

Bulletin d'information sur l'archivage des documents électroniques n° 15, juin 2004

Des changements au département de l'innovation et de la normalisation

Françoise Banat-Berger a pris, le 17 mars dernier, ses fonctions de chef du département de l'innovation technologique et de la normalisation, en remplacement de Catherine Dhérent.

Olivier de Solan, arrivé le 1^{er} avril dernier au bureau du traitement des archives et de l'informatisation, reprend les dossiers suivis par Joël Poivre.

Claire Sibille, qui devient chef du bureau du traitement des archives et de l'informatisation, reste chargée du visa des instruments de recherche et du suivi des normes de description archivistique (ISAD/G, EAD, EAC, etc.).

Rappelons les noms des autres membres du département : Henri-Luc Camplo, responsable de la cellule technologique, Elsa Marguin-Hamon, chef du bureau de la conservation matérielle et des constructions, France Saie-Belaisch, chargée des questions de bâtiments dans ce même bureau.

Edmond Fernandez, affecté au Centre des archives d'outre-mer, réalise pour la DAF des missions de conseil technique dans le domaine du microfilmage et de la numérisation.

Annie-Claude de François assure le secrétariat du département.

Leurs coordonnées peuvent être retrouvées sur le site de la DAF :
<http://www.archivesdefrance.culture.gouv.fr/fr/annuaire/index.html>.

Le programme de travail de la direction des archives de France sur l'archivage électronique en 2004

Parmi les nombreux chantiers ouverts sur l'archivage électronique, deux urgences sont apparues plus particulièrement pour les prochains mois :

- Etablir des recommandations sur la gravure, le stockage et l'évaluation des CD-R.

Alors que les services d'archives ont produit, lors d'opérations de numérisation, de très nombreux CD-R, et au moment où de nombreux services administratifs gravent des données sur CD-R en vue de leur archivage, il est apparu urgent de diffuser des conseils sur la gravure, le stockage et l'évaluation des CD-R. A quoi servirait-il, en effet, de réfléchir à l'archivage des données électroniques (migrations de supports, de formats, question de l'intégrité, de l'organisation territoriale...) si ces données sont perdues ou endommagées à l'origine ?

- Accompagner les premières expériences de dématérialisation de procédures administratives.

Dans le cadre du plan stratégique pour l'administration électronique (programme ADELE), plusieurs procédures administratives sont en cours de dématérialisation ou le seront prochainement. C'est le cas du contrôle de légalité des actes des collectivités territoriales, des marchés publics, etc.

Il est indispensable que la direction des Archives de France puisse se faire entendre dès la conception de ces projets, afin que l'archivage soit pris en compte. D'autant que les procédures dématérialisées peuvent concerner toutes les collectivités territoriales, donc toutes les archives territoriales. Les expérimentations menées à cette occasion, en partenariat avec le Conseil général des Yvelines par exemple, devraient permettre de définir, à terme, un modèle général de versement d'archives électroniques, avec des exigences précises en matière de procédures, de métadonnées, de formats de fichiers...

Dans un domaine voisin, il est prévu de diffuser des conseils sur la numérisation de documents (cahiers des charges, contrôle des prestations, stockage et diffusion des images, etc.).

La direction des archives de France reste évidemment à votre disposition pour prendre en compte toutes vos questions.

Archivage numérique : organisation, retour d'expérience (par Claude Huc, Centre national d'études spatiales, Service valorisation des données et outils d'ingénierie)

Introduction

Le développement de l'information numérique pose un premier problème qui est celui de la réduction des échelles de temps : ainsi, pour les documents bureautiques, la **chaîne de compatibilité est souvent rompue en moins de 10 ans** : cela signifie que des documents produits avec un outil de traitement de texte en 1993, peuvent ne plus être lisibles aujourd'hui ! Or, ce mouvement ne fait que s'accélérer (au moins 5 versions de MS Word sous Windows de 1995 à aujourd'hui).

On constate une très grande diversité des objets d'information numérique : toute l'information est créée, manipulée, stockée et souvent diffusée sous forme numérique (bureautique, image (photographie, images de synthèse), son, vidéo, observations scientifiques, documents multimédia...

Cette nouvelle forme de représentation de l'information est à la fois incroyablement puissante et vulnérable au temps, mais c'est cette vulnérabilité qui constitue le cœur du problème.

Par ailleurs, l'information numérique est immatérielle

● **sa perte ou son altération ne se voit que si on en a besoin**

● **la non prise en charge du besoin de pérennité conduit inéluctablement à la perte de l'information**

Aujourd'hui, pratiquement tout le monde est concerné

administration : état civil...

secteur de la santé,

caisses de retraite,

l'industrie : pétrole, aéronautique...

la recherche scientifique, le domaine spatial

la Défense,

le nucléaire,

et aussi les particuliers

Les causes de cette vulnérabilité sont multiples :

Des facteurs techniques : obsolescence des technologies de stockage, des logiciels et des systèmes, dépendances entre données créées et l'environnement de création, données ou documents non décrits.

Des facteurs organisationnels : la pérennisation de l'information constitue une activité en soi. l'organisation du travail, le partage des responsabilités, la mise en place des bonnes compétences au bon endroit sont à repenser
Des facteurs normatifs, juridiques, industriels, psychologiques, liés à l'absence de formation...

1- La diversité des besoins : par exemple pour les documents textuels

Ils peuvent être par exemple de pérenniser l'apparence visuelle et le contenu du document, dans un but juridique, ou de mémoire ou de témoignage ; ou encore de pérenniser le contenu : certains documents n'évoluent plus ; d'autres au contraire peuvent pouvoir être modifiés, enrichis, corrigés ; ou encore pour certains documents, on peut vouloir faire évoluer et diversifier l'apparence visuelle.

Dans tous les cas, il convient de pérenniser les métadonnées : afin de pouvoir retrouver le document au sein d'un ensemble, avec la possibilité d'extraire automatiquement une partie des métadonnées à partir du document, avec la possibilité de pouvoir s'adapter dans le futur à des techniques de recherche d'information de plus en plus performantes, et de pouvoir diffuser le document au sein de la communauté des utilisateurs.

2- Comprendre le problème posé pour le résoudre

C'est l'objet de la norme ISO 14721:2002 'Reference Model for an Open Archival Information System (OAIS)' Issue 1. January 2002

<http://www.ccsds.org/CCSDS/recommandereports.jsp#interchange>

Référence CCSDS : CCSDS 650.0-B-1 (gratuit)

On peut également se procurer ce même document auprès de l'ISO (<http://www.iso.org/>) pour ISO 206 francs suisses.

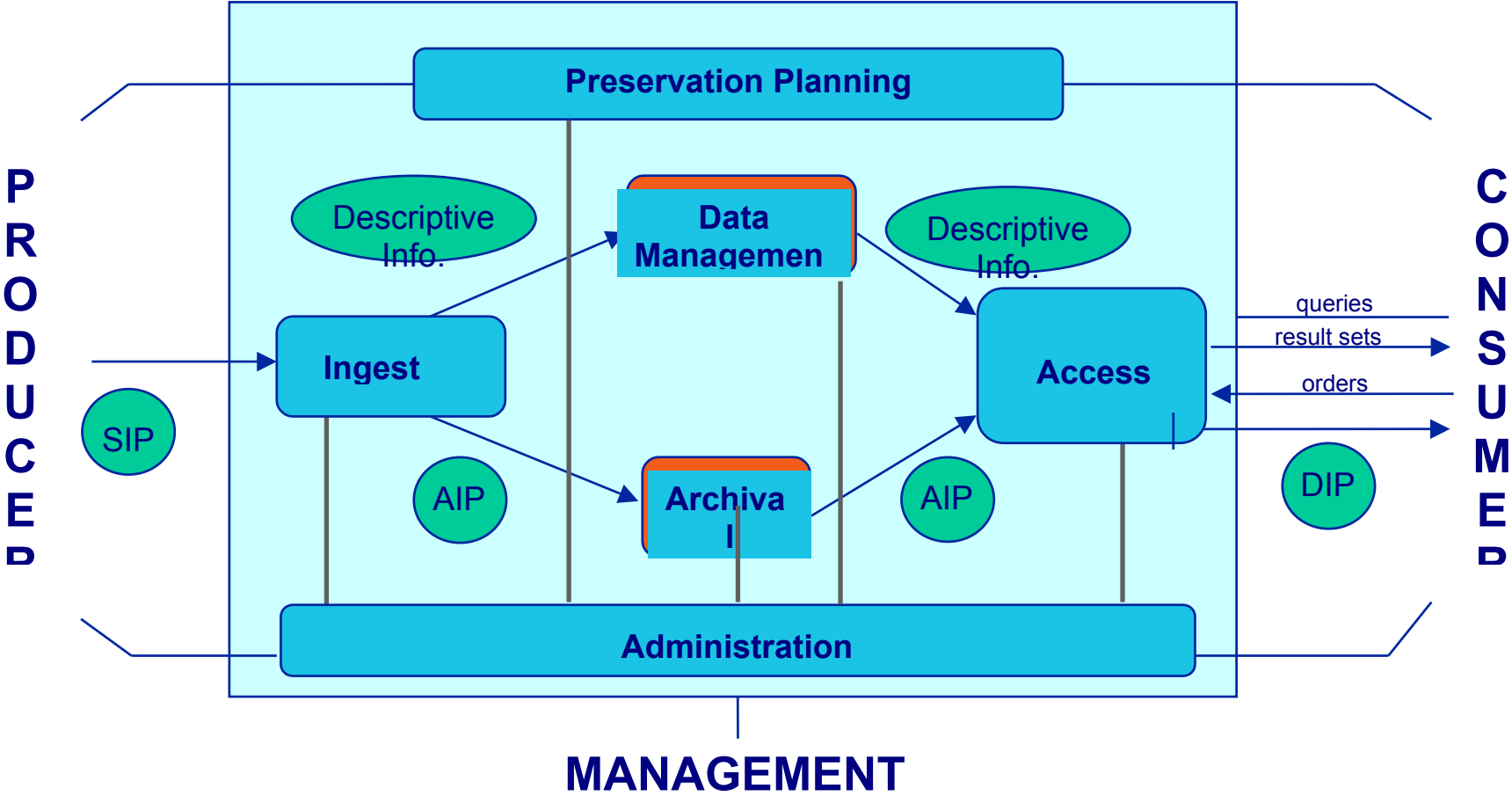
Il s'agit d'une analyse détaillée, d'une définition des concepts, d'un modèle fonctionnel et d'un modèle d'information pour comprendre l'ensemble des spécificités de l'archivage de l'information sous forme numérique

Qu'est ce qu'une archive ?

C'est une organisation dont la vocation est de préserver l'information pour permettre à une Communauté Définie d'Utilisateurs d'y accéder et de l'utiliser, ce qui sous-entend : la pérennité des données, un accès pérenne aux données ainsi que la préservation avec les données, de toutes les informations nécessaires à leur compréhension et utilisation

Par conséquent, l'archivage n'est ni une sauvegarde, ni un backup système, ni un rangement définitif des données quand on pense qu'elles ne serviront plus.

OAIS : entités fonctionnelles



SIP = Submission Information Package
AIP = Archival Information Package
DIP = Dissemination Information Package

Une politique en matière de préservation de données numériques implique par conséquent de mettre en place trois services qui doivent être coordonnés :

- un service s'occupant de la collecte
- un service s'occupant du stockage des données
- un service s'occupant de leur accès.

Le coordonnateur est le chef d'orchestre et le véritable responsable de l'archivage (au sens OAIS) : suivant le contexte, on parlera de gestionnaire des données, de gestionnaire du patrimoine technique, d'archiviste, d'archiviste principal,...

Son rôle est d'organiser le partage du travail entre les différents services pour assurer la clarté des interfaces entre ces services et coordonner le travail pour les domaines de compétence communs : le modèle d'information, le dictionnaire des objets numériques livrables.

3- Collecte - Préparation des données et métadonnées :

Fonctions

En entrée, nous avons les données et métadonnées transmises par le(s) producteurs (services versants).

En sortie, les fichiers à préserver ont été archivés dans une arborescence mise en place dans le système de stockage d'archive en charge de la pérennisation de ces fichiers . Quant aux métadonnées, elles ont été également transmises au service de stockage et en parallèle livrées au service 'gestion des données et accès des utilisateurs'.

Le service 'Préparation des données...'prend en charge l'ensemble des tâches identifiées dans le projet de norme ISO 'Producer-Archive interface abstract standard' et négocie sur ce que le producteur peut et ne peut pas faire, opère les transformations nécessaires sur les données et métadonnées restant à la charge du service d'archive

Les données numériques (documents issus de la bureautique, observations scientifiques, images...) doivent être :

- sous une forme (un format)de préférence normalisé,
- indépendantes des logiciels mis en œuvre pour les créer,
- décrites (syntaxe et sémantique) de façon exhaustive.

Les métadonnées doivent être normalisées

Nous avons rejeté les voies s'appuyant sur des migrations de format régulières et incertaines des données.

Des exemples de formats de données

Un objet numérique quelconque, composé de 0 et de 1 :

```
0010000000100000001000000011000100110010001100000010000000100000001000000011
010100111000001000000010000000110101001000000010000000100000001100010011100100111000
001100010010110100110000001110010010110100110011001100000101010000110001001101000011
101000110010001100100011101000110000001100010010111000110110001100100011100101011010
00100000001000000010000000110011001110000011100100111000001011100011011001100000010
000000100000001000000011001000110011001101010011001000101110001101000011010100100000
001000000010000000100000001000000010110100110001001011100011011000110110001000000010
0000001000000010110100110111001011100110011001000100000001000000101101001100010
011100001011100011011000110010010000001000000010110100111000001100000101100011000
```

000110000001000000010000000101101001110000011000000101110001100000011000000100000001
000000010000000110001001000000010000000110100001100110010000000100000001000000011100
100100000001000000010000000100000001110010011100000111001001011100011010000110010001
000000010000000100000001000000010000000110001001101100010111000111000001100100010000
000100000001000000010000000100000001101010011001100101110001110000011001100100000001
000000010000000100000001000000011010100110110001011100011011000110010001000000010000
000100000001000000010000000100000001101000010111000111000001100010010000000100000001
000000010000000110111001110010011001100101110001101010011100000100000001000000010000
000100000001000000010000000110011001011100011001100111000000101000100000001000000010
000001100010011001000110000001000000010000000100000001101010011100000110000010000
001000000011010100100000001000000010000000110001001110000011000001100000110001001010011
00000011100100101101010011001100100000101010000110001001101000011101000110010001100100
011101000110000001101010010111000110111001110000011100001011010001000000010000000100
000001100110011100000111001001101000010111000110110001101000010000000100000001000000
011001000110101001101000011011000101110001110000011100000100000001000000010000000100
000001000000010110100110001001011100011010000111001001000000010000000100000001011010
011100100101110001101110011011000100000001000000010110100110001001110000010111000111
001001101100010000000100000001011010011100000110000001011100011000000110000001000000
010000000101101001110000011000000101110001100000011000000100000001000000010000000110
001001000000010000000110100001101000010000000100000001000000011100100100000001000000
010000000100000001110010011100100110010001011100011000100110001001000000010000000100
000001000000010000000110001001101100010111000110000001100110010000000100000001000000
010000000100000001101010011010000101110001100000011100000100000001000000010000000100
000001000000011010100110110001011100011100000110111001000000010000000100000001000000
010000000100000001101000010111000111001001100000010000000100000001000000010000000110
111001110010011000000101110001101100011100000100000001000000010000000100000001000000
010000000110011001011100011001100111000000010100010000000100000001000000011000100110
010001100000010000000100000001000000011010100111000001000000010000000100000001101010
010000000100000001000000011000100111001001110000011000100101101001100000011100100101
101001100110011000001010100001100010011010000111010001100100011001000111010001100010
011000000101110001100000011001000110111010110100010000000100000001000000011001100111
001001100100011010100101110001100000011010100100000001000000010000000110010001101000
01100100011100100101110001100010011010100100000010000000100000001000000010000000101
101001100010010111000110100001110010010000000100000001000000010110100111000001011100
011011100111000001000000010000000101101001100010011100000101110000101110001100100110100
000001000000010110100111000001100000010111000110000001100000010000000100000001011010
011100000110000001011100011000000110000001000000010000000100000001100010010000000100
000001101000011010000100000001000000010000000111001001000000010000000100000001000000
0111001001110010011010000101110001101110011100100

L'interprétation de cet objet numérique en ASCII (ISO 646)

9429221140.1217.6867.9671.2815.09634.628.071205851981-09-30T14:26:17.736Z2537.991
884.430.11-80.00-80.00-14.78-17.0229221145.0617.7668.4271.7715.75629.758.361205851
981-09-30T14:26:21.975Z3902.583974.89-1.46-18.69-25.60-80.00-80.0013391147.5217.81
68.6572.0116.09627.348.511205851981-09-30T14:26:26.134Z3553.843160.87-1.35-15.79-2
3.01-80.00-80.0013391149.9917.8568.8872.2516.44624.938.671205851981-09-30T14:26:34
.533Z3606.403068.59-1.13-15.65-21.63-80.00-80.0013391154.9217.9469.3472.7417.18620
.148.991205851981-09-30T14:26:38.772Z3729.722871.54-1.10-13.96-19.98-80.00-80.0013
391157.3817.9969.5772.9817.57617.779.161205851981-09-30T14:26:42.932Z3715.963019.5
2-1.10-14.96-20.79-80.00-80.0013391159.8518.0369.8073.2217.96615.409.331205851981
09-30T14:26:51.330Z4047.072315.92-0.46-11.32-13.37-80.00-80.0013391164.7718.1370.26
73.7018.79610.709.681205851981-09-30T14:26:55.570Z4073.732350.72-0.27-12.11-12.78-8
0.00-80.0013391167.2318.1870.4973.9319.23608.369.861205851981-09-30T14:26:59.

Il convient de séparer les différents champs d'information

120	58	5	1981-09-30T14:22:01.629Z	3898.70	2352.45	-1.66	-7.72	-18.61	-80.00	-80.00
120	58	5	1981-09-30T14:22:05.788Z	3894.64	2546.88	-1.49	-9.76	-18.96	-80.00	-80.00
120	58	5	1981-09-30T14:22:10.027Z	3925.05	2429.15	-1.49	-8.78	-18.27	-80.00	-80.00
120	58	5	1981-09-30T14:22:18.426Z	3946.38	2500.82	-1.71	-8.42	-19.33	-80.00	-80.00
120	58	5	1981-09-30T14:22:22.585Z	3964.01	2591.92	-1.76	-8.83	-19.86	-80.00	-80.00
120	58	5	1981-09-30T14:22:35.222Z	4021.69	2406.25	-1.62	-7.75	-18.24	-80.00	-80.00
120	58	5	1981-09-30T14:22:39.381Z	4007.80	2474.73	-1.69	-8.05	-18.88	-80.00	-80.00
120	58	5	1981-09-30T14:22:43.620Z	4037.11	2329.44	-1.74	-6.70	-18.23	-80.00	-80.00
120	58	5	1981-09-30T14:22:52.018Z	4099.12	2399.84	-1.33	-8.46	-16.84	-80.00	-80.00
120	58	5	1981-09-30T14:22:56.177Z	4120.44	2447.08	-1.46	-8.25	-17.44	-80.00	-80.00
120	58	5	1981-09-30T14:23:00.417Z	4114.44	2291.40	-1.57	-6.73	-17.11	-80.00	-80.00
120	58	5	1981-09-30T14:23:08.815Z	4140.36	2347.36	-1.47	-7.41	-16.91	-80.00	-80.00
120	58	5	1981-09-30T14:23:12.974Z	4172.42	2302.13	-1.27	-7.70	-15.82	-80.00	-80.00
120	58	5	1981-09-30T14:23:17.213Z	4198.63	2434.23	-1.38	-8.11	-16.80	-80.00	-80.00
120	58	5	1981-09-30T14:23:25.611Z	4182.57	2381.31	-1.30	-8.10	-16.30	-80.00	-80.00
120	58	5	1981-09-30T14:23:29.770Z	4214.61	2112.57	-1.14	-6.63	-14.20	-80.00	-80.00
120	58	5	1981-09-30T14:23:34.009Z	4245.87	2113.51	-0.96	-7.24	-13.35	-80.00	-80.00

Puis de définir la signification de chaque champ

Les compétences

Il apparaît clairement ici un besoin de double compétence :

celle de l'archiviste capable de définir, en relation avec le Producteur, les informations à préserver, de vérifier l'intelligibilité de ces informations et leur complétude.

celle de l'informaticien spécialisé dans la **gestion des données** et la représentation de l'information sous forme numérique, afin de définir les formats de données et de métadonnées acceptables pour la pérennisation, de vérifier la conformité, de mettre en œuvre si nécessaire un processus de transformation de formats, de spécifier le développement des outils informatiques nécessaires à ce service, de les développer et de les exploiter.

Ces compétences spécialisées sur la représentation numérique présupposent également une connaissance généraliste en Informatique.

Les deux compétences sont réunies dans un métier nouveau qualifié de gestionnaire de données numériques s'appuyant très fortement sur les normes et standards de représentation de données et de métadonnées.

En outre, il apparaît la nécessité de pouvoir dialoguer, négocier avec les entités productrices de données et de documents, soit un travail itératif de longue haleine.

4- le Service de Stockage - au CNES le STAF

Fonctions vue du côté ' client '

Le 'client' du service de stockage est par exemple le responsable de l'archivage de données scientifiques, images, documents.... Ces données sont des ensembles de fichiers

c'est à dire des trains de bits

dont il connaît le format et le contenu

qu'il sait manipuler et présenter aux utilisateurs finaux sous une forme intelligible.

Ce que le client attend d'un service de stockage d'archive, c'est :

- la prise en charge de ces fichiers en vue de leur conservation à long terme
- la garantie de l'intégrité de ces fichiers
- la capacité à lui restituer ces fichiers dans le délai convenu par le contrat de service
- la disponibilité d'une **'interface technique' stable** me permettant de faire appel à ses services (archivage de fichier, restitution, renommage, création d'arborescence virtuelle,...)

- la capacité à prendre en compte les évolutions de la technologie (migrations de supports de stockage....) sans aucun impact sur l'interface et donc sans impact sur ses applications.

Ce concept permet : une organisation du service de stockage d'archive totalement indépendante des autres services, ainsi qu'une réutilisation de ce service dans de multiples contextes au sein de l'organisme concerné.

Compétences

Il s'agit cette fois de compétences d'informaticiens spécialisés dans :

- **la gestion de grands ensembles de fichiers stockés, dupliqués sur différents types de supports,**
- **les technologies réseau à haut débit permettant de communiquer avec les ' clients ' du service**
- **les technologies de stockage à grande capacité, robots de stockage..., les supports de stockage, leurs caractéristiques, leur fiabilité**
- **les moyens de surveillance de l'état des supports, mise en œuvre de ces moyens**
- **la capacité à maintenir en fonctionnement opérationnel un système ouvert 24 heures sur 24 et à faire évoluer le système en fonction des évolutions de la technologie et des montées en charge.**

5- Gestion des données et accès des utilisateurs

Fonctions

Il s'agit de mettre en place et maintenir en fonctionnement, un système informatique permettant aux utilisateurs d'accéder à distance - via une interface conviviale - à un ensemble de fonctions

- **connaître le contenu de l'archive,**
- **rechercher les données qui les intéressent (critères de sélection basés sur les métadonnées par exemple)**
- **sélectionner les données qui correspondent à leur besoin**
- **commander et récupérer ces données**
- **éventuellement transformer les données archivées avant de les fournir à l'utilisateur (changements de format, Services à Valeur Ajoutée....)**

La recherche des données utiles s'appuie sur les métadonnées mais aussi sur différentes techniques complémentaires (feuilletage, data mining...)

Les moyens de récupération peuvent être le réseau ou la recopie sur un support de diffusion courant (CD-rom, DVD, DLT...) en fonction du volume : un service dédié au Centre informatique du CNES (SEM : Service d'Echange de Media) a été mis en place pour la diffusion et la réception de données sur supports physiques

Il s'agit également de gérer les relations avec la communauté des utilisateurs.

Compétences

Il s'agit d'une part de compétences d'informaticien spécialisé dans :

- **la modélisation des données**
- **les processus de recherche d'information**
- **les technologies de base de données**
- **les technologies et langages de l'Internet (Interface Homme-machine sur navigateur,...)**
- **le maintien en fonctionnement opérationnel de systèmes ouverts à des communautés d'utilisateurs plus ou moins vastes.**

Il s'agit d'autre part d'avoir une connaissance générale de la problématique de l'archivage :

- **connaissance des catégories de données manipulées**
- **connaissance des métadonnées et des critères de sélection des données adaptés aux besoins des utilisateurs**

6- Retour d'expérience

Le concept du STAF - Service d'Archivage de Fichiers (stockage d'archive) a totalement fait ses preuves: avec dix ans d'expérience, on peut constater qu'aucune donnée n'a été perdue, un nombre croissant de clients et un volume croissant de données stockées :

- plus d'1,5 millions de fichiers et 6 TO de données¹ dans le service STAF d'origine,
- 0,3 millions de fichiers et 18 TO de données² pour un nouveau service réservé aux gros fichiers,
- 1,7 millions de fichiers et 200 TO de données³ dans le service dédié aux données issues du satellite SPOT.

Le concept permet aussi la mutualisation du stockage d'archive entre plusieurs établissements d'une même institution et entre plusieurs institutions distinctes.

Une masse critique minimale est indispensable pour réduire les coûts (des moyens matériels, logiciels et humains).

Les systèmes génériques pour la gestion et l'accès aux données basés sur le principe que la recherche d'information passe le plus souvent par les mêmes mécanismes : mots clés, critères numériques (temps, espace.....), avec des terminologies différentes. C'est le rôle actuel du SIPAD⁴ et du SIPAD- Nouvelle Génération en cours de développement. Ceci est un point plus difficile à résoudre (que le stockage d'archive) pour aboutir à une solution à la fois générique, évolutive et adaptables aux besoins des différents clients

Toutefois, le point critique essentiel est celui de la collecte de l'information et de l'ensemble des activités conduisant à la constitution de fichiers dont le format est acceptable pour la préservation à long terme et de métadonnées 'normalisées'. Ce point critique concerne à la fois le contenu (complétude, exactitude, authenticité) ainsi que le format (ouvert, normalisé,...).

Ce point critique n'est pas sans rapport avec la politique technique ou la politique bureaucratique de l'entreprise.

Un versement d'archives électroniques : la nouvelle chaîne pénale de Paris et de la région parisienne, par Françoise Banat-Berger

Le projet d'archivage de la nouvelle chaîne pénale de Paris et de la région parisienne, qui s'applique pour le moment au registre des affaires classées sans suite avant le 1^{er} janvier 1994, a débuté en 2000.

Plusieurs fois évoqué dans ce bulletin, notamment à l'occasion des tests effectués aux archives de Paris en 2001, il vient de trouver un premier aboutissement.

Le 26 mai 2004, en effet, il a été procédé au versement par le ministère de la Justice aux archives départementales concernées des CD-R contenant les données archivées.

Les données stockées à plat devront être remontées par les différents services d'archives dans leurs systèmes d'information, afin de pouvoir y accéder, le ministère de la Justice envisageant d'éliminer définitivement les données de ses propres serveurs. La direction des Archives de France, étant donné le

¹ Y compris les sauvegardes.

² Y compris les sauvegardes.

³ Y compris les sauvegardes.

⁴ Système générique pour la gestion et l'accès aux données.

caractère novateur de cette opération, propose également de faire développer, durant le mois de septembre prochain, une application permettant de remonter les données, si les services d'archives concernées n'ont pas eu la possibilité de réaliser cet outil de consultation elles-mêmes. Avant leur envoi, les CD-R ont été confiés à Edmond Fernandez (CAOM), qui les a soumis à un testeur de CD : s'étant avéré de très médiocre qualité, ils ont dû être re-gravés.

Voici le bordereau de versement qui a précédé l'envoi des CD-R :

N° de versement

BORDEREAU DE VERSEMENT D'ARCHIVES

aux Archives Départementales de Paris

Adresse : 18 boulevard Sérurier, 75019 Paris
tel. : 01.53.72.41.23 ; fax : 01.53.72.41.34

Service producteur :		Tribunal de grande instance de Paris (tribunal correctionnel et Parquet) 4 boulevard du Palais 75055 PARIS R.P. tel : 01.44.32.51.51
Statut du service producteur :		Service déconcentré de l'Etat
Service versant :		Ministère de la justice / direction de l'administration générale et de l'équipement / service des archives
Agent responsable du versement		Mme Françoise Banat-Berger, chef du service
Téléphone :		01.44.77.64.18
Nombre total d'articles :		1
Espace mémoire occupé :		619 Mo
Dates extrêmes du versement :		1975 – 1993
Conservation intégrale :		oui
Délai de communicabilité :		100 ans
Date de communicabilité de la totalité du versement :		2094
N° des articles	Description sommaire	Dates extrêmes
1	Tribunal de grande instance de Paris (tribunal correctionnel et Parquet) Affaires pénales arrêtées au niveau du bureau d'ordre : données enregistrées dans la base Nouvelle Chaîne Pénale (NCP) ou reprises de l'ancienne application Bureau d'ordre pénal (BOP).	1975-1993
Bordereau de 4 pages (y		Pris en charge le

compris la présente)
Le secrétaire en chef du parquet du Tribunal de grande instance de Paris	Le directeur du service des archives départementales de Paris
M. Jean-Bernard DESJARDINS

Cote	Analyse						
.....	<p><u>Un document papier</u> La couverture fonctionnelle de la Nouvelle Chaîne Pénale / Ministère de la Justice, Direction des services judiciaires, Bureau de l'informatisation des juridictions (AB4 - maîtrise d'ouvrage pénale.- janvier 2001</p> <p><u>Un CR-Rom (et un duplicata) contenant les fichiers suivants :</u> <i>Nb : pour visualiser ces fichiers, utiliser Wordpad.</i></p> <p>- 2 fichiers de documentation sur le présent archivage NCP :</p> <ul style="list-style-type: none"> - Le fichier <u>extraction.pdf</u> qui présente la structure de l'archivage NCP. Pour les tables de référence, il donne la structure des tables fournies. Pour chacun des 3 fichiers de données, il précise l'organisation des informations fournies, notamment : <ul style="list-style-type: none"> l'ordre d'apparition des champs dans les étiquettes la longueur maximale de chaque champ (exemple C10) la nature de l'information contenue dans le champ (il est précisé si le champ est géré par une table de référence) <p>A noter :</p> <ul style="list-style-type: none"> Dans l'étiquette Affaire, les valeurs contenues dans le champ « Numéro microfiche » n'apportent aucune information utile en l'absence des microfiches. Ces microfiches pourront faire l'objet d'un archivage ultérieur. Dans l'étiquette Personne, si le champ « Nature du nom » d'une personne physique n'est pas renseignée, la valeur par défaut est P (soit NOM DE PERSONNE) <ul style="list-style-type: none"> - Le fichier <u>rapport_200401_paris.txt</u> qui récapitule les informations contenues dans le CD : la date de l'extraction, le nom des fichiers, la nature des informations et le nombre d'occurrences. <p>- 3 fichiers de données de format ASCII délimité (séparateur ;) :</p> <table style="margin-left: 20px;"> <tr> <td>Fichier : affaires.dat</td> <td>2 442 617 affaires</td> </tr> <tr> <td>Fichier : personne.dat</td> <td>4 086 993 personnes</td> </tr> <tr> <td>Fichier : textes.dat</td> <td>4 734 377 lignes de textes</td> </tr> </table> <p>- 9 tables de référence de format ASCII délimité (séparateur ;) :</p> <ul style="list-style-type: none"> <u>nataff.txt</u> : les natures d'affaires <u>coraff.txt</u> : la concordance entre ancienne et nouvelle nomenclature des natures d'affaires <u>elst.txt</u> : les éléments de structure nationaux et locaux (dans cette table, le code des éléments de structure locaux débute par 999) <u>evt.txt</u> : les types d'événement déclenchant l'action publique <u>motevt.txt</u> : les motifs des événements <u>natinf.txt</u> : les natures des infractions <u>role.txt</u> : les rôles des parties <u>service.txt</u> : les noms des services ayant traité les affaires (table locale) <u>stajuri.txt</u> : les statuts juridiques des personnes morales 	Fichier : affaires.dat	2 442 617 affaires	Fichier : personne.dat	4 086 993 personnes	Fichier : textes.dat	4 734 377 lignes de textes
Fichier : affaires.dat	2 442 617 affaires						
Fichier : personne.dat	4 086 993 personnes						
Fichier : textes.dat	4 734 377 lignes de textes						

METADONNEES

Métadonnées de description

Contexte du versement : En 1999, dans le cadre d'expérimentations sur l'archivage électronique, la Direction des archives de France a décidé, sur la suggestion du service des archives du ministère de la justice, de procéder à un test avec une application développée au ministère de la justice, dans le secteur pénal, appelée « nouvelle chaîne pénale » (NCP) et utilisée dans les tribunaux de grande instance de Paris, Nanterre, Bobigny, Créteil, Pontoise, Evry et Versailles. Les archives de Paris ont pris en charge ce premier versement expérimental (TGI Paris) en décembre 2001.

En janvier 2003, la procédure d'extraction des données de la NCP et leur versement aux AD est généralisée à l'ensemble des tribunaux concernés. Cette procédure interviendra annuellement.

A noter : pour que les données soient interrogeables à tout moment (à la demande des producteurs ou des chercheurs), la Direction des archives de France a développé une application sous Access, permettant aux sept services d'archives concernées par l'archivage de la NCP, d'intégrer les données qui lui sont transmises sur CD-Rom.

Si le service des AD souhaite utiliser cette application, elle doit s'adresser à la DAF, département de l'innovation technologique et de la normalisation :

Henri-Luc Camplo, tel : 01.40.27.63.85 ; mel : henri-luc.camplo@culture.gouv.fr

Description du versement : Enregistrements des affaires pénales classées sans suite ou assimilées (cf. règles d'extraction ci-dessous) avant 1994, au niveau du bureau d'ordre, extraits de la nouvelle chaîne pénale (NCP) et des affaires récupérées de l'ancienne application Bureau d'ordre pénal (BOP) dans la NCP.

Règles d'extraction appliquées pour ce versement : Ont été extraites les affaires dont le dernier événement attaché à l'affaire est d'un des motifs suivants, daté d'avant 1994 :

<u>Code</u>	<u>Intitulé</u>
-------------	-----------------

1020	Classement en attente
------	-----------------------

1140	Mise en attente
------	-----------------

25000	Classement sans suite
-------	-----------------------

25100	Classement
-------	------------

41000	Nullité totale de la procédure
-------	--------------------------------

74200	Avis de classement sans suite ou de dessaisissement
-------	---

Précisions :

Lorsque la date de dernier événement attaché à l'affaire n'était pas renseignée (reprise BOP de Nanterre et Paris), l'année contenue dans le numéro d'affaire a été retenue.

Lorsque le dernier événement attaché à l'affaire n'était pas renseigné, on a recherché le dernier événement attaché aux personnes dans la liste ci-dessus, et daté d'avant 1994. Par définition, pour une même affaire, l'événement personne peut être différent d'une personne à l'autre : il ne devait pas y avoir pour une personne un événement autre que

ceux de la liste ci-dessus.

Lorsque la date de l'événement attaché à la personne n'était pas renseignée, on a pris l'année contenue dans le numéro d'affaire.

Evaluation du versement et lacunes

Pour le TGI de Paris : Les affaires classées sans suite Ab initio d'avant 1986 n'avaient pas fait l'objet en 1992 d'une reprise de données du BOP dans la NCP. En conséquence, le présent archivage présente d'importantes **lacunes qui peuvent être évaluées à environ 2 994 500 affaires manquantes**.

Date de création de la NCP : 1992 (premier site implanté à Pontoise)

Système d'exploitation et SGBDR : les programmes NPC tournent sur Gcos 7 ; bases de données Oracle version 7 sur serveurs Unix

Logiciel de capture : programme sous Unix d'extraction des données de la base Oracle, requêtes SQL, traitement par batch

Périodicité de l'extraction : annuelle

Référence : Circulaire n° SJ. 03-013-DSJ/10.09.2003 – Gestion des archives des juridictions - 050TGI

Métadonnées de gestion

Format des données : ASCII délimité

Supports : CD-Rom (données à plat et tables de référence)

Volumétrie : 619 Mo

Conservation : intégrale et illimitée

Délai de communicabilité : 100 ans

ATTENTION

Une fois remontées dans le système d'information des Archives Départementales, la Justice procédera à l'élimination définitive, sur les serveurs du Centre de prestation régional de Grigny, des données ayant fait l'objet de l'archivage.

Par conséquent, les Archives Départementales pourront être amenées comme pour tout autre versement à communiquer ces dossiers au Parquet concerné.

Les crédits de numérisation de la MRT en 2004

Lors de sa réunion du 4 mai, le comité scientifique pour la documentation informatisée et le multimédia (CSDIM), chargé notamment de la répartition des crédits de numérisation de la mission de la recherche et de la technologie (MRT), a annoncé la mise à disposition de crédits en 2004.

Cependant, étant donné l'avancement de l'année, il n'est pas possible de lancer un nouvel appel à projets.

Il a donc été prévu de financer des projets présentés l'an dernier et dont la valeur scientifique avait été reconnue mais qui n'avaient pas pu être pris en compte faute de crédits. Les services d'archives concernés ont en principe été contactés.

Les modalités d'utilisation des crédits de 2004 ne sont pas encore arrêtées : en particulier, il n'est pas certain que de nouveaux marchés nationaux puissent être lancés, auquel cas les services retenus devraient chercher eux-mêmes leurs prestataires.

Rappel :
vos contributions sont les bienvenues, notamment des récits de visites ou d'expériences concrètes.