

### **Advanced full text research**

With PDF/A, it is possible to save documents in text mode. This permits search functions on a word, on an expression or on figures. Image file formats like JPEG, TIFF or others need an attached file with the OCR text. (Optical Character Recognition) in order to be able to provide the same search functions.

## **Metadata**

Metadata, like the title, the author, the document creation date, the document modification date, the key words,... can directly be stored in the PDF/A file. The PDF/A files could be automatically sorted and classified by using the metadata without manual actions.

## **Specifications available**

The published norm permits the guaranteed possibility for each software engineer to access to the format specifications. This means more freedom for the development of software solutions (readers, converters...) to support PDF/A.

## **The digital signature**

The PDF/A norm accepts digital signatures. This feature is important for archiving due to the need to define the author and to guarantee that the document is not modified.

## **■ PDF/A in an archiving project:**

Key points for PDF/A production:

### **Archiving strategy**

PDF/A is only a file format. In consequence, the use of this format in an archiving project leads to consider some further questions:

What are the documents concerned by the archiving project ? What will be the future use of the archived documents? What is the type of the documents? What is the legal period of archiving? What are the actual and future volumes for storage?

If the originals are paper documents, it is important to consider the time to spend for the scanning process, the packaging and the carriage conditions. Depending of the needs, an OCR function has to be added to the process in order to permit full text search or automatic indexing.

Another example: Production of PDF/A files from emails should be analysed and studied with attention. The processing could be complex due to the format of attached files, and due to the fact that an email can contain an email which contains an email....which contains a compressed attached file...

PDF/A will be a good archiving file format only if the archiving strategy is clearly defined and if the characteristic of the norm are considered.

### **Protected files**

Protected files, with password (read, print, or modify) will be difficult or impossible to convert into PDF/A. In addition, protection is prohibited in PDF/A. It uses encryption methods and encryption is not allowed within the norm. In consequence, the authorisation concerning the access to the files has to be managed in the archiving system but not in the file. A level of security is possible with the integration of digital signature (see digital signature).

### **Fonts**

The fonts used in the documents have to be available in the operating system where the PDF/A file is produced. This is required, because in PDF/A it is mandatory that all used fonts are embedded. If the PDF/A file is produced by a conversion process, it might happen, that not all fonts of the original document are available on the computer where the conversion is running. In this case a substitution is done by the conversion application. This could compromise the faithful conversion to PDF/A. One possibility is to create a process



where the author of the document generates the PDF/A file in its own software environment. The second solution is to make sure that users will not use special fonts and to put all used fonts on the conversion systems platform.

## Colours

The colours of an illustration or a graphics in a document must normally be the same whatever equipment will be used for displaying or printing. PDF/A answers to this requirement. PDF/A manages the colours by using different profiles (RGB, SRGB, Adobe RGB, ISO coated...). Within the scope of the PDF/A Norm the used colours must be defined by the appropriate profiles or an output intend.

## Metadata

The metadata have to be integrated in the PDF/A file to support systems for search functions. In the case of archiving solutions, they are used for a better organisation in the document management. For PDF/A, metadata must be defined respecting the XMP (eXtensible Metadata Platform) standard base on XML. This standard was developed by Adobe. But only one metadata is mandatory in PDF/A: "PDF/A identifier".

## File size

An increase of the file size could be possible after a conversion from PDF to PDF/A. This is due to the embedded fonts in PDF/A files. To avoid this, an alternative is merging the PDF files into one, if this is possible in the dedicated context. This permits to share the fonts. An other alternative, is to create a process which embeds the fonts only when the file is checked out from the archiving systems. But be careful, the files are not considered PDF/A compliant in this case. This second alternative is not possible for documents with a digital signature. The easiest and secure solution is to choose the best compression method.

## Compression

The PDF/A Norm do not allow compression or even encryption of the whole file or the structure information. The compression for PDF/A files is only allowed for pictures. But certain compression methods are not allowed, like LZW, or JPEG2000.

It is important to ensure that the compression methods used during the generating process are compliant to the norm.

## Forms

Forms are not prohibited in the norm, but certain kind of fields used in forms are using actions which are not compatible with long term archiving :

Fields with actions like «Submit form» «get data» «Reset form». Those actions are forbidden. They could change the visual presentation of the document. JavaScript is not allowed. This is also able to modify the document during the displaying or use.

## PDF/A and digital signature

Sales contracts that are exchanged between administrations, are often done online. How to guarantee the authenticity of the exchanged documents and the partners of the contract? As it was done since thousands years, the signature (manual, stamps...) is still the solution. The difference here is that the signature is digital. Three functions are associated to the digital signature: identification of the person who signed the document, guarantee that the document was not modified and a time information. PDF/A and digital signature are compatible. The main question is how to sequence the steps in the document archiving workflow. First, the PDF/A file has to be produced afterwards signed. If a document is modified after signature, the signature is no longer valid.

## **Checks before archiving: Is it really a PDF/A file?**

It's important to make sure that all PDF/A files generated are really compliant with the norm. Errors in generating PDF/A in an archiving process could lead to files which are not legible in 5, 10 or 20 years... For this reasons, a checking process before archiving is recommended. The files which are not compliant will be rejected or adjusted when possible (non protected, without special fonts...). This shows that it is recommended to define the workflow, to analyse the origins of the document production (from users computers, on central servers...).

## **■ Outlook to the future of PDF/A: PDF/A-2**

The ISO norm 19005-2( PDF/A-2) is discussed by the Technical Working Group of the ISO. As the ISO norm 19005-1, the future norm will concern only static documents. A third part (ISO 19005-3 - PDF/A-3) will consider interactive documents( documents with forms...) as well.

The PDF/A-2 norm will be based on the specifications of the latest version of the PDF file format (1.7 specified in the ISO Norm 32000). The ascending compatibility is guaranteed: if a PDF file is conform to PDF/A-1, it will be conform to PDF/A-2 too. Many functionalities are discussed for the version 2 of the norm (19005-2 / PDF/A-2), some of them are listed here:

### **JPEG2000:**

This new compression method is more efficient than JPEG.

### **Transparency:**

Until the version PDF 1.4, the transparency was not clearly defined in the PDF specifications. This function was prohibited in PDF/A-1. As this function is now clearly described, it will be included in the new version of the norm.

### **PDF Layers:**

Different views of a document will be possible, but only one will be displayed and printed in one time (the printed view is the displayed one). This evolution is important, especially, for CAD files, it will be possible to store different layers in the same document.

### **PDF Packages/Collections:**

This is a new functionality introduced with the version PDF 1.7 which permits to embed others files in the PDF file. A PDF file package will be conform PDF/A-2 only if all embedded PDF files are conform to PDF/A-2.

### **Compressed XRefs and Streams:**

This is also a new functionality introduced with the PDF version 1.6. The compressed XRefs are used for the « big files » (containing many pages and others elements) , it permits a more efficient compression for this kind of files.

### **Tags and accessibility improvements:**

PDF/A-2 will permit new TAGS for a better description of the content and improved accessibility.

### **Metadata management improvements:**

The embedded metadata are richer than the previous version of PDF/A and will optimize the access to documents.

### **Digital Signatures improvements:**

A better digital signature management will be available

### **Summary:**

The future version of the norm will mainly permit to :

- Manage bulky files with the collections/packages functions.
- Improved the image compression (JPEG 2000).
- Better support for PowerPoint and other complex graphical file formats (Transparency).
- Better support of digital signatures
- ...

The norm ISO 19005-2 should be published during the first quarter of 2010.

## **■ Conclusion:**

Our world becomes more and more digital, the technologies and the new regulations regarding archiving, leads administrations and companies to consider the preservation of digital documents over the time. PDF/A-1 fulfils the requirements for long term archiving of static documents. The most important issue in a long term archiving project is to define the nature and characteristics of existing document (format, volumes, digital signature, metadata..), and the process for their conversion. This will permit to measure the effort needed and to define the best path to achieve this goal. A validation process before archiving is also recommended.

# **Atelier 4**

## **Les stratégies de conservation à long terme**

---

---

***Workshop 4***  
***Long-term***  
***preservation strategies***

---

---

# DIGITAL PRESERVATION POLICIES CREATED BY THE ATON PROJECT

---

Osmo Palonen

Project Manager, Mikkeli University of Applied Sciences, Finland

## *Digital preservation and migration policies created in project ATON*

*In a two-year project ATON Mikkeli University of Applied Sciences improved and extended the university's and its partner's knowledge in migration and digital preservation. The goals were to document current procedures of digital repository and define the guidelines for the future. As a result there was lot of material from PREMIS data dictionary translation in Finnish to the external evaluation of repository by Luleå University of Technology.*

*The preservation and migration strategy is based on the overall approach in the OAIS-based repository with some variations:*

- 1. The native format of content is also stored. It can be considered as permanent SIP version or as an appendix to the content sent in SIP package.*
- 2. The AIP's are in formats accepted in the policy of the repository. Those formats have to carry enough data for migration and to create DIP's for the future use. The accepted formats should be sustainable and are estimated to last 10-30 years. ISO standards are preferred.*
- 3. There are separate DIP's generated from AIP's to provide fast and up-to-date service for the users. These DIP's have lower resolution and are based on widely used industry standards. The life-cycle of the DIP-versions of the content is shorter than AIP's, only a couple of years. As examples of this are MP3, Flash video and low res PDF.*
- 4. The AIP's will be in the separate storage system, the RDF-based metadata is common for all presentations. Before the new storage now under planning is available, the AIP's from the historical archives are on tape in two different locations.*

*Since MiUAS digital repository's short history, there has not been a big demand of migration until now. The repository policy is keep DIP's on disk and AIP's on LTO tape that has to be checked once in 2.5 years. The plan to move the contents to a new tape format is between 5 to 7.5 years. The decision will be made on the latest when the LTO4 drives, which can read LTO2-4 tapes, will lose the support. In keeping authenticity and integrity of the contents on tape it is important to recognise possible main problems in both approaches:*

- in copying to new format some data can be lost or*
- when keeping the existing tape the condition check-up might not always find all possible faults*

*The solution model created in ATON project is to have the AIP's in more than one format and use WORM-media. Two copies in the same media and format is not enough, at least one additional copy in another media is needed. It was studied if the authenticity and integrity of the contents can be improved by using digital signatures or complex checksums, but the results were not promising. The digital signatures may be used in contents with limited life cycle as in the medical sector, but both of the technologies are too complex in the long term digital preservation.*

*Migration between the formats was tested in the project by collecting all contents of certain type and dropped those into a patch process. The new versions of the contents are then automatically ingested again into the archive and complete process is documented into the metadata.*

*The digital archiving and digitization activities in MiUAS are part of the R&D work for the region. In the City of Mikkeli there are national and provincial operations that all need digital repository services. The local grid model developed in the project Aton was designed for the local cooperation. The grid consists of a fast fibre optics network and four main nodes: two nodes close to digitisation storages and service infrastructures, one node that is a mirror and service backup of those two sites and finally one separate dark archive. The project studied that the next step required is long distance replication, which can be also realised in cooperation between European countries*

[See the presentation.](#)

## **La conservation à long terme. La pérennisation du numérique et les politiques de migration au sein du projet ATON**

Dans le cadre du projet ATON mené sur deux ans, l'Université des sciences appliquées de Mikkeli a amélioré et développé les connaissances de l'université et de ses partenaires en matière de migration et de conservation numérique. L'objectif du projet était de documenter les procédures actuelles des plates-formes d'archivage numériques et de définir les lignes directrices pour l'avenir. En conséquence, de nombreux documents sont apparus, variant du dictionnaire des données PREMIS traduit en finnois à l'évaluation externe de la plate-forme d'archivage numérique par l'Université de technologie de Luleå.

La stratégie de conservation et de migration s'appuie sur l'approche globale adoptée par la plate-forme basée sur la norme OAIS avec quelques variations :

1. Le format original du contenu est également conservé. Il peut être considéré comme une version permanente SIP ou comme une annexe au contenu transférée sous la forme d'un paquet SIP.
2. Les AIP se présentent sous des formats acceptés par la politique de la plate-forme. Ces formats doivent transporter suffisamment de données pour permettre la migration et créer des DIP pour l'utilisation future. Les formats acceptés devraient être durables et leur durée de vie est estimée à 10-30 ans. Les normes ISO sont privilégiées.
3. Il existe des DIP distincts créés à partir des AIP afin d'offrir aux utilisateurs un service rapide et à jour. Ces DIP ont une résolution inférieure et sont basés sur des normes industrielles couramment utilisées. La durée de vie des versions DIP du contenu est plus courte que celle des AIP, deux ans seulement (ex. formats MP3, Flash video et PDF basse résolution).
4. Les AIP seront localisés dans le système de conservation séparé, les métadonnées en RDF sont communes à toutes les présentations. En attendant que le nouveau système de conservation prévu soit disponible, les AIP issus des archives historiques sont stockés sur des cassettes localisées dans deux endroits différents.

Depuis la courte histoire de la plate-forme d'archivage numérique MiUAS, la demande de migration n'a pas été significative à ce jour. La politique de la plate-forme est de conserver les DIP sur disque et les AIP sur cassette LTO devant être vérifiée une fois tous les 2,5 ans. Le transfert des contenus vers un nouveau format de cassette est prévu au bout de 5 à 7,5 ans. La décision sera prise au plus tard lorsque les lecteurs LTO4, capables de lire les cassettes LTO2, perdront le support. En conservant l'authenticité et l'intégrité des contenus sur cassette, il est important d'identifier les problèmes majeurs que peuvent poser les deux approches :

- lorsque l'on copie des données sous un nouveau format, certaines données peuvent être perdues ou
- lorsque l'on conserve la cassette actuelle, il se peut qu'en vérifiant l'état, tous les défauts ne soient pas décelés.

Le modèle de solution créé dans le cadre du projet ATON consiste à disposer des AIP sous plus d'un seul format et d'utiliser des médias WORM. Il ne suffit pas de conserver deux copies dans le même média et sous le même format. Il est nécessaire de disposer d'au minimum une copie supplémentaire dans un autre média. On a étudié la possibilité d'améliorer l'authenticité et l'intégrité des contenus en utilisant des signatures numériques ou des sommes de contrôle complexes, mais les résultats n'ont pas été concluants. Les signatures numériques peuvent être utilisées pour des contenus dont le cycle de vie est limité, dans le secteur médical par exemple, mais les deux technologies sont trop complexes pour être appliquées à la conservation numérique à long terme.

La migration entre les formats a été testée dans le cadre du projet en recueillant tous les contenus d'un certain type et elle a déposé ces derniers dans un processus patch. Les nouvelles versions des contenus sont alors réintroduites dans les archives et le processus complet est documenté dans les métadonnées.

L'archivage numérique et les activités de numérisation de la plate-forme MiUAS font partie du travail de recherche et développement réalisé pour le compte de la région. Dans la ville de Mikkeli, il existe des opérations au niveau national et provincial nécessitant toutes des services de plate-forme numérique. Le modèle de grille local développé dans le cadre du projet Aton a été conçu pour la coopération locale. La grille est composée d'un réseau de fibres optiques à haut débit et de quatre nœuds principaux : deux nœuds près des conteneurs de stockage pour la numérisation et des infrastructures de service, un nœud servant de miroir et de service de secours aux deux sites, et enfin une archive noire. Selon le projet, la prochaine étape nécessaire sera la duplication à longue distance, pouvant également être réalisée en coopération avec des pays européens.

[Consulter la présentation.](#)

Archives and libraries are progressing step by step towards digital content management and eServices as the mainstream of their public service method. At the moment there is a great need for new understanding and knowledge. The national, regional and local archivists and librarians need tools, skills and also a theoretical approach as to how to make sure that the records and publications created in the digital mainstream of today's content production will also be available after seven, seventeen, seventy or seven hundred years.

Mikkeli University of Applied Sciences is trying to satisfy the growing demand of new archival knowledge by creating educational programmes for the community. In this it has been important to have R&D projects for research and knowledge transfer as well as collecting practical experience from the customer service in the university's trusted repository. MiUAS started in 2005 as a post-graduate programme with 30 ECTS credits for archivists as continuing education and in the beginning of this year a new master's programme was launched in eServices and digital archiving.

## ■ 1. The ATON project

With the ATON project, carried out in 2006-2007, Mikkeli University of Applied Sciences improved and extended the university's and its partners' knowledge in digital preservation and migration. The goals of the project were also to document current systems and practices of the MiUAS digital repository and define guidelines for the future. As the outcome, there was lot of material starting from the PREMIS data dictionary translation in Finnish and the complete project report on the external evaluation of repository by Luleå University of Technology and the LDP Centre of Sweden.

As described before, the background of the ATON project was both educational and business development. MiUAS as a forerunner in digital archiving and eServices needed to update the technical and archival knowledge from international sources and its own research. When that know-how retrieval was fully up and running, studies on how to improve digital archiving services could also start. In addition the knowledge for education was continuously growing; one important part of the project was the development and planning of the new master's degree programme.

The digital archiving and digitisation activities in MiUAS are part of the R&D work for the region. In the City of Mikkeli, there are national and provincial operations that all need digital repository services. An outcome in the project was to create a model that can be used as a technical backbone for the digitisation and digital content management services in the region. It was not assumed that the model would be implemented immediately, but it gives the roadmap, as to how a common infrastructure can be built for memory organisations and other digital content management activities. As an R&D the ATON project just provided the information; the operations in the region will decide if this model will still be in use when the next generation of mass digitisation equipment and storage technology is installed. For the archival community what is more important than to see this regional model in operation is the conclusion as to what kind of structures a trustworthy digital repository will need and how the cooperation between the institutions can be implemented.

## ■ 2. Model for the common service of memory organisations

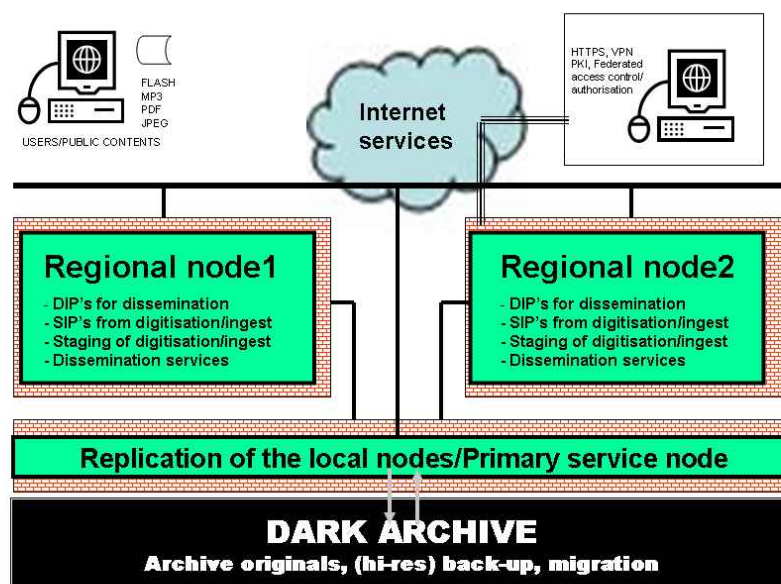
The first part of this study describes how to build a common data management infrastructure for a group of memory organisations when they produce a vast amount of data for digitisation or for ingesting the large born digital collections. The problem the operations should tackle is on one hand the growing need for local short term data storage for digitisation and on the other hand the enormous growth in the long term storage requirements for high resolution digital contents, not depending if the source is digitisation or the born digital ingest. The common data management infrastructure gives the operations undeniable benefits especially when they need to provide their customers a reliable 24h/7d/week service and trusted long-term preservation at the same time as with reasonable costs.

The idea of a regional grid consists of a fibre optics network that is connected to all operations which produce masses of data. The local nodes must have sufficient storage space on high-speed disks for the



digitisation and ingest requirements<sup>1</sup> and tools to parse and collect the primary metadata into the databases. When building reliable services the local node can be the primary server room for the clustered service; the other site with other cluster nodes and mirrored data is somewhere else in the cooperative grid. It is also possible that there is a central service platform, in which the local production nodes can serve as a backup service provider, especially when the server rooms are close to the production are not built using the strictest security standards. Distributed high availability solutions are common in big business but they have not been widely used by the memory organisations. Archives, libraries and museums will find the same kind of challenges of continuity as will the companies providing the services, when digital networked services are or will be the mainstream in regards to customer contacts.

## 2.1 Interpretation of OAIS – where are the packages?



Using the functional model of separate dark archive, dissemination is separated from archival storage and a new interpretation of the OAIS reference model is developed. Where ingest and staging is done can vary; the implementation depends on the needs of the operations. In the ATON model these are connected to the local node. Also the metadata can, in this case, be automatically parsed from the content files in the first steps of the metadata management. The transfer to the dark archive can occur when all versions needed in distribution as well as the metadata are completed.

Through this approach several benefits are gained. The first is a common Dark Archive that can isolate the data archiving from the daily service level. The long-term preservation of information held in the data packages will not have those problems that are common in the distribution: the size or format of the file is not a problem as it could be in the distribution, for example. When the original contents are not available through information systems, there can be much simpler methods in protecting the authenticity and integrity of the files. A starting point of this model is to be able to provide dissemination files in those formats that are the same as the current developments of web services. As an example of this MiUAS has already converted the DIPs of all digital videos to Flash streaming technology instead of MPEG-1 download, that was the only independent operating system option four years ago. The hi-res digital video files in the dark archive level are still unchanged in the MPEG-2/AVI format preserved by using the 50 or 25 Mbit/s bit rate and it can be estimated that there is no need to change this format within 10 years. Also the roadmap of digital video can be considered from the preservation point of view. The tests using lossless Motion JPEG2000 instead of AVIs are going to start even when it is not easy to distribute any moving images in this format via web services as currently Motion JPEG2000 needs special players.<sup>2</sup>

The isolation of the long-term preservation from the web-based information services will also give remarkable benefits in distribution or dissemination. When the preservation of the data and information is not

<sup>1</sup> Mass digitisation of video can create over 300Mbit/s continuously. That is 2.5 terabytes in 24h production. See also: <http://www.sammassystems.com/?q=node/98>

<sup>2</sup> About the methods in the digitisation of the video Wright, Richard: *Archivists of the World, You Have nothing to Lose but Your Shelves!* IS&T Archiving 2008, Bern 24.26.6.2008, conference publication.

the central part of the archiving information service, this part of the digital archiving service can be implemented in a much more efficient and flexible way than traditional finding aids. It even makes it possible to outsource the services when there is no connection to the preservation. In the model created in the ATON project it was thought that there would at least be commercial partners in implementing the information service, if complete outsourcing seems to be out of the question or more expensive than own production.

It is worth pointing out that the OAIS model (ISO 14721) is a reference model. There are few binding details for the implementation of a trusted repository. The model created in the ATON project can be interpreted as OAIS-compliant, even with some developments. One has to remember that the foundation for OAIS came from the late 1990s and the space industry. As written above, the ATON model has two separate systems whereas the OAIS model can be understood to include only one. However, OAIS has separate functionality for ingest, staging, preservation, data management and dissemination. In the ATON model ingest can be either in the upper service level, or in the lower dark archive level, but the main difference is in how to manage the DIPs. The OAIS standard assumes that the DIPs are created by request from AIPs; in ATON the experience from the repository operations in MiUAS demanded another approach. In most cases an efficient dissemination or distribution service requires that the DIPs are different versions of AIPs and already exist. There is no time to convert hi-res-versions of digital video to consumer formats on the fly. As another example the hi-resolution digitised prints or manuscripts loaded by a mobile phone will not display if there are TIFFs and take a long time to download even when the phone can manage PDFs. Personally, I don't think this will be solved in the near future by new innovations in telecommunication. Better formats for originals and higher resolutions will challenge the communication: we will simply need separate distribution copies in the future also.

These differences still do not mean that the OAIS principles are broken in the repository. It is only a new interpretation of the standard that has to be made when the type and nature of the contents in the archive are different to those in existence when the OAIS standard was written.

## **2.2 Long distance replication**

The project also considered long distance replication even when it was not in the primary scope of the project. In this part of the study a question was raised about if the long distance replication for the dark archive could be put into effect by cooperation between European countries. The development of data communications already has tools for distant data replication when the transfer is not time-critical as in the dissemination service. The open question still is, how to manage the contents from different sources: projects like Distarnet<sup>3</sup> at Basel University have given an inkling into this.

The archiving of the national heritage is a national task in the EU. However, it is a waste of scarce resources if all countries of the union create their own solutions for the preservation of their digital heritage. If there are laws prohibiting rational cooperation between the countries, I'd like to quote what the other Finnish speaker at this conference, Dr. Markku Nenonen often reminds his audience: new laws are made every Thursday in Finland.

## **■ 3. Formats and media used in the repository**

The ATON project defined three different types of formats: original, preservation version and dissemination version. Original formats are those file formats that are included in the SIP package, not only dependent on if those are real originals of the process. As an example, the real originals of municipal records are in most cases in paper format in the local municipality. The paper documents are scanned or digitised by a camera and converted locally as a PDF-file. Those PDFs are transmitted to a repository and saved on tape. In the future PDFs will be converted into PDF/A in the repository. When the PDFs files are too big for distribution there are smaller files created by splitting the complete Archival Unit files on a monthly basis and/or by decreasing the resolution. There is also a need to take some files as created originals: spreadsheet files in most cases should be kept as originals even when a PDF from the representation is created. Original files will be available at least as long as there are applications that can open those either by emulation or conversion. The wider the application is used the bigger is the probability that you can still use those files. Originals can also be needed when there are problems in the migration process in the future. After sequential migrations some information

---

<sup>3</sup> [http://www.forschungsdb.unibas.ch/ProjectDetailShort.cfm?project\\_id=3001](http://www.forschungsdb.unibas.ch/ProjectDetailShort.cfm?project_id=3001)

might have been lost and a good option can be to find or build a migration tool from the original format into the up-to-date application.

### 3.1 Archive originals

The archive original formats accepted by MiUAS are the following:

**Moving image:** AVI (MPEG-2, Pinnacle codec) 50 or 25 Mbit/s; M2V (MPEG-2, Pinnacle codec) 25 Mbit/s (no sound, digitised narrow films). Testing: MXF with Motion JPEG2000 (Samma codec)

**Sound:** BWF (Broadcast wave) 48 kHz 24 bit PCM. Testing: FLAC (Free Lossless Audio Codec)

**Digitised paper documents:** ->200 dpi colour or grayscale, TIFF (technical drawings), PDF 1.4 ->, PDF/A (ISO 19005 a&b)

**Born digital paper documents:** PDF 1.4 ->, PDF/A (ISO 19005 a&b)

**Photographs & slides:** Resolution 300 dpi/A4 -> TIFF 24 bit colour, JPEG (high quality) Testing: JPEG 2000 (lossless and lossy)

To verify the authenticity of the archive original the dark archive tapes will be WORM-tapes when those will be converted to LTO 4. There is an option to use Blu-ray WORM-disks as well. The integrity of the files will be verified in the tape storage software.

### 3.2 Dissemination copies

The dissemination copies formats accepted by MiUAS are the following:

**Moving image:** MPEG-1, Already replaced by Flash video streaming

**Sound:** MP3 96 Kbits/sec ->

**Digitised paper documents:** PDF 1.4 -> (resolution can be decreased); PDF/A (ISO 19005 a&b)

**Born digital paper documents:** PDF 1.4 ->; PDF/A (ISO 19005 a&b)

**Photographs & slides:** Resolution 72 dpi -> JPEG or PNG; Testing: JPEG 2000 (lossy)

### 3.3 Media and systems

Primary media for DIPs is hard disks in SAN storage with tape backup. Fiber channel - SCSI disks are used for fast access like databases and Serial-ATA disks to hold the contents. The AIPs are normally on tape. The service level for the storage system used in the DIP service has to be 24/7 with 4 hrs delay. All the software used in search and management functions has to have reliable support as well.

Dark Archive as a separate entity can be based mainly on tapes; the disk storage used there does not need the same kind of service level as above, when the response time for the queries can be days. Open source solutions are preferred when possible, to limit the costs to memory organisations. The main question is the reliability when storing great amounts of data. This can be done by replication. Energy costs can be saved by using MAID (Massive array of inactive disks) technology. However, the MAID archive can be one media besides that of tape, especially if cheap consumer disks can be used.

## ■ 4. MiUAS as an example of a service provider

The MiUAS repository has material from various sources. It has customers from memory organisations, companies and the medical sector. The policy concerning the formats created in the ATON project is partly described above, but I still need to clarify those principles in detail. The major part of data, at the moment 75 Terabytes on disk and over 200 on tape, held in the MiUAS repository is from the medical sector, but as this data has a short preservation time - ten to twenty years - it doesn't need any special plan for the format

management or migration, especially when long-term preservation of all medical data in Finland will be taken care of by the National Medical Archive by 2011.

From the memory organisations' point of view it is much more interesting how MiUAS is managing the contents from historical archives. As described above, the content versions in the repository are different depending on if those are used in dissemination or long-term preservation; also the processes that take care of contents are entirely different depending on if those are used in the service or in the dark archive.

Historical archives using the MiUAS repository services are ELKA (Central Archives for Finnish Business Records), City Archives of Espoo, Tampere and Turku, Municipal Archives of Pieksämäki town, Archives of UPM Kymmene, some archiving services for the Finnish Broadcasting Company YLE, Työväen Arkisto (Labour Archive) as well as the National Library of Estonia. MiUAS has also just started a YKSA archiving service for civil society and ordinary families and citizens, in which some associations are already depositing their records, documents and photos.

The customer services have tailored functions to the different types of customers, but there is a common technology in the background. The software for ingest, search and dissemination is in most cases based on RDF<sup>4</sup> modelling.

#### **4.1 Preservation and migration process**

Preservation and migration process in MiUAS is as follows:

All contents have correct and required metadata in RDF operated by archiving applications. At the moment the Profium Metadata Server<sup>5</sup> is used in all archive applications based on metadata search. As a back-up there is an option to take all RDF including location information and move the information into a new RDF-engine. In the project MiUAS carried out a small study with open source products like Jena and Fedora. The former is RDF-based and the latter RDF-compatible.

During the MiUAS digital repository's short history, there has not been a big demand for migration until now. The repository policy is to keep DIPs on disk and AIPs on LTO tape that has to be checked once every 2.5 years. The plan is to move the contents to a new tape format after 5 to 7.5 years. The decision will be made, at the latest, when the LTO4 drives, which can read LTO2-4 tapes, will lose the support. Tape storage management is currently done by Tivoli software. By keeping authenticity and integrity of the contents on tape it is important to recognise the possible main problems in both approaches:

- in copying to a new format some data can be lost or
- when keeping the existing tape the condition check-up might not always find all possible faults

The model created in the ATON project is to have the AIPs in more than one format and use WORM-media. Two copies in the same media and format is not enough, at least one additional copy in another media will be needed. It was studied to establish if the authenticity and integrity of the contents can be improved by using digital signatures or complex checksums, but the results were not promising. Digital signatures may be used in contents with a limited life cycle as in the medical sector of data transfer, but both of these technologies are too complex for long-term digital preservation.

Migration between the formats was tested in the project by collecting all contents of a certain type and these were dropped into a patch process. The new versions of the contents were then automatically ingested again into the archive and the complete process is then documented into the metadata. The first process was to create new Flash video distribution copies of videos from archive originals. The previous distribution copies can not be used in migration.

The ATON project material in English: <http://www.mikpoli.fi/showattachment.asp?ID=2347&DocID=1271> and the evaluation report <http://www.mikpoli.fi/showattachment.asp?ID=2348&DocID=1271>

The ATON project was co-funded by the European Social Fund and the Province of Eastern Finland.

---

<sup>4</sup> RDF = Resource Description Framework a W3C standard for metadata management.

<sup>5</sup> [www.profium.com](http://www.profium.com)

# CONVERSION OF THE DANISH NATIONAL ARCHIVES' COLLECTION OF DIGITAL RECORDS

---

Kirsten Villadsen Kristmar

Head of the Department of Appraisal and Transfer

*As the first archival institution in the world The Danish State Archives is about to finish the conversion of its collection of electronic records. The main results are:*

- *that the conversion strategy also works in real life*
- *that conversion can be carried out semi-automatically*

*The Danish strategy for preserving electronic records is based on the transferral of electronic records in system independent formats and the occasional conversion of data to contemporary formats. Public authorities have transferred electronic records to The Danish National Archives since the beginning of the 1970s. In 2005 The Department of Culture supplied The Danish National Archives with an extraordinary grant of 1.960.000 EUR for the conversion of all electronic records, transferred before 2000 to the present preservation standard. This part of the collection consists of extracts of different kinds of database systems where data are structured in files and fields. The present preservation standard implies that data, structured in files and fields are converted into*

- *a relational structure*
  - *a standardized character set*
  - *standardized representations of e.g. dates*
- and digitally documented.*

*The grant was also supposed to cover the digitalization of The Danish State Archives' collection of sound and moving pictures. To be converted were 1.292 extracts of database systems, including record management systems, created by public authorities. Before conversion the collection amounted to 1.024 gigabytes. Data had been transferred in system independent formats but with different character sets and in different structures, including hierarchical databases. Some of the data were packed.*

*To be converted were also 1.800 sound recordings and 200 video sequences.*

*It has taken 4 years to finish the project. In 2005 the project was planned in detail and a binding budget was drawn up. In 2006 the preparations for the conversion were carried out. The preparations included*

- *Restructuring data*
- *Registering data in Daisy, The Danish State Archives' archival database. Daisy can be reached on [www.sa.dk](http://www.sa.dk)*
- *Digitalization of the documentation transferred together with the data*
- *Creation of a XML-schema for describing the original data, including structure, character sets and so on*
- *Creation of a computer program called KonVold. On the basis of the XML-schema, KonVold can convert most of the original data into the present preservation format. However, in some cases it has been necessary to create small computer programs for the preconversion of data.*

*The actual conversion began at the end of 2006 and was carried out by 6 temporary employees with IT-backgrounds but without programming skills. They were supported by 1-2 programmers and an archivist with more than 30 years experience in the field. The management of the project was handled by a full time project manager.*

*At the beginning of June 2008 almost 80 percent of the data were converted, and as planned the project is expected to be finished at the end of the year. The project has evolved almost as planned. However, it has been possible to convert more data than expected, using KonVold. At the moment, it also looks as if it will be possible to convert more data than expected. Data that cannot be converted includes data without documentation or very badly documented.*

[See the presentation.](#)



## La conversion des archives électroniques des Archives nationales danoises

*Ce résumé concerne le thème central de la conservation à long terme.*

*Première institution d'archivage au monde, les Archives nationales danoises sont sur le point d'achever la conversion de leurs archives électroniques. Les résultats principaux sont les suivants :*

- *la stratégie de conversion fonctionne également dans la vie réelle*
- *la conversion peut être réalisée de façon semi-automatique.*

*La stratégie danoise en matière d'archivage électronique est basée sur le transfert d'archives électroniques sous des formats indépendants du système et sur la conversion occasionnelle de données vers les formats actuels. Les autorités publiques transfèrent des archives électroniques vers les Archives nationales danoises depuis le début des années 1970. En 2005, le Ministère de la culture a accordé aux Archives nationales danoises une subvention exceptionnelle de 1 960 000 euros pour convertir la totalité des archives électroniques, transférées avant 2000, à la norme de conservation actuelle. Cette partie de la collection est composée de différentes sortes de systèmes de bases de données où les données sont organisées en fichiers et en champs. La norme de conservation actuelle suppose que les données, organisées en fichiers et en champs, sont converties en :*

- *une structure relationnelle*
- *un jeu de caractères normalisés*
- *des représentations normalisées, telles que les dates*

*et sont documentées de façon numérique.*

*La subvention devait couvrir également la numérisation de la collection de documents sonores et de films.*

*Il s'agissait de convertir 1 292 fichiers extraits des systèmes de bases de données, y compris des systèmes de record management, créés par les autorités publiques. Avant sa conversion, le volume de la collection était de 1 024 gibabits. Les données avaient été transférées sous des formats indépendants du système mais dotées de jeux de caractères différents et sous différentes structures, y compris les bases de données hiérarchiques. Certaines données ont été compressées.*

*Il convenait également de convertir 1 800 enregistrements sonores et 200 séquences vidéo.*

*Il a fallu 4 ans pour parvenir au terme du projet. En 2005, le projet était planifié en détail et un budget était engagé. En 2006, les prévisions de conversion ont été introduites. Ces prévisions concernaient :*

- *la restructuration des données ;*
- *l'enregistrement des données sur Daisy, la base de données d'archivage des Archives nationales danoises. Daisy est accessible sur [www.sa.dk](http://www.sa.dk) ;*
- *la numérisation de la documentation transférée en même temps que les données ;*
- *la création d'un schéma XML pour décrire les données initiales, y compris la structure, les jeux de caractères, etc. ;*
- *la création d'un programme informatique dénommé KonVold. Sur la base du schéma XML, KonVold peut convertir la plupart des données initiales au format de conservation actuel. Dans certains cas, il a été toutefois nécessaire de développer des petits programmes informatiques destinés à la pré-conversion des données.*

*La conversion actuelle qui a débuté fin 2006 a été réalisée par 6 intérimaires ayant de l'expérience en informatique mais sans compétences de programmation. Ils étaient assistés de 1-2 programmeurs et d'un archiviste doté de plus de 30 ans d'expérience dans le domaine. La gestion du projet a été assurée par un chef de projet à plein temps.*

*Début juin 2008, près de 80 % des données étaient converties, et comme prévu, le projet devrait s'achever à la fin de l'année. Le projet a évolué quasiment selon les prévisions. Cependant, en utilisant KonVold, il a été possible de convertir plus de données que prévu, et aujourd'hui également, il semblerait que le volume de données converties dépasse les prévisions. Les données ne pouvant être converties sont celles qui ne sont pas ou qui sont peu documentées.*

[Consulter la présentation.](#)

## ■ Introduction

Danish public authorities have transferred digital records to The Danish National Archives since the beginning of the 1970'ies. In 2000, The Danish National Archives issued its first standard for the transfer of digital records. The standard is also used for long-time preservation. Until then digital records had been transferred in system independent formats, but in a variety of data structures and data formats.

When in 2005 the Ministry of Culture supplied the Danish National Archives with an extraordinary grant of 1,960,000 EUR for the conversion of the collection of digital records transferred before 2000, it had for many years been evident that the collection was in danger of becoming obsolete. Data could not be read by ordinary office or database applications or by applications for statistical use. In order to make use of data, potential users therefore had to have a thorough knowledge of obsolete data structures and data formats and had to be able to develop their own applications to convert them. Even if you had these skills, you could not be certain that you could actually read the data.

The purpose of this paper is to describe the challenge that the Danish National Archives was met with and how the challenge was undertaken. As far as we know, it is the first time that a national archive has gone through all steps of the conversion strategy. Since 2005 the Danish National Archives has not only converted digital records older than 2000 but has also developed a standard application called Sofia that gives access to all electronic records in the present preservation standard.

## ■ Digital Records: The Danish Strategy for Transfer, Preservation and Access

Since the beginning of the 1970'ies the Danish National Archives has based the transferral and long-term preservation of digital records on the conversation strategy. It means:

- that digital records are transferred in system independent formats
- that the collection of digital records is to be converted into new system independent formats before the formats become obsolete
- that the collection of digital records is migrated from one storage medium to another before the medium becomes obsolete

According to the Danish Archives Act, today, all digital records created by the Danish Public Authorities must be transferred according to a single standard. Because all digital records comply with the same standard, the Danish State Archives can develop standard software for converting and creating access to the collection of digital records.

The present standard was taken into use in 2004 but has not been altered much since 2000. The starting point for the standard is that all electronic records are databases where data are structured in files and fields. To some databases digital documents are attached. These systems are e.g. digital record keeping systems. The present preservation standard implies that a system independent extract of data, structured in files and fields are converted into

- a relational structure
- a standardized character set
- standardized representations of e.g. dates
- and digitally documented.

The documentation is both syntactic and semantic. The syntactic part of the documentation is an xml-file where the structure of data is described in a way that makes it possible to recreate a database. The semantic

part of the documentation consists of an over all description of the system and how it was used and a description of the contents of every single file and every single field.

## ■ The Collection of Digital Records

To be converted were 1,292 extracts of database systems, including both business systems and record management systems, created by public authorities. The extracts from business systems were e.g. transferred from the tax authorities, the statistical department, The Ministry of the Interior and the municipalities. They contain a large amount of information about Danish citizens and Danish businesses and their interaction with the Danish public sector. Before conversion the collection amounted to 1,024 gigabytes. Data was transferred in system independent formats but with different character sets and in different structures, including hierarchical databases. Some of the data were packed. To be converted were also 1,800 sound recordings and 200 video sequences.

## ■ Preparations for the Project

The first year was dedicated to the planning of the project. Three experienced members of staff and three members of staff hired for the project analyzed the collection of digital records to establish how much the individual data extracts varied as to formats and structures and a small amount of data was converted. Based on this, it was concluded that it would be possible to develop a computer programme that could perform a semi-automatic conversion of most data. Finally, a budget and a plan were drawn up.

- At the same time as the project was planned, the preparations for it began. They involved:
- A restructuring of the old datasets so that data describing individual years was split out into separate datasets.
- Recording of the datasets in Daisy, a database that contain information about the holdings of The Danish State Archives. Daisy can be accessed on [www.sa.dk](http://www.sa.dk).
- Scanning of the paper documentation that accompanied the original transfers of data.
- Development of the necessary computer programs.
- Hiring of the rest of the necessary staff. The project was organised with its own full time project manager. At the end of the project ten people worked full time to finish it. They all had an IT background, in some cases dating back to the 1970'ies, but only two of them were actually system developers.

## ■ Conversion of data

The computer programme that performs the actual conversion is called KonVold and is based on the creation of an xml-file that describes the data to be converted. For every data set the description contains information about:

- the record creator
- the date of the creation of data
- the character set used

For every file in a dataset, the description contains information about

- the name of the file



- a semantic description of the contents of the file
- the individual record types contained in the file

For every record type the following is described:

- the length of the record
- the starting position of every field
- the length of every field
- type of data in the individual field e.g. string, numerical, date and if packed what sort of compression is used
- a semantic description of the contents of every field
- a description of the codes used in encoded fields
- primary key
- foreign keys

Most of the information can be found in or detected from the paper documentation that accompanied every transfer but some of the information, especially about character sets, has to be detected from an analysis of the hex values of the data in the data set.

When the data has been described, KonVold can validate the description by showing the contents of individual records so that it can be determined if the data has been correctly described. If that is the case, KonVold perform the actual conversion of data by

- splitting the files into new files, one for every record type
- converting the contents of every field into the correct character set and the correct data type
- creating a new xml-file that describes the new structure

Finally KonVold attaches the scanned version of the original paper documentation and the new xml-file to the converted data. It can now be tested if the converted data and their documentation comply with the present standard using the same test programme that is used for testing transfers of data.

As its first task the project set out to convert all data extracts from record management systems. The reason for this was that it was assumed that these data sets would be fairly simple and that they would resemble each other. This assumption proved to be correct.

At this phase of the project a log of decisions taken in connection with the conversion process was begun. The decisions could for example concern description of missing data parts or guidelines for correcting data. The log was maintained right to the end of the project and it is therefore a valuable documentation of the result of the conversion.

It proved possible to convert most data using KonVold. However, new conversion routines had to be added to KonVold when a new character set was discovered and when other new issues had to be handled. KonVold has therefore been updated several times. There have also been problems that KonVold couldn't handle e.g. header information that had to be removed before conversion could take place. In these cases a small routine was developed that removed the necessary data so that the rest of the conversion process could be carried out in KonVold.

## ■ Conversion of sound and video

Part of the project was also the conversion or rather digitalization of sound and video found in the collections of The Danish State Archives. The oldest part of the collection dated back to the 1920'ies and had typically been transferred together with paper records from public and private record creators such as political

parties. The collection consisted of a substantial number of different media types some of which were rapidly beginning to decay. The first part of the project was an appraisal of which part of the collection had to be digitalized. As foreseen a considerable number of recordings were also part of the holdings of The State Library that is responsible for the long term preservation of radio and television broadcasts. Other recordings were not considered valuable enough to digitalize. Finally, it was decided not to digitalize films because they are better preserved on the media where they were initially created. At the next stage it was decided to preserve sound in Wave 48 kHz 24 bit and video in MPEG2. The actual digitalization was carried out by two private firms. Hein IT in Denmark took care of all sound recordings and Moving Media in Ireland took care of all video recordings. When the digitalized sound and video was returned to The Danish State Archives, it was tested and finally documented in the same way as digitally born records thereby making it possible to preserve and create access to digitalized sound and video in the same way as digitally born records.

## ■ Experiences Gained

The most important result of the project is of course that it has been possible to carry out a semi-automatic conversion of data, transferred during a period of more than twenty years, thereby proving that the conversion strategy actual works. It is also an important result that the strategy not only works for digitally born records but also for digitalized records.

But the project has also given valuable input to the continuous work with transfer, preservation and access. Among the experiences are

- A confirmation of the importance of securing the necessary documentation at the time of transfer. If you don't, it may be very difficult to understand data at a later time.
- A confirmation of the importance of testing that data transferred to the archive complies with the standard. If you don't, it will be much more expensive to convert.
- The necessity to establish the right IT-infrastructure right from the beginning. The project would have been finished some months earlier and much frustration would have been avoided if it had been realized how much processing power is needed for the conversion process.

Thus, the project has produced valuable input both to the ongoing revision of the standard for transfer and preservation and to future conversions.

Finally, it has been established that it was the right decision to organize the conversion as a project though the costs should not be overlooked. It has been expensive for the basic organisation to hire and train so many people for such a short time period. At the same time, there is no doubt that some of the knowledge acquired by the temporary staff will disappear together with them. On the other hand, the conversion would never have been finished if it wasn't possible to dedicate a number of full time posts to it.

# DIOSCURI: EMULATION IN PRACTICE<sup>1</sup>

---

Remco Verdegem

Senior Advisor Digital Longevity, National Archives of the Netherlands

For a long time migration was the only viable preservation approach for most cultural heritage institutions. Since a few years the digital preservation community is starting to recognize the need of applying emulation as well as migration in their preservation policy. Migration is focusing on the digital object itself, changing the object in such a way that software and hardware developments will not affect its original representation. By converting the file format of a digital object, it is possible to render these objects on current platforms. Emulation on the other hand does not focus on the digital object, but on the hardware and software environment in which the digital object is rendered. It aims at recreating the environment in which the digital object was originally created. Emulation is still looked upon with scepticism due to its technical complexity and initial costs. In 2005, the National Archives of the Netherlands and the National Library of the Netherlands acknowledged the need for emulation, especially for rendering complex digital objects in the future without affecting their authenticity and integrity. For these reasons, a project was started to investigate the feasibility of emulation as a preservation approach by developing and thoroughly testing an emulator specifically designed for digital preservation purposes. The objective was to develop an emulator capable of recreating a modern x86 computer environment, while remaining durable and easy to configure. In July 2007 the first version of this emulator was delivered: Dioscuri. Dioscuri is available as open source software<sup>2</sup>. Since July 2007 Dioscuri officially became part of the European project Planets<sup>3</sup>.

Dioscuri has two distinct features:

- **Modularity:** Each module is a small emulator that reproduces the functional behaviour of its related hardware component (e.g. hard disk, processor, memory). By applying such a component based architecture, modules can be arranged in all kinds of configurations, similar to real hardware. Based on the requirements of operating system and applications, a customized emulator can be created that fits these requirements.
- **Durability:** Running software indefinitely is wishful thinking. Every computer application is dependent on its underlying platform consisting of hardware and system software. Changes in the platform hold implications for the software depending on it. Porting software to multiple computer platforms reduces the risk that none of these instances work over time. To achieve this an intermediate layer between the host platform and the emulator is realized, called a Virtual Machine (VM). As a result only the VM has to be maintained over time, while the emulator that runs on the VM remains untouched.

Although developing an emulator is not an easy task, the joint project of KB and Nationaal Archief of the Netherlands has shown that emulation is a feasible preservation approach. The current version of Dioscuri is able to authentically represent obsolete digital objects like WordPerfect 4.2 and 5.1 files.

At the moment the current processor (16 bit) is extended to support 32 bit computing, allowing executing of a larger range of software like modern versions of Microsoft Windows and Linux. Support will be added for new modules like mouse, sound and improved graphics.

The presentation will place emulation within the context of digital preservation and will indicate the strengths and weaknesses of emulation as compared to migration. It will focus on the key attributes of Dioscuri and the results so far. The presentation will end with a live demonstration of Dioscuri, the modular emulator for the digital preservation community.

[See the presentation: part 1; part 2.](#)

---

<sup>1</sup> Based on the paper "Emulation for digital preservation in practice: the results", by Jeffrey van der Hoeven (National Library of the Netherlands), Bram Lohman (Tessella Support Services plc) and Remco Verdegem (National Archives of the Netherlands), published in the International Journal of Digital Curation, Issue 2, Volume 2, 2007

<sup>2</sup> <http://sourceforge.net/projects/dioscuri/>

<sup>3</sup> <http://www.planets-project.eu/>

## **DIOSCURI : l'émulation en pratique**

*La migration a longtemps été la seule approche viable en matière de conservation pour la plupart des institutions du patrimoine culturel. Depuis quelques années, la communauté de la conservation numérique reconnaît la nécessité d'appliquer l'émulation et la migration dans sa politique de conservation. La migration concerne l'objet numérique en lui-même, en changeant l'objet de telle façon que les développements de logiciels et de matériel n'affectent pas sa présentation originale. En convertissant le format de fichier d'un objet numérique, il est possible de restituer ces objets sur les plates-formes actuelles. Quant à l'émulation, elle ne concerne pas l'objet numérique mais l'environnement matériel et logiciel où l'objet numérique est restitué. Elle vise à recréer l'environnement où l'objet numérique a été initialement créé. Cependant, l'émulation est encore considérée avec scepticisme en raison de sa complexité technique et de son coût initial. En 2005, les Archives nationales et la Bibliothèque nationale des Pays-Bas ont reconnu la nécessité de l'émulation, notamment en ce qui concerne la restitution future d'objets numériques complexes sans impact sur leur authenticité et leur intégrité. Pour ces raisons, un projet a été lancé afin d'étudier la faisabilité de l'émulation en développant et en testant de façon exhaustive un émulateur conçu à des fins spécifiques de conservation numérique. L'objectif était de développer un émulateur capable de recréer un environnement informatique moderne x86 tout en étant pérenne et facile à configurer. En juillet 2007, la première version de cet émulateur a été livrée : Dioscuri. Dioscuri est disponible en logiciel libre<sup>4</sup>. Depuis juillet 2007, Dioscuri fait officiellement partie du projet européen Planets<sup>5</sup>.*

*Dioscuri présente deux caractéristiques distinctes :*

- *Modularité : chaque module est un petit émulateur reproduisant le comportement fonctionnel du composant matériel correspondant (par ex. disque dur, processeur, mémoire). En s'appuyant sur une telle architecture à base de composants, les modules peuvent être utilisés sous tous les types de configurations, similaires au matériel physique. Basé sur les exigences du système d'exploitation, un émulateur sur mesure peut être créé de façon à répondre à ces exigences.*
- *Pérennité : utiliser un logiciel indéfiniment serait un vœu pieux. Chaque application informatique dépend de sa plateforme sous-jacente composée du matériel et du logiciel système. Les modifications de la plateforme ont des conséquences sur le logiciel qui en dépend. Une configuration faisant appel à des plateformes informatiques multiples réduit le risque de défaillance au fil du temps. Pour y parvenir, une couche intermédiaire entre la plateforme d'accueil et l'émulateur est réalisée, sous le nom de « machine virtuelle ». En conséquence, seule la machine virtuelle doit être maintenue à long terme, alors que l'émulateur qui fonctionne sur la machine virtuelle reste intact.*

*Bien que le développement d'un émulateur ne soit pas une tâche aisée, le projet conjoint de KB et des Archives nationales des Pays-Bas a montré qu'une émulation est une approche réalisable en matière de conservation. La version actuelle de Dioscuri est capable de représenter de façon authentique des objets numériques désuets tels que des fichiers WordPerfect 4.2 et 5.1.*

*Aujourd'hui, le processeur en place (16 bits) est étendu pour supporter 32 bits afin de permettre l'exploitation d'une plus grande gamme de logiciels telles que les versions modernes de Microsoft Windows et Linux. Le support sera rajouté pour de nouveaux modules tels que la souris, le son et des graphiques améliorés.*

*Cette présentation décrira l'émulation dans le contexte de la conservation numérique et analysera les forces et les faiblesses de l'émulation par rapport à la migration. Elle concernera les caractéristiques clés de Dioscuri et les résultats obtenus à ce jour. Nous terminerons la présentation par une démonstration de Dioscuri, l'émulateur modulaire destiné à la communauté de la conservation numérique.*

[Consulter la présentation: partie 1; partie 2.](#)

---

<sup>4</sup> <http://sourceforge.net/projects/dioscuri/>

<sup>5</sup> <http://www.planets-project.eu/>

## ■ Introduction

With the ongoing paradigm shift from traditional, physical documents moving towards digital representation, the term preservation acquires a whole new dimension. No one, not even software companies themselves, seems to be prepared for the impact of this massive movement to digital in our daily life. Each day new inventions and technologies introduce new capabilities in the digital world, quickly superseding old and obsolete versions.

Today, every individual or organisation builds up a digital case file which grows in volume as well as in complexity. Losing digital information can really harm an organisation and it can be an expensive operation to repair the damage done. As successful caretakers of old documents, artefacts, music and much more, cultural heritage organisations were the first to be aware of this fragility. With current projects on digital preservation, they are trying to prevent a digital black hole in history.

The National Archives of the Netherlands is responsible for the long-term preservation of Dutch cultural heritage and governmental information and has the legal obligation to preserve and give access to archival records of Dutch governmental organisations. With an increasing amount of digitized and born digital documents and applications, new protocols and strategies for preservation and access are required.

Digital objects are dependent on their environment. They rely upon the hardware and software that were originally intended to interpret them. When the environment becomes obsolete, sometimes within the space of a few years, the problem how to read that object without the original hardware and software arises. It is unlikely that different versions of a software application will interpret the digital object in the same way and this may well result in a change of the rendered object (the visible or available view of the computer file(s)) affecting its (archival) integrity. Some data may be lost altogether; in other areas, data may be gained. It is not always possible or feasible to compare a new rendition with the original, so changes may go unnoticed. Any changes to the digital object may affect its authenticity and integrity, which in turn may affect its (archival and legal) status. Depending on the nature of the digital object and its use, this can cause problems, not least that of losing or misrepresenting history.

Long-term preservation of digital objects requires more than secure storage of the objects. It also includes the development and execution of preservation strategies to access and render these objects, now and in the future. Such strategies or approaches can roughly be divided into two groups: migration and emulation. Migration is focussing on the digital object itself, changing the object in such a way that current software and hardware developments will be able to render the object authentically. By converting the format of an object, it is possible to render these objects on modern platforms. Emulation on the other hand is focussing on the hardware and software environment in which the digital object was originally created and rendered, leaving the digital object intact. Emulation aims at (re)creating the environment in which the digital object was originally created and used.

The wide variety of digital formats and applications makes it impossible to choose a so called “one-size-fits-all” solution for preservation. The choice for a specific preservation strategy should be determined by the retention period of the object, the essential characteristics of the object, the requirements of the user that would like to access the object and the policy of the institution responsible for the preservation of the object. The National Archives of the Netherlands consider migration and emulation to be complementary.

## ■ Emulation in practice

The choice for emulation as a preservation strategy is not undisputed, although its benefits are recognized. Emulation tools like MS Virtual PC<sup>6</sup>, QEMU<sup>7</sup> and Bochs<sup>8</sup>, and virtualisation tools like VMware<sup>9</sup> have become very

---

<sup>6</sup> Microsoft Virtual PC. Available at <http://www.microsoft.com/windows/virtualpc> (accessed 10 November 2008).

<sup>7</sup> QEMU. Available at: <http://bellard.org/qemu> (accessed 10 November 2008)

<sup>8</sup> Bochs (2007). Available at: <http://bochs.sourceforge.net/> (accessed 10 November 2008)

<sup>9</sup> VMware (2007). Available at: <http://www.vmware.com/> (accessed 10 November 2008)

advanced and are capable of running complete operating systems like Microsoft Windows, Apple's MacOS or GNU/Linux, supporting a wide array of device peripherals. However, none of these solutions have been designed specifically for digital preservation. Current emulation and virtualisation tools provide no guarantee that they will operate on new computer platforms in the future. Various techniques have been proposed to overcome this problem of obsolescence (e.g. by chaining or migrating emulators<sup>10</sup>), but these techniques introduce a new risk of degrading functionality and performance each time the emulation process has to adapt to new circumstances.

From a preservation perspective this is an undesired situation. The need for a digital preservation-proof emulator is evident, but requires high initial effort. There are claims that developing such an emulation solution is far too complex and expensive<sup>11</sup>. Another drawback mentioned is that the lack of data exchange between the emulated and real environment is too much of a disadvantage to make it a worthwhile solution<sup>12</sup>.

The fact remains that until now emulation has never been developed and tested within an operational digital archiving environment. The National Archives of the Netherlands and the National Library believe that emulation provides a good solution for long-term access to digital objects without affecting their authenticity and integrity. These institutions agree that this strategy has to be developed and tested first, before the potentials and limitations of emulation can be assessed.

That is why in 2005 these institutions started a two year project to develop a preservation strategy based on emulation. The objective of this project was to develop an emulator that is capable of recreating a modern x86 computer environment, while remaining durable and easy to configure. Furthermore, it should support a mechanism to extract and insert data from the emulated to the real environment.

## Design

Before the start of the development of the emulator, a preliminary study into emulation-based preservation was conducted. The current state of affair was explored and several possibilities to create a preservation-proof emulation strategy were researched<sup>13</sup>. In cooperation with emulation advocate Jeff Rothenberg, this resulted in a new design for an emulation strategy: modular emulation<sup>14</sup>. The principles of this strategy are based on earlier ideas about an Emulation Virtual Machine (EVM) of Jeff Rothenberg<sup>15</sup> and the Universal Virtual Computer (UVC)-based preservation method of Raymond Lorie<sup>16</sup>. Two aspects distinguish this design from any other emulation approach: modularity and durability.

## Modularity

The modular emulation strategy is strongly based on the basic architecture of today's hardware, known as the Von Neumann architecture<sup>17</sup>. Modular emulation can be defined as:

*“Emulation of a hardware environment by emulating the components of the hardware architecture as individual emulators and interconnecting them in order to create a full emulation process. In this, each distinct module is a small emulator that reproduces the functional behaviour of its related hardware component, forming part of the total emulation process.”*

<sup>10</sup> R. Verdegem, J.R. van der Hoeven, Emulation: To be or not to be (IS&T Archiving conference, Ottawa, Canada, 2006) pgs. 56 - 60

<sup>11</sup> S. Granger, Emulation as a Digital Preservation Strategy (D-Lib Magazine, Vol. 6 No. 10, 2000). Available at: <http://www.dlib.org/dlib/october00/granger/10granger.html> (accessed 10 November 2008)

<sup>12</sup> T.A. Phelps, P.B. Watry, A No-Compromises Architecture for Digital Document Preservation (ECDL2005, Vienna, Austria, 2005) pg. 266

<sup>13</sup> J.R. van der Hoeven et al, Emulation - a viable preservation strategy (The Hague, The Netherlands, 2005). Available at: [http://www.kb.nl/hrd/dd/dd\\_projecten/Emulation\\_research\\_KB\\_NA\\_2005.pdf](http://www.kb.nl/hrd/dd/dd_projecten/Emulation_research_KB_NA_2005.pdf) (accessed 10 November 2008)

<sup>14</sup> J.R. van der Hoeven, H.N. van Wijngaarden, Modular emulation as a long-term preservation strategy for digital objects (IWAW, Vienna, Austria, 2005). Available at: <http://www.iwaw.net/05/papers/iwaw05-hoeven.pdf> (accessed 10 November 2008)

<sup>15</sup> J. Rothenberg, An experiment in using emulation to preserve digital publications, Koninklijke Bibliotheek, The Hague, The Netherlands, pg. 8. (2000)

<sup>16</sup> R.A. Lorie, Long-term archiving of digital information, IBM Research report, IBM Almaden Research Center, San Jose, Almaden (2000)

<sup>17</sup> Von Neumann computer architecture. Available at: [http://en.wikipedia.org/wiki/Von\\_Neumann\\_architecture](http://en.wikipedia.org/wiki/Von_Neumann_architecture) (accessed 10 November 2008)



By applying such a component-based architecture, modules can be arranged in all kinds of configurations, similar to real hardware. Based on the requirements of operating system and applications, a customized emulator can be created that fits these requirements.

Of course, a lot of effort is required when recreating every hardware piece into a software module. But emulating hardware is preferred over other levels of emulation such as emulating an operating system or application<sup>18</sup>. By emulating hardware, the original operating system, applications, drivers and configuration settings, which guarantee authenticity of the original software environment, are retained. Moreover, as hardware has to be manufactured, well-defined specifications are required. Although these are not always publicly available they offer a better understanding of the hardware's functionality. Similarly, applied industry standards are well-described and form a good starting point for learning about hardware interfaces.

## Durability

Running software indefinitely is wishful thinking. Every computer application is dependent on its underlying platform, consisting of system software and hardware. Changes in the platform hold implications for software depending on it. Porting software to multiple computer platforms reduces the risk that none of these instances work over time. This line of thought forms the basis for creating a sustainable emulator and has been addressed in ideas of Jeff Rothenberg's EVM-approach, Raymond Lorie's UVC and Olonys VM from Vincent Joguín<sup>19</sup>. These concepts propose an intermediate layer between host platform and emulator, called a virtual machine (VM). A VM allows running the same application on different computer platforms without the need to change that application. The only restriction is that the VM's interface to this application remains stable over time, while the interface between the VM and underlying host platform is adapted each time that platform changes. As a consequence only the VM needs to be maintained over time, while any application (emulator, migration tool, etc.) that runs on the VM remains untouched. For this reason, the modular emulation strategy includes the use of a VM on which the emulator can run.

## Modular emulation strategy

In Figure 1 the system design for the modular emulation strategy is depicted. There are five main elements:

- Universal Virtual Machine (UVM)
- Modular emulator
- Component Library
- Controller
- Emulator specification document (ESD)

The UVM ensures that the modular emulator can operate on many different computer systems, now and in the future, ranging from a PC or Mac to a mobile device or embedded hardware. Configuration of the modules in the emulator is done via a controller. This component reads the desired configuration from an Emulator Specification Document (ESD) which could be an XML-structured input file provided by a user or automated service. Using the information in the ESD the controller selects the requested modules from a component library and starts the emulation process.

---

<sup>18</sup> Digital Preservation Testbed White Paper (pg 20-22), [http://www.digitaleduurzaamheid.nl/bibliotheek/docs/white\\_paper\\_emulatie\\_EN.pdf](http://www.digitaleduurzaamheid.nl/bibliotheek/docs/white_paper_emulatie_EN.pdf) (accessed 10 November 2008)

<sup>19</sup> V. Joguín, Emulating emulators for long-term digital objects preservation: the need for a universal machine (Emulation Expert Meeting 2006, The Hague, The Netherlands, 2006) Available at: [http://www.kb.nl/hrd/dd/dd\\_projecten/slides/eem\\_aconit\\_vjoguin.pdf](http://www.kb.nl/hrd/dd/dd_projecten/slides/eem_aconit_vjoguin.pdf) (accessed 10 November 2008)

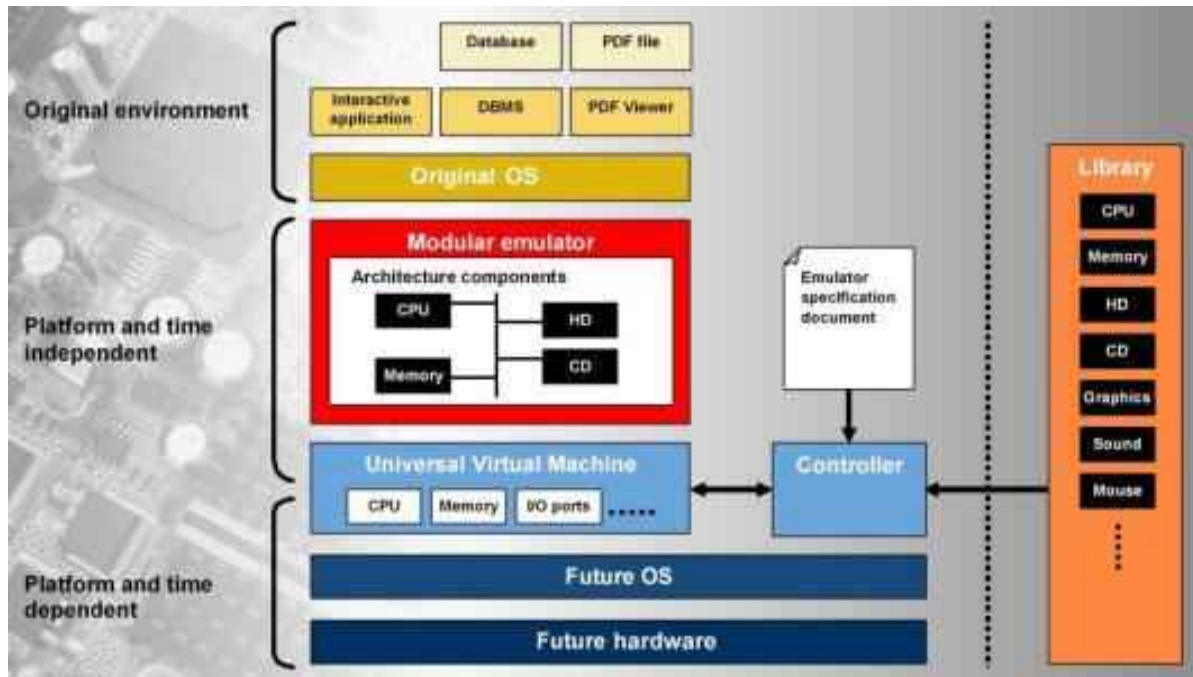


Figure 4: design of modular emulation strategy

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

## ■ Building the emulator

Building a personal computer from scratch in software, component after component, is a daunting task. The amount of information required for building each piece is overwhelming, not to mention the inter-component messaging. To make the process more manageable, a twofold simplification of the development path was applied: the number of components was limited to include only the essentials, and the complexity of each component was minimized. New components and functionality were added once the existing components were working as required. Also, when necessary, the complexity of the existing components was expanded at this time.

As mentioned before, most of the current computer systems are based on the Von Neumann architecture [Figure 2]. This architecture consists of memory, a control unit, an Arithmetic Logic Unit (ALU), and input and output devices. The simplest emulator that adheres to this architecture consists of Random Access Memory (RAM) and a Central Processing Unit (CPU) for control and arithmetic/logic. Input and output are only necessary for interaction with the user, so simple files were used as input and output, respectively. The implementation of this initial concept offers roughly fifty percent of the functionality of a current computer, as the CPU and memory subsystems make up the majority of the system.



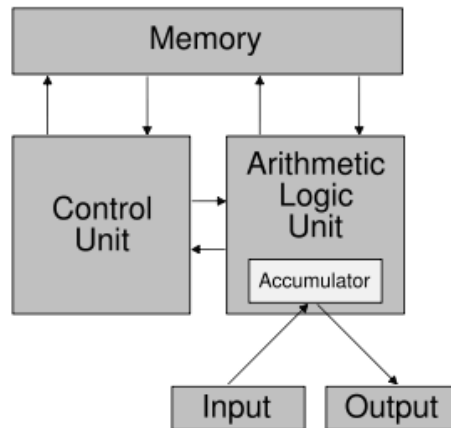


Figure 2: Von Neumann architecture

In terms of basic functionality, it was decided to start by emulating an Intel 8086 CPU, the first of the x86 architecture - which is still the most popular today. Following the original development of the x86 architecture makes sense because all subsequent generations of the x86 architecture have been built on the groundwork of the 8086. Every x86 processor is backwards compatible with the original 8086 as the core of today's x86 processors is still the same as that of the 8086 released in 1978 [Figure 3].

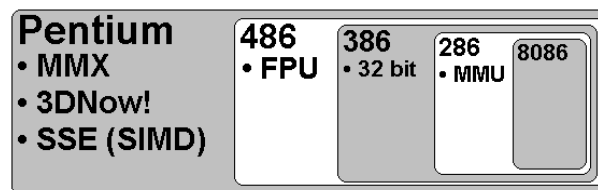


Figure 3: incremental generations of CPUs

This method of prototyping has another advantage: it allows a rapid development of the emulator, delivering several different models of a CPU while at the same time continuing development in the right direction.

Once the basic Von Neumann architecture was in place, other peripherals were added to interface with these components. The addition of proper input in the form of a keyboard and mouse module, as well as proper output through a video card, along with a 'screen' to display the output, offers a huge amount of functionality to the emulator. Other, more indirect components such as an interrupt mechanism, interval timers, floppy and hard disk support, have served to enhance the user experience of the emulator, and allowed for a more faithful rendition of the original system.

The development of these components naturally leads to certain modularity, since each software module of a component is a small emulator by itself. However, with future iterations of the emulator in mind, careful selection of boundaries of each component must be taken into account, as the goal of the modularity is to provide the ability to interchange similar software components and still allow the emulator to run, within physical limits of course.

This requires a careful implementation of each component. Java was chosen as programming language because it is Object Oriented (OO), which supports the modular design and it provides a Virtual Machine, making the emulator portable to many Java-enabled platforms. For each hardware component, the logical switches and electronic pulses were translated into software objects with associated functions and interfaces. This allows each component from the same family to adhere to certain methods and calls which will assure it is interchangeable.

Using abstract classes and a modular implementation, where each component extends the parent class ('Module') and its inherited methods, Java guarantees that interchanging components will adhere to the standards set in the Module class [Figure 4]. This way, any future module can be placed in an existing emulator and still be able to interface with other modules.

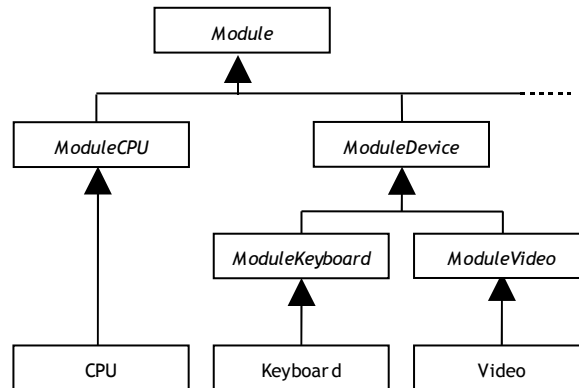


Figure 4: modular structure in Java

### Benefits and limitations of the virtual framework

The choice of portability played a huge part in the development process. However, the decision to develop for the Java Virtual Machine (JVM), also limited the software possibilities.

The obvious benefit of using the JVM as an interface layer to the underlying platform is that this layer provides the specific calls for each architecture. Generic methods can be written that are guaranteed to work on every architecture, as the JVM provides the necessary interface. This removes any system specific implementations and calls necessary to interface with the underlying OS and hardware. It also takes away the need for recompilation on any computer platform. Instead, it allows the development to focus on the implementation of the functionality.

Unfortunately, choosing to work on top of an intermediate layer also has its disadvantages. The choice of programming language immediately forces developers to use the language's syntax, type system, and libraries. However, if these prove to be insufficient to perform certain tasks, workarounds have to be implemented to ensure all functionality of the underlying components can be reliably emulated. An obvious example in Java is that it uses only signed primitive types. A Java byte therefore has different values (-128 to +127) than hardware bytes (0 to +255), although the range is the same. Memory addressing as well as computational logic requires careful implementation to adhere to the intended function.

Another drawback of developing on an intermediate layer is that only high-level functionality is supported. In Java, there is no direct primitive for a bit, and although bit-wise operators exist, these can be cumbersome to work with when using (signed!) bytes. This lack of low-level hardware functionality, which in the case of an emulator is necessary, can be difficult, and is detrimental to the speed and efficiency of the emulator, and makes it sensitive to mistakes.

### The need for reference material

Building an emulator depends on the availability of accurate documentation, which is not always easy to come by. This problem is twofold: on one hand, some of the components that are currently being emulated are more than twenty years old, and so a lot of the documentation has disappeared over the years. The other problem is that not all components were designed based on industry standards, and this information was never published. A current-day example is the video card. Most major video cards are produced by companies for which commercial interest is of the utmost importance. This means that no specifications are published, for fear of competitors learning trade secrets. This makes it especially difficult to emulate these components at a low level for which the inner workings are necessary to know.

Fortunately, other components that were developed according to official or de facto standards, and thus properly documented, can be easily emulated. The x86 architecture and ATA storage device interface are both widely published, and this information has made it fairly straightforward to build software components that fulfil the requirements.

## ■ Results so far

In July 2007 the first version of the modular emulator was released as open source software under the name Dioscuri. It takes its name from the Greek myth of the twins Castor and Pollux: one is mortal while the other becomes immortal. This is a symbolic representation of the idea behind emulation and long-term preservation: giving mortal digital objects their immortal equivalents.

Version 0.2.0 can emulate the following hardware components:

- Intel 8086 based (16-bit) Central Processing Unit (CPU)
- Random Access Memory (RAM)
- VGA graphics adapter
- Text and graphics display
- Floppy disk support
- Hard disk support
- Interrupt handling
- Timing mechanism (clock)
- BIOS and CMOS settings
- Motherboard supporting I/O address space
- Direct Memory Access (DMA) for fast data transfer
- AT/XT/PS2 compatible keyboard supporting multiple keyboard layouts

At the time of writing the latest release is version 0.3.0. This version supports 16-bit and 32-bit processing and is capable of running MS Windows 3.0 and other 32-bit applications. The emulated 32-bit CPU is taken from the JPC project<sup>20</sup> with whom the Dioscuri project team has been cooperating. The new features of this release are:

- 16-bit and 32-bit X86-based CPU (real-address and protected mode)
- Added serial mouse support (experimental)
- Added serial port (UART 16550A with COM-ports 1 to 4)
- Fixed minor bugs in modules CPU, DMA, PIC
- Updated GUI

## ■ Conclusions

Although developing an emulator has proven not to be an easy task, the joint project of National Archives of the Netherlands and the National Library has shown that it is feasible, even with limited resources. The total effort to develop Dioscuri is roughly two man years. Much work still has to be done, but the current version of

---

<sup>20</sup> <http://www-jpc.physics.ox.ac.uk/>

Dioscuri already shows its value by executing old applications more accurately than is done by a modern computer platform. Dioscuri is more durable than other emulators as it is portable to a great variety of computer platforms without extra effort. Due to its modular design, Dioscuri can be configured into any target platform based on the available modules.

Having reached this milestone, next steps are already in progress. Since July 2007 Dioscuri officially became part of the European project Planets, a four-year project with the primary goal to build practical services and tools to help ensure long-term access to digital cultural and scientific assets. Within this context development of Dioscuri is continued to extend its functionality. The processor has been extended to support 32-bit computing, allowing execution of a greater range of software like modern versions of Microsoft Windows and Linux. Support will be added for new modules like mouse, sound and improved graphics. Also, a module library will be designed and the configuration and invocation of the emulator itself will be further automated. By making Dioscuri open source and supporting a dedicated emulation development platform, any interested individual or organisation can use Dioscuri for their personal needs or integrate Dioscuri into their preservation process. Moreover, it stimulates joint development to take Dioscuri to the next level.



# **Table-ronde**

## **Les Archives sont-elles prêtes pour affronter le monde numérique ?**

### **Réflexions sur le métier**

---

---

Sous la présidence d'Evelyn WAREHAM

Archives New Zealand

***Round-table***  
***Are archivists ready for***  
***the digital world?***  
***Reflection on the***  
***profession and its***  
***evolution***

---

---

*Chairing Evelyn WAREHAM*

*Archives New Zealand*

# Introduction

## **BREAKING THE BARRIERS OF TRADITIONAL RECORDS MANAGEMENT. RECORDS, RECORDS MANAGEMENT, ENTERPRISE CONTENT MANAGEMENT AND MoReq2**

Dr. Ulrich Kampffmeyer

Managing Director of Project Consult (Germania, Hamburg)

*The presentation covers current trends in integrating records management functionality in standard software and web 2.0 applications. Based on the enhanced scope of MoReq2 topics include :*

- *expanding RM into ECM*
- *RM starts when generating content*
- *taxonomy and ontology vs. folksonomy*
- *integrating RM in Office and ERP environments*
- *enterprise search vs. metadata retrieval*
- *RFID linking the paper world into ERM*
- *automated format and metadata conversion*
- *intelligent objects require new management strategies*
- *knowledge protection vs. open access to information*
- *audit trails as source for business process management*
- *from the records manager to the information manager*

[See the presentation.](#)

### **Faire sauter les barrières du records management traditionnel**

*La présentation couvrira les tendances actuelles en matière d'intégration des fonctions de records management dans les logiciels standard et applications web 2.0.*

*Sur la base de la version élargie de MoReq2, les thèmes suivants seront abordés :*

- *L'évolution du records management vers la gestion de contenu*
- *Le records management commence avec la création de contenus*
- *La taxonomie et l'ontologie par rapport à la 'folksonomie'*
- *L'intégration du records management dans des outils bureautiques et de planification des ressources*
- *Les besoins de recherche de documents face à la récupération de métadonnées*
- *La radio-identification en lien avec le monde papier dans le cadre de l'archivage électronique*
- *La conversion automatisée des formats et des métadonnées*
- *Les objets intelligents exigent de nouvelles stratégies de gestion*
- *La protection du savoir face au libre accès à l'information*
- *Les pistes des audits comme source de gestion des processus de l'entreprise*
- *La transition du métier de records manager vers celui d'« information manager »*

[Consulter la présentation.](#)



## ■ 1. Records and Records Management

The perception of records and records management has changed over time. In the transition from a world of paper and folders to the digital age, the definition and understanding of records and records management had to be adjusted. The most widely accepted definition of records and records management are given by the ISO standard 14589:

### *“Record*

*Information created, received, and maintained as evidence and information by an organisation or person, in pursuance of legal obligations or in the transaction of business.”*

### *“Records Management*

*Field of management responsible for the efficient and systematic control of the creation, receipt, maintenance, use and disposition of records, including processes for capturing and maintaining evidence of and information about business activities and transactions in the form of records.”*

But different users have different interpretations of the terminology and what records management is:

- The archivist sees it as “pre-preservation”
- The public sector records manager sees it as result of a “bureaucratic process”
- The chief compliance officer sees it as “legal evidence” of his business activities
- The IT manager sees it as “application” of his ERP-solution
- The programmer sees it as “structure and fields” in his legacy application ...
- The database manager sees it as “contents of tables” ...
- The music fan sees it as the “directory functionality” of his LastFM records ...

Accordingly, there is a need for clearer and more general definition of the term “record” before we can talk about records management. Records can easily be defined by their attributes like business value, legal value, context, integrity, persistence, authenticity, defined metadata, consistency, classification, and more. These attributes separate records from other types of documents and information objects. This leads to the following five step approach to define the term “record:”

A record is

- Defined by its legal and business value,
- A stable, authoritative and consistent information object,
- Independent of its physical format,
- Made up of its contents enriched with context and descriptive information,
- Identified by a unique identifier and its classification.

In the electronic world, records can originate from web content, office documents, scanned paper, emails, media assets, forms, files, images, videos, blogs, audios, twitters, PDFs, datasets, etc. Some propose general use of the term “information objects” only. But there is still a need for the concept of a record. Not all information objects will become a record, and not all information objects, or records for that matter, will end up in an archive. Records management is not digital preservation. Digital preservation is a support function of records management to ensure the integrity and persistence of the records. And we should not forget that we live in a hybrid environment. Records can also be paper-based and should be managed by the database of the records management system combined and synchronized with electronic information objects.

One question is whether records management will stay a discipline of its own or if it will become a general feature of all business activities. In the context of electronic records management systems, this leads to the next question, whether an electronic records management solution is stand-alone or part of the IT infrastructure serving other applications.

## ■ 2. Comparing different Records Management approaches

The differences in regard to approach records management systems can be shown using the industry term ECM Enterprise Content Management by comparison with the MoReq2 concept.

### The ECM Enterprise Content Management concept

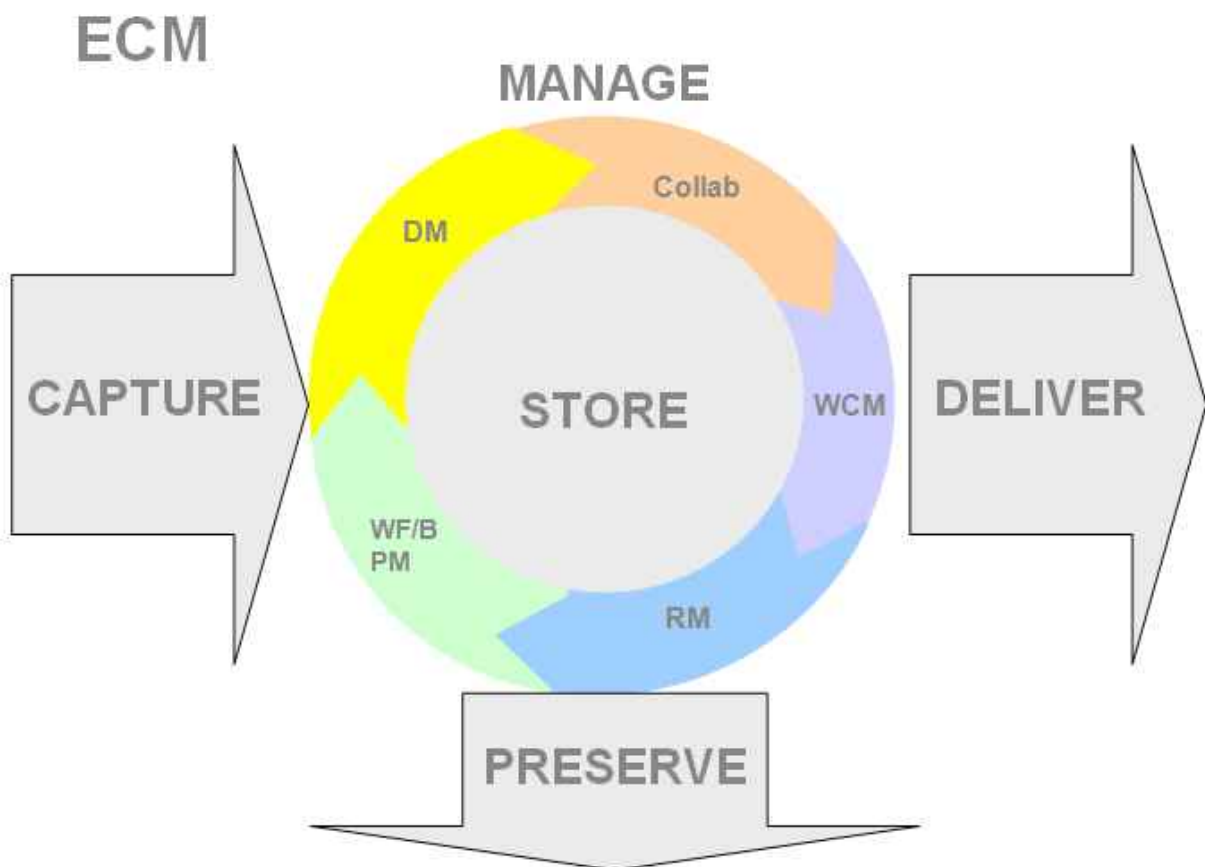


Fig.1: The AIIM ECM model with the five major components capture, manage, store, deliver and preserve. The manage component contains the modules document management, collaboration, web content management, records management, and workflow / business process management.

The definition of ECM by the international industry organization AIIM international in 2008:

*“Enterprise Content Management (ECM) is the strategies, methods and tools used to capture, manage, store, preserve, and deliver content and documents related to organizational processes. ECM tools and strategies allow the management of an organization’s unstructured information, wherever that information exists.”*

In the ECM model, RM records management is an integrated module which is part of the process of managing information and interacts with the other components of the ECM model like workflow, document management, digital preservation, input management etc.

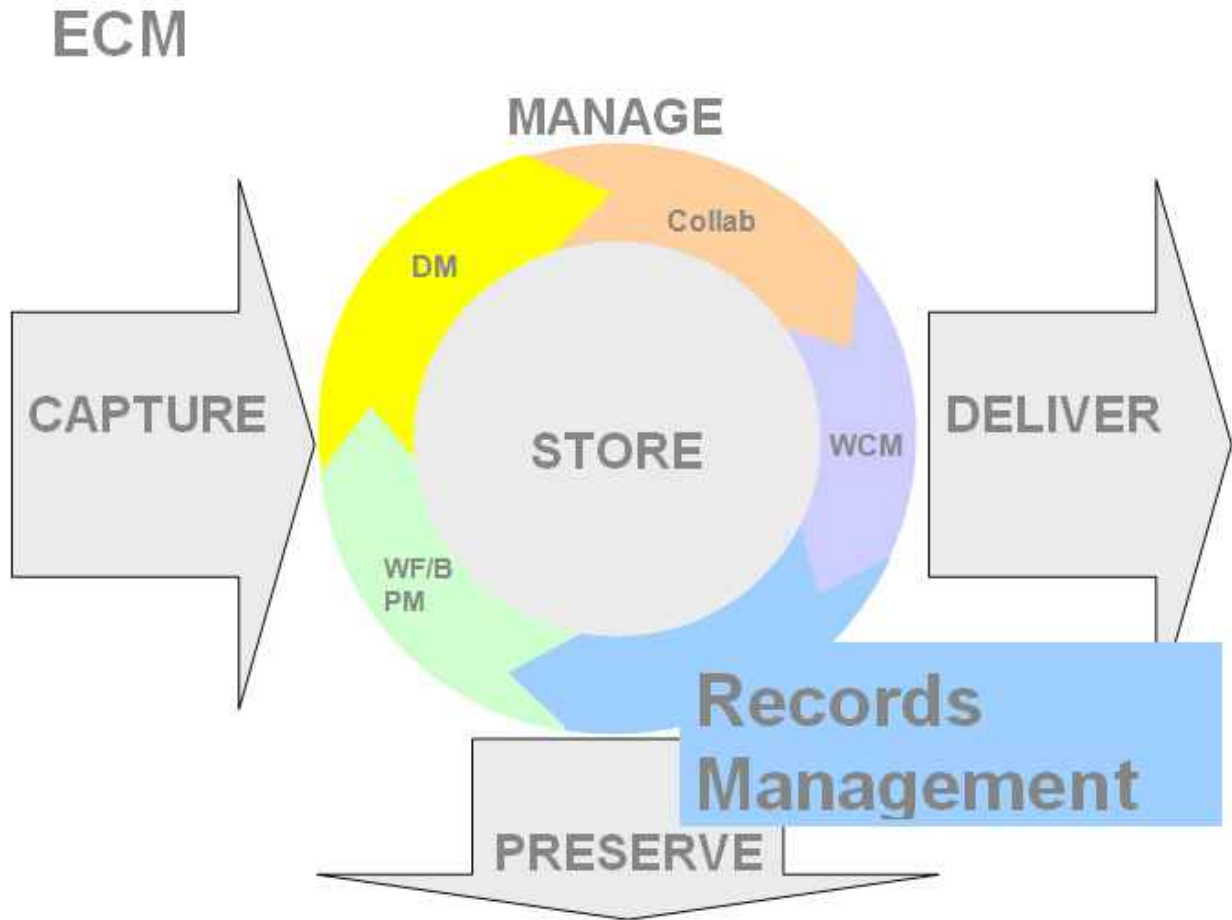


Fig.2: In the AIIM ECM model records management is an integral part of the manage modules.

### The MoReq2 Records Management concept

When we compare this model with the MoReq2 approach, we find similar terminology. MoReq2 also describes capture, store, preserve and other components of the AIIM ECM model.

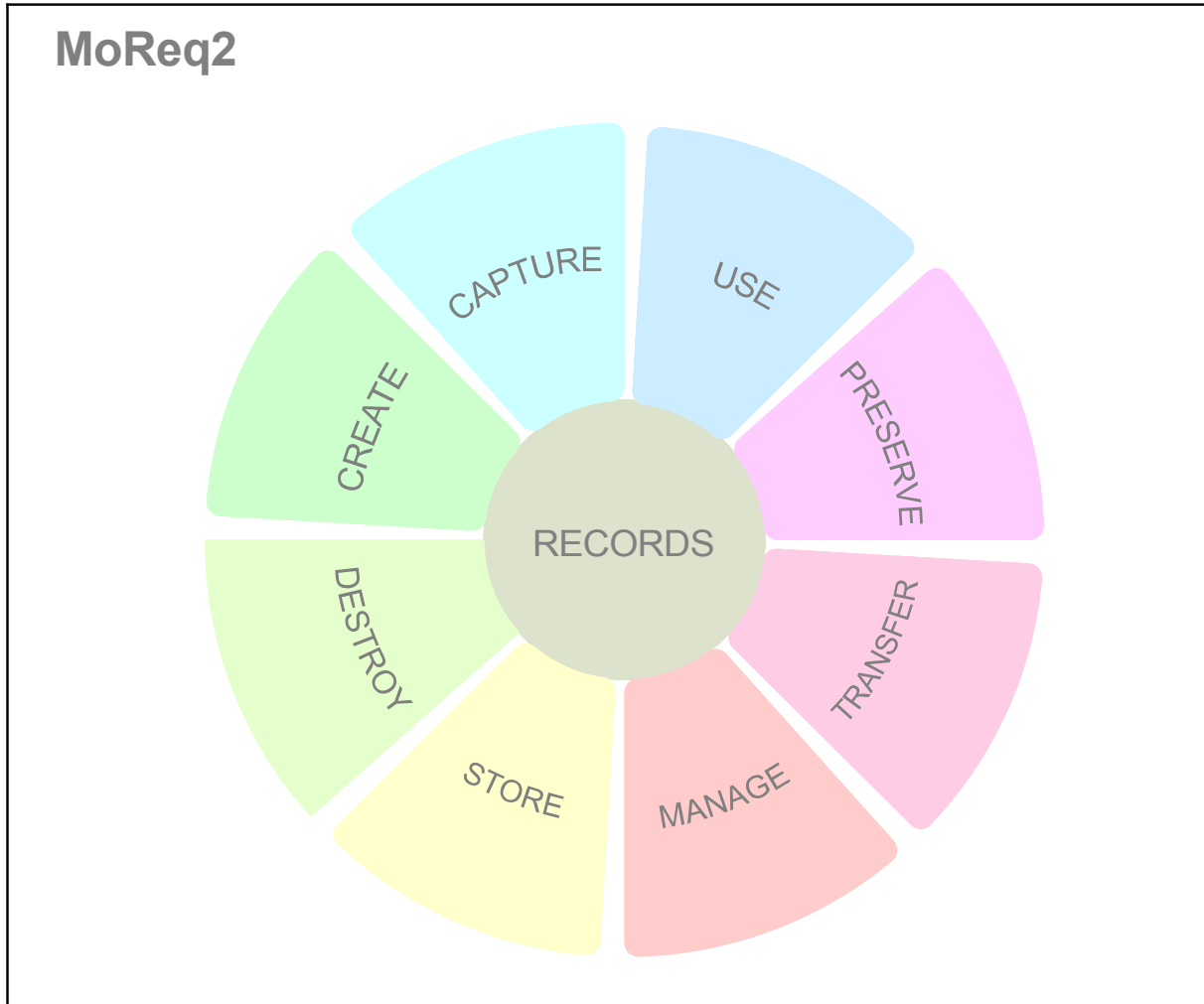


Fig.3: The MoReq2 functionality model.

The principal difference becomes even more obvious when we compare the ECM model with the MoReq2 structure model.

# MoReq2 Structure Model

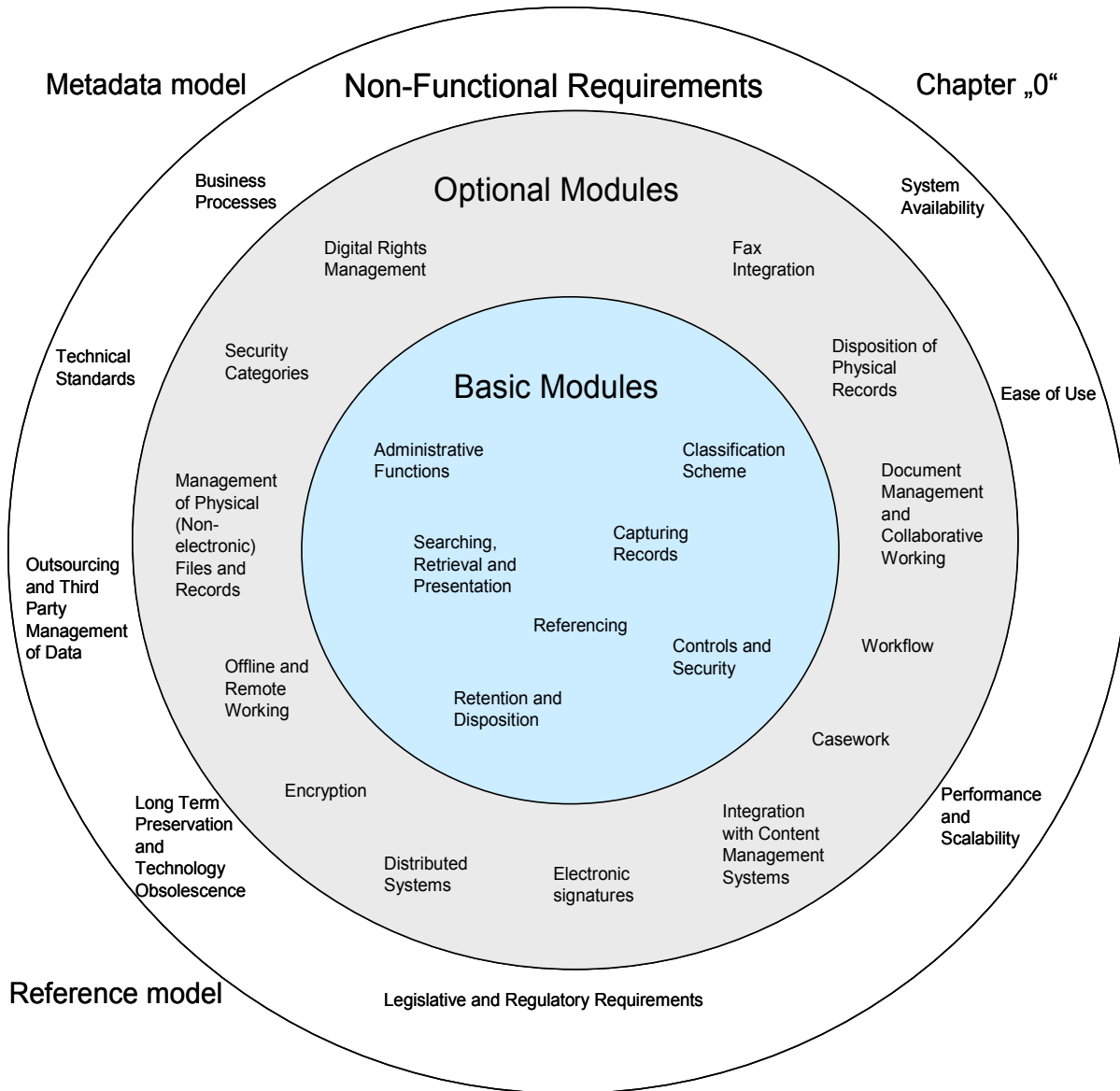


Fig.4: The MoReq2 structure model.

Most ECM components are only satellites in the MoReq2 records management model. The focus is different. In the future, records management can no longer be positioned as the “centre of the world” but must become an integral part of every business application. The ECM approach to records management is more realistic and more advanced.

This can be demonstrated as well in regard to the adoption of new technologies. ECM was enhanced in recent years to come to grips with the challenges of “2.0” whilst Records Management and Electronic Archival are just in an “experimental phase”. The wave of digital records, the information overflow of the last decade, has not yet arrived at the gates of archives. But first indicators are visible, i.e. the challenge of managing and archiving the presidential records of the Bush administration, which caused substantial difficulty for the NARA National Archives and Records Administration. Millions of emails are possibly lost and the rest of the overwhelming mass of electronic documents from the Bush administration is in bad order.

## ■ 3. Records Management challenges in the era of 2.0

New technologies provide many challenges to records management.

### **The technology challenge**

Most “new” Technologies do not natively provide objects suitable to become consistent records or to be archived properly. This especially refers to personalisation & individualisation of web site content, media like blogs, twitters, instant messages and wikis, proprietary formats for multimedia objects, mashups and distributed resources, non-persistent URLs, links and content, etc. We might even have to repeat the discussion, if we need a definition of something like a “digital original.”

### **The cultural challenge**

New perceptions on how to generate and use information. The value of information is often disregarded or cannot be defined when information is originally created. Developments like social communities, digital immigrants versus digital natives, “always online,” new ways of communication and interaction, rights management and data protection issues etc. change the landscape for records management. Records management today has to start with the creation of information, otherwise information cannot be managed when it arrives in records management and archive systems at a later stage. It is not only about the perception of records management and records managers’ tasks; we might even have to discuss, again, the information society and social implications of what we are doing.

### **The information growth challenge**

We face exponential growth of information and no sufficient “management” of it. It is not only the sheer volume of bits and bytes growing to petabytes and exabytes, it also about uncontrolled redundancy, the identification of the value of information, the growing dependency on the availability and accuracy of information, etc. Records management has to address the complete lifecycle of records and is in no way only a “pre-preservation” issue. Records management is essential to fight the uncontrolled growth of information. There also needs to be a discussion on whether records management should be part of every software application which generates, processes or receives information that might become a record. This would mean no more special records management solutions, but instead records management as part of every software and IT infrastructure.

### **The role of records managers and of records management users in the era of 2.0**

To demonstrate some of the current challenges I use some provocative theses with possible positions of traditional records managers and archivists on one side, and on the other side the users who grew up with Nintendo, iPhones and Web 2.0.

*“Records managers are excellent in structured search! They are used to expert faceted search forms.*

*2.0-users hate complex data entry forms! They want to trust in automatic categorization.”*

*“Records managers are aware of the importance of context information! They are used to spending hours to create orderly structures.*

*2.0-users hate complex search forms! They love ‘single field search’ à la ‘Google’ finding everything everywhere.”*

*“Records managers still believe they own their records and really care for them.*

*2.0-users are just using information, it's their daily life."*

*"Records managers know perfectly how to navigate by tree views and folder-sub-folder-volume-hierarchies! It's just about having fun.*

*2.0-users are lazy! They don't care where they store their stuff."*

*"Records managers know about the importance of selection and disposal! They work until late in the night to keep their repositories clean.*

*2.0-users will lose their temper when the interface doesn't work like a media library or Youtube! Or they get bored and just stop to search what they have been looking for."*

*"Records managers are specialists in determining the value of records! They are working busily to keep information unique, sober, accessible and available for all time.*

*2.0-users know, that a 1-TB-harddisk at the discount grocery store is now only 100 €! They believe that there will always be enough storage space for everything."*

*"Records managers are fond of their fileplans, thesaurus, taxonomies, ontologies and classification schemes! They'll go through 7 levels of hierarchy just to find the right spot for a note.*

*2.0-users have no idea about a record and who might be responsible to store it safely! Everybody gets copies of everything, so somebody will take care of it."*

*"Records managers are prepared to work hard to learn every function of their records management solution! With extensive training every software interface can be turned into a usable tool.*

*2.0-users believe in easy-to-use tagging and folksonomy! If there are enough users choosing the same tag, it can't be wrong."*

*"Records managers know by heart (or possibly by their retrieval system) the place of every object! They are in control of everything in the electronic realm and in the world of paper documents.*

*2.0-users love their iPod! If the user interface is not sexy and doesn't work like Facebook, they will just shut the application down.*

*"Records managers still have an academic view of their work and often forget about the end user - who is not a trained records manager.*

*2.0-users are just users who want to have easy access to the correct information at any time."*

*"2.0-users, the digital natives, started life in a virtual world ...and the systems are no longer jailed in their silicon cages. With RFID, robotics, cameras, and sensors they are entering the real world and taking control.*

*Will the profession of the traditional records manager be replaced by the user or even by systems doing their job of declaring, classifying, ordering and archiving records?"*

The role of records managers and records management has to be redefined, including all processes of when and how an information object becomes a record. We are already running out of time, and future historians might name our era “The dark age of the early information society.”

## ■ 4. Where do we stand now with MoReq2?

MoReq2 might be an answer to the questions of how to handle and manage records in a professional way. In fact everything is now in place:

The Moreq2 requirements are online and printed in English,

- The test material is in electronic form, including test cases and available test data. The test descriptions substantiate the text of the requirements
- The XML schema is defined and allows the creation of interfaces to exchange records and classification schemes to define the structures of the records management system
- The first contract with a test centre for the certification of MoReq2 compliant software products is signed, and the first test candidates are lining up.
- The MoReq Governance Board for the maintenance, dissemination and administration of the MoReq2 specification is established,
- First translations and “Chapter 0s” are coming in. The French translation and chapter 0 was published in time for the DLM Forum Toulouse conference.

This is a good starting point to gather experiences, initiate improvements and generate acceptance. There has been a lot of criticism of MoReq2 - too long, too many test cases, too restrictive, too complex. There is still much to do to make MoReq2 work and to demonstrate the value of MoReq2 for the public sector as well for private sector organizations, both on a small scale and for huge enterprise implementations.

There is an urgent need to demonstrate the usefulness and usability of MoReq2:

- Even a complex structure can be easily used with modern software architecture, optimized applications and ergonomic user interfaces.
- Used in the right way, MoReq2 will help save money and increase efficiency for all applications related to records management.
- Not everybody will need every function. We have to make the modular approach of MoReq2 work.
- Systems must not only be capable of MoReq2 functionality, they must also be able to support even office workers with their daily work. Records management applications must not be designed solely for the professional records manager and archivist.
- MoReq2 must also be used as an educational tool, in the academic world as well in day-to-day life.

The good news is that MoReq2 is prepared to serve these purposes.

### **Some ideas on DLM future activities**

With regard to the future development of MoReq2, the DLM Forum must initiate further activities. The DLM Forum should concentrate on the Forum’s assets, combining the competences of archives and records management. Future action plans should consider the following:

#### **- Archiving**

Developing strategies for the challenge of new and ever-changing formats, defining interfaces between records management and digital preservation, standardization of a unique access interface for long-term archival systems, methods for lossless continuous migration, and other related archive-centric topics.



- **Connectivity**

Testing the interoperability of MoReq2 solutions and the records interchange using the MoReq2 XML scheme in practice. The sponsoring and monitoring of such pilots with the dissemination of the results to public will boost the use of MoReq2.

- **Optimizing and Expanding**

Putting more focus on the optional modules of MoReq2, expanding its scope to new technologies, and creating feasible productive subsets of MoReq2 will encourage the use of the standard by the ECM industry as well by the user organisations.

- **Dissemination**

The dissemination of MoReq2 and related results from certifications, first practice and real-life applications must be combined with educational programmes, training, events, publications and other activities for building reputation and bringing the message to the public as well to the private sector.

- **Maintenance**

Although the MoReq Governance Board is now established, we cannot just do maintenance of the MoReq2 standard, and especially not “indefinitely.” Technology and the use of technology are changing rapidly! This will lead us to the need by 2010 at the latest of having to do a re-scoping of MoReq2 or even something like MoReq3. A new (or substantially updated) version - let’s call it MoReq3 - will be necessary for the year 2013 by the latest.

### **A paradigm change is needed**

Last but not least, we need a paradigm change in records management to adapt to technical and cultural change. Otherwise we will see a paradigm shift to a place where the records management and digital preservation community does not want to be! With the next version of MoReq, we really have to break the barriers of traditional records management.

# L'expérience française

## COMMENT TRANSFORMER UN ARCHIVISTE EN SPÉCIALISTE DE L'ARCHIVAGE ÉLECTRONIQUE

---

**Carole Gragez**

conservateur du patrimoine, responsable du Service historique de la Défense antenne marine de Toulon

**Edouard Vasseur**

conservateur responsable des archives au Ministère de la Culture.

*En France, l'intégration des problématiques relatives à l'archivage électronique dans la formation des records managers et des archivistes est relativement récente tant à l'Institut national du patrimoine, qu'à la Direction des Archives de France, ainsi que dans les formations proposées par les universités ou encore par les associations professionnelles.*

*Les deux intervenants, ayant une pratique dans le domaine des archives numériques, se sont interrogés sur la pertinence et l'aspect opérationnel de ces formations dans la mise en œuvre effective des stratégies d'archivage électronique.*

*Les principales questions sont les suivantes : comment se dégager d'une réflexion issue du domaine de l'archivage papier, dont l'évaluation scientifique est moins soumise à l'urgence due à l'obsolescence technique que l'archivage des documents électroniques ? Comment former techniquement des archivistes qui basiquement ne sont pas informaticiens ? Comment faire comprendre des normes parfois complexes quand on ne possède pas les fondamentaux ?*

*Une commande du Centre national de la fonction publique territoriale (CNFPT) délégation de Bourgogne, au printemps de l'année 2007, leur a fourni l'opportunité de passer de la réflexion à l'action et de monter un programme de formation pratique pour une session de deux jours à destination d'archivistes employés dans des services municipaux, départementaux et régionaux.*

*La session aura pour objectif de livrer au public les réflexions croisées des deux intervenants, leur cheminement pour élaborer la formation test qui s'est déroulée en septembre 2007 au CNFPT ainsi que d'expliquer les suites qui ont été données à cette expérience.*

## **Setting up electronic archiving procedures: the central role of training for records managers and archivists**

*In France, including issues related to electronic archiving in the training of records managers and archivists is relatively recent, whether at the National Heritage Institute or at the French Archives authority, or in courses organised by universities and professional associations.*

*There will be two speakers, both of whom have experience in the field of digital archives and who have analysed the relevance and the operational aspects of these courses for the practical implementation of electronic archiving strategies.*

*The main questions are as follows: How can we move beyond a mind-set appropriate for paper-based archives, where scientific evaluation is less threatened by technical obsolescence than for electronic archives? How can we provide technical training for archivists who are basically not IT experts? How can we explain fairly complex standards to people who do not understand the basics?*

*In the Spring of 2007, a commission from the Burgundy branch of the National Centre for the regional civil service (CNFPT) provided an opportunity to move beyond theory into practice and to set up a practical training programme with a two-day session designed for archivists employed by municipal, Departmental and Regional services.*

*The presentation will aim to provide the audience with the comparative views of the two organisers, explain how they developed the experimental training course organised for the CNFPT in September 2007 and report on the follow-up resulting from the experience.*

## ■ Introduction

La problématique de la conservation pérenne des données électroniques révolutionne l'archivistique. Spécialiste de l'organisation de la production documentaire, l'archiviste a donc logiquement un rôle important à jouer dans les années à venir. Mais, habitué à l'univers « papier » comment peut-il s'approprier une démarche qui semble au premier abord plus proche de la technologie que de sa réflexion traditionnelle ? Et comment l'aider à reprendre confiance et à mettre en place des procédures adaptées ? En d'autres termes, quelle peut être sa place dorénavant ?

La question de la formation est cruciale pour permettre à l'archiviste de s'adapter à son nouvel environnement et aux informaticiens qu'il a en face de lui et dont la technicité est tout aussi développée que la sienne. Suffisamment technique pour permettre à l'archiviste de comprendre ses interlocuteurs et les stratégies qu'ils sont en train de mettre en œuvre pour résoudre les problèmes abordés par leurs maîtres d'ouvrage, elle doit néanmoins lui permettre de prendre conscience des ressources et des points forts qui sont les siens et qui constituent son apport essentiel.

L'article suivant résulte d'une pratique professionnelle et d'une réflexion issues de plusieurs expériences.

Le service des archives électroniques du site de Fontainebleau (archives contemporaines, Service à Compétence Nationale Archives nationales) a créé à partir de 2007 une formation pour les archivistes des ministères. L'objectif était, entre autres orientations, de leur permettre de pouvoir coopérer aux actions des services informatiques de leur institution et de pouvoir maîtriser la sélection et la préparation des données électroniques. Des tests de collectes ont démarré conjointement. Par la suite, une procédure d'évaluation et de collecte a été réalisée par ces mêmes personnes, ce qui a prouvé l'intérêt pratique de la démarche. Toutes ces actions ont permis de montrer quelles étaient les étapes à franchir pour un archiviste.

Parallèlement à ces opérations qui se poursuivent, le service des archives électroniques de Fontainebleau avait été contacté par le CNFPT (Centre national de la Fonction Publique Territoriale) puis par la DAF (Direction des Archives de France) qui souhaitaient mettre en place ou renforcer des formations concernant l'archivage électronique. La préparation de ces programmes a été l'occasion d'échanges féconds et critiques avec des archivistes. Puis les remarques des stagiaires ont permis de prolonger la réflexion.

## ■ 1. Une adaptation nécessaire

### 1.1. Un changement d'univers

La définition française des archives (loi de 1979 reprise et abrogée par la loi du 1er juillet 2008) indique que, quel que soit son support, un document est une archive. Mais l'archiviste est encore principalement confronté au papier. C'est un univers rassurant, bien connu, composé en général de pièces, dossiers et cartons que l'on sait analyser et répertorier en articles afin d'en assurer la gestion et la communication ultérieures.

Le domaine des données électroniques est bien différent. Avec l'irruption des ordinateurs individuels et la disparition des secrétariats autrefois chargés d'assurer une structuration minimale de la production, on assiste désormais à une prolifération de versions et, alors que les connaissances informatiques s'améliorent, il n'est plus rare de voir des fichiers ou logiciels « traficotés ». D'immenses vrac de fichiers hétérogènes sont ainsi créés, stockés par les informaticiens et souvent éliminés, sans visa par manque de connaissance, pour gagner de la place. Il est parfois proposé en urgence à l'archiviste des disquettes, CD-Rom ou DVD, voire des disques durs, comportant des données ni analysées ni sélectionnées.

Or, ces systèmes informatiques, formats et supports connaissent une obsolescence souvent rapide, qui nécessite des migrations régulières et fréquentes. Se tenir au courant des innovations et connaître les technologies retenues par les autres spécialistes de la gestion des données devient une obligation. Mais l'informatique ne peut à elle seule assurer l'archivage électronique et la pensée archivistique continue d'être prépondérante. En effet, il est toujours et peut-être plus qu'autrefois nécessaire d'opérer une sélection

drastique des documents, de définir rapidement leur cycle de vie et de bien les repérer pour pouvoir les décrire et les indexer afin d'assurer leur communication présente et future. Un archiviste doit donc être aidé par un informaticien et apprendre à dialoguer avec ce nouvel interlocuteur.

## 1.2. Les difficultés liées à la description des données électroniques

Il est cependant difficile d'appréhender cette nouvelle production électronique et ceci gêne énormément l'archiviste qui doit pourtant prévoir la gestion des documents pendant tout leur cycle de vie<sup>1</sup>. Les confusions entre les termes « archivage électronique », « numérisation » et « gestion électronique de documents » ou GED sont encore fréquentes.

On peut également noter plus particulièrement des problèmes récurrents, notamment sur les questions de typologie et d'analyse. Quand le papier était le support quasiment unique, les typologies dépendaient principalement de la nature des documents, de leur fonction, rarement de leur aspect matériel et technique<sup>2</sup>. Les documents électroniques, en plus de leur nature diplomatique, se distinguent par leur forme technique, qu'il s'agisse par exemple de bases de données, de fichiers de texte, de fichiers d'images, etc... La numérisation des documents papier complexifie la question puisqu'un compte-rendu de réunion, numérisé en mode texte ou en mode image, peut être intégré dans une base de données via une GED. Or il s'avère que l'aspect technique des documents électroniques a un impact très fort sur la mise en œuvre de l'archivage. Il faut notamment repérer les documents dont les données sont structurées (bases de données, structuration en XML), ce qui détermine le mode de préparation des données et de leurs métadonnées. Enfin, l'archiviste doit également toujours se demander comment les données ont été créées, avec quels logiciels et si on est encore capable de les recevoir, de les lire et de les traiter (problèmes d'obsolescence, d'interopérabilité).

La description archivistique des données électroniques pose un nouveau problème. Il est habituel dans l'univers papier de faire coïncider l'article avec le contenant. Or, un support informatique peut contenir plusieurs versements et plusieurs centaines de fichiers peuvent très bien constituer un seul élément intellectuel. Par ailleurs, il est techniquement faux de comparer un fichier ou un répertoire avec un dossier contenant des documents. L'archiviste doit cependant impérativement se poser la question de la communication future des données avant d'établir des articles intellectuels qui ne correspondent pas à des contenants physiques.

Ce faisant, on peut même parfois procéder à une description au pièce à pièce de fichiers, voire de données, contrairement à l'évolution en matière de gestion des archives contemporaines. Pour une base de données structurée, il devient en effet nécessaire d'écrire un dictionnaire de données et de détailler plusieurs métadonnées liées.

## 1.3. Un nouveau cadre mental

Mais ce n'est pas tout. La théorie des trois âges (courant, intermédiaire, historique) devient obsolète dans son découpage strict : des données peuvent être transférées à un service d'archives définitives dès la période intermédiaire, voire courante dans certains cas. Par ailleurs, l'habitude du versement aux archives peut être remise en cause par la longueur de certaines durées d'utilité administrative entraînant dès le départ un important effort de conservation pérenne de la part du producteur lui-même. Enfin, alors que l'archiviste français pense traditionnellement le versement en fin de vie des documents, quand il s'agira de données électroniques il devra se positionner dès l'amont<sup>3</sup>, afin de garantir une bonne fabrication des données, leur possible extraction et préparation pour une future conservation.

« L'habitus »<sup>4</sup> de l'archiviste basé sur un monde - celui du papier - relativement stable et cadré ne fonctionne donc plus dans un univers en constante évolution et demandant une adaptabilité permanente.

L'archiviste a donc besoin d'un nouveau cadre mental lui permettant d'appréhender et de penser ce nouveau rapport à un objet en cours de transformation. Il est cependant capital de faire comprendre aux

---

<sup>1</sup> Ages courant, intermédiaire et définitif. Cette difficulté à appréhender les besoins peut entraîner le risque d'une externalisation coûteuse du stockage des données dans le cadre de projets ne résolvant pas le problème de la conservation intermédiaire longue et pérenne.

<sup>2</sup> Par exemple les registres qui connaissent un traitement spécifique.

<sup>3</sup> Voir à ce sujet la norme MOREQ (version 2 en 2008).

<sup>4</sup> Pour reprendre le terme du sociologue Pierre Bourdieu.

professionnels que leur expérience est primordiale mais doit être revisitée pour pouvoir avancer dans le domaine de l'archivage électronique.

### **1.4. Travailler en amont**

Dans l'univers des données électroniques, si une donnée est mal produite, sa conservation est difficile voire impossible. Un fichier ou un groupe de fichiers mal nommé et repéré est à terme un document ou un dossier si ce n'est perdu, tout au moins très difficile d'accès<sup>5</sup>.

L'archiviste s'oriente donc, quand il travaille dans l'univers des données électroniques, vers une démarche impactant l'amont. Il est, de fait, un des rares professionnels de l'organisation de la production documentaire capable de mettre en place, en accord avec le producteur, des procédures claires et partagées non seulement de création mais également de préparation de ces données. Ceci est le seul moyen d'assurer à terme leur gestion (gestion électronique de documents par exemple), leur éventuelle élimination automatique, ainsi que leur versement en tant qu'archives définitives.

Une des problématiques à affronter en France réside dans le « retard » pris quant au positionnement à avoir par rapport au records-management. Il convient donc de prendre rapidement des décisions afin de garantir une évolution de l'ensemble de la profession.

### **1.5. De la collecte traditionnelle à la construction d'un projet**

L'archiviste va désormais plus loin que la simple réalisation d'un tableau de gestion et passe à la création d'un véritable projet.

Il doit envisager directement l'avenir non seulement pour les problèmes de communications futurs<sup>6</sup> des fichiers (nommage, indexation, récupération des métadonnées), mais également pour ce qui concerne les changements à prévoir pour la gestion des documents produits. Il se pose, par exemple, la question de savoir comment faire coïncider les règles de nommage des fichiers, les formats, comment assurer l'interopérabilité technique entre les différents interlocuteurs.

L'archiviste n'est donc plus en position d'attendre les versements en les anticipant plus ou moins. Il doit impérativement construire très en amont le projet de versement qu'il intègre dans une dimension globale d'appréhension de son service et de ses rapports avec les services producteurs et versants.

## **■ 2. Comment former les archivistes ?**

Dans les années 1980, le service des archives électroniques des Archives nationales a fait partie des pionniers en matière de méthodologie d'archivage électronique. A l'époque, les documents à conserver étaient principalement des bases de données concernant des informations comptables ou de gestion du personnel. Un travail mené en commun avec des ingénieurs informaticiens de l'École des Mines et avec des chercheurs de l'INSEE (Institut National de la Statistique et des Etudes Economiques) a permis de mettre au point une méthodologie adaptée et qui a servi plus tard à écrire la norme OAIS. Les procédures concernant la préparation de données structurées ont donc été parfaitement rodées et ont permis, par exemple, de mettre au point des modules d'automatisation chez certains producteurs comme le Ministère de l'Agriculture. Cependant, avec l'explosion de la bureautique et la diversification des formats et logiciels, le service s'est retrouvé dans une situation très difficile. Il devenait prépondérant de sélectionner et préparer au maximum les données en amont afin de simplifier le travail. Seul, le responsable du service ne pouvait répondre à la demande de tous les ministères. Il était donc évident que les archivistes des ministères devaient être formés. Mais aucune formation pratique courte n'existant alors, il a fallu réfléchir pour pouvoir créer une méthodologie adaptée et

<sup>5</sup> On pense souvent à la recherche plein texte mais ce mode de recherche par mot de manière libre donne des résultats aléatoires et avec beaucoup de « bruits ». Ceci peut être comparé à une recherche sur Internet quand les documents n'ont pas été indexés correctement.

<sup>6</sup> Il faut par ailleurs noter qu'outre le fait que la nature même des documents évolue -les blogs doivent-t-ils par exemple être conservés?-et permet dorénavant de nouveaux traitements dans le cadre d'une recherche (croisement des données), la notion de chercheur est également en voie de changement. En effet l'archiviste a en face de lui des spécialistes du traitement informatique, parfois des experts (cas des logiciels Autocad dans le domaine de l'architecture et de la construction).

correspondant aux interrogations des professionnels. Répondre dans le même temps aux demandes du CNFPT et de la DAF a également été l'occasion de poser la problématique de ce que pourrait être une démarche d'accompagnement de professionnels dans le domaine de l'archivage électronique. Cette réflexion a été menée par le service des archives électroniques des Archives nationales en relation avec M. Edouard Vasseur archiviste du ministère de la Culture.

## 2.1. Les questionnements

En France, les organismes de formation des archivistes, tant initiale que continue, intègrent depuis quelques années les problématiques de l'archivage électronique dans leurs programmes. Mais ces offres sont restées longtemps très théoriques. Quand le projet de création d'une formation pour les archivistes des ministères ainsi que pour la DAF et le CNFPT est intervenu, les professionnels se plaignaient alors souvent de l'aspect non pratique des formations et de leur difficulté technique.

Par ailleurs, les livres publiés sont en faible nombre<sup>7</sup>, alors que la documentation existante sur le réseau est pléthorique et il est difficile de s'y retrouver sans formation initiale. Outre les divers sites et articles susceptibles d'être trouvés sur Internet, nous évoquerons ceux de l'Association des Archivistes de France (AAF) et le Portail International des Archivistes (PIAF) qui proposent des fiches techniques ou des modules. Quant à la DAF, son site propose une abondante documentation touchant à la fois aux aspects légaux et techniques.

Les archivistes sont donc confrontés à une masse documentaire importante et à des formations en majeure partie théoriques et/ou peu adaptées à leur univers mental, car parfois trop techniques.

Il faut cependant dire que la demande récurrente concerne le plus souvent la technique. Les professionnels se sentent en effet réconfortés par un discours technologique... qui est cependant le plus souvent incompréhensible pour eux. De plus, ceci ne résout aucunement le problème de l'approche méthodologique. Il est également courant de présenter quelques expériences en les accompagnant d'un discours explicatif et terminologique. Cependant, comme un projet d'archivage électronique ne peut être calqué mécaniquement d'une institution à l'autre, ce mode d'approche n'est pas approprié. Enfin, on peut aussi partir d'une démarche projet et faire analyser des situations pour en retirer des procédures mais ceci nécessite une expertise très pointue et le recours à une méthodologie archivistique encore en devenir.

## 2.2. Comment créer une formation adaptée ?

Donner aux stagiaires un référentiel pérenne dans le temps constitue l'enjeu essentiel pour tout formateur en archivage électronique.

La rapidité avec laquelle les solutions électroniques proposées par le marché évoluent et se modifient en rend en effet inopérante une présentation détaillée, même s'il est indispensable de donner quelques outils permettant de comprendre l'environnement. La question est donc de trouver un mode d'appréhension pérenne des documents électroniques et de leur gestion, indépendamment de leur évolution technologique et de l'organisation dans laquelle travaille l'archiviste. Il faut également apprendre aux archivistes à mettre en valeur leurs compétences propres, sans leur donner l'illusion d'acquérir des compétences techniques qui, de toutes façons, ne feront jamais d'eux des informaticiens. La question de l'adaptabilité des démarches présentées est encore plus importante puisque les archivistes n'ont pas tous les mêmes conditions de travail ni les mêmes outils mis à disposition par leur organisme. La présentation de solutions techniques doit donc être exceptionnelle et illustrer une démarche mais ne doit pas être une fin en soi. L'équilibre est toutefois difficile à trouver, car les archivistes sont souvent en attente de solutions « clé en main » immédiatement réutilisables. L'objectif est donc avant tout de faire comprendre que la démarche prime sur les outils technologiques.

La connaissance de la circulation de l'information, du cycle de vie des documents, des délais de rétention (délais d'utilité administrative) font en effet des archivistes des partenaires dotés de compétences utiles pour améliorer la gestion des données. Il faut donc leur prouver que leur démarche professionnelle peut aider les

---

<sup>7</sup> Citons par exemple les publications en français les plus connues :

*Les archives électroniques*. Manuel collectif pratique sous la direction de Catherine Dhérent publié par la Direction des Archives de France (éditions Documentation française, 2002). Cet ouvrage est complété par un manuel sur la collecte des fichiers bureautiques.

Ainsi que : *Dématérialisation et archivage électronique*, ouvrage collectif de Jean-Marce Rietsch, Marie-Anne Chabin et Eric Caprioli (éditions Dunod, 2006).



services informatiques des organisations dans lesquelles ils travaillent. Aussi, bien plus que d'expliquer les techniques des informaticiens, est-il indispensable d'expliquer aux stagiaires comment la méthodologie de l'archiviste peut se mettre au service du bon fonctionnement des organisations.

### **2.3. Quelle formation pour quels objectifs ?**

Par ailleurs, il est nécessaire de créer une formation relativement courte pour des professionnels très occupés. Etant donnée la complexité du domaine couvert, il est donc obligatoire de se concentrer sur l'essentiel.

Dans le même temps, il est évident que l'objectif n'est pas de faire des projets d'archivage électronique à la place des archivistes, mais plutôt de donner aux personnes un cadre mental leur permettant de s'adapter et de créer leur propre cahier des charges. Il s'agit donc d'apporter aux professionnels les moyens de penser les documents électroniques et leur univers, d'acquérir un vocabulaire pour pouvoir discuter avec des informaticiens et avoir une approche plus pratique.

Il faut surtout montrer aux stagiaires qu'ils disposent déjà des outils intellectuels de compréhension de la problématique. L'idée est donc de partir de la méthodologie traditionnelle pour en arriver à l'archivage électronique tout en évitant les confusions et les « faux amis »<sup>8</sup>.

## **■ 3. La formation**

### **3.1. Comment procéder : une démarche par petits pas**

Tout d'abord, il convient de donner aux personnes formées une idée de ce que peut être un service d'archives ne s'occupant que d'archives électroniques car, très souvent, les archivistes n'en ont quasiment aucune idée. Une présentation du service des Archives nationales fournit l'occasion d'expliquer pas à pas quelles sont les fonctions qui ont été modélisées par la norme OAIS : collecte, préparation des données, intégration pour stockage, etc...

Puis, en reprenant des réflexions déjà très abouties sur la question, l'idée est d'insister sur la notion de typologie de documents électroniques, en faisant des parallèles avec les documents papier<sup>9</sup>. Ceci permet aux archivistes de se familiariser avec un univers qui leur paraît souvent très étranger et de leur montrer qu'ils sont capables d'établir une évaluation et une sélection comme ils savent le faire dans l'univers du papier.

Il paraît également très important de se concentrer sur les métadonnées des données électroniques. L'expérience a maintes fois prouvé que cette dernière notion est très complexe à saisir pour un archiviste traditionnel. Ce dernier reste ainsi souvent « bloqué » à la compréhension de métadonnées correspondant à l'analyse de premier niveau, et donc à la norme de description ISAD-G (et à la DTD-EAD). Mais il ne saisit pas l'existence de plusieurs autres niveaux de métadonnées, en particulier celles qui ont été décrites par la norme OAIS et qui permettent de conserver la structure et la sémantique d'un document. Pour ne pas déstabiliser les stagiaires, des exemples très simples sont retenus comme celui des numéros de sécurité sociale français qui ont un mode de codage et de structuration très spécifique. Ceci permet de faire comprendre facilement ce que peut être un dictionnaire des données, un dessin de fichier et autres concepts qui peuvent sembler complexes au premier abord.

Il devient alors possible d'aborder sereinement la norme OAIS qui apparaît souvent barbare aux archivistes. En effet, cette norme dont il est question dans tous les forums ou publications relatives à l'archivage électronique, est souvent présentée sans explication, ce qui en rend la compréhension assez complexe. A l'inverse, dans le cadre d'un stage ainsi structuré selon une démarche logique et adaptée à l'univers mental d'un archiviste, il apparaît à chaque fois que la modélisation d'un service d'archives électroniques et le modèle de description des données et métadonnées sont tout à fait évidents pour les stagiaires.

---

<sup>8</sup> Par exemple, il est important de faire comprendre aux archivistes qu'il ne faut pas comparer les contenants physiques des dossiers et cartons d'archives avec des fichiers et répertoires de données électroniques, puisque ceux-ci sont en quelque sorte virtuels, même si le concept intellectuel reste le même.

<sup>9</sup> Une base de données documentant des dossiers de personnel peut, par exemple, être comparée à un registre traditionnel.



Ces bases posées et les personnes étant familiarisées avec les grandes notions de l'archivage électronique, on peut alors les amener à s'interroger sur une intervention très en amont dans la préparation et la prise en charge des documents électroniques. La préparation concrète d'un versement d'archives électroniques - une enquête statistique réalisée par le ministère de l'Agriculture au début des années 2000 ou une enquête du Ministère de la Culture concernant les bibliothèques en France - permet de voir concrètement ce qui rapproche et ce qui différencie le traitement des archives papier de celui des archives électroniques. Les difficultés inhérentes à la préparation des versements d'archives électroniques, et notamment la difficulté à reconstituer *a posteriori* des métadonnées qui ne sont plus à jour, que les créateurs ont oublié ou qui n'ont jamais existé, constituent une transition efficace pour convaincre de l'absolue nécessité d'une intervention le plus en amont possible. L'évaluation archivistique, la création des métadonnées lors de la production des données et la préparation des modules d'archivage sont des problématiques qui peuvent alors être abordées.

A ce stade de la formation, on présente alors les normes sur le Records management (ISO 15 489 et 23 081) et MOREQ comme des outils destinés à aider les archivistes dans leur intervention en amont.

### **3.2. Le constat après avoir formé plusieurs stagiaires**

Ce stage a déjà été réalisé quatre fois, en des versions à chaque fois améliorées, mais reste encore largement expérimental. Plusieurs bilans dynamiques ont été faits avec les participants dans l'objectif de savoir comment améliorer ce module de formation. Les personnes interrogées ont ainsi l'impression de participer au montage d'une méthodologie encore en devenir.

#### ***Réactions des stagiaires***

Globalement, les stagiaires se déclarent en majorité très satisfaits de cette formation qu'ils jugent pratique, adaptée et leur permettant de commencer à travailler et à faire des expériences concrètes. Il faut signaler que ce stage a déjà été suivi par des informaticiens travaillant avec des archivistes et que cela a été enrichissant pour tous.

Malgré la densité du propos et l'attention que ce dernier nécessite de la part des stagiaires, il faut noter qu'il est rare qu'une personne décroche en cours de formation, exceptés les cas de personnes qui n'étaient ni archivistes ni informaticiens. L'enchaînement logique des présentations et le propos volontairement très clair des informations permettent à tous de suivre facilement. Les explications concernant les normes paraissent absolument nécessaires même si elles sont jugées rébarbatives. Pratiquement tous les stagiaires déclarent avoir acquis rapidement et aisément le cadre mental leur permettant de progresser. La présentation détaillée des métadonnées est très souvent plus particulièrement appréciée car elle correspond, selon les participants, à un thème traité de manière assez flou dans les publications.

En cours de route, des éléments complémentaires ont été intégrés à la demande ou pour suivre l'actualité. Ainsi, il a été jugé absolument nécessaire d'avoir une information concernant les formats et les supports. De même, une présentation du protocole standard d'échange (PSE), répondant aux besoins d'automatisation des échanges, était indispensable. Les choses évoluant très vite depuis quelques années, il est également devenu courant que les demandes des stagiaires deviennent beaucoup plus techniques et portent sur le cryptage des données, la signature électronique, les empreintes et la problématique d'un stockage dans un coffre-fort propriétaire.

Un certain nombre de questions connexes sont cependant encore posées : terminologie croisée des archivistes et des informaticiens ; éléments de diplomatique.

Les échanges de messages qui suivent chaque formation confirment toujours la même impression. Les participants paraissent en général décomplexés par rapport au sujet et plusieurs, dès la semaine du retour, entament des démarches à destination de leurs collègues des services informatiques. Ils semblent prêts à sensibiliser leur hiérarchie et leurs collègues des autres services sur la question et déclarent être plus sereins dans l'appréhension de l'abondante documentation disponible.

## ■ Conclusion après une année et demie de formation

Ces expériences permettent de progresser lentement avec les stagiaires vers la mise en place d'une méthodologie de formation archivistique adaptée.

Les formations en matière d'archivage électronique sont en pleine évolution. Il sera nécessaire de réfléchir pour intégrer plusieurs problématiques listées non exhaustivement ci-après :

- la question des formats et de leur migration ;
- la création de méthodologies adaptées aux documents spécifiques comme les images fixes et animées, les documents d'architecture (et formats complexes liés), les mails, les sites internet et intranet ;
- la problématique de la différenciation entre le stockage et l'archivage : quels besoins ? quels avantages et inconvénients pour ces deux méthodes ? ;
- la diffusion des données : la différence avec l'archivage et les aspects économiques et/ou légaux liés ;
- la méthodologie de description des données électroniques et la question des instruments de recherches ;
- la formation au traitement des archives électroniques ;
- l'action en amont, lors de la création des données ;
- etc...

L'équilibre est difficile à trouver, et les premières tentatives négligeaient certainement trop le côté technique, insuffisamment développé, notamment en raison du temps limité. Cependant, une approche davantage méthodologique que technique, visant à donner aux archivistes un cadre mental utilisable dans le maximum de situations possibles et insistant sur l'intervention en amont, permet d'être mieux armé dans un domaine en constante évolution.

Un ancien stagiaire a pu, dans le cadre d'une formation commune avec des informaticiens et d'autres gestionnaires de l'information, animer le groupe, faire le lien entre les différentes spécialités et orienter les discussions sur les modalités d'une meilleure collaboration entre tous les professionnels de l'information. Ceci montre qu'un archiviste, conscient de ses atouts techniques et correctement formé, est capable de nouer un dialogue interprofessionnel constructif pour assurer une meilleure gestion de l'information produite. Experts de l'organisation de la production documentaire, les archivistes doivent donc désormais prendre leur juste place dans ce nouvel univers électronique.

# LE MÉTIER D'ARCHIVISTE À L'ÉPREUVE DE L'ARCHIVAGE ÉLECTRONIQUE : ENTRE PERMANENCE ET RENOUVELLEMENT DES CONCEPTS ET DES PRATIQUES. L'EXPÉRIENCE DU SERVICE NATIONAL DES ARCHIVES DU GROUPE LA POSTE

---

Anne Burnel

directrice du Service national des archives du Groupe La Poste<sup>1</sup>

*Dans le cadre de la mise en place de sa plateforme d'archivage électronique et de la mise en place de procédures de versement d'archives électroniques, le Service national des archives de La Poste adapte ses pratiques de collecte et de collaboration avec les services producteurs. Cette expérience amène le service à confronter les procédures traditionnelles de traitement des archives intermédiaires aux particularités posées par la production documentaire électronique.*

- *La notion de versement garde-t-elle du sens et le numéro de versement en tant qu'outil de gestion est-il toujours opérationnel ? Pour des raisons techniques, certains versements peuvent être alimentés par plusieurs transferts de données. En outre, la capture régulière des données au plus près de leur date de création tend à fragmenter les versements. Ces nouvelles pratiques annoncent-elles pour les archives contemporaines l'abandon de la cotation continue et un retour au cadre de classement avec un système de cotation inspiré du records management ?*
- *La notion de service versant est-elle encore opérante quand les versements d'archives électroniques sont réalisés par la maîtrise d'œuvre des applications pour des raisons purement techniques ?*
- *La DUC et la DUI ont-elles encore du sens ? Les versements d'archives électroniques peuvent s'effectuer dès la création des données. Cela permet de sécuriser les données dès leur création, voire de pallier les limites de certains applications, incapables de « rejouer » la création des données. La question se pose alors de la DUC dans l'application productrice : sont-ce les coûts de stockage qui détermineront la purge des données par ailleurs déjà collectées par le SAE ?*
- *La notion de territorialité des archives est-elle encore pertinente pour la collecte et la conservation des archives ? De nombreuses applications sont conçues et administrées au niveau central. Il est alors possible de dissocier le ressort territorial de leurs utilisateurs du SAE, qui lui peut être centralisé ou mutualisé par grandes plaques géographiques.*
- *Les dossiers hybrides. Certaines procédures ne sont que partiellement dématérialisées. L'archiviste doit alors inventer une méthodologie et des outils adaptés pour articuler les versements d'archives électroniques et papier.*
- *Une exigence renforcée de mise en œuvre des pratiques archivistiques. Le fait que les données électroniques ne soient pas d'un accès immédiat impose à l'archiviste d'intervenir en amont du cycle de vie des archives pour collecter les métadonnées de contenu et de gestion nécessaires pour documenter les versements. Cette « remontée à la source » permet d'ailleurs à l'archiviste d'approfondir sa connaissance du SI et des processus de travail de son organisme et d'en tirer profit pour documenter les versements d'archives papier en obtenant des extractions automatisées de métadonnées du SI.*
- *Le savoir-faire de l'archiviste créateur de valeur dans la chaîne documentaire. Pour consolider les données à archiver, l'archiviste collabore désormais avec les administrateurs et les utilisateurs des applications pour normaliser la saisie de manière à récupérer des données fiables et exploitables sur le très long terme. De surcroît, pour certaines productions documentaires, l'archiviste est amené à intervenir dans le processus de création des données ou de gestion des dossiers courants, contribuant ainsi au développement de la qualité.*

*Ces pistes de réflexion s'appuieront sur des cas concrets.*

---

<sup>1</sup> Avec la collaboration de Sandrine Soulas, chargée du département projets, processus, NTIC au Service national des archives du groupe La Poste.

Avec mes remerciements à Hélène Servant, chef du Bureau des politiques de collecte des archives au Département de la politique archivistique et de la coordination interministérielle de la Direction des archives de France, pour son amicale et constructive relecture.

*Cette approche renouvelée du rôle de l'archiviste conduit à une meilleure prise en compte des besoins des producteurs et à une plus grande maîtrise des ressources documentaires. N'est-ce pas une opportunité formidable à saisir par notre profession ?*

[Consulter la présentation.](#)

## **The profession of archivist is being sorely tested by electronic archiving: staying put or renewing concepts and practices. The experience of the Post Office (Groupe La Poste) National Archive Service**

*Within the framework of setting up its digital repository and establishing procedures for transferring digital records, the Post Office national archives service adapts its methods of information gathering and collaboration to suit service producers. The experience leads the service to compare traditional methods of processing intermediary archives with the specific issues involved in the production of electronic documents.*

- *Does the notion of transfer, in its traditional meaning, still make sense and is the transfer number still useful as a management tool? For technical reasons, some transfers can be made up of several data transfers. In addition, the regular capture of information close to the date of creation tends to fragment these deposits. Do these new practices mean that contemporary archives will abandon continuous listing and return to a framework of classification with a listings system inspired by records management?*
- *Does the notion of transferring service still apply when deposits of digital records are carried out under the supervision of applications for purely technical reasons?*
- *Does it still make sense to distinguish between current usefulness and intermediate usefulness? Digital records can be deposited as data is created, thus enabling the data to be secured immediately and compensating for the limitations of some applications, unable to "reconstruct" the data creation. We may therefore ask how to determine the duration of current usefulness in the producer application: is it the storage costs that will decide when data collected by the ERMS is to be purged?*
- *Is the notion of an archive's territoriality still relevant for collecting and preserving archives? Many applications are created and administered centrally. It is therefore possible to dissociate territoriality from users of the system, which can be centralised or grouped into large geographical zones.*
- *Hybrid files. Certain procedures are only partially dematerialised. The archivist must therefore invent a methodology and suitably adapted tools to facilitate the joint depositing of digital and paper archives.*
- *A strong need to fully utilise the skills of an archivist. The fact that electronic data may not be immediately accessible forces the archivist to intervene early in the life-cycle of the archives to collect metadata of contents and management necessary for documenting these deposits. This 'return to the source' allows archivists to deepen their knowledge of information systems and their organisation's work methods and to take advantage of this for documenting deposits of paper archives by obtaining automated mining of metadata from the IS.*
- *The knowledge that allows archivists to inject added value into the document-processing chain. In order to consolidate records data, archivists must now collaborate with application administrators and users to standardise data acquisition so that reliable and readable data can still be recovered far into the future. In addition, for production of some documents, the archivist is required to take part into the data creation process or the management of ongoing files, contributing to quality development.*

*We shall illustrate these creative processes with concrete examples.*

*This new approach to the role of the archivist leads to a better understanding of the needs of producers and better management of documentary resources. Isn't this a marvellous opportunity for our profession?*

[See the presentation.](#)

Dans le cadre de la mise en place de sa plateforme d'archivage électronique et de procédures de versement d'archives électroniques, le Service national des archives du Groupe La Poste (SNA) adapte ses pratiques de collecte et de collaboration avec les services producteurs. Cette expérience amène le service à confronter les procédures traditionnelles de traitement des archives intermédiaires aux particularités posées par la production documentaire électronique. Dans un environnement caractérisé par des évolutions organisationnelles et technologiques rapides, la gestion et la description des archives se complexifient et nécessitent de mener une réflexion archivistique approfondie pour concevoir des solutions d'archivage adaptées aux besoins des services producteurs et des utilisateurs.

## ■ 1. Le contexte de production et les acteurs

Les évolutions techniques et organisationnelles contribuent à distendre le rattachement traditionnel entre le ressort territorial du service producteur et les archives produites, voire à modifier la nature de leur lien.

### 1.1. L'affaiblissement du lien entre archives et territoire

Aujourd'hui de nombreuses applications informatiques sont conçues et administrées au niveau central. Une application nationale peut ainsi être utilisée par de nombreux utilisateurs répartis dans plusieurs services et / ou sur de multiples points du territoire. Les applications métier contribuent souvent à recentraliser une partie des processus de travail et de l'information générée. Les archives produites sont de ce fait de moins en moins liées à un territoire. L'archivage est alors pensé et mis en œuvre au niveau central par l'interfaçage des applications productrices des données et le système d'archivage électronique (SAE).

La pratique de plus en plus courante qui consiste à confier des processus d'activité à des plateformes mutualisées spécialisées concourt également à dissocier les activités et donc les archives produites d'un ressort territorial donné. Ces plateformes ont un périmètre fonctionnel sans lien nécessaire avec le territoire des structures pour lesquelles elles assurent des prestations. Ce modèle d'organisation est favorisé par la dématérialisation des procédures, l'éloignement géographique ne constituant plus un frein au travail à distance et à l'échange rapide de documents. Les SAE peuvent eux aussi être centralisés ou bien mutualisés par grandes plaques géographiques, compte tenu des investissements requis pour la mise en place d'une plateforme d'archivage électronique.

Ce type d'organisation conduit à s'interroger plus largement sur la compétence territoriale des services publics d'archives. Le périmètre géographique constitue-t-il l'approche la plus pertinente pour certains fonds ? La recentralisation des archives de certaines institutions n'est-elle pas envisageable ? On pourrait imaginer, par exemple, des services publics d'archives spécialisés par grand domaine administratif pour la collecte des archives produites par les services territoriaux de l'Etat. En tout état de cause, les archivistes devront prendre en compte les transformations profondes du contexte de production des documents pour adapter leur organisation et la description des archives.

### 1.2. La diversification des acteurs

Le développement des pratiques de mutualisation et de sous-traitance amène aussi à s'interroger sur la notion de producteur d'archives. Des entreprises recourent par exemple à des sociétés spécialisées pour la gestion administrative et de la paie de leur personnel. Une telle société, qui agit en tant que prestataire pour son client, produit certes les documents liés à la gestion des salariés de son client, mais c'est ce dernier qui est responsable et propriétaire des archives, car c'est à lui qu'il incombe de rendre des comptes à ses interlocuteurs (personnel, autorités de contrôle, organismes de sécurité sociale....) Le client peut d'ailleurs à tout moment changer de prestataire et récupérer ses archives. Ce mode d'organisation se développe également au sein des entreprises et des administrations, certains services jouant le rôle de prestataire interne pour le compte de structures commanditaires. Dans ce cas aussi, le prestataire est service producteur du point de vue fonctionnel, un producteur par délégation en quelque sorte, mais l'entité commanditaire doit être considérée comme propriétaire des archives. Dans ce cas de figure, on ne peut se contenter de décrire le service producteur, simple opérateur interchangeable ; il convient d'introduire la notion de service propriétaire, dont la description est fondamentale pour contextualiser les documents.

Ces interrogations concernent également le service versant de données d'applications informatiques. Le versement d'archives électroniques est réalisé, dans la plupart des cas, par les services informatiques parce que ceux-ci assurent ou pilotent l'hébergement des données et disposent des moyens techniques pour réaliser l'opération de transfert des données des applications vers le SAE. Le rôle de service versant assuré par le service informatique est alors celui d'un opérateur technique, la responsabilité juridique du versement revenant au service propriétaire de l'application.

La production des données électroniques et les nouveaux modes d'organisation que facilite le développement de la dématérialisation, amènent ainsi à réévaluer le rôle des acteurs du processus d'archivage et à prendre en compte de nouveaux intervenants.

## ■ 2. Les processus de collecte et de gestion des archives électroniques

La dématérialisation de la production documentaire affranchit les producteurs, les archivistes et les utilisateurs des contraintes matérielles posées par les documents papier. Cette souplesse permet de collecter les archives électroniques bien plus en amont de leur cycle de vie que ce n'est possible pour les archives physiques, ce qui entraîne une évolution des pratiques archivistiques portée en particulier par la prise en compte de nouveaux besoins des producteurs et des utilisateurs.

### 2.1. L'assouplissement de la mise en œuvre de la théorie des trois âges

Avec la dématérialisation des procédures de travail et des documents qu'elles génèrent, les versements d'archives peuvent s'effectuer avant l'échéance de la durée d'utilité administrative, voire, dans certains cas, dès la création des données et non à l'issue de la durée d'utilité courante. Cela permet notamment de sécuriser les données dès leur création en évitant tout risque d'altération, voire de pallier les limites de certaines applications, incapables de « rejouer » la création des données. Cette collecte précoce n'interdit pas au service producteur d'accéder aux mêmes données dans son application métier, les données dupliquées pouvant y demeurer en parallèle. Ce point soulève d'ailleurs la question de l'original et du double, qui n'a pas forcément de sens sur le plan technique, mais qui, sur le plan organisationnel, est généralement tranchée en faveur de l'exemplaire transmis au SAE, chargé de garantir son intégrité et sa pérennité aussi bien à court qu'à long termes ou de manière illimitée. La question se pose alors de la durée de conservation à appliquer dans l'application productrice : outre les besoins des utilisateurs, sont-ce les coûts de stockage et la préoccupation de ne pas conserver d'inutiles doublons qui détermineront la purge des données par ailleurs déjà collectées par le SAE ? Ou le changement de logiciel ou de socle technique rendant les données précédemment produites inexploitable, sinon à un coût rédhibitoire, dans le nouvel environnement ? Ou encore la capacité du SAE à bâtir un système cohérent et pérenne de métadonnées et à offrir des modalités commodes et efficaces de recherche et de consultation, ce qui n'est pas la finalité principale des applications productrices ?

Dans cette perspective, la coupure entre âge courant et âge intermédiaire qu'a définie la pratique archivistique française paraît moins opérante que la distinction anglo-saxonne faite entre *documents* et *records*. En effet, la première théorie, qui a privilégié une approche pragmatique déterminée par l'usage des dossiers et la fréquence du recours à ceux-ci, est sous-tendue par des contraintes logistiques propres aux archives physiques, qui n'ont plus lieu d'être avec des archives électroniques. La seconde, plus théorique, est basée sur la valeur de preuve *a priori* que revêt un document suite à sa validation officielle. Même si ce concept n'est pas applicable à l'ensemble de la production documentaire, il n'en reste pas moins que l'archivage électronique facilite sa mise en œuvre en permettant de capturer les données à la source. Il facilite également la collecte et la conservation dans le SAE des *records* rattachés intellectuellement à des dossiers par ailleurs encore actifs et maintenus de leur côté dans un outil de GED jusqu'à la clôture de ces derniers et leur versement à leur tour dans le SAE.

### 2.2. L'abandon du numéro de versement ?

Pour des raisons techniques, certains versements peuvent être alimentés par plusieurs transferts de données. En outre, la capture régulière des données au plus près de leur date de création tend à fragmenter



les versements. Dans ces conditions, l'on peut s'interroger sur le caractère opérationnel de la notion de versement, dans l'acception d'un ensemble de dossiers transférés à un service d'archives en une seule fois à une date donnée. Pour la collecte des archives contemporaines, ce concept permet d'identifier des ensembles documentaires généralement volumineux et souvent cohérents sur le plan intellectuel et d'alimenter des répertoires méthodiques ouverts. Outil de gestion adapté aux contraintes logistiques posées par les archives physiques, qui ne peuvent être transférées aux archives qu'à l'issue de leur durée d'utilité courante, et pour lesquelles la constitution de versements importants et peu fréquents est commode, le versement ne répond plus aux mêmes nécessités et n'est plus soumis aux mêmes contraintes dans un monde documentaire dématérialisé. On peut en effet envisager un versement à « flux tendus » dans le SAE de données dès leur création, les contraintes de volume étant alors liées aux capacités des réseaux informatiques. On aboutit ainsi à un flux régulier, voire permanent, de données organisées en « paquets » dont l'organisation intellectuelle est reconstituée dans le SAE. On peut alors s'interroger sur l'intérêt d'attribuer un numéro de versement à un flux continu de données, sauf à déterminer arbitrairement la limite entre le précédent et le nouveau versement.

### **2.3. La cotation est-elle pertinente pour les archives électroniques ?**

Dans le monde physique, la cote est un système commode de signalisation et de repérage visuel des documents dans les magasins d'archives et remplace avantageusement le libellé de la série et l'intitulé de l'article. Pour les archives contemporaines, l'adoption de la cotation en continu a certes permis, entre autres, d'assouplir les contraintes de rangement dans les magasins en brisant le lien entre la cote et sa signification méthodique, mais la cote reste indispensable pour les archives physiques. La cote est également un identifiant pratique pour assurer en particulier la gestion des communications physiques. En revanche, son intérêt pour l'identification de fichiers électroniques est plus discutable, car les archives électroniques permettent de s'affranchir des contraintes matérielles que posent les archives papier. Certes le SAE attribue un identifiant à chaque fichier électronique archivé, mais il peut sans inconvénient être transparent pour l'utilisateur, à condition toutefois d'avoir été pensé par l'archiviste pour garantir son unicité et sa pérennité. La recherche dans le SAE s'effectue par l'intermédiaire d'un instrument de recherche, comme pour les archives physiques, mais il suffit ensuite de cliquer sur le lien de la description du fichier pour télécharger le document que l'on souhaite consulter. Les instruments de recherche n'ont pas à renvoyer à une cote, puisque le lien avec les fichiers électroniques archivés est immédiat.

### **2.4. Le recours accru à l'indexation et le retour aux cadres de classement pour les archives intermédiaires et définitives**

La collecte d'archives électroniques très récentes accroît la fréquence de consultation des données pour des besoins administratifs. Aussi convient-il d'adapter l'organisation et la description des données à ce type d'utilisation après analyse des types de requête possible et sélection des métadonnées pertinentes. Des masques de requête par domaine ou typologies documentaires peuvent être élaborés, qui s'appuient sur un jeu de métadonnées, préalablement fiabilisées et normalisées par le service d'archives, et des listes d'autorités et thesauri adaptés. Ce mode de description n'exclut évidemment pas une description contextualisée et hiérarchisée avec rattachement au fonds du producteur pour assurer notamment la gestion des fonds et le « traçage » des responsabilités sur ceux-ci ainsi que leur exploitation à moyen et long termes. Mais il permet d'effectuer des recherches suivant une logique documentaire.

On peut aussi organiser intellectuellement les données par fonction administrative descendant, le cas échéant, jusqu'à des séries typologiques (ex : données de paie, dossiers de maintenance immobilière, dossiers client...) pour accueillir des flux d'archives électroniques dans un cadre de classement prédéterminé, composé de séries ouvertes alimentées de manière régulière, voire permanente. Ces outils de gestion des archives courantes s'avèrent particulièrement efficaces pour les archives électroniques dont on a vu qu'elles contribuaient parfois à faire remonter à l'âge courant leur prise en charge par le service d'archives. Basé sur les fonctions et non sur les organisations, qui sont de plus en plus mouvantes, le cadre de classement peut rester opérant pour les archives intermédiaires et définitives. Il doit évidemment être combiné à des instruments de recherche décrivant le producteur ou propriétaire et la composition du fonds.

Ces nouvelles pratiques préfigurent peut-être pour les archives contemporaines électroniques l'abandon de la cotation en continu et le retour, non exclusif d'autres outils, au cadre de classement avec des séries ouvertes alimentées de manière régulière, voire permanente. Il pourrait s'agir là du prolongement de la réflexion amorcée avec la série W, qui avait entrepris de désacraliser la cote en la réduisant à une adresse

physique et apporté de ce fait une véritable souplesse descriptive en mettant l'accent sur la description des fonds. L'affranchissement des contraintes matérielles qu'offrent les archives électroniques permettent de franchir un nouveau pas dans le perfectionnement de la description archivistique en réconciliant les notions de fonds et de cadre de classement et en démontrant le bien-fondé de la combinaison des méthodes de l'analyse et de l'indexation.

## ■ 3. Contraintes et atouts posés par la dématérialisation

### 3.1. La confidentialité : une préoccupation beaucoup plus prégnante

Contrairement aux archives historiques, dont l'objectif pour l'archiviste est d'assurer la mise en valeur et la communication au plus large public, les archives électroniques, du moins dans leurs âges courant et intermédiaire, doivent bénéficier des mesures de sécurité requises pour les protéger de tout accès non autorisé. La mise en œuvre des règles d'accès et le contrôle des habilitations font partie des missions habituelles de l'archiviste. Cependant, les risques liés au piratage informatique, au croisement de fichiers de données personnelles et à la collecte très en amont d'archives « chaudes », la sensibilité des données étant inversement proportionnelle à leur âge, nécessitent la mise en place de dispositifs de sécurité adaptés pour garantir la confidentialité des données archivées, y compris en permettant de gérer des niveaux et des secteurs d'habilitation au sein des services d'archives.

Cette préoccupation rejoint la réflexion qui se développe aujourd'hui sur la nécessité de mettre en place des systèmes de classification des documents produits par un organisme.

### 3.2. Les archives hybrides

Le caractère hybride des données à archiver est dû à la dématérialisation partielle de certaines procédures. Le SNA est confronté à cette problématique de dossiers constitués pour partie de données électroniques et pour partie de documents papier, impliquant le plus souvent le versement des unes depuis une application centralisée et des autres par des services implantés sur l'ensemble du territoire. Cet éclatement du dossier complexifie la mise en œuvre de la collecte notamment parce que les deux systèmes documentaires répondent souvent à des logiques différentes (par exemple, un système informatique basé sur un identifiant unique s'incrémentant à chaque nouveau dossier électronique créé, mais des dossiers papier tenus dans des services opérationnels ayant instauré un mode de classement répondant à des logiques fonctionnelles différentes). Il faut donc concevoir des systèmes de référencement et de classement physique et logique suffisamment souples pour garantir le lien entre dossier électronique et dossier papier. L'archiviste doit mettre en place une méthodologie et des outils adaptés pour articuler les versements d'archives électroniques et papier. L'organisation de la communication des dossiers hybrides se trouve également compliquée, puisqu'il convient de gérer deux procédures distinctes, la communication physique étant généralement plus contraignante.

La numérisation de flux papier entrants, réalisée le plus souvent pour alimenter une GED ou un *workflow* est un autre cas de figure. Cette pratique, si elle n'est pas pensée dans un schéma documentaire global, contribue à faire éclater l'unité du dossier papier, dont les pièces sont pourtant généralement conservées en raison de leur qualité d'originaux. Dans cette situation, deux écueils sont à éviter. D'une part, le risque est la dispersion du dossier papier causée par un traitement logistique industriel qui ne gère pas ou mal la constitution ou la reconstitution du dossier après numérisation, les documents étant considérés comme des objets unitaires et non comme des supports d'information complémentaires les uns des autres. Ils sont alors stockés comme des archives intermédiaires alors que le dossier est actif. Il arrive même que des dossiers papier n'appartenant pas à la même entité soient mélangés dans le même carton. Dans un tel modèle de fonctionnement, la désorganisation du dossier est aggravée lorsque de nouvelles pièces arrivent ponctuellement pour alimenter le dossier initial. Elles sont alors numérisées et stockées au fur et à mesure de leur arrivée et non rangées dans le dossier papier qu'elles complètent pourtant. D'autre part, le lien avec le dossier électronique créé et les documents sortants du système n'est pas nécessairement mis en place en termes d'organisation et de gestion. On aboutit ainsi à la constitution de vracs, isolés des dossiers électroniques. Outre que l'on peut s'interroger sur l'intérêt qu'il y a à conserver pour leur valeur soit disant



probante des dossiers papier dont on a brisé le caractère organique qui en faisait toute la valeur en cas de contentieux, la gestion rationnelle des durées de conservation et des éliminations devient irréalisable puisque sont mélangés dans le même conteneur des pièces ressortissant à des dossiers dont les dates de clôture peuvent être très éloignées les unes des autres. Cela génère un surcoût non maîtrisé, amusant paradoxe pour des projets le plus souvent portés par des préoccupations d'efficacité et de rentabilité économique. Deux solutions sont alors envisageables. Soit l'on considère, après analyse des risques, que la conservation des originaux n'est pas justifiée. Soit ceux-ci présentent une réelle valeur, auquel cas les dossiers originaux doivent être gérés comme des dossiers courants notamment par leur reconstitution soigneuse après numérisation et en y versant et classant les nouvelles pièces. Par mesure d'économie et par souci d'efficacité, il est recommandé de ne numériser que les documents dont la consultation est fréquente et correspond à un véritable besoin fonctionnel. Le référencement, les métadonnées et le classement des dossiers originaux et de leurs doubles électroniques doivent enfin être définis conjointement et non isolément.

### **3.3. Les métadonnées électroniques**

Dans un univers documentaire certainement hybride pour encore de longues années compte tenu, entre autres, de la faible et lente généralisation de la signature électronique, les *métadonnées* électroniques constituent une part importante du matériau à archiver. En effet, dans de nombreux cas, ce n'est pas le dossier lui-même qui a été dématérialisé mais les fichiers d'enregistrement et de suivi des affaires. Ces données relatives aux dossiers papier et qui permettent d'y accéder et de les interpréter doivent évidemment être collectées et conservées avec les dossiers qu'elles documentent, comme les registres dans le monde papier. Ces métadonnées peuvent en outre être exploitées efficacement par l'archiviste qui peut les capturer pour enrichir l'instrument de recherche des dossiers correspondants.

Le fait que les données électroniques ne soient pas d'un accès immédiat en raison de contraintes techniques et leurs métadonnées évanescences, impose à l'archiviste d'intervenir en amont du cycle de vie des archives pour collecter les métadonnées de contenu et de gestion indispensables pour documenter les versements. Cette « remontée à la source » permet d'ailleurs à l'archiviste d'approfondir sa connaissance des systèmes d'information et des processus de travail de son organisme et d'en tirer profit pour documenter les versements d'archives papier en obtenant des extractions automatisées de métadonnées des SI. La récupération des référentiels métier au plus près de leur création relève de la même démarche. De même, comme pour les archives intermédiaires papier, il est intéressant d'appliquer aux archives électroniques, dès leur création, les règles de gestion et de description archivistiques. Cela permet de rentabiliser la fonction archives, car il serait dommage d'attendre l'échéance de la durée d'utilité administrative pour effectuer cet investissement.

## **■ 4. Quelle place pour l'archiviste dans la chaîne documentaire ?**

Pour consolider les données à archiver, l'archiviste doit désormais collaborer avec les administrateurs et les utilisateurs des applications. De surcroît, pour certaines productions documentaires, l'archiviste peut être amené à intervenir dans le processus de création des données ou de gestion des dossiers courants, contribuant ainsi au développement de la qualité au sein des organisations.

### **4.1. Une intervention sur le processus de création des données et des métadonnées**

La collaboration avec la maîtrise d'ouvrage des applications productrices de données est indispensable dès l'amont d'un projet pour intégrer les contraintes de l'archivage dans la définition de leurs fonctionnalités. C'est une évidence et elle est à juste titre régulièrement rappelée par notre profession. Cela implique de reporter vers l'amont le contrôle des versements en collaborant à la définition du contenu et de l'organisation des données, en déterminant la périodicité des versements... Mais le travail ne s'arrête pas là. Pour consolider les données à archiver, l'archiviste peut aussi travailler avec les utilisateurs des applications pour normaliser les règles de saisie de manière à récupérer des données fiables et exploitables sur le long terme. Cette

collaboration avec les services producteurs permet de déployer une collecte systématisée et programmée de données offrant une bonne qualité documentaire.

Il est même envisageable dans certains cas de pousser plus loin l'intervention de l'archiviste qui peut alors contribuer au traitement des données. Ainsi, dans le cadre de l'archivage des enregistrements sonores des débats du conseil d'administration du Groupe, le SNA récupère le fichier numérique quelques heures après la fin du conseil. Il traite le fichier : écoute, amélioration de la qualité du son, séquençage, description et découpage en plusieurs fichiers. Après ce travail, la séance traitée est transmise au secrétaire du conseil d'administration en vue de la rédaction du procès-verbal, qui s'en trouve accélérée grâce à un accès facilité aux informations enregistrées. Dans ce cas, le SNA participe à l'organisation des données qu'il archive. Le savoir-faire de l'archiviste est alors créateur de valeur dans la chaîne documentaire.

#### **4.2. L'archivage électronique n'est pas une problématique technique, mais une démarche documentaire**

Au fur et à mesure de l'élargissement de son champ d'intervention, notre profession a su inventer des concepts et des méthodes parfaitement opérantes qu'il convient d'appliquer aux archives électroniques. Pour toute application informatique, c'est au métier utilisateur d'assurer la maîtrise d'ouvrage. Il appartient aux archivistes de concevoir le cahier des charges des SAE que devront mettre en œuvre les informaticiens. L'approche technique posée en préalable est une voie sans issue, car elle ne permet pas de définir la méthodologie archivistique à mettre en place et impose des contraintes techniques qui n'auraient pas lieu d'être si le cahier des charges documentaire avait été posé d'emblée. Comme le répètent à l'envi nos collègues informaticiens, ils savent faire tout ce que leurs clients demandent. Par conséquent, osons demander et ne nous laissons pas enfermer dans des solutions techniques réductrices, que nous n'aurions pas définies.

Notre expérience nous conduit par conséquent à penser que mis à l'épreuve des problématiques posées par l'archivage électronique, les concepts archivistiques restent indéniablement valides, même si leur mise en œuvre requiert réflexion et ajustement.

La collecte des archives électroniques entraîne nécessairement une collaboration plus étroite avec les services producteurs et une meilleure prise en compte de leurs besoins. Elle permet aussi de développer une plus grande maîtrise des ressources documentaires, enjeu grandissant dans les organismes publics et privés aujourd'hui.

Enfin, l'archivage électronique recentre l'action de l'archiviste sur la valeur primaire des archives et sur la satisfaction des besoins des producteurs et des utilisateurs administratifs (clients, salariés, fournisseurs, organismes de contrôle, usagers, citoyens...), dont la demande est majoritaire et impérative, plutôt que de satisfaire prioritairement les usagers de la valeur secondaire des archives (historiens, généalogistes...), minoritaires. Sans ignorer la demande de ces derniers, les archivistes se doivent d'accompagner les services producteurs dans une gestion documentaire efficace, gage de plus-value globale pour la collectivité. Nous pourrions par là réaffirmer l'utilité sociale de notre métier.

# LA COLLECTE DES DONNÉES NUMÉRIQUES PRODUITES DANS LE CADRE DES MISSIONS D'AIDE SOCIALE DES DÉPARTEMENTS

---

Emilie Goubin

responsable des archives contemporaines, Archives départementales du Finistère

Nathalie Regagnon

chef du service des relations avec les administrations, Archives départementales de la Haute-Garonne

*Une des spécificités de l'archivage électronique est la diversité et la pluralité des intervenants avec lesquels il est nécessaire de travailler pour parvenir à la mise en œuvre concrète d'un versement d'archivage électronique, dans le contexte réglementaire des archives publiques françaises :*

- services de la collectivité : services fonctionnels, producteurs des données à archiver, tant sur le plan juridique que sur le plan historique ; direction des systèmes d'information ; archives départementales,
- réseau professionnel : Direction des archives de France,
- intervenants du secteur privé : éditeurs de logiciels, fournisseurs de matériels.

*La nature diverse de ces différents interlocuteurs et des liens qui sont les leurs, juridiquement différents (liens institutionnels, fonctionnels, réglementaires ou commerciaux) nécessitent une forte coordination, qui peut trouver différents modes de mise en œuvre.*

*En effet, l'intervention des différents interlocuteurs est plus ou moins importante, suivant le cas :*

- choix politique de la collectivité,
- type et ancienneté du progiciel,
- phase d'expérimentation/d'automatisation.

*Il nous est donc apparu intéressant de vous présenter le travail réalisé par un groupe de travail d'archivistes départementaux, relatif à l'archivage des bases de données produites dans le cadre des compétences en matière d'aide sociale des Conseils généraux. Ce groupe de travail, réuni à compter de mars 2006, a travaillé à la rédaction de profils, permettant l'export de données textuelles conformément au standard d'échanges de données pour l'archivage à partir des progiciels utilisés par les services sociaux des départements. Ont été publiés pour le moment les profils relatifs à l'Aide sociale à l'enfance et à l'Allocation personnalisée d'autonomie ; sont en cours de finalisation les profils relatifs à l'Aide sociale générale et au Revenu minimum d'insertion. Depuis, une nouvelle problématique est à l'étude, à savoir les modalités de numérisation et l'archivage des dossiers numériques produits dans les systèmes de GED des Maisons départementales pour les personnes handicapées (et non plus seulement des données).*

*Il s'est agi en la matière de coordonner les interventions et les apports des différents interlocuteurs, et de définir et formaliser le rôle et les responsabilités de chacun dans le cadre d'un projet d'archivage électronique.*

*Par ailleurs, plusieurs possibilités existent quant à la mise en œuvre technique de tels projets : dans le groupe de travail qui nous occupe, deux approches techniques ont été privilégiées, du fait du contexte :*

- développement d'un module d'export spécifique par la Direction des systèmes d'information, s'agissant d'un progiciel pour lequel aucun module d'archivage n'avait été prévu dans le cahier des charges d'origine, dans un contexte d'appropriation et d'expérimentation d'un projet d'archives électroniques,
- développement d'un module d'archivage électronique par l'éditeur du progiciel.

*Ces expériences ont permis de définir quelques grands principes...*

- mutualisation de la réflexion nécessaire : les spécifications sont génériques, et doivent ensuite être adaptées à chaque application, même si le cadre général est le même,
- adaptations locales, soit avec la DSI, soit avec l'éditeur, des spécifications, en fonction du contexte de chaque collectivité (plate-forme d'archivage ou non, ...)
- coordination nécessaire avec l'archivage des dossiers papier.

*... et de poser de nouvelles réflexions propres au métier d'archiviste :*

- devoir de mémoire et droit à l'oubli : est-il pertinent d'anonymiser les données électroniques ?

- *la pertinence du tri des données électroniques : doit-on préserver les mêmes principes que pour le papier ?*

[Consulter la présentation.](#)

### **Managing and distributing digital information. A presentation of the results of the working group on collection of digital data generated during social welfare missions in French Départements**

*A particular characteristic of electronic archiving is the diversity and multiplicity of stakeholders we need to work with in order to effectively carry out the accession of an electronic archive, within the regulatory framework covering French public archives:*

- *public authority services: functional services, bodies producing data to be archived, both legal and historical; information systems (IS) services; Département archives,*
- *professional network: French archives authority,*
- *private-sector organisations: software companies, equipment suppliers.*

*The diverse nature of these various stakeholders, as well as each one's particular, legally disparate relationships (institutional, functional, regulatory or commercial links) mean that close coordination is essential. This can be achieved in a number of ways.*

*Each stakeholder's involvement is relatively important, according to the circumstances:*

- *public authority's choice of policy,*
- *type and age of software package,*
- *experimentation/automation phase.*

*We therefore decided that it might be interesting to present the work conducted by a working group of Département archivists, on the archiving of databases generated in connection with the responsibilities of the Conseil Général concerning social welfare. This working group, which was formed in March 2006, focused on writing profiles that enable textual data conforming to the data exchange format to be exported for archiving from software packages used by the Département's social services. To date, the profiles relating to child welfare assistance (Aide sociale à l'enfance) and personal care allowance (Allocation personnalisée d'autonomie) have been published; those for general welfare assistance (Aide sociale générale) and income support benefits (Revenu minimum d'insertion) are currently being finalised. More recently, another issue has come under study, namely the digitising methods and archiving of digital files (and no longer just data) generated for disabled people in the EDM systems of Département institutions.*

*This involved coordinating the involvement and contributions of the various stakeholders, and defining and formalising each one's role and responsibilities within the scope of an electronic archiving project.*

*Furthermore, there are several possibilities with regard to the technical implementation of such projects: in our working group, two technical approaches were favoured, depending on the context:*

- *development by the IS service of a special export module, a software package for which no archiving module was planned in the original specifications, in a context of appropriation and experimentation of an electronic archive project,*
- *development by the software company of an electronic archiving module.*

*These experiments enabled us to define a few major principles...*

1. *the need for a collective debate: the specifications are generic and must then be adapted to each application, even if the general framework is identical,*
2. *local adaptation of specifications, either with IS services or the software supplier, depending on each public authority's context (whether or not there is an archive platform, etc.),*
3. *coordination needed with archiving of paper files.*

*...and to raise some new issues which relate specifically to the function of archivist:*

1. *the duty to remember and the right to forget: should electronic data be anonymous?*
2. *the relevance of selective sorting of electronic data: should we maintain the same principles as for paper?*

[See the presentation.](#)

Une des caractéristiques de l'archivage électronique est la diversité et la pluralité des intervenants avec lesquels il est nécessaire de travailler pour parvenir à la mise en œuvre concrète d'un versement d'archivage électronique, dans le contexte réglementaire des archives publiques françaises :

- services de la collectivité : services fonctionnels, producteurs des données à archiver, tant sur le plan légal que sur le plan historique ; direction des systèmes d'information ; Archives départementales,
- réseau professionnel : Direction des archives de France, Association des archivistes français,
- intervenants du secteur privé : éditeurs de logiciels (métiers du service producteur et des archives), fournisseurs de matériels.

La nature diverse de ces différents interlocuteurs et des liens qui sont les leurs, juridiquement différents (liens institutionnels, fonctionnels, réglementaires ou commerciaux) nécessitent une forte coordination, qui peut trouver différents modes de mise en œuvre.

En effet, l'intervention des différents interlocuteurs est plus ou moins importante, suivant le cas :

- choix politique de la collectivité,
- type et ancienneté du logiciel,
- phase d'expérimentation/d'automatisation.

Il nous est donc apparu intéressant de vous présenter le travail réalisé par un groupe de travail d'archivistes départementaux, relatif à l'archivage des données produites dans le cadre des compétences en matière d'aide sociale des Conseils généraux.

## ■ 1. Présentation du groupe de travail relatif à la collecte des données numériques produites dans le cadre des missions d'aide sociale des départements

Sur l'initiative de plusieurs départements<sup>1</sup>, un groupe de travail rassemblant à la fois des archivistes, des travailleurs sociaux et des chefs de projets informatiques s'est réuni en vue de définir les modalités d'archivage des données numériques produites dans le cadre des missions d'aide sociale des Départements. La réflexion a été suivie de près et encouragée par la Direction des Archives de France (DITN et DPACI).

La volonté de partager des informations autour d'une problématique commune - en l'occurrence l'archivage des données numériques de l'aide sociale à l'enfance dans un premier temps - a permis d'aboutir à des résultats rapides : les premiers travaux, portant sur l'aide sociale à l'enfance et sur l'allocation personnalisée d'autonomie, ont démarré en mars 2006 pour aboutir à la publication d'une note d'information et d'une instruction de la Direction des Archives de France<sup>2</sup> en juin 2007.

Les travaux se sont ensuite poursuivis avec l'étude des dispositifs de l'aide sociale générale (aides accordées aux personnes âgées et handicapées) et du revenu minimum d'insertion (RMI). Les instructions et profils afférents sont en cours de finalisation.

Dans le cadre des premiers travaux (ASE et APA), le groupe a cherché à définir et à établir une méthodologie de travail adaptable et réutilisable par tous pour la mise en œuvre des versements et des

---

<sup>1</sup> Participants : Département et ville de Paris, Départements de l'Ain, de l'Aube, des Bouches-du-Rhône, de la Dordogne, de la Drôme, de l'Eure, du Finistère, de la Haute-Garonne, de la Haute-Savoie, de la Loire, du Maine-et-Loire, du Nord, du Pas-de-Calais, des Pyrénées orientales, du Rhône, de la Saône-et-Loire, de la Seine-Maritime, de la Seine et Marne, des Yvelines. Coordination : Nathalie Regagnon (Archives départementales de la Haute-Garonne) et Céline Guyon (Archives départementales de l'Aube).

<sup>2</sup> Note d'information DITN/DPACI/2007/001 relative au groupe de travail sur la collecte des données numériques produites dans le cadre des missions d'aide sociale des Départements ; Instruction DITN/RES/2007/003 relative aux profils du standard d'échange de données pour l'archivage des données numériques de l'aide sociale à l'enfance (ASE) et de l'allocation personnalisée d'autonomie (APA).

éliminations des données issues des bases de données utilisées dans les services des Conseils généraux chargés de l'aide sociale.

Pour cela, il s'est attaché à définir le cycle de vie des données, à proposer une sélection des données présentant un intérêt historique, pour chaque type de dispositif, et surtout à mettre en œuvre le standard d'échange de données pour l'archivage (dont la version 0.1 a paru en mars 2006) en rédigeant les profils idoines.

Le groupe a, par ailleurs, entamé un travail d'information en direction des éditeurs des progiciels métiers du social et métier des archives pour la mise en œuvre de ces préconisations.

Les travaux suivants, et notamment ceux portant sur le RMI, ont permis d'aller plus loin et de travailler sur la mise en œuvre technique d'un versement d'archives électroniques.

Plusieurs possibilités existent quant à la mise en œuvre technique de tels projets ; dans le groupe de travail qui nous occupe, deux approches techniques ont été privilégiées, du fait du contexte :

- développement d'un module d'export spécifique par la Direction des systèmes d'information, s'agissant d'un progiciel pour lequel aucun module d'archivage n'avait été prévu dans le cahier des charges d'origine, dans un contexte d'appropriation et d'expérimentation d'un projet d'archives électroniques,
- développement d'un module d'archivage électronique par l'éditeur du progiciel.

## ■ 2. Définition et rédaction des spécifications

### 2.1. Le contexte : pourquoi l'aide sociale ?

L'aide sociale constitue une mission essentielle des Départements. Cette mission donne lieu à une production abondante d'archives, chaque situation donnant lieu à la création d'un dossier individuel, voire de plusieurs, suivant l'organisation administrative de l'aide sociale mise en place par les Conseils généraux. Ces dossiers individuels, lorsqu'ils sont versés aux Archives départementales, le sont sous forme papier.

Dans le contexte actuel, l'environnement papier reste donc largement prédominant en terme de collecte archivistique.

Néanmoins, sous l'effet du développement des technologies, de l'évolution des dispositifs d'aide sociale (multiplication des dispositifs et des partenariats), l'administration a informatisé ses procédures : le suivi et la gestion des droits des usagers tout au long de leur présence au sein d'un dispositif sont dorénavant en partie assurés via des bases de données de plus en plus riches et complexes. Ces bases de données sont développées par la Direction des systèmes d'information de la collectivité ou, le plus souvent, dans des progiciels métiers.

Elles ont la plupart du temps remplacé les anciens fichiers, répertoires ou registres d'enregistrement. Les informations récapitulatives et synthétiques donnant accès aux dossiers papier n'existent donc plus sous forme papier.

L'archivage de ces bases de données se pose aujourd'hui avec acuité.

D'une part, du point de vue réglementaire, les informations à caractère personnel renseignées dans ces bases de données sont soumises à la loi informatique et liberté : de ce fait, elles ne doivent être conservées par les services gestionnaires que pendant le temps strictement nécessaire à l'exercice de leurs compétences. Dans le domaine de l'aide sociale, la commission nationale informatique et liberté préconise de ne pas conserver ces données plus de 24 mois après la sortie du dispositif du bénéficiaire. A l'issue de ce délai, déjà atteint dans la plupart des collectivités, ces données numériques ne doivent plus rester dans la base de production ou base métier mais doivent être, selon les cas, éliminées ou transférées dans une base d'archivage intermédiaire.

D'autre part, du point de vue archivistique, le lien existant entre les données numériques et les dossiers papiers, permettant leur identification, est bien souvent perdu au moment du versement aux Archives départementales, voire au moment du préarchivage des dossiers : le dossier est en effet sorti de la



classothèque pour être rangé dans le local de préarchivage, selon une autre logique et les données numériques sont soit détruites soit conservées mais inexploitable, comme en déshérence dans le système de gestion.

Or, dans le cas d'une conservation intégrale de la collection des dossiers papier, la récupération automatique de ces informations sous forme électronique permettrait la rédaction de bordereaux de versement plus détaillés et éviterait des ressaisies inutiles, tant pour le service versant au moment de la préparation du versement des dossiers papiers que pour le service d'archives pour l'intégration de ces données descriptives dans son système de gestion.

Par ailleurs, dans le cas où il est procédé à l'échantillonnage des dossiers papier, l'archivage des données individuelles, en tant qu'objet d'archives, et non plus en tant que données descriptives, permet également d'assurer une collecte archivistique plus représentative des nouveaux modes de fonctionnement de l'administration et de chaque dispositif.

Dès lors, il apparaît nécessaire d'intervenir dans le système de production du producteur pour transposer, dans l'environnement informatique, les notions de préarchivage et d'archivage, en définissant précisément le cycle de vie des données numériques, et en coordonnant l'archivage des données électroniques et des dossiers papier.

## **2.2. La méthodologie**

### **2.2.1. La définition du cycle de vie des données numériques**

Un dossier individuel numérique est réputé clos lorsqu'une mesure concernant un bénéficiaire arrive à son terme mais il peut être réactualisé pendant toute sa durée d'utilité courante. A la fin de cette durée d'utilité courante (par exemple les 18 ans de l'enfant, son décès ou son adoption, dans le cadre de l'aide sociale à l'enfance), le dossier est préarchivé, c'est-à-dire, transféré dans une base d'archivage intermédiaire, où il sera conservé pendant toute sa durée d'utilité administrative.

Plusieurs scénarii sont envisageables :

- une base d'archivage intermédiaire développée par l'éditeur dans le progiciel métier,
- une base d'archivage intermédiaire commune à plusieurs outils,
- le transfert des données numériques vers une plate-forme d'archivage.

Pour certains dispositifs, la durée d'utilité courante peut être courte et ne nécessite pas un transfert dans une base d'archivage intermédiaire : les données produites dans le cadre du RMI par exemple peuvent ainsi être exportées et versées sous la forme d'un échantillon directement depuis le progiciel métier au bout de deux ans à compter de la sortie du dispositif, moyennant la production de bordereaux d'élimination nominatifs, permettant la recherche des dossiers papier durant la fin de leur durée d'utilité administrative (6 ans).

### **2.2.2. L'évaluation et la sélection des données**

Le premier travail à entreprendre lorsqu'on souhaite archiver des données numériques issues de bases de données est sensiblement le même à fournir que pour le papier.

En premier lieu, le groupe a travaillé dispositif par dispositif et a veillé ainsi à respecter l'unité d'un fonds (aide sociale à l'enfance, aide sociale aux personnes âgées, insertion, etc.).

L'évaluation et la sélection des données présentant un intérêt historique s'est appuyée sur une étude détaillée du fonctionnement des applications et sur un recensement des données enregistrées et de la façon dont elles sont enregistrées (présence de tables, d'outils de contrôle de la saisie, de champs obligatoires, etc.). La déclaration à la CNIL, les chartes d'écriture et les guides réalisés par les administrateurs constituent des outils précieux pour réaliser cette étude. Une démonstration approfondie de la base par ses utilisateurs, ou, mieux, la participation à une formation globale destinée aux utilisateurs, est une bonne manière également de comprendre le fonctionnement d'un progiciel.

Il convient d'utiliser, chaque fois que possible, les outils déjà réalisés, même s'ils ont été conçus pour le papier. Ainsi, les circulaires de tri et d'élimination de la direction des archives de France et les tableaux de gestion sont le point de départ incontournable de ce travail. Dans le cas de l'aide sociale à l'enfance, la circulaire AD 98-6 du 6 juillet 1998 relative au traitement des archives produits dans le cadre de l'aide sociale en faveur des mineurs a servi de base au travail du groupe, qui a ainsi fait l'économie de la réflexion sur les grands principes de tri (conservation de tous les dossiers). Dans le cas de l'allocation personnalisée d'autonomie, en revanche, il n'existait pas de telle circulaire, ce qui a amené le groupe de travail à travailler sur la DUA des documents produits dans le cadre de ce dispositif. Par ailleurs, pour le RMI, la circulaire existante (AD99-1) a fait l'objet d'un réexamen, afin de prendre en compte les évolutions réglementaires intervenues depuis sa parution.

Le groupe de travail a dû également définir les grands principes de tri - notamment l'échantillonnage ou non des dossiers. La présence du DPACI était alors nécessaire. Plusieurs questions ont fait débat : profite-t-on du support électronique pour garder une trace de tous les individus et ainsi laisser le choix aux chercheurs de se constituer leur propre échantillon ? Ou applique-t-on les mêmes critères de tri que pour le papier en partant du principe que la sélection des archives est une mission incontournable de l'archiviste ?

Une fois les grands principes de tri définis, il est nécessaire d'affiner la sélection de manière détaillée, champ par champ le cas échéant et de préciser, pour chaque type de données, le sort qui lui est réservé (conservation ou élimination).

Deux scénarios d'archivage se sont dessinés :

- cas n° 1 (ASE) : ont été retenus les éléments permettant d'identifier de manière sûre et certaine un enfant et les éléments permettant de contextualiser l'aide sociale à l'enfance et laissant ouvert les possibilités d'une recherche à caractère sociologique. Ce choix s'explique notamment par le fait que les dossiers des enfants sont versés intégralement aux Archives départementales sous forme papier ;
- cas n° 2 (APA, ASG, RMI) : les informations saisies dans la base de données pouvant se substituer, au terme de la DUA, aux dossiers papier (présence des mêmes catégories d'information sur les deux supports), ont été retenus d'une part, les éléments permettant de compléter le bordereau de versement, d'autre part, le contenu de données (équivalent du contenu du dossier papier). La conservation d'un échantillon représentatif (conservation des années 0 et 5) a été proposée pour l'électronique ; le groupe a proposé de conserver en parallèle un échantillon papier (5% des dossiers).

### **2.3. L'implémentation du standard d'échange de données**

Conformément à l'instruction DITN/RES/2006/001 du 8 mars 2006, l'export des données à verser aux Archives départementales ou à éliminer doit être conforme, techniquement, au standard d'échange de données pour l'archivage, dont la version 0.1 a été publiée par la DAF et la DGME en mars 2006. Les services publics d'archives seront ainsi en mesure d'inclure ces données dans leur propre système d'information, en vue d'en assurer la bonne conservation.

Le standard d'échange de données pour l'archivage spécifie, sous la forme d'un ensemble de schémas XML, les messages nécessaires aux différents échanges entre le service versant (service à l'origine du transfert des données ou de la demande de visa d'élimination) et le service public d'archives. Sont en particulier décrits les bordereaux de versement et d'élimination eux-mêmes, mais aussi d'autres messages nécessaires au cas où le dialogue entre le service versant et le service public d'archives se poursuit sous forme électronique (accusés de réception, messages d'anomalie, etc.).

Le groupe de travail s'est donc attaché à établir des profils ou « mapping » pour l'ensemble des dispositifs de l'aide sociale. Un profil est un mode d'emploi du standard d'échange de données pour l'archivage, pour une catégorie de données. Il établit la correspondance entre les champs du progiciel métier et les balises XML du standard d'échange. Le choix du groupe de travail s'est porté sur l'écriture d'un profil par dispositif afin de rester fidèle aux processus administratifs.

Les profils comportent :

- la nature des données à verser ou à éliminer,



- la structure du bordereau de versement et du bordereau d'élimination, c'est-à-dire la description des dossiers à verser ou à éliminer,
- les fonctions d'export souhaitées.

Pour la description des dossiers des bénéficiaires de l'aide sociale à verser, deux niveaux ont été retenus :

- le niveau supérieur, identifié par l'élément « Archive » dans le standard d'échange de données pour l'archivage, correspondant à un ensemble de dossiers individuels clos pendant le même intervalle de dates ;
- le niveau inférieur, identifié par l'élément « Objet d'archive » dans le standard d'échange de données pour l'archivage, correspondant à un dossier individuel ; il y a autant d'« Objets d'archives » que de dossiers individuels versés.

A chacun de ces niveaux de description sont associées des métadonnées qui portent sur le contenu (informations de description qui proviennent de la base de données), la gestion et le format des données et des mots-clefs, en vue de faciliter le repérage, la gestion et la préservation des données versées. Pour la description des dossiers ou des données à éliminer, seul l'élément « Archive » a été retenu, car une description plus poussée ne paraissait pas s'imposer.

## **2.4. La rédaction de modèle de contrats tri-partites**

Les transferts et demandes d'élimination obéissent aux principes définis par le décret n° 79-1037 du 3 décembre 1979. Mais, en raison des contraintes techniques liées à tout projet informatique et au nombre des partenaires impliqués, il est souhaitable de définir, dans un contrat de service les rôles et responsabilités des différents intervenants, en particulier du service public d'archives, du service d'aide sociale et du service informatique.

Le contrat de service définit les caractéristiques des données numériques à verser et/ou éliminer (le format des transferts, leur volume maximal, leur fréquence, les actions précises à réaliser par chacun, les outils utilisés), encadre les obligations et responsabilités de chaque acteur (autorité d'archivage et opérateur d'archivage) et modélise les processus de versement et d'élimination.

## **■ 3. La mise en œuvre des versements**

Suite aux réflexions menées par le groupe de travail, des départements ont souhaité mettre en œuvre techniquement les projets de versements de données aux Archives départementales, soit avec les éditeurs de progiciels, soit en interne, en étroite collaboration avec le service informatique.

### **3.1. Le développement d'un module d'archivage électronique par l'éditeur du progiciel**

Le groupe avait déjà souhaité aller plus loin que le profil et a proposé une liste de fonctionnalités d'export attendues, dans le cas où l'éditeur du progiciel serait prêt à faire évoluer son outil : il faudrait bâtir un module d'export au sein même du progiciel, sous la forme d'un écran comprenant des champs spécifiques par exemple, prévoir la traduction des données en clair au moment de l'export, etc.

Une fois les spécifications achevées, le groupe a souhaité les présenter aux éditeurs de progiciels sociaux et a organisé une réunion en ce sens.

L'un d'entre eux avait déjà travaillé avec un nombre restreint de département (Aube notamment) pour développer un module de gestion des données intermédiaires au sein du logiciel. Ce module répond aux contraintes posées par la commission nationale Informatique et Libertés et permet de ne pas rendre accessibles à tous les travailleurs sociaux des données de plus de vingt-quatre mois après clôture d'un dossier. Il n'est accessible qu'à un nombre restreint de personnes, et permet de retrouver facilement un dossier ; en cas de réactivation du dossier, les données sont automatiquement réinjectées dans le système courant. Les versements de données définitives aux Archives départementales ne sont pas encore conformes au standard

d'échange puisque les travaux ont débuté avant la parution de celui-ci : les données sont transférées sous forme de fichiers Excel en attendant l'implémentation du standard. Ce module est actuellement optionnel, c'est-à-dire que les collectivités doivent en faire la demande et l'acquisition spécifiques.

Le second éditeur a souhaité apporter des réponses à la problématique de l'archivage suite aux travaux du groupe. L'évolution du progiciel et la mise en œuvre d'un module d'archivage intermédiaire et d'exports conformes au standard d'échange s'avèrent beaucoup plus longues que dans le premier cas, l'éditeur ne faisant évoluer son produit qu'en fonction des attentes exprimées par les départements clients lors des journées annuelles du club utilisateurs. Après présentation de la réflexion nationale au club, l'éditeur a fait voter une étude sur le sujet de l'archivage : la première réunion avec des départements pilotes (Rhône, Haute-Garonne, Nord, etc.) a permis de leur expliciter les tenants et les aboutissants d'un tel projet et de définir un schéma fonctionnel pour le développement d'une base d'archivage intermédiaire prenant en compte le standard d'échange. Les premiers résultats de l'étude n'ont été présentés au club utilisateur que plusieurs mois après. Actuellement, de nouvelles journées d'études ont été soumises à leur vote et approuvées. Mais le processus est long et plusieurs départements réfléchissent à une solution intermédiaire en attendant un outil plus performant.

### **3.2. Le développement d'un module d'export spécifique par la Direction des systèmes d'information du département**

Dans certains cas, il est cependant difficile de prévoir le développement d'un module par les éditeurs :

- progiciel pour lequel la maintenance n'est plus assurée ou pour lequel la collectivité n'est plus sous contrat avec l'éditeur,
- progiciel « en fin de vie », dont le remplacement est prévu à court ou moyen terme,
- urgence de se conformer à la réglementation en vigueur, à savoir suppression des données de la base de production 24 mois à compter de la sortie du dispositif, sans qu'un module d'archivage ait été prévu au marché.

Le département du Finistère s'est retrouvé confronté au dernier cas cité pour ses données produites dans le cadre du RMI, la DSI étant régulièrement interpellée par les travailleurs sociaux pour que la situation soit régularisée.

Par ailleurs, il est apparu intéressant, tant à la DSI qu'aux Archives départementales, de travailler sur une expérimentation concrète, appuyée sur les travaux du groupe national, afin de développer une expertise interne à la collectivité en matière d'archivage électronique, et d'entamer, à partir de cette expérimentation, une réflexion quant à la mise en œuvre d'un Système d'archivage électronique à l'échelle du Conseil général.

Cette expérimentation a été menée à moyens constants, avec l'appui du DPACI : un export des données a été réalisé en juillet 2008, et a été intégré à la plate-forme PILAE du Ministère de la culture pour tester sa conformité.

Le travail a porté sur :

- l'export de données descriptives conformes au standard d'échange de données pour les bordereaux de versement et d'élimination,
- le contenu de données archivées à proprement parlé, à savoir les données individuelles elles-mêmes : en la matière, il s'est agi de définir le format des données (choix du xml, compte tenu du besoin d'historisation des données, et donc de leur structure complexe), d'affiner la sélection des données définie dans le profil en l'adaptant à la structure de la base de données du progiciel (définition de la table d'origine de chaque donnée sélectionnée par exemple), de définir l'organisation des données entre elles (structure du contenu de données à proprement parlé, ce qui a abouti à la génération d'un schéma xml, dans lequel sont également définies les noms des balises),
- la programmation d'un module d'export de données,
- la rédaction d'une procédure commune à l'archivage des données électroniques et des dossiers papier.

## ■ Conclusion

La méthodologie employée par le groupe de travail a vocation à être réutilisée et améliorée par d'autres archivistes pour d'autres types de dispositifs : l'archivage électronique des dossiers de demande de titres de séjour, des dossiers des bénéficiaires de la CAF, etc.

Par ailleurs, le groupe de travail a abordé la problématique liée à la restitution des données : un travail avec les éditeurs des progiciels métiers des archives reste à entreprendre pour que les données collectées aujourd'hui puissent être communiquées en salle de lecture des services d'archives demain.

# Séance plénière de clôture

---

---

sous la co-présidence de Toivo JULLINEN  
président du DLM-Forum  
et d'Hubert SZLASZEWSKI  
directeur au secrétariat général de la Commission européenne

# ***Plenary closing session***

---

---

*co-chairing Toivo JULLINEN*

*DLM-Forum President*

*and Hubert SZLASZEWSKI*

*Director, Secretariat General, European Commission*

## **ACHIEVEMENTS AND NEW DIRECTIONS: CONCLUSIONS OF THE *DLM-FORUM CONFERENCE 2008* IN *TOULOUSE* ADOPTED BY THE CONFERENCE PARTICIPANTS**

---

*The successful DLM Forum conference in Toulouse in December 2008, organised in the framework of the French EU presidency, re-affirmed the leading role of the DLM Forum as a platform for multi-disciplinary co-operation on electronic records and archives management in Europe. Since its creation in 1996, the DLM Forum has fostered cooperation between public administration, research, records and information managers and the ICT industry. In recent years it has successfully developed into a partnership bringing together organisations from all sectors that have an interest in electronic records and archives management.*

*The Council Recommendation of November 2005 on closer co-operation between archives in the EU<sup>1</sup> asked for an update and extension of the Model Requirements for the management of electronic records as one of five priority measures for cooperation between archives in Europe. The DLM Forum adopted an extensive scoping report for the development of MoReq2 and, with support from the European Commission, oversaw the development of MoReq2 as a testable specification for electronic records management systems.*

*In its progress report to the Council of the European Union of August 2008<sup>2</sup>, the European Archives Group (EAG) encourages the DLM Forum to pursue its work in the field of electronic records and archives management. In particular the EAG recommends that the DLM Forum:*

*maintains a viable governance regime for MoReq2, including notably the monitoring of the translation of MoReq2 into the official languages of the EU and the addition of national chapters zero to take account of the national traditions and legislative framework, as was done in France during the French presidency of the EU;*

*sets up and monitors a European conformance testing and certification regime for MoReq2;*

*continues to promote interdisciplinary co-operation in the area of electronic records and archives management and to hold a members meeting during each EU presidency;*

*continues to organise large DLM Forum conferences once every three years.*

*The participants at the 5<sup>th</sup> DLM Forum conference in Toulouse call on the DLM Forum to implement these recommendations and to regularly report on its activities to the European Archives Group. The DLM Forum conference 2008 calls on the European Commission to continue to support the DLM Forum in the accomplishment of these objectives.*

*The DLM Forum welcomes the active participation of the new member states of the EU in the activities of the forum. It takes note of the use of MoReq2 beyond the boundaries of the EU, resulting in translations<sup>3</sup> and increased interest from abroad, and of the already initiated projects using MoReq2 for electronic records management solutions. The participants of the DLM Forum conference 2008 ask that the DLM Forum continues to broaden its appeal to all interested communities, and that it promotes the use of MoReq2 across Europe and elsewhere in the world.*

*The conference encourages the DLM Forum to look in new directions and notably to:*

*consider issues surrounding the transfer of records from the originating bodies to archival services, in particular by encouraging the use of the MoReq2 XML schema in joint pilot projects for interoperability between records management and archive applications;*

*evaluate the impact of new software technologies on records management and archives systems in practice;*

---

<sup>1</sup> Council Recommendation of 14 November 2005 on priority actions to increase cooperation in the field of archives in Europe (2005/835/EC), OJ, L312 of 29.11.2005

<sup>2</sup> COM(2008)500; SEC(2008)2364.

<sup>3</sup> Into Russian and Korean.

*provide guidance for developing electronic records management skills and competencies within user communities across Europe;*

*establish strategies for influencing decision-makers to implement ERDMS, notably by carrying out cost-benefit analyses.*

*provide the DLM members a set of practical case studies and concrete solutions.*

*The 2008 DLM Forum conference has brought forth new practical approaches and solutions as well as strategies for the future. The DLM Forum should continue to set the agenda and provide guidance for user oriented solutions and models for best practice with a view to maintaining the trustworthiness and authenticity of the digital knowledge of our society for now and for the future.*



## **RÉALISATIONS ET NOUVELLES DIRECTIONS : CONCLUSIONS DE LA CONFÉRENCE 2008 DU DLM-FORUM, TOULOUSE, ADOPTÉES PAR LES PARTICIPANTS À LA CONFÉRENCE**

---

Le succès de la conférence du DLM-Forum tenue à Toulouse en décembre 2008, organisée dans le cadre de la présidence française de l'Union Européenne, a réaffirmé le rôle moteur du DLM-Forum en tant que plateforme de coopération multidisciplinaire sur l'archivage électronique en Europe. Depuis sa création en 1996, le DLM-Forum a encouragé la coopération entre les administrations publiques, la recherche, les gestionnaires des documents et de l'information et l'industrie des TIC (Technologies de l'Information et de la Communication). Ces dernières années, il a développé avec succès un partenariat regroupant ensemble des organismes concernés par la gestion des documents et des archives électroniques.

La Recommandation du Conseil de Novembre 2005 pour une coopération étroite entre les services d'archives dans l'Union Européenne<sup>4</sup> a souhaité que la mise à jour et l'extension du Modèle de Spécifications pour la gestion des documents électroniques soient inscrites comme l'une des cinq mesures prioritaires pour la coopération entre les services d'archives en Europe. Le DLM-Forum a adopté une étude détaillée pour le développement de MoReq2, et avec le soutien de la Commission Européenne, a prévu le développement de MoReq2 comme une norme de test des systèmes d'archivage électronique.

Dans son rapport d'étape d'août 2008<sup>5</sup> au Conseil de l'Union Européenne, le Groupe Européen des Archives (EAG) encourage le DLM-Forum à poursuivre son travail dans le domaine de l'archivage électronique. L'EAG recommande en particulier que le DLM-Forum :

Assure un régime viable de gouvernance de MoReq2, incluant notamment le contrôle de la traduction de MoReq2 dans les langues officielles de l'Union Européenne et l'ajout de chapitres zéro pour chaque pays, prenant en compte les pratiques et le cadre juridique nationaux, comme cela a été le cas en France, à l'occasion de sa présidence de l'Union Européenne ;

Mette en place et pilote un dispositif européen de test de conformité à MoReq2 et de certification ;

Continue à promouvoir la coopération interdisciplinaire dans le domaine de l'archivage électronique et tienne une réunion des membres sous chaque présidence de l'Union Européenne ;

Continue à organiser des grandes conférences du DLM-Forum tous les trois ans. Les participants à la 5<sup>e</sup> Conférence de DLM-Forum à Toulouse engagent le DLM-Forum à mettre en œuvre ces recommandations et à présenter régulièrement au Groupe Européen des Archives des rapports sur ses activités. La Conférence 2008 du DLM-Forum engage la Commission Européenne à poursuivre son soutien au DLM-Forum pour la réalisation de ces objectifs.

Le DLM-Forum salue la participation active des nouveaux Etats membres de l'Union Européenne aux activités du Forum et prend bonne note de l'utilisation de MoReq2 au-delà des frontières de l'Union Européenne, qui engendre de nouvelles traductions<sup>6</sup> et un intérêt accru de l'extérieur, ainsi que des projets déjà lancés qui utilisent MoReq2 pour le développement de solutions de gestion des documents électroniques. Les participants à la Conférence demandent que le DLM-Forum continue à élargir son ouverture à toutes les communautés concernées, et qu'il promeuve l'utilisation de MoReq2 auprès de ses voisins en Europe et ailleurs dans le monde.

La Conférence encourage le DLM-Forum à réfléchir à de nouvelles orientations, notamment :

---

<sup>4</sup> Recommandation du Conseil du 14 Novembre 2005 sur les actions prioritaires pour développer la coopération dans le domaine des archives en Europe (2005/835/EC), OJ, L312 of 29.11.2005.

<sup>5</sup> COM(2008)500; SEC(2008)2364.

<sup>6</sup> En russe et en coréen.

Prendre en considération les problématiques liées au transfert de données des organismes producteurs vers les services d'archives, en particulier en encourageant l'utilisation du schéma XML MoReq2 dans le cadre de projets pilotes pour l'interopérabilité entre des applications de gestion de documents et des applications d'archivage.

Evaluer l'impact pratique de nouvelles technologies logicielles sur les systèmes d'archivage.

Fournir des conseils pour le développement d'aptitudes et de compétences pour l'archivage électronique au sein des communautés d'utilisateurs à travers l'Europe.

Mettre en place des stratégies pour convaincre les décideurs, notamment en menant des études de coût.

Fournir aux membres du DLM un ensemble d'études de cas pratiques et de solutions concrètes.

La Conférence 2008 du DLM-Forum a mis en avant de nouvelles approches et solutions pratiques ainsi que des stratégies pour le futur. Le DLM-Forum doit continuer à proposer un calendrier de travail et à fournir des conseils pour développer des solutions à l'attention des usagers et des modèles de bonne pratique dans un souci de respect de l'intégrité et de l'authenticité du patrimoine numérique de notre société, pour aujourd'hui et demain.

# **FROM PROJECTS TO REALISATIONS, CONCLUSIONS: THE DLM-FORUM, MoREQ2 AND BEYOND...**

---

**Dr. Seamus Ross**

professor of Humanities Informatics and Digital Curation, University of Glasgow

[\*See the presentation.\*](#)

## DISCOURS DE CLÔTURE

---

Martine de Boisdeffre

Directrice des Archives de France

J'ai souhaité m'exprimer après la Commission européenne, car la V<sup>e</sup> Conférence du DLM-Forum est l'événement de clôture de la Présidence française en ce qui concerne les archives.

Je souhaite avant tout remercier tous ceux qui l'ont rendu possible, en tout premier lieu la Commission européenne, très bien représentée ici, ainsi que le DLM-Forum et son président, Toivo Jullinen. Le comité scientifique de la conférence, les intervenants et les modérateurs nous ont offert un programme professionnel de haut niveau, relayé par l'équipe des interprètes : qu'ils en soient tous vivement remerciés. L'organisation matérielle était à la charge de la direction des Archives de France : elle n'aurait pu réussir sans le savoir-faire de la société Carte blanche et l'appui des étudiants du Département Archives et Médiathèques de l'Université de Toulouse-le-Mirail. Christine Martinez, Jean-François Moufflet et Dimitri Sarris ont coordonné l'ensemble pour la direction des Archives de France, et je tiens à les en remercier.

Je veux également, après ces remerciements, souligner tout ce que retirera la Présidence française de cette conférence, et tout d'abord l'intérêt des échanges dans un cadre comparatif. Plus de cinquante intervenants venus d'Europe et d'ailleurs nous ont présenté de multiples expériences et points de vue qui reflètent tous, cependant, l'universalité des questions et des problèmes auxquels les professionnels, d'où qu'ils viennent, sont confrontés.

Intérêts et préoccupations communs, mais aussi convergence des approches -et c'est mon deuxième point- : qu'il s'agisse de la nécessité de travailler en amont, de la mise en place d'équipes mixtes d'archivistes et d'informaticiens, des questions de sécurité, de la mutualisation et des niveaux de celle-ci, ou des relations nouvelles avec les usagers.

Le troisième point que je souhaite souligner est celui de l'équation entre tradition et modernité : nombre de présentations de cette conférence ont rappelé que les concepts traditionnels restent valables et doivent continuer à inspirer notre travail et les choix de solutions, même si les modalités techniques et d'organisation sont différentes. Ces concepts peuvent être dans une certaine mesure repensés ou revus, mais les mécanismes intellectuels qui sous-tendent nos missions doivent être préservés.

Il faut également saluer la multiplicité des initiatives, des projets, des expériences lancés partout dans le monde et à tous les niveaux. Nous en avons eu ici un échantillon qui montre une formidable créativité. Ces initiatives et tentatives sont nécessaires, elles sont un pas indispensable dans la recherche de solutions concrètes.

Enfin, et c'est mon dernier point, je voudrais mettre l'accent sur la formation professionnelle, en liaison avec les remarques faites ci-dessus : elle doit tenir compte en effet des rôles respectifs des uns et des autres dans les processus de traitement des archives dans un environnement électronique, de la nécessité du dialogue interprofessionnel, de la capacité à manipuler ou à concevoir de nouveaux outils, de la relation nouvelle avec les usagers, etc. Formation qui peut d'ailleurs s'adresser à l'utilisateur lui-même dans un contexte différent d'accès à l'information.

La V<sup>e</sup> conférence du DLM-Forum aura constitué une étape essentielle dans notre réflexion et dans la recherche de solutions concrètes. Il reste cependant beaucoup à faire : nous avons commencé cette conférence avec La Fontaine, c'est avec La Fontaine que je souhaiterais la terminer, mais en citant cette fois la morale de la fable *Le lion et le rat* : « patience et longueur de temps font plus que force ni que rage ».

Je déclare close la V<sup>e</sup> conférence du DLM-Forum de Toulouse.

# Intervenants

---

---

# ***Speakers***

---

---

## ■ Johan Bengsston

Johan E. Bengtsson has 17 years of R&D work experience in the areas of Internet services, software technology and telecommunications. Extensive experience in international project management and R&D strategy. Proposal Editor and Advisor for the FP7 PROTAGE project. Acting Scientific Coordinator in the ambitious FP6 COGNOW. Technical Coordinator of the successful Esprit ROXY project 1998-2000; played an active role in starting Testbed Botnia, now the leading Living Lab Botnia. 7 years background as a private entrepreneur, developing software applications for the energy sector, and being a lecturer. From 1991 programmer, system engineer, international project leader, manager, at Telia Research. 2000-2003 Project Area Manager and Strategic Advisor. Since 2004 private entrepreneur and Senior Project Manager at CDT, Luleå University of Technology.

[johan@internit.se](mailto:johan@internit.se)

## ■ 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 oeuvre. 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.



## ■ Filip Boudrez

Filip Boudrez studied history at the university of Louvain. In 1997 he obtained the degree of Archivist and Records Manager and the same year he followed training in software engineering and computer programming. He worked from October 2000 until December 2003 in the DAVID-project. In the DAVID-project, he developed electronic record keeping strategies, practical guidelines and best practices for all kind of electronic records (e-mail, office documents, websites, databases, content management systems, GIS, etc). He creates tools for the implementation of the record keeping procedures for which he uses new technologies like XML and XSLT.

He currently works for the City Archives of Antwerp where he puts the (e)DAVID findings into practice. He continues the research on electronic record-keeping within the Expertisecentrum DAVID, the successor of the DAVID project.

## ■ Anne Burnel

Archiviste-paléographe, conservateur en chef du patrimoine, Anne Burnel a été responsable du pôle « communication des documents » au Centre d'accueil et de recherche des Archives nationales, puis chargée de mission auprès du directeur du Centre historique des Archives nationales.

Depuis 1997, elle dirige le Service national des archives du groupe La Poste. Le service oriente de plus en plus son action vers l'organisation, la collecte, la description et le traitement des archives courantes et intermédiaires pour répondre aux besoins des différentes entités du Groupe. Cette orientation de son activité le conduit notamment à développer des processus et des outils d'archivage électronique. Dans le cadre de la refonte globale de son SI, le SNA met ainsi en place une plateforme d'archivage électronique opérationnelle fin 2008. Ce projet s'appuie sur l'interfaçage progressif des applications métiers avec le système d'information Archives impliquant l'analyse en amont des processus de travail et du contenu et de la structure des données enregistrées, l'élaboration de profils de versement adaptés, la capture sélective de métadonnées et la conception de référentiels fiables et pérennes.

## ■ 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<sup>TM</sup>, 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, OAIIS et MoReq2.

## ■ Conni Christensen

Conni is a founding Partner and Director of Synercon Management Consulting. A New Zealand native, Conni moved to Australia in 1988. Her substantial industry and ECRM software product knowledge is founded upon more than 15 years of public and private sector experience as a records management consultant and trainer.

Conni is in high demand in Australia, New Zealand, Canada, and the United States as a trainer, conference and workshop presenter on records and information management issues. She contributes to a wide range of industry publications including InfoRMAA Quarterly and Image and Data Manager.

She is regularly consulted by industry and government on the change management issues relating to electronic records and document management. Conni is a member of ARMA International, and a former board member of the Records Management Association of Australasia.

Conni is also the designer and product manager for a.k.a.® records classification and retention software - a product now used by over 300 organizations world wide including the Australian National Archives, British Petroleum, DuPont, and British and American Tobacco. a.k.a.® was selected in 2005 by the QLD government for the whole of government development of thesauri and disposition schedules.

### Contact Details:

PO Box 952 Rozelle NSW 2039

Phone 02 9698 8888

Mobile 0438 849 319

Email: [conni@synercon.com.au](mailto:conni@synercon.com.au)

## ■ Saïd El-Morabiti

Saïd EL Morabiti, represents the PDF/A Competence Center ([www.pdfa.org](http://www.pdfa.org)) in France. He organized the first participation of the association ADDS (Association for Digital Document Standards) at a French Fair in March 2008 (Documation). Because of this successfull event, PDF/A CC board decided to take part of the FAN 2008 fair in October. Seven members from Germany and France were participating to the event. Active in promoting the PDF/A standard, Saïd EL Morabiti collaborates with Fedisa, in order to write a “ PDF/A whitepaper”.

This document is available free of charge for download at : <http://www.pdfa.fr>

## ■ Jean-Marc Fontaine

Ingénieur de recherche au Ministère de la Culture, effectue des études et recherches sur la conservation des enregistrements sonores et audiovisuels dans le cadre de l'Université Pierre-et-Marie Curie (Institut Jean le Rond d'Alembert - LAM), du Ministère de la Culture et du CNRS.

De 1980 à 2000, est en charge des programmes de recherche sur la conservation des documents sonores, audiovisuels et multimédia de la Bibliothèque nationale de France. Il étudie les conditions de préservation des différents supports d'enregistrements : disques noirs, bandes magnétiques, disques optiques pressés et enregistrables une fois,... Les études portent également sur l'extraction et la numérisation de l'information sonore issue des supports analogiques et enfin sur les outils de restauration des enregistrements.

2000-2005 : participe à l'élaboration des programmes de numérisation des documents audio du Ministère de la Culture et de la Communication et au contrôle de qualité des documents résultant des transferts.

Depuis 2005, poursuit ses travaux sur les thèmes mentionnés, développe notamment les travaux sur la restauration du son : évaluation cognitive et description acoustique et participe au renforcement des études sur les disques optiques dans le cadre d'un groupement d'intérêt scientifique GIS-DON

Contribue à l'élaboration de recommandations et de normes en matière de conservation des données numériques notamment au sein du comité technique de l'IASA (International Association of Sound and Audiovisual Archives) et de commissions nationales et internationales de normalisation : AFNOR, ISO, AES.

## ■ Marc Fresko

Marc Fresko is a highly experienced consultant, having provided consultancy and advice since 1979. Since 1992, he has specialised solely in electronic document management and electronic records management. He is recognised as an authority on electronic document and records management (EDRM); he has contributed significantly to the de facto standard for EDRM in the public sector, the “Functional Requirements for ERMS” published by The National Archives (then the Public Record Office); and was the project leader and architect of the European Commission’s “MoReq” (Model Requirements for ERMS - see <http://www.cornwell.co.uk/moreq.html>) and MoReq2 (<http://www.moreq2.eu>) specifications. As Serco’s EDM & ERM Consulting Services Director, he now leads a team of some 30 consultants in delivering expert advice to many public and private sector organisations, including financial sector companies, government departments, local authorities and others. Associated with his expertise in EDM and ERM, Marc has specialist knowledge about, and a particular interest in, Digital Preservation - namely the actions necessary to ensure that electronic information remains accessible in the medium and long term - a topic which clearly is increasing in importance and immediacy. After graduating from Oxford University in 1976, Marc spent a short spell in manufacturing industry. From 1979 to 1987 he was with Arthur Andersen & Co. (now Accenture), from 1984 to 1986 working in the U.S.A., based in Chicago, IL. on a wide range of assignments including the implementation of office systems. He worked for KPMG Management Consulting from 1988 to 1992, where he initiated, launched and coordinated KPMG’s marketing initiative in electronic document management. He first worked with Cornwell in 1992, and was appointed EDM & ERM Consulting Services Director in April 2001. Cornwell was acquired by Serco in 2007. His work both for clients and in standards setting has earned him an international reputation, and he is frequently invited to speak at or chair industry conferences and seminars on EDRM, as well as being widely published in a variety of media and publications.

He is now Director and founder of Inforesight Limited, an independent consultancy for advice and services in Information management.

## ■ Pavel Golob

Pavel has more than 20 years of experience in the IT systems implementation, audit and ICT security implementation. His current position is at Eurojust where he is Head of the Security, Facility Management, General Services and Events Unit ND ICT Security Officer and Risk Manager. With regard to the ICT security, Pavel is responsible for:

Develops, implements security standards, procedures and guidelines for multiple platforms and diverse systems environments;

Reviews the development, testing and implementation of security plans, products and controls techniques;

Identifies and assesses ICT security risk/ exposure on new and existing infrastructure, investigates and recommends appropriate corrective actions for ICT security incidents;

Develops and maintains security policies, controls and their compliance;

Analyzes security incidents and escalation of security events;

Liaises with Eurojust post holders with regards to information security incidents;

Develops cutting-edge, innovative solutions for ICT network security;

Studies the proliferation of viruses; monitors developments in hacker intrusion methods;

Conducts active penetration tests; discovers vulnerabilities in information systems.

Pavel poses also vast experience in audit of information systems, information and communication security and internal audits. When he was employed at Slovenian Budget Supervision Office, he was responsible for Cohesion EU fund audits of spending centres participating in the processing and use of the EU Cohesion fund and the national co-financing, ICT security audit and assessment of information security maturity within the agency based on the ISO -IEC 17799:2002 standard.

Pavel was also involved in consulting in the field of Information security controls implementation, control assurance reviews of information security control environment, IT security audits, Control assurance audits, due diligence reviews, consulting in IT systems implementation and procurement.

Pavel Golob,

Acting Head of the Security, Facility Management, General Services and Events Unit

EUROJUST

Maanweg 174

2516 AB The Hague

The Netherlands

Office Phone +31 70 412 5731

Mobile +31 646 203 755

Fax +31 70 412 5745



## ■ **Carole Gragez**

Conservateur du patrimoine (archiviste).

Responsable du service des archives électroniques aux Archives nationales (juillet 2005 - juillet 2008).

Intervenante en formation dans le domaine de l'archivage électronique (Centre National de la Fonction Publique Territoriale, Direction des Archives de France).

Actuellement responsable de l'antenne Marine (Toulon) du Service historique des armées (SHD).

## ■ Céline Guyon

Maîtrise d'histoire moderne et DESS information-communication, option archives

Attaché de conservation du patrimoine

A exercé aux Archives municipales d'Aulnay-sous-Bois et aux Archives départementales de l'Aisne

Adjointe du directeur des Archives départementales de l'Aube

Co-animatrice du groupe de travail national sur la collecte des données numériques produites dans le cadre des missions d'aide sociale des Départements (l'exemple de l'aide sociale à l'enfance)

Intervenante sur la thématique de l'archivage électronique dans le cadre de la formation continue des archivistes (DAF, AAF) et lors de journées d'études organisées par le réseau des archives (colloque national de Clermont-Ferrand, journée d'études de la région Grand Sud-Ouest, etc.) ou la Direction générale de la modernisation de l'Etat (matinées Adèle)

## ■ Dr Henk Harmsen

Henk Harmsen (1958) studied computer linguistics at the University of Amsterdam (UvA). He wrote a PH.D. at the Vrije Universiteit in Amsterdam (VU) about automatic syntactical- and semantic text-parsing. He was head of the IT department (1993-1995) and interim manager (1996-1997) of an institute of the Netherlands Royal Academy of Arts and Sciences (KNAW). In 1998 he went to the UvA as a librarian and operational manager at the Faculty of Science. In 2000 he returned to the KNAW to become operational director at the Netherlands Institute of Scientific Information (NIWI-KNAW). He has lead a few reorganizations and he is the primary 'booster' of various innovative IT-projects.

Henk is also a musician and a 'self-made' expert on digital audio recording.

Dr. Henk Harmsen

Deputy Director of DANS

[henk.harmse@dans.knaw.nl](mailto:henk.harmse@dans.knaw.nl)

## ■ Tatjana Hatjnik

Tatjana Hajtnik, MA has over 22 years of experience in the field of information science. Her expertise experience evolved over the years through her work in business companies, public sector and state administration alike. She began her professional career as a mainframe system engineer, and continued her work at the Health Insurance Institute of Slovenia as an active participant in the process of information system reorganisation in OE Krško. She was one of contributing participants in the pilot project for health insurance card in OE Krško, where she was in person in charge of adjusting infrastructure for health insurance card operations in health centres. In the years between 1999 and 2006 she has been intensely involved in dealing with the issue of information security in state administration (Head of Security Service at the Government Centre for Informatics) and has been the one who encouraged implementation of recent standards in the field of information security. In October 2006 she became the Head of Division on electronic archives and computer support at the Archives of the Republic of Slovenia, devoting most of her time to implement provisions of PDAAIA (Protection of Archives and Documents and Archival Institutions Act) in relation to e-archiving of documents and archives, as well as to establish e-archives.

Tatjana Hajtnik, MA

Secretary - Head of the Division for e-archiving

Archives of the Republic of Slovenia

Zvezdarska 1, SI-1000 Ljubljana, Slovenia

tel.: +386 1 241 42 24

E-mail: [tatjana.hajtnik@gov.si](mailto:tatjana.hajtnik@gov.si)

## ■ Hans Hofman

Hans Hofman is senior advisor on digital longevity at the Nationaal Archief of the Netherlands. He represents the Nationaal Archief in Planets research project ([www.planets-project.eu](http://www.planets-project.eu)) and the Digital Preservation Europe coordination action ([www.digitalpreservationeurope.eu](http://www.digitalpreservationeurope.eu)) and he is since 2000 representing the Netherlands in ISO TC46/SC11 on Records Management, in which committee he is chairing the Working Group on RM metadata. He has acted as co-director of ERPANET (2001-2004, [www.erpanet.org](http://www.erpanet.org)) and was co-investigator in InterPares project (1999-2006).

## ■ Michael Hollman

Michael Hollmann is Director of the modern archives of the Federal Archives. Previously, he was head of the image archives of the Federal Archives of the Federal Republic of Germany. In this role, he is responsible for the digitization and web-presentation of one of the most significant archives for historical photographs in Europe. During the last fifteen years he was involved in several IT-development projects. Since 2000, he has been a member of the ICA Committee on Information Technologies.

Dr. Michael Hollmann, Bundesarchiv, D-56064 Koblenz, Tel.: 0049261505300,

Email: [m.hollmann@barch.bund.de](mailto:m.hollmann@barch.bund.de)

## ■ Perla Innocenti

Perla Innocenti is Co-Principal Investigator for Requirements Analysis and Identification of User Scenarios in the project Sustaining Heritage Access through Multivalent Archiving (SHAMAN). She is also involved in repository design, audit research (e.g. DRAMBORA) as part of DigitalPreservationEurope (DPE) and has contributed to usage models research within the project Preservation and Long-term Access through NETWORKED Services (Planets). Perla has a background as a researcher specialising in information systems for industrial design and as a digital librarian at Politecnico di Milano, Italy, where she coordinated digital library, digitization and library portal projects. She has worked as a consultant and collaborator in digital libraries and e-learning projects.

## ■ Richard Jeffrey-Cook

Richard Jeffrey-Cook, CITP, is a Director In-Form Consult Ltd. Richard has over 20 years of implementing IT solutions to both public sector and private sector organisations and has supplied information management consultancy to, and implemented systems on behalf of, organisations including the European Central Bank, The UK National Archives and the Office of Government Commerce, Invest Northern Ireland, Nursing & Midwifery Council, Honda UK, local authorities, NHS and emergency services. His articles have appeared in RM bulletin and other leading magazines.

In-Form Consult Ltd is a leading European Information Management Consultancy helping organisations to introduce better information management practices and implement electronic document, content, knowledge and record management systems.

Richard Jeffrey-Cook

Managing Director

In-Form Consult Ltd

West Clayton Business Centre

CHORLEYWOOD

WD3 5EX

Tel: 07810 422879

Email: [richard.jeffrey-cook@inform-consult.com](mailto:richard.jeffrey-cook@inform-consult.com)



## ■ Dr. Ulrich Kampffmeyer

Dr. Ulrich Kampffmeyer has been working in the field of document, records and archive management since 1978. He is the managing director of the consulting company PROJECT CONSULT in Germany which focuses on Enterprise Content Management and Document related Technologies. Since 1998 he has been active for the DLM Forum as keynote speaker, editor of the AIIIM/DLM whitepapers, co-founder and manager of the DLM Network EEIG, and member of different DLM committees. Dr. Kampffmeyer was involved in the development of the ISO standards 15489 and 19005 and worked as a member of the Editorial Board for the MoReq2 Requirements and the MoReq2 Test Framework. He is author of several books on Electronic Archival, Document Management and Enterprise Content Management.

References for DLM speaking engagements:

Keynote DLM Forum 1999, Brussels

Keynote DLM Forum 2002, Barcelona

Books and DLM publications:

<http://www.project-consult.net/portal.asp?UR=26>

MoReq2 activities:

<http://www.MoReq2.de>

Articles:

<http://www.project-consult.net/portal.asp?UR=24>

Selection of presentations (not current):

<http://www.project-consult.net/portal.asp?sr=500>

Newsletter contributions:

<http://www.project-consult.net/portal.asp?SR=221>

## ■ Livii Karpistsenko

Liivi Karpistsenko works as an advisor at the Department of Records Management, the State Chancellery of the Republic of Estonia. The department is responsible for the development of records management and archiving in the public sector of Estonia, and Liivi's area of responsibility is electronic records management. She is involved in various national projects on standardising XML documents and metadata packages to enhance interoperability between different records management systems, and she works closely with the Document Exchange Centre to promote and simplify paperless document exchange between and with public sector agencies. Liivi also has long-time experience with electronic databases and services, having been actively involved in the process of transition to "paperless office" at pension departments and the Social Insurance Board of Estonia.

[Liivi.Karpistsenko@riigikantselei.ee](mailto:Liivi.Karpistsenko@riigikantselei.ee)

State Chancellery of the Republic of Estonia

Rahukohtu 3

15161 Tallinn, Estonia

## ■ Andrew McHugh

**Andrew McHugh** earned a degree in Scots Law from Glasgow University (2000) and went on to complete his MSc in Information Technology (2001). Since then he has been employed within HATII (the Humanities Advanced Technology and Information Institute at the University of Glasgow) in various capacities including taking responsibility for revolutionizing the information infrastructure in the Department of Music. In late 2004 he joined the Digital Curation Centre in the position of Advisory Services Manager, leading a world-class team of digital curation practitioners in offering leading-edge expertise and insight in a range of issues to a primarily higher and further education audience. His most recent work at the DCC has involved leading its work in trusted repository Audit and Certification. In Spring 2008 he joined the PLANETS project, researching intelligent object technologies to support information preservation. He also lectures on multimedia systems and design on the MSc in Information Technology run by the Computing Science Department at Glasgow.

## ■ Philippe Martin

Philippe Martin, s'exprimant au nom de l'Aproged.

Philippe Martin est directeur associé du Bureau van Dijk où il est entré en 1982. Spécialiste de la GED et de l'archivage électronique, il a mené de nombreux projets dans ce domaine, en particulier pour les Archives Nationales de France et pour plusieurs services d'archives départementales ainsi que pour des entreprises du secteur privé.

Philippe Martin est administrateur de l'Aproged où il anime le Pôle Normalisation. Dans ce cadre, il participe activement aux travaux de révision de la norme française Z 42013. Il anime également un groupe de travail international au sein de l'ISO TC 171 pour la préparation d'une norme de spécification d'un système d'archivage électronique.

Philippe Martin anime de nombreux séminaires sur le thème de la dématérialisation et notamment les séminaires État de l'art de Caggémini sur la GED et l'archivage électronique.

Contact : téléphone 01 45 24 49 10, courriel : [phm@bvdic.com](mailto:phm@bvdic.com)

## ■ Osmo Palonen

Osmo Palonen is Project Manager in MiUAS with responsibilities to lead the digital archiving projects and services. Palonen has been working for MiUAS from 2003 and has background as journalist, IT and industrial automation. Palonen graduated BA from University of Tampere in 2008.

Address: Mikkeli UAS, P.O. Box 181, Patteristonkatu 2, FIN-50101 Mikkeli, Finland

E-mail address: [osmo.palonen@mamk.fi](mailto:osmo.palonen@mamk.fi)

## ■ Jean-Louis Pascon

Ingénieur diplômé de l'Ecole Nationale Supérieure de Chimie et Physique de Bordeaux en 1979, *Jean-Louis Pascon* a plus de 20 années d'expérience en Gestion Electronique de Documents.

Il a notamment mis en place l'une des premières bases de données d'images numériques en Europe pour l'ouverture du musée d'Orsay.

Aujourd'hui, consultant au sein d'Hénon Conseil, il assure des audits en matière d'archivage et de dématérialisation pour des sociétés ou des administrations. Par ailleurs, il conseille des entreprises qui souhaitent mettre en place des systèmes d'archivage électronique.

Il a été responsable de groupes de travail au sein de l'AFNOR et de chef de projet de plusieurs normes, aussi bien au plan national qu'international dans le domaine de l'archivage électronique.

Il est également expert auprès de la Direction Générale Société de l'Information (DG INFSO) de l'Union Européenne dans le domaine des Nouvelles Technologies de l'Information.

Il est membre du conseil d'administration du COREF et trésorier de cette organisation.

Il est membre fondateur de FEDISA ([www.fedisa.eu](http://www.fedisa.eu)) association professionnelle regroupant fournisseurs et utilisateurs d'outils de stockage et d'archivage.

Jean-Louis Pascon

### Adresse postale :

Hénon Conseil  
57 Rue Hénon  
69004 Lyon (France)

Adresse électronique : [jean-louis.pascon@wanadoo.fr](mailto:jean-louis.pascon@wanadoo.fr)

Téléphone : +33 6 27 65 05 75

## ■ Jacques Perdereau

Responsable de la Recherche et Développement sur les produits multimédia au Laboratoire National de Métrologie et d'Essais (LNE). Ces recherches sont principalement axées sur l'étude des performances des systèmes audio vidéo, de traitement automatique de la parole et d'enregistrement de l'information numérique (papier et disques optiques).

Dans le domaine audio vidéo, depuis les années 80 sont étudiées des méthodes de caractérisation des systèmes d'acquisition, d'enregistrement et de restitution des images et du son. Ces méthodes concernent aussi bien les systèmes analogiques que numériques. Les travaux ont permis de contribuer à l'élaboration des normes AFNOR et CEI.

Dans le domaine de l'enregistrement de l'information, depuis 1990 les études portent principalement sur la qualité et la pérennité de l'information enregistrée sur les supports papier utilisés pour la bureautique et la photo et sur les disques optiques Worm (CD et DVD).

Les études sur les disques optiques Worm ont été menées pour le Ministère de la Culture et de la Communication, le Ministère de l'Industrie ainsi que dans le cadre du groupement d'intérêt scientifique GIS-DON. Elles permettent de contribuer à l'élaboration des normes sur la conservation des données numériques dans le cadre des commissions de normalisation AFNOR et ISO.

## ■ **Bodgan-Florin Popovici**

Bogdan-Florin Popovici graduated from the Faculty of Archival Studies (1999) in Bucharest, and was awarded a Ph.D. in History in 2004. He is currently working as an archivist at the Brasov County Archives in Brasov, Romania. He is also a part time lecturer at the University of Bucharest, lecturing at Master level on Records and Archives management. He was the main coordinator of the translation of the ICA standards into the Romanian language. Currently, he is working with Lucia Stefan on the Romanian translation of Moreq 2.



## ■ **Nathalie Regagnon**

Nathalie Regagnon est responsable du service des relations avec les administrations aux Archives départementales de la Haute-Garonne (France) depuis le 1er janvier 2008. Son équipe est chargée du contrôle et de la collecte des archives publiques (tous supports) produites par les services du conseil général, les services déconcentrés de l'Etat, les établissements publics et les organismes privés chargés de mission de service public ayant leur siège dans le département. Avant 2008 -et depuis le 1er novembre 2004- Nathalie était chargée de mission pour les archives électroniques aux AD31 : elle traitait de la collecte des archives électroniques tous secteurs confondus, de la collecte des archives de la préfecture, préparait les marchés de numérisation des archives et était chargée de développer le nouveau site Internet. Elle a obtenu son diplôme de maîtrise IUP Ingénierie documentaire (spécialité archives) et le titre d'ingénieur-maître en 2004 ; auparavant, elle a effectué plusieurs stages et vacations aux Archives départementales du Nord, au Centre des archives du monde du Travail (Roubaix) et au laboratoire régional des Ponts et Chaussées de Toulouse (bureau des archives intermédiaires). C'est suite à l'obtention du concours d'attaché de conservation du patrimoine en 2004 qu'elle est entrée aux Archives départementales de la Haute-Garonne.

## ■ Kadi Riisma

PhD, head of the Records Management Department of the State Chancellery of the Republic of Estonia and the lecturer of Tallinn University. The main responsibility is to coordinate development of records management and archives in the public sector.

[Kadi.Riismaa@riigikantselei.ee](mailto:Kadi.Riismaa@riigikantselei.ee)

State Chancellery of the Republic of Estonia

Rahukohtu 3

15161 Tallinn, Estonia

## ■ Dr. Seamus Ross

Dr Seamus Ross, Professor of Humanities Informatics and Digital Curation, is founding Director of the Humanities Advanced Technology and Information Institute (HATII).

His research focuses on digital preservation including work on preservation, repository design, digital library design and services, ingest, and semantic metadata extraction.

He is also Associate Director of the Digital Curation Centre in the UK (<http://www.dcc.ac.uk>, since 2004), Principal Director of Digital Preservation Europe (DPE) (<http://www.digitalpreservationeurope.eu>) a project partner and member of the management boards of Cultural, Artistic and Scientific knowledge for Preservation, Access and Retrieval (CASPAR, <http://www.casparpreserves.eu>) and Preservation and Long-term Access through NETworked Services (Planets, <http://www.planets-project.eu>).

He was a co-principal investigator in the DELOS Digital Libraries Network of Excellence (<http://www.dpc.delos.ac.uk>) (2002-8) and Principal Investigator of the AHDS-Performing Arts (2005-8), both of which projects made contributions to preservation of digital materials.

He was Principal Director of ERPANET a European Commission activity to enhance the preservation of cultural heritage and scientific digital objects (<http://www.erpanet.org>), and a key player in The Digital Culture Forum (DigiCULT Forum) which worked to improve the take-up of cutting edge research and technology by the cultural heritage sector (<http://www.digicult.info>).

He has more than twenty years experience as a practitioner and researcher in ICT and has for fifteen years been investigating issues associated with digital preservation and curation; Before joining the University of Glasgow he was Head of ICT at the British Academy and a technologist at a company specialising in knowledge engineering.

He co-Chaired with Prof Margaret Hedstrom (Uni of Michigan) the EU(DELOS)/NSF Workgroup on Digital Preservation and Archiving which proposed an innovative research agenda in its report /Invest to Save: Report and Recommendations of the NSF-DELOS Working Group on Digital Archiving and Preservation / (2003).

He earned a doctorate from the University of Oxford. Some of his publications are available at <http://eprints.erpanet.org>. From 1 January 2009 he will take up the post of Dean of the Faculty of Information at the University of Toronto.

## ■ 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)

## ■ Lucia Stefan

Lucia Stefan was born and educated in Bucharest, Romania. She studied Chemical Engineering to postgraduate level at the Technical University of Bucharest, but the collapse of communism put an end to her short career as an engineer. In 1992 she moved to London and studied Information Science to Master degree level (MSc Computing and Information Science, 1998). Her interest in electronic records management started after an EDRM system implementation and continued after being involved in EDRM certification testing at the National Archives of UK. Now working as a consultant, Lucia has been involved in projects in UK, Italy (the Joint Research Centre of the European Commission) and Ireland.

## ■ Remco Verdegem

Remco Verdegem began his professional career in the area of information technology in 1989. In October 1998, he joined the Dutch State Archives' Service, where he was among other things responsible for the functional maintenance of the archival system for paper records. From October 2000 till July 2003 he was the project manager of the Digital Preservation Testbed project, which was sponsored by the Dutch State Archive Service and the Ministry of the Interior and Kingdom Relations. Since April 2005 Remco is working as Senior Advisor Digital Longevity at the Nationaal Archief of the Netherlands.

Remco Verdegem

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 5543

[remco.verdegem@nationaalarchief.nl](mailto:remco.verdegem@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)

## ■ Kirsten Villadsen Kristmar

Kirsten Villadsen Kristmar is head of the Department of Appraisal and Transfer at The Danish National Archives. She has studied history and informational science at the University of Aarhus. At The Danish National Archives she is responsible for the transfer and long time preservation of digital records, created by Danish public authorities. In this capacity, she has played an active role in the development of the Danish strategy for long time preservation of digital records which is based on conversion. The implementation of the strategy has included transferral of several thousand system independent extracts from the Danish public authorities, migration of data from obsolete media to modern media, conversion of obsolete data formats and data structures to the present preservation standard and the creation of tools that makes it possible for the users of The Danish State Archives to gain access to data.

She has also participated in the development of Daisy, an archival database containing information on the holdings of The Danish State Archives.

Mrs. Kirsten Villadsen Kristmar

Head of the Department of Appraisal and Transfer

The Danish National Archives

Rigsdagsgården 9

1218 København K

Tel. +45 33 92 23 10

E-mail: [kvk@ra.sa.dk](mailto:kvk@ra.sa.dk)

## ■ Dr. Uwe Wächter

Dr. rer. nat. Uwe Wächter studied Physics at the Technical University of Rostock and finished with his PhD in 1990. From 1990 to 1996 he was working as external consultant in the software development department from Siemens PG.

Since 1996 Dr. Uwe Wächter is working for SEAL Systems AG. SEAL Systems is the international leading developer of information and document distribution solutions. Dr. Wächter is responsible for the specification of many document distribution and output systems of german customers in the branches automotive and supplier, plant and mechanical engineering.

As product manager for the PDF solutions he was involved into PDF/A long before the official norm was published. Dr. Wächter is an admitted expert for PDF/A. As an active member of the PDF/A Competence Center he is doing presentations about PDF/A since many years.



## ■ Gérard Weisz

Consultant en ingénierie documentaire depuis plus de quinze ans au sein de différentes organisations Gérard WEISZ dirige aujourd'hui Sirius Systems France, société spécialisée dans la dématérialisation des documents au travers de l'étude, du conseil et de l'assistance à la mise en oeuvre de systèmes de GEIDE et d'archivage électronique à valeur probante. Il intervient également en tant qu'expert dans le cadre d'audit de conformité des systèmes d'archivage à la norme AFNOR NFZ42-013.

Membre de l'APROGED, il participe aux travaux de la commission de normalisation CN 171 de l'AFNOR et intervient au sein de la délégation française au comité ISO TC171 dont les travaux sont consacrés aux domaines de l'imagerie et du document électronique. A ce titre il a conduit en tant qu'éditeur de projet les travaux de révision 2008 de la norme NFZ42-013.

Ayant élaboré une approche méthodologique juridique et technique des systèmes d'archivage électronique, il est co-auteur d'une étude sur le principe de précaution en matière d'écrit électronique avec Sabine LIPOVETSKY (Cabinet d'avocats Kahn & Associés).

En 1998 et 2002, la certification CDIA+ (Certified Document Imaging Architect) lui a été attribuée par la Computing Technology Industry Association (CompTIA - USA) et un « International Award » lui a été décerné lors de l'AIM Show 2000 à New York.

Gérard Weisz est membre fondateur et ancien secrétaire général de la Fédération Nationale des Tiers de Confiance (F.N.T.C.).