

# Guide

---

de

---

**l'information**

---

**numérique**

Comment traiter les données lisibles  
par machine et les documents numériques

**Édition mise à jour et amplifiée**

Ce guide multidisciplinaire a été conçu à l'origine par Jean-Michel Cornu, consultant, en étroite coopération avec les archives historiques de la Commission européenne et des experts des États membres. Une version provisoire avait été rédigée pour être présentée au forum DLM sur les archives électroniques (Bruxelles, 18 au 20 décembre 1996). ISBN 92-827-9130-0 (EN, FR, DE)  
Les commentaires et annotations reçus pendant le forum et transmis peu après aux archives historiques de la Commission européenne ont été inclus par Syllis SA (Lille, France)

Cette version du guide, mise à jour et amplifiée, est un des résultats concrets du forum DLM. Celle-ci est le reflet des efforts continus de coopération faits par différentes disciplines (administration publique, archives, industrie, communauté des chercheurs) concernées par le problème des données électroniques et de la gestion de l'information dans la société de l'information en rapide mutation.

De nombreuses autres informations sur l'Union européenne sont disponibles sur Internet via le serveur EUROPA (<http://europa.eu.int>).

Une fiche bibliographique figure à la fin de l'ouvrage.

Luxembourg: Office des publications officielles des Communautés européennes, 1997

ISBN 92-828-2286-9

© Communautés européennes, 1998

Reproduction autorisée, moyennant mention de la source

*Printed in Italy*

# Sommaire

Préface	5
1. Introduction	7
1.1. Présentation du guide multidisciplinaire	7
1.2. Comment lire ce guide multidisciplinaire?	9
1.3. De la production personnelle à l'utilisation par tous	10
1.4. Les données et le temps: Les trois phases	10
2. De la donnée à l'information numérique structurée	11
2.1. Qu'est-ce que l'information?	11
2.2. Qu'est-ce qu'une donnée?	11
2.3. Pourquoi l'information numérique est-elle différente?	11
2.4. Création des données électroniques	12
2.5. Deux façons de structurer les données	13
2.5.1. Les documents	14
2.5.2. Les bases de données	15
2.5.3. Acquisition de matériels et de logiciels pour l'information numérique	15
3. Cycle de vie et répartition des responsabilités	17
3.1. Vue d'ensemble du cycle de vie	17
3.2. Définir les responsabilités pour chaque phase	18
4. Conception, création et maintenance de l'information numérique	20
4.1. La phase de conception	20
4.2. Création de l'information numérique	20
4.3. Intégration, migration et évaluation de l'information numérique	21
4.3.1. Intégration de l'information	21
4.3.2. Migration de l'information	22
4.3.3. Du papier ou microfilm à l'image scannée	24
4.3.4. De l'image scannée aux formats codés	24
4.3.5. D'un format numérique à l'autre	26
4.3.6. Évaluation de l'information électronique	27
4.4. Gestion et classement de l'information numérique	30
4.5. Transfert	32

34	<b>5. Conservation à long et à court terme de l'information numérique</b>
34	<b>5.1. Supports de données</b>
37	<b>5.2. Formats de fichiers</b>
38	5.2.1. Images en mode point
39	5.2.2. Graphiques vectoriels
39	5.2.3. Textes
41	5.2.4. Données et programmes
43	<b>6. Accès et diffusion de l'information</b>
43	<b>6.1. Vers la société de l'information</b>
44	<b>6.2. Normes et standards pour l'accès aux données</b>
45	<b>6.3. La sécurité</b>
45	6.3.1. Droits d'accès
46	6.3.2. Cryptage et authentification
47	<b>6.4. L'accès aux données</b>
48	<b>7. Conclusions</b>
49	<b>8. Annexes</b>
49	<b>8.1. Terminologie</b>
50	<b>8.2. Questions en suspens</b>
50	<b>8.3. Pour comprendre la normalisation</b>
53	<b>8.4. Check-list pour les DLM</b>
55	<b>8.5. Prototype: quelles métadonnées faut-il créer?</b>
57	<b>8.6. Prototype: Comment choisir les bons standards?</b>
58	<b>8.7. Index</b>
60	<b>8.8. Table des figures</b>
60	<b>8.9. Bibliographie</b>

# PRÉFACE



Le forum DLM 96, qui a été organisé conjointement par les États membres de l'Union européenne et la Commission européenne en décembre 1996 à Bruxelles, a réuni des experts de l'industrie, de la recherche, de l'administration et des archives autour d'un thème central, de plus en plus souvent appelé la «mémoire de la société de l'information». Il y a quelques années encore, nul n'aurait imaginé que les innovations technologiques bouleverseraient le secteur des archives au point que l'authenticité et la conservation durable d'informations indispensables s'en trouveraient gravement menacées à court terme si les mesures appropriées n'étaient pas prises à temps.

Les services d'archives font partie intégrante de la société de l'information. Depuis la publication du rapport intitulé «L'Europe et la société de l'information planétaire» (rapport Bangemann) en 1994, les programmes d'action et de soutien de la Commission européenne leur accordent une place accrue. Elles joueront aussi un rôle croissant dans la gestion moderne de l'information. Le forum DLM 96 a planté un jalon important dans ce contexte tout en ouvrant la voie à de nouveaux changements.

Ce Guide de l'information numérique est l'un des résultats les plus marquants du forum DLM. Il a été distribué aux participants sous forme d'avant-projet et a servi de base de travail et de discussion. La nouvelle version de ce guide, qui a été considérablement révisé et étendu, a été élaborée grâce aux quelque 300 propositions de modification et d'amélioration, généralement fondamentales, que les experts nationaux ont présentées aux services de la Commission pendant et après le forum DLM.

Je suis convaincu que ce guide rédigé en commun et de manière interdisciplinaire contribuera à résoudre les problèmes urgents que posent la gestion et l'archivage des informations sauvegardées sur support électronique, en proposant des stratégies à court et à moyen terme. Désormais, c'est aux décideurs de l'administration et de l'industrie de prendre les mesures nécessaires, en concertation avec les spécialistes de l'archivage électronique, et de concevoir des solutions pratiques dans le cadre d'une coopération européenne. La confiance des entreprises et des personnes privées dans la sécurité des nouveaux médias en sortirait renforcée.

Grâce à sa structure, ce guide constitue une source d'informations d'une grande richesse, tant pour les experts que pour les amateurs. Il représente un pas important vers la préservation durable de la «mémoire de la société de l'information» et la promotion active de la transparence, pour tous les citoyens, de l'activité ainsi que des processus de décision des administrations nationales et des institutions de l'Union européenne.

**D' Martin Bangemann**

Membre de la Commission européenne chargé des affaires industrielles, des technologies de l'information et des télécommunications



# 1. Introduction

## 1.1. Présentation du guide multidisciplinaire

---

Aujourd'hui, la grande majorité des documents se présente sous forme papier. Dans un très grand nombre de cas cependant, ces documents sont produits à l'origine sur des équipements bureautiques ou informatiques. C'est le cas, par exemple, des messages électroniques, des mémos ou encore des comptes rendus de réunion qui rythment la vie professionnelle. Avec l'essor du courrier électronique, les documents numériques commencent même à remplacer le support papier et le nombre de documents électroniques augmente rapidement dans les administrations et les entreprises. Il devient nécessaire d'analyser l'impact de ce phénomène sur les pratiques ainsi que les règles d'utilisation et d'archivage.

Par ailleurs, les organisations ont développé des bases de données. Il est souvent nécessaire de conserver les informations contenues dans ces bases un certain temps après leur utilisation courante, pour des raisons légales ou pour permettre des recherches ultérieures. Il faut alors prendre en compte la conservation à long terme et l'accessibilité des données ainsi que leur capacité à produire de nouvelles informations.

Ce guide multidisciplinaire ne fournit pas une réponse complète et définitive dans tous ces domaines. En effet, ces questions nécessitent plus de discussion entre les différentes disciplines concernées, à savoir les administrations publiques, les archives, les industries et la communauté des chercheurs. Ce guide fournit quelques suggestions et offre des conseils pour aider chaque organisation à définir sa propre stratégie dans le domaine de l'information électronique. Il ne s'agit pas de présenter une façon européenne unique de traiter ce sujet, mais plutôt de rassembler l'expérience acquise dans les organisations nationales, régionales et européennes pour le bénéfice de chacun.

Un projet de ce guide multidisciplinaire a été distribué par la Commission européenne aux participants du forum DLM sur les archives électroniques qui s'est tenu à Bruxelles du 18 au 20 décembre 1996. Le suivi a consisté en l'envoi de commentaires et d'annotations écrits par des experts nationaux et en discussions consacrées à la préparation de cette version mise à jour et amplifiée.

Ce guide peut également être utilisé en complément du guide du Conseil international des archives (CIA) sur les enregistrements électroniques, qui se limite toutefois au point de vue d'une seule discipline.

Les actions de suivi du forum DLM qui sont précisées dans ce que l'on appelle les «dix points» (\*) (voir les cercles dans les graphiques qui suivent) sont liées à toute la chaîne de production et de maintenance des enregistrements numériques. Elles sont contrôlées par le comité de surveillance DLM qui a été constitué comme une des actions de suivi. Elles comprennent:

- une aide fournie aux utilisateurs sous la forme d'une formation et de ce guide multidisciplinaire;
- une étude détaillée sur les relations entre l'administration publique et les archives;
- des indications spécifiques sur les fonctionnalités requises, sur les spécifications DLM, sur les normes TI pour les fournisseurs de logiciels et matériels informatiques et les organismes de normalisation, sur les implications juridiques pour les gestionnaires DLM ainsi que sur l'accès à l'information pour le citoyen et le monde de la recherche.

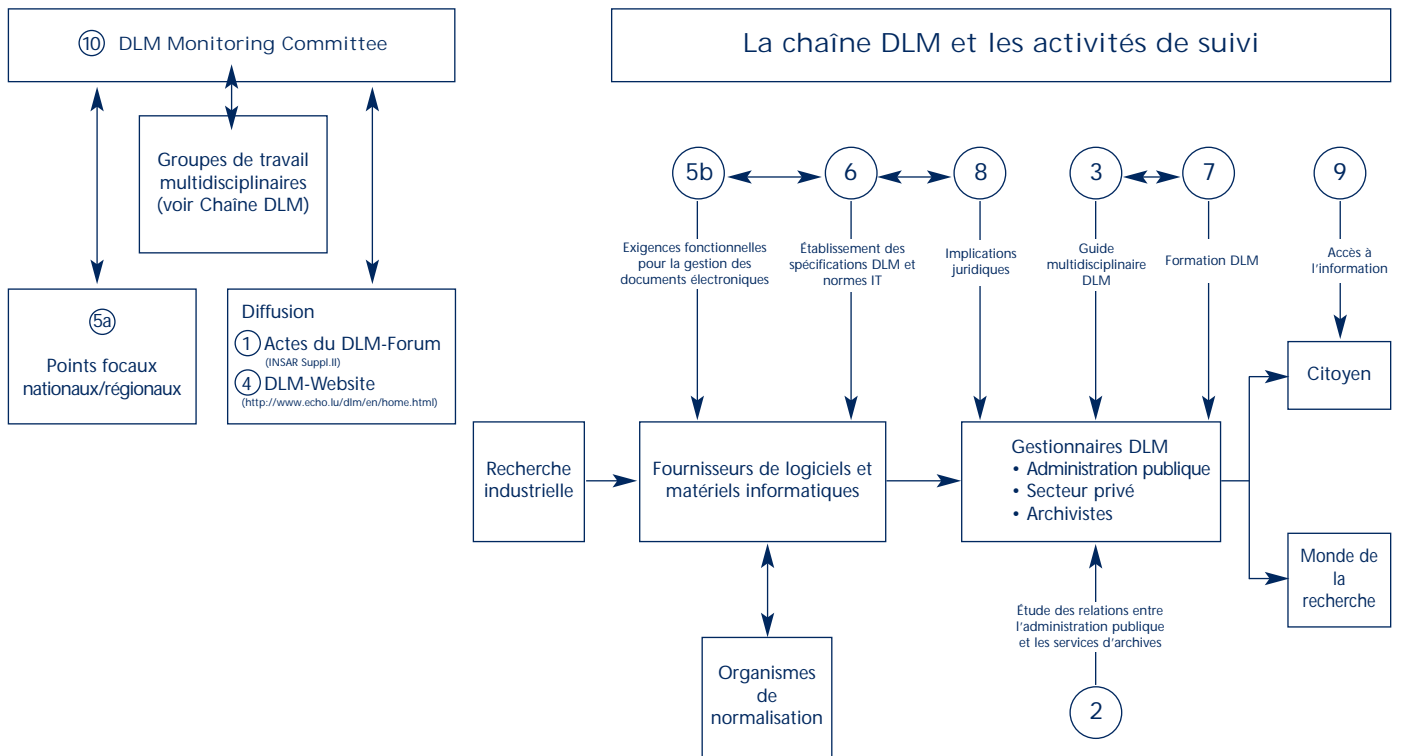
◆ **Documents et bases de données** sont de plus en plus utilisés sur des supports électroniques

◆ **Des suggestions** pour aider à définir sa stratégie dans le domaine de l'information numérique.

◆ **La chaîne DLM** Elle comprend plusieurs types d'action:

- des indications spécifiques,
- des études,
- une aide aux utilisateurs.

(\*) Voir: *Actes du DLM-Forum sur les données lisibles par machine, Bruxelles, 18 au 20 décembre 1996, INSAR – Courrier européen des archives, supplément II, EUR-OP, Luxembourg, 1997, p. 353* et sur Internet: <http://www.echo.lu/dlm/en/home.html>



Une approche multidisciplinaire pour prendre en compte les besoins des divers acteurs

Ce guide multidisciplinaire s'adresse à des personnes qui n'ont qu'une connaissance rudimentaire de la gestion des enregistrements électroniques ainsi qu'à celles qui en ont une connaissance plus poussée, en particulier:

- les administrations publiques;
- les services d'archives;
- l'industrie;
- la communauté de la recherche

dans les pays membres et les institutions de l'Union européenne.



## 1.2. Comment lire ce guide multidisciplinaire ?

---

Le corps du texte s'adresse à tous les publics.

Quatre types d'encadrés permettent à chacun de trouver les informations qu'il recherche en fonction de ses besoins et de son niveau de connaissance.



### Le CD-ROM

Notions de base

(Encadrés présentant les notions de base.)

#### ◆ Notions de base

Ces encadrés présentent des rappels sur les diverses notions de base utilisées par chaque profession.



### Fichiers graphiques

Notions avancées

(Encadrés présentant les notions avancées.)

#### ◆ Notions avancées

Ces encadrés présentent des notions plus techniques pour ceux qui souhaitent entrer dans les détails



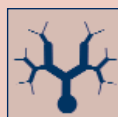
### Allemagne

Exemple

(Encadrés présentant les exemples.)

#### ◆ Exemples

Divers exemples permettent de se faire une idée des différentes solutions utilisées.



### Scanner un document

Pour choisir

(Encadrés présentant l'assistance aux choix.)

#### ◆ Pour choisir

Ces encadrés présentent des conseils ou des arbres de décisions pour faire les bons choix.

### Spécifications

La maturité de chaque norme ou standard n'est pas toujours la même.

Les normes et standards indiqués dans ce guide n'ont pas tous le même niveau de maturité. Des feux de signalisation permettent de repérer rapidement le degré de confiance que l'on peut accorder à chaque spécification.



«Feu vert» — Spécifications stables et reconnues



«Feu orange» — Spécifications en cours ou avec peu de fournisseurs



«Feu rouge» — Spécifications propriétaires sans garantie de pérennité

### 1.3. De la production personnelle à l'utilisation par tous

Un document ou une base de données créé par une personne dans le cadre de ses fonctions est un élément du système d'information global de l'organisation.

Pourquoi est-il si important d'avoir une stratégie à long terme lorsque l'on crée, met à jour ou fournit de l'information?

En fait, même si des documents ou des bases de données sont créés par une ou quelques personnes, ils ont souvent un intérêt pour beaucoup plus de personnes qu'on ne pourrait le supposer:

- ils peuvent être utilisés ou mis à jour par des personnes qui n'ont pas participé à leur création;
- ils sont souvent utilisés en liaison avec beaucoup d'autres enregistrements, dont certains ne sont pas même connus du créateur;
- l'information contenue peut être réutilisée longtemps après la fin de leur utilisation courante (par exemple, pour des raisons légales ou historiques).

Cela veut dire qu'un document ou une base de données doit être pensé comme un élément d'un système d'information global. Il est fondamental de prendre en compte, dès les premières étapes, les aspects de conservation, d'accès à l'information et de protection de la vie privée.

### 1.4. Les données et le temps: le trois phases

- 3 phases:
- la conception,
  - la création,
  - la maintenance.

Il existe trois phases principales dans la vie de l'information électronique:

la phase de conception: cette phase est celle pendant laquelle une stratégie globale est définie;

la phase de création: cette phase est celle de la création effective. Elle implique en général un nombre limité de personnes;

la phase de maintenance: cette phase inclut l'utilisation et la conservation des données.

Ces trois phases ne sont pas équivalentes. Des différences existent à divers égards:

- mise à jour de l'information;
- nombre d'accès à l'information;
- responsabilité;
- etc.

## 2. De la donnée à l'information numérique structurée

### 2.1. Qu'est-ce que l'information?

---

Une information est une indication ou un événement porté à la connaissance d'une personne ou d'un groupe. Il est possible de créer, de maintenir, de conserver ou de transmettre une information.

L'information est à la base de l'organisation du déroulement des opérations. Ce concept est devenu tellement important que la société de l'information est aujourd'hui considérée comme l'étape faisant suite à la société industrielle.

#### ◆ L'information

Elle est une indication ou un événement porté à la connaissance d'une personne ou d'un groupe.

### 2.2. Qu'est-ce qu'une donnée?

---

Une donnée est une unité élémentaire d'information.

Dans un document, par exemple, de nombreuses données sont regroupées pour présenter un raisonnement ou pour rendre compte d'une action. La plupart des données étaient jusqu'à présent conservées et transmises à l'aide de documents papier (ou, parfois, pour les plus anciennes, sur d'autres supports comme des pierres taillées).

Parfois, nous trouvons des listes de données inscrites sur du papier comme c'est le cas pour les répertoires téléphoniques. Il ne s'agit plus de présenter un raisonnement, mais plutôt de fournir la matière première pour une action future (quel est le numéro de téléphone de Monsieur Dupont?). La classification des données est, dans ce cas, cruciale pour permettre de retrouver facilement les données recherchées (par exemple, les noms sont classés par ville puis par ordre alphabétique dans un répertoire téléphonique).

Il est possible de conserver des données sur d'autres supports que le papier. C'est le cas, par exemple, des informations qui peuvent être conservées sur des supports électroniques pour faciliter leur traitement.

Du fait de l'évolution rapide des technologies, il devient de plus en plus complexe de garantir une bonne pérennité dans la conservation des données. Les formats de fichiers et les supports électroniques évoluent rapidement et leur longévité est encore beaucoup plus faible que celle du papier.

De plus, la production d'informations est en très forte augmentation, en particulier dans les administrations. Cela implique qu'il devient de plus en plus complexe de classer et de structurer ces données pour y accéder longtemps après leur production.

#### ◆ Les données

Elles peuvent être regroupées sous forme de document ou de catalogue.

#### ◆ La gestion des données

Il est aujourd'hui de plus en plus complexe de gérer les données.

### 2.3. Pourquoi l'information numérique est-elle différente?

---

Quand les données sont conservées sur un support numérique, il n'est plus possible de les lire sans un outil supplémentaire, une «machine» (en général un ordinateur).

En fait, le «support n'est pas le message» comme c'était le cas avec les documents papier. Dans ces lignes directrices, nous appellerons information numérique, les données qui sont conservées dans un format qui permet directement un traitement automatisé. Il s'agit, en règle générale, de supports numériques.

#### ◆ Le support n'est pas le message

Il faut un outil pour lire l'information numérique.

## L'information numérique

ajoute des avantages et des contraintes dans la gestion des données.

Il y a plusieurs avantages à placer des données sur un support directement lisible par une machine:

- il est bien plus facile de traiter ces données avec une machine. Il n'est plus nécessaire de les ressaisir intégralement lorsqu'une seule partie des données doit être modifiée;
- les supports électroniques permettent en général de placer plus de données dans un volume plus petit;
- il est plus facile de copier un enregistrement complet;
- il est plus facile et plus rapide de transférer de l'information d'une place à une autre;
- un tel support permet une utilisation plus élaborée en recourant à une structure de traitement électronique.

Pourtant, l'utilisation de supports numériques ajoute de nouvelles contraintes:

- il est nécessaire d'utiliser un outil pour qu'un opérateur puisse lire les données;
- les supports numériques ont en général une vie plus courte que le papier ou les microfilms;
- il est plus facile de dupliquer ou de modifier un original (cela pose des problèmes de preuve et d'authentification);
- l'évolution rapide des technologies et du marché de l'information pose des problèmes de stabilité et de pérennité des formats à utiliser.

## 2.4. Création des données électroniques

---

Les données sont conservées sur un support. Dans le cas de l'information numérique, les données peuvent être traitées, communiquées et interprétées par ordinateur. Un ensemble de données peut parfois constituer un enregistrement.

### Qu'est-ce qu'un enregistrement?

Définition du Conseil international des archives

Un enregistrement est «un morceau particulier d'information enregistrée générée, collectée ou reçue lors du commencement, de la conduite ou de l'achèvement d'une activité institutionnelle ou personnelle, et qui comprend un contenu, un contexte ainsi qu'une structure suffisants pour constituer une preuve ou une évidence de cette activité» (Comité CIA sur les enregistrements électroniques).

Un enregistrement dans une administration publique a une valeur légale, d'information ou de preuve. Le terme enregistrement est pris dans ce guide dans le sens administratif et archivistique, et non dans le sens technique habituellement utilisé en informatique.

Le fait qu'un enregistrement numérique est distinct de son support a un certain nombre de conséquences. Par exemple, il est facile de dupliquer un enregistrement numérique d'un support à un autre. Cela veut dire qu'il est plus facile de dupliquer et de diffuser de l'information, mais d'un autre côté la notion même d'original devient plus difficile à cerner bien que non moins cruciale (voir en particulier l'annexe 8.2 sur les questions en suspens).

Les données placées dans un enregistrement numérique doivent constituer un ensemble cohérent et consistant d'information. Définir un système de classement est une des tâches majeures lorsque l'on souhaite définir des enregistrements. Il y a souvent plusieurs possibilités pour rassembler des données dans un enregistrement suivant le niveau de granularité souhaité. Par exemple, dans le cas d'une base de données, la totalité de la base ou un sous-ensemble cohérent peuvent former un enregistrement.



## Enregistrement numérique

Notions avancées

Un enregistrement numérique est constitué de 4 éléments principaux. Les trois premiers doivent être conservés:

1) Le contenu de l'enregistrement lui-même, qui peut inclure plusieurs types de données:

- du texte (pages, paragraphes, mots);
- des nombres (entiers, flottants...);
- des tableaux (tableau complet ou cellules);
- des images, des graphiques, du son et de la vidéo;
- des liens hypertexte.

2) La structure logique de l'enregistrement (cette structure peut être incluse dans le document ou la base de données, ou encore être ajoutée comme une entité séparée. Dans ce dernier cas, la même structure peut être utilisée pour plusieurs enregistrements). La structure logique peut être très différente de la structure physique de l'enregistrement.

3) Le contexte décrit dans un document associé. Il peut inclure des informations telles que:

- les métadonnées techniques (environnement matériel et logiciel, avec les numéros de version, structure du fichier, description des données, un historique des liens avec d'autres enregistrements);
- la description du contexte administratif.

Le contexte décrit dans la documentation peut devenir complexe lorsque l'enregistrement est intégré dans une architecture en réseau.

4) La présentation (en particulier pour un document). Ce dernier aspect est de plus en plus séparé de l'enregistrement. Cette séparation permet d'abstraire l'information de la façon dont elle sera présentée. La distribution de l'information sur des supports variés (CD-ROM, accès en ligne, papier...) est appelée la gestion intersupport. L'outil qui permettra dans plusieurs années de visualiser les données enregistrées aujourd'hui n'est pas encore inventé.

### ◆ L'enregistrement numérique

Il est composé:

- d'un contenu,
- d'une structure,
- d'un contexte,
- d'une présentation.

Comme cette dernière dépend fortement du support de visualisation utilisé, il se peut qu'elle soit impossible à conserver.

## 2.5. Deux façons de structurer les données

Pour être en mesure de retrouver une information particulière, il est nécessaire de la structurer. Suivant le but de l'information, il existe principalement deux façons de structurer les données:

- la base de données: dans cette structure, les données sont regroupées dans un «réservoir» d'information. Les données peuvent ensuite être extraites ou mises à jour.
- le document: cette structure est utilisée lorsque les données ont été regroupées pour rendre compte d'un raisonnement ou d'une action. Ce document peut servir de preuve d'une activité particulière (acte juridique...). Dans ce cas, il doit être saisi dans un enregistrement.

### ◆ Bases de données et documents

Deux types de structures pour les données qui sont de plus en plus souvent amalgamées pour former un ensemble composite.

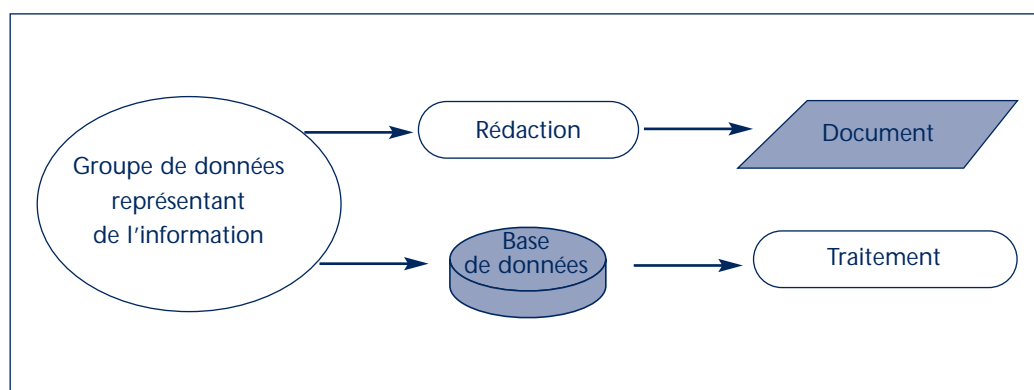


Figure 1 — Document ou base de données?

Bien sûr, il existe de nombreuses façons de structurer des données. L'importance grandissante de l'information dans les organisations pousse à des structures plus complexes, à la fois de type document et de type base de données (par exemple, une page sur un CD-ROM ou un service en ligne remplie automatiquement avec des données extraites d'une base de données).

De plus, il existe souvent une relation étroite entre une base de données et plusieurs documents. Par exemple, une mise à jour de données dans la base peut induire la production d'un document défini par la procédure administrative. Ce type de solution est de plus en plus fréquent avec les outils de *Workflow* (flux des tâches) qui régulent le déroulement des opérations dans l'entreprise.

Les points suivants énumèrent quelques différences entre l'orientation document ou base de données d'un enregistrement:

- la conservation des divers types d'enregistrement (bases de données et documents) n'induit pas les mêmes besoins;
- un document doit être consistant et persistant (une mise à jour conduit à produire un nouveau document);
- une base de données peut être mise à jour régulièrement;
- l'information composite (mi-base de données, mi-documents) pose de nouvelles questions face à ces approches antagonistes.

### 2.5.1. Les documents

Il existe plusieurs étapes entre le bureau fondé sur les documents papier et celui fondé sur les documents numériques:

#### Documents papiers et numériques

Une coexistence s'installe.

Il est nécessaire de prendre en compte les spécificités de chacun.

- 1) le bureau traditionnel fondé sur les documents papier;
- 2) le bureau mixte comprenant des documents papier et numériques;
- 3) la conversion des documents papier (par scannage);
- 4) le bureau tout numérique (tous les documents produits, reçus ou distribués sont numériques).

Ce guide porte principalement sur les étapes 2 et 3 étant donné que celles-ci correspondent à la situation actuelle. Il fournit des renseignements sur les relations entre les documents numériques et les documents traditionnels (principalement les documents papier et les microfiches). L'étape 4 est une direction future plutôt qu'une étape en elle-même. Les documents papier et numériques continueront probablement à coexister pendant de nombreuses années.

Il existe de nombreuses formes de documents qui nécessitent des traitements spécifiques: lettres, notes, mémos, formulaires, rapports...

Les documents peuvent être regroupés en dossiers pour former un ensemble cohérent d'information. La classification des documents est un point important pour permettre de retrouver aisément une information.



Notions de base

## Classification des documents

Il existe de nombreuses façons de classer les documents. Les deux principales sont:

- le classement chronologique (un numéro est donné à chaque document au moment où il est enregistré);
- le classement par thème (par exemple, un numéro est attribué en fonction d'un plan de classement).

La deuxième solution est souvent la plus efficace pour retrouver facilement un enregistrement. Il est nécessaire de définir au préalable un plan de classement efficace.

Une autre solution consiste à retrouver l'information grâce à des mots clés identifiés ou à une recherche dans le texte même.

La relation entre le plan de classement et le système de mots clés ressemble à la relation entre la table des matières et l'index d'un livre. Les deux sont des moyens efficaces de recherche, mais aucun ne remplace l'autre.

Il se peut que les technologies modernes de l'information offrent de nouveaux moyens d'accès et de classement.

◆ Plan de classement et mots clés  
Deux outils complémentaires pour faciliter la recherche d'information

### 2.5.2. Les bases de données

Les bases de données posent un problème lorsqu'il s'agit d'accéder aux données longtemps après la vie normale de la base (par exemple, pour des raisons légales ou pour des recherches). En fait, il existe aujourd'hui très peu de formats standard pour les bases de données. Il n'existe souvent que deux solutions de rechange possibles:

- copier la base de données dans un format de plus bas niveau (en texte plat ou en format ISAM — méthode d'accès séquentiel indexé — par exemple);
- conserver l'application qui a généré la base de données, y compris la documentation (système de gestion de la base de données, application comptable...).

La première solution peut conduire à perdre des éléments structurels de la base. La seconde, en revanche, implique souvent la nécessité de conserver, en plus de l'application, la plate-forme informatique capable de la faire tourner. Il est de même nécessaire de conserver un savoir-faire sur le logiciel et le matériel (ce qui n'est pas évident après plusieurs années).

Dans beaucoup de cas, la base de données est totalement intégrée dans une application propriétaire, comme souvent dans les programmes de gestion.

### 2.5.3. Acquisition de matériels et de logiciels pour l'information numérique

L'acquisition de matériels, de logiciels et de services est un aspect important de l'utilisation et de l'archivage des données numériques. Il existe aujourd'hui de nombreux types de matériels et de logiciels ainsi que beaucoup de services faisant appel à différentes technologies, et aussi de nombreux standards, normes et spécifications. Cependant, l'évolution rapide des technologies de l'information et des communications font craindre que la survie des données électroniques ne dépendent indûment de leur format d'enregistrement, qui pourrait ne pas être accepté par les futurs produits de la technologie de l'information.

◆ Conserver une base de données  
Un problème complexe à cause du manque de formats standard pouvant assurer la pérennité

### EPHOS ▶

Le programme de l'Union européenne qui vise à aider les acheteurs du secteur public.

Le secteur des marchés publics de l'Union européenne est régi par la décision du Conseil de ministres 87/95/CEE, qui oblige tous les acheteurs publics à se référer à des normes officielles (c'est-à-dire des normes approuvées par les instances officielles de normalisation) dans leurs procédures d'appel d'offres. EPHOS (le manuel européen d'achat des systèmes ouverts) est un programme européen qui vise à aider les acheteurs du secteur public à appliquer la décision du Conseil 87/95/CEE en leur fournissant une série de manuels contenant des conseils stratégiques sur la technologie des systèmes ouverts. EPHOS fait notamment référence aux profils internationaux standardisés de normes internationales, quand ils sont d'application. Les modules EPHOS actuels portent exclusivement sur les normes officielles bien que certains modules en cours de préparation abordent l'utilisation de spécifications disponibles pour le grand public.

### Achats avec des normes officielles ▶

Ils posent des problèmes pratiques dus à la non-disponibilité ou au coût des produits et à des standards de fait en concurrence.

Cependant, l'expérience a montré que dans certains domaines de l'information et des communications, notamment des domaines du secteur de la technologie de l'information, les dispositions de la décision du Conseil 87/95/CEE posent des problèmes pratiques. Cela est dû :

- à la non-disponibilité des produits qui soutiennent des normes officielles;
- au coût élevé des produits qui soutiennent des normes officielles;
- au danger que ces normes ne deviennent obsolètes à l'avenir et ne soient plus soutenues par des produits futurs;
- aux standards de fait en concurrence (qui peuvent être des spécifications disponibles pour le grand public) qui sont largement appliqués à des produits, les produits étant eux-mêmes largement déployés par les utilisateurs (par exemple, les spécifications d'Internet).

Il faut toutefois noter que le problème du contrôle futur s'applique aussi bien – et certains experts diraient plus particulièrement – au déploiement de produits fondés sur des normes de facto qu'à des produits fondés sur des normes officielles.

### Achat de produits ▶

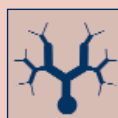
L'achat basé sur des normes stables et ouvertes est un des principaux critères d'évaluation des produits.

L'analyse ci-dessus suggère, par conséquent, que les achats publics soient fondés sur des produits plutôt que sur des normes. Les utilisateurs qui produisent ou maintiennent des données lisibles par machine s'assurent qu'ils disposent des matériels, des logiciels et de la documentation nécessaires pour récupérer à long terme les données et les documents générés par leurs applications. Néanmoins, il est admis qu'il n'est pas toujours pratique ou réaliste de conserver tous les matériels, les logiciels et la documentation pendant une longue période. Il est évident que l'achat de produits qui sont fondés sur des normes stables et ouvertes est un des principaux critères d'évaluation des produits.

Plus précisément, les utilisateurs doivent avoir une idée claire des normes que les produits de la technologie de l'information soutiennent, y compris une évaluation de la stabilité et de l'ouverture de ces normes. Cela indique la nécessité d'une politique d'achat à long terme. Il se peut qu'il faille élaborer un ensemble de lignes directrices communes d'achat pour la communauté d'intérêt participant à la conservation des données lisibles par machine.

### Documentation sur les formats ▶

Il est bon d'imposer aux fournisseurs de procurer tous les moyens qui permettront de récupérer les données générées avec leurs applications pour faciliter leur conservation et leur utilisation à long terme.



Pour choisir

## Clause d'achat

Entre-temps, il est proposé d'inclure la clause standard suivante dans les appels d'offre.

Clause type à insérer dans les appels d'offre

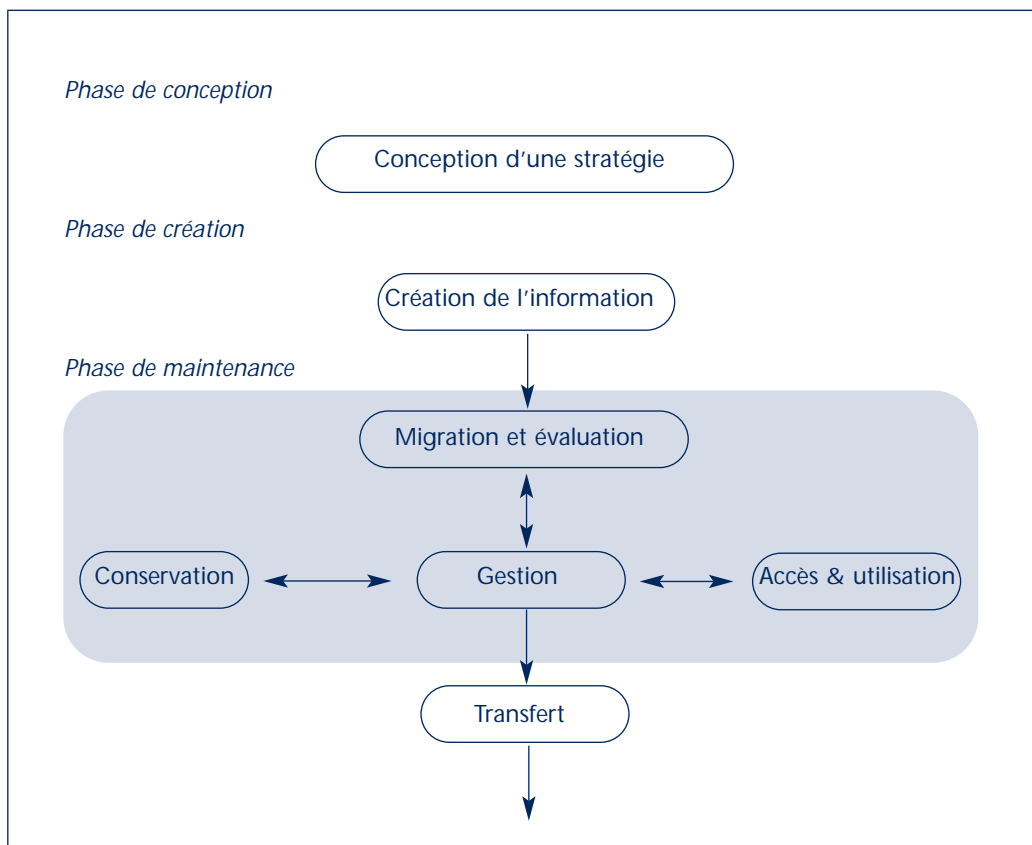
«Afin de garantir la conservation et l'accès à long terme des données du ou des organismes adjudicateurs, le fournisseur s'engage à fournir tout matériel, logiciel et documentation permettant la récupération des données générées par ses applications et leur exportation vers d'autres environnements et d'autres formats.»



## 3. Cycle de vie et répartition des responsabilités

### 3.1. Vue d'ensemble du cycle de vie

L'information numérique passe par plusieurs phases pendant son cycle de vie. Différentes tâches peuvent être définies lors de ces diverses phases. Il est important de définir les règles et les standards qui permettent d'assurer la synchronisation entre toutes les personnes impliquées dans ces tâches.



♦ **Le cycle de vie**  
Un ensemble de phases qu'il faut coordonner au mieux

Figure 2 — Cycle de vie de l'information électronique

Il peut exister plusieurs phases de maintenance où la responsabilité de l'information numérique est transférée d'un organisme à un autre.

Le chapitre 4 présente les tâches qui ont trait à la création et à la gestion des enregistrements avec des suggestions pour aider à définir une stratégie cohérente.

Le chapitre 5 présente plusieurs suggestions pour faciliter la conservation des enregistrements numériques (en particulier sur les différents types de supports et de formats de fichiers).

Le chapitre 6 présente plusieurs façons d'organiser l'accès et l'utilisation de l'information numérique en prenant en compte les aspects organisationnels et de sécurité ainsi que les standards d'échanges de données.

### 3.2. Définir les responsabilités pour chaque phase

**Définir les responsabilités** ▶  
Quels que soient les choix de chaque organisation, il est nécessaire de bien définir les responsabilités pour chaque phase.

Les documents et les bases de données sont d'une importance vitale pour une administration publique comme pour toute organisation. L'information constitue la matière première pour le fonctionnement des organisations. La définition d'une stratégie à court, à moyen et à long terme pour le traitement, la conservation et l'accès à l'information constitue un des enjeux majeurs qui conditionnent la réussite de la mission d'un organisme.

La répartition des responsabilités est différente d'un pays à l'autre et d'une organisation à l'autre. En fait, quels que soient les choix opérés, il est crucial qu'une définition claire des responsabilités résulte d'une stratégie globale plutôt que du hasard et de l'historique des rapports entre les services.



#### Répartition des responsabilités

Exemple

Les administrations nationales ou communautaires ont fait des choix différents en terme de répartition des responsabilités:

- l'Unesco préconise «*d'utiliser l'expertise des archivistes pour leur capacité à évaluer la valeur de l'information, mais de conserver la source de l'information comme le gardien physique de l'enregistrement*» (étude Unesco/RAMP);
- aux États-Unis, les enregistrements sont conservés par l'administration des archives nationales (NARA) dans plus de vingt services répartis à travers les États-Unis ainsi que par des «archives affiliées» autorisées par la NARA. Ces enregistrements relèvent de la responsabilité de la NARA tandis que les archives affiliées sont responsables de leur conservation, de leur gestion et de leur accès;
- à la Commission européenne, certains services/bureaux des archives sont responsables de la fourniture des bordereaux de diffusion pour le courrier postal dans certaines directions générales, tandis que d'autres directions générales utilisent une organisation différente.

Une bonne répartition devra prendre en compte chaque tâche, mais également la culture et le savoir-faire des organismes. Elle doit impliquer dès le départ les différents services concernés de l'administration ainsi que les archivistes tout en prenant en compte les informations en provenance du marché et de la recherche.

Lorsque l'information numérique est produite et maintenue par plusieurs organisations, l'identification des responsabilités doit se faire grâce à un dialogue entre les différentes organisations et les services d'archives.

**À quoi servent les archivistes?** ▶  
Ils disposent d'un savoir-faire précieux dans l'évaluation des enregistrements et peuvent être utiles dès les débuts de leur cycle de vie.

L'archiviste est responsable de la conservation des enregistrements. Il peut aussi fournir un savoir-faire très précieux dans l'évaluation de la valeur d'un enregistrement. Le rôle de l'archiviste passe de la réception passive des enregistrements en fin de vie à une participation active dès les premières étapes.

Une coopération plus étroite est nécessaire entre les archivistes et les personnes s'occupant de l'information numérique dans l'administration publique et dans le secteur privé.



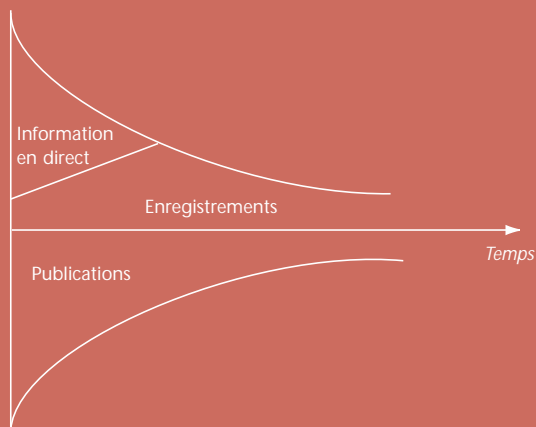
## Trois types d'information

Notions avancées

Il existe trois types d'information :

- l'information en direct. Elle est produite dans l'organisation comme information non-enregistrée. Elle peut être éphémère ou elle peut être placée dans un enregistrement;
- les enregistrements. Les archivistes sont les experts qui aident à gérer les enregistrements pendant tout leur cycle de vie. Une information peut être présentée sous la forme d'un enregistrement ou une information en direct peut être enregistrée. Les archivistes disent «saisir un enregistrement». Dans ce cas, l'information est «gelée» et le contexte devient un problème primordial pour faciliter l'accès et l'utilisation des enregistrements.;
- les publications. Les bibliothécaires sont les experts chargés de traiter ce type d'information.

Alors que le guide multidisciplinaire présente l'information et les enregistrements numériques du point de vue du déroulement des opérations pour différentes disciplines, le guide CIA sur les enregistrements électroniques (voir annexe 8.9) donne davantage de détails sur la conservation des enregistrements du point de vue des archivistes.



- ◆ Les différents types d'information requièrent des savoir-faire différents:
- l'information en direct est souvent gérée par l'auteur du document ou le gestionnaire de la base de données;
- les archivistes ont des compétences en gestion des enregistrements;
- les bibliothécaires ont des compétences en gestion des publications.

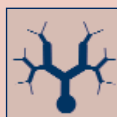
## 4. Conception, création et maintenance de l'information numérique

Chaque tâche représentée dans le cycle de vie a ses particularités. Ce chapitre décrit celles qui ont trait à la conception, à la création et à la maintenance de l'information numérique.

### 4.1. La phase de conception

**Groupe stratégique** ▶  
La mise en place dès le départ d'un groupe multidisciplinaire pour définir et suivre la stratégie de l'information numérique permet d'optimiser les résultats.

La meilleure façon d'optimiser la gestion de l'information numérique est de définir une stratégie globale et cohérente dès le départ. Une telle stratégie doit être développée en impliquant tous les acteurs concernés. Une des solutions consiste à mettre en place un groupe multidisciplinaire qui définit et maintient cette stratégie. Bien que ce ne soit pas une pratique très répandue, on peut considérer cette approche comme la plus souhaitable.



#### Définir une stratégie

Pour choisir

Un groupe multidisciplinaire responsable de la stratégie de l'information numérique peut prendre en charge les points décrits ci-dessous:

- la prise en compte des besoins des utilisateurs (y compris les aspects légaux);
- l'identification des enregistrements importants;
- l'établissement de règles pour définir un plan de classification efficace;
- la définition des normes et spécifications à utiliser pour assurer l'indépendance des données par rapport aux supports et garantir leur pérennité;
- la mise en place d'un plan d'évaluation de l'information numérique;
- l'identification des responsables pour chaque tâche du cycle de vie;
- la définition d'une politique de formation et d'information dans la ou les organisations impliquées (\*);
- le suivi de la mise en place des nouveaux systèmes.

Voir également la liste de contrôle pour la stratégie de l'information numérique à l'annexe 8.4.

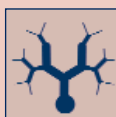
**Aspects légaux** ▶  
Il s'agit d'un des points majeurs à prendre en compte dans toute stratégie de l'information numérique.

Les aspects légaux font partie des points les plus importants dans le contexte d'une administration publique. Dans le cas de l'information numérique, un certain nombre de points sont encore mal connus et il est nécessaire d'y accorder toute son attention (voir en particulier la liste des questions en suspens à l'annexe 8.2).

### 4.2. Création de l'information numérique

Il est important de prendre en compte chaque tâche décrite dans le cycle de vie le plus tôt possible. C'est pourquoi il est nécessaire d'utiliser des règles communes lors de la création de l'information numérique. Cela permet de faciliter grandement les échanges entre les divers acteurs qui auront à les traiter par la suite.

(\*) Il serait souhaitable que l'information numérique et les enregistrements numériques deviennent un élément de la formation des archivistes et des agents des administrations publiques.



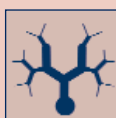
## Création de l'information numérique

Pour choisir

Les règles suivantes s'appliquent dès la création de l'information numérique:

- chaque document ou base de données doit être clairement identifié par l'organisation en charge de la gestion de l'information numérique;
- une documentation contextuelle doit être produite avec toute information numérique. Elle est ensuite maintenue par le responsable de la gestion de l'information numérique de l'étape en cours du cycle de vie;
- une procédure doit être mise en place pour traiter l'information numérique de provenance douteuse;
- aucune destruction ou modification ne doit être effectuée sans procédure d'approbation. Cela permet de garantir la conservation de l'historique (la destruction ou la mise à jour inclut la perte de possibilité de combiner les données, la perte de possibilité de reconnaître ou de retrouver une donnée et la perte de possibilité d'identification des données).

L'information peut être produite initialement sous forme de papier (puis être numérisée) ou bien directement dans un format numérique (traitement de texte, base de données, messagerie électronique, etc). Dans tous les cas de figure, elle doit être produite ou convertie dans un format standard (à moins que son authenticité ne soit mise en danger). Le chapitre 5.2 donne des détails sur les formats les plus adaptés.



## Information numérique ou documents papier?

Pour choisir

- Généralement, l'information sur papier est conservée sous forme papier. Les outils de recherche électronique peuvent aider à la gestion et à la récupération.
- Les enregistrements numériques doivent être conservés sur des supports numériques.

Les documents papier ou une partie de leur information contextuelle peuvent être scannés pour produire des enregistrements numériques afin de faciliter les recherches et leur utilisation.

### 4.3. Intégration, migration et évaluation de l'information numérique

#### 4.3.1. Intégration de l'information

La réorganisation, la sélection et le regroupement de données pour créer un ensemble plus compact est une tâche délicate car il faut respecter le droit à la vie privée (voir section 6.3 sur l'anonymat).

Cependant, il peut parfois être nécessaire de consolider les données provenant de différentes organisations: les futurs chercheurs qui utiliseront des données publiques veulent pouvoir chercher par sujet aussi bien que par organisation d'origine.

#### ◆ Règles de création

Il faut respecter certaines règles dès la création de l'information numérique pour faciliter toutes les étapes ultérieures du cycle de vie.

#### ◆ Le regroupement de données

Il peut s'avérer utile pour des recherches ultérieures, mais aussi délicat.

Que doit-on inclure dans un enregistrement?

La documentation, la structure, et un ensemble cohérent de textes et données.

Il existe en effet toujours plus d'une façon de regrouper l'information. Faut-il regrouper un rapport avec ses corrections ultérieures? Avec quels documents doit être sauvegardée telle base de données? Quoi qu'il en soit, l'enregistrement doit contenir, en plus du document ou de la base de données, sa propre documentation (y compris les métadonnées) et sa structure (sous forme d'un fichier séparé ou inclus dans le fichier principal).



Exemple

### Documents composites créés à partir d'une base de données

Dans le cas d'un document créé dynamiquement par une base de données, il existe plusieurs solutions pour le conserver à long terme:

- soit «geler» la base de données avec une requête spécifique pour produire un document classique. Dans ce cas, il existe une perte de combinaisons possibles;
- soit exporter la totalité de la base de données quand le système n'est plus opérationnel, pour autant que l'application comporte une analyse rétrospective complète;
- soit conserver la base de données, mais également l'application qui génère le document dynamiquement.

Les solutions à adopter avec des documents composites, qui sont plus complexes que de simples documents, ne sont pas toujours évidentes

La taille de l'unité d'information peut varier en fonction des choix effectués. Actuellement, la multiplicité des références et des liens hypertexte entre les documents pousserait à sauvegarder l'ensemble des données de la planète pour former un seul enregistrement! D'un autre côté, les unités d'information doivent conserver une taille acceptable pour être exploitables.

#### 4.3.2. Migration de l'information

Migration du papier vers le numérique:

Deux solutions:

- scanner seulement,
- scanner et coder

Lorsque l'on souhaite faire «migrer» un document papier vers un format numérique, il existe deux solutions principales:

- scanner le document pour en obtenir une image;
- scanner le document puis le coder dans un format électronique (par exemple, avec la reconnaissance optique des caractères ou la vectorisation des graphiques comme décrit dans la section 4.3.4).

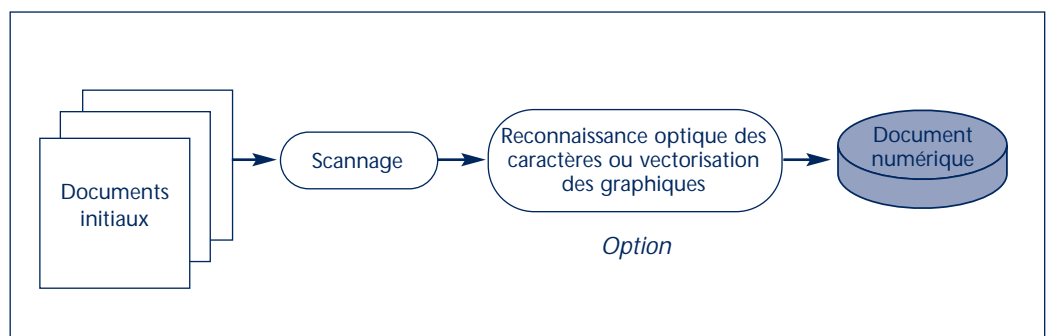
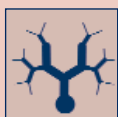


Figure 3 — Migration d'un document papier vers un document numérique



## Coder ou ne pas coder?

Pour choisir

Pour les textes, le scannage simple est plus facile et ne demande pas de traitement supplémentaire comme c'est le cas pour la reconnaissance optique des caractères.

Cependant, la taille du fichier à conserver est bien plus grande (plus de 50 kilooctets par page à comparer à quelques kilooctets dans le second cas). De plus, il est bien plus facile d'éditer un texte composé de caractères qu'une image représentant ce texte. L'indexation pour établir des références croisées est également bien plus simple.

Une autre possibilité consiste à mixer les deux solutions. Il est ainsi possible de conserver les images dans leur format initial tout en passant les parties de texte dans un système de reconnaissance optique des caractères. Des indications détaillées sont données sur les formats adaptés à la conservation à court et à long terme (chapitre 5 – Conservation de l'information numérique) ainsi qu'à l'accès aux données (chapitre 6 – Accès et diffusion de l'information).

Il existe un troisième type de migration qui transforme un format numérique en un autre. Ce changement peut intervenir seul lorsqu'il s'agit de transformer le format d'un enregistrement existant en un format plus standard et plus pérenne ou de le transformer dans un standard qui offre plus de possibilités (par exemple, ajouter une structure à un fichier non structuré). Il peut également intervenir comme une troisième étape permettant d'obtenir depuis un document papier un enregistrement dans un format numérique plus structuré (par exemple, un document avec une structure explicite ou une base de données).

Bien que la migration des enregistrements papier vers des formats numériques concerne principalement les documents, cette troisième étape permet dans certains cas de donner une structure de base de données à l'enregistrement numérique obtenu (voir section 4.3.3).

### ◆ Migration numérique

Les changements de format numérique concernent tous les types d'enregistrements.

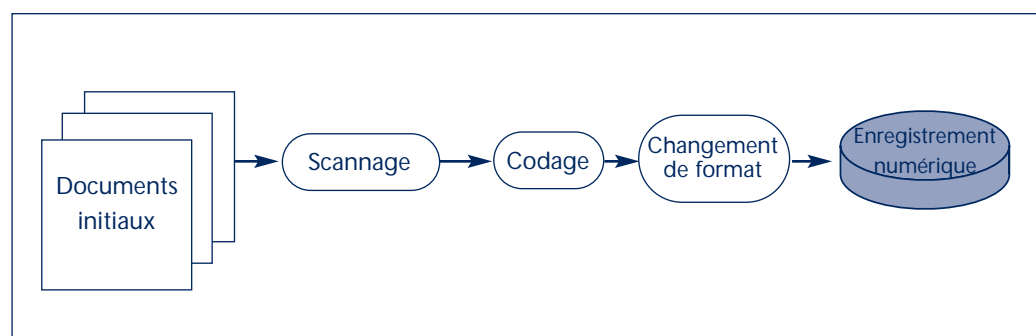


Figure 4 — Changement de format numérique

#### 4.3.3. Du papier ou microfilm à l'image scannée



### Scanneurs et télécopieurs

Notions de base

Le scannage de document permet de retranscrire le contenu de pages de papier sous la forme d'un fichier informatique. Le fichier obtenu représente une image du document initial avec une résolution plus ou moins grande.

La résolution d'un scanner se compte en points par pouce (dpi). À l'heure actuelle, les scanners permettent facilement une résolution de 300 ou 600 dpi en couleur.

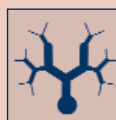
Un télécopieur est constitué d'un scanner, d'un système de transmission des données sur la ligne téléphonique et d'une imprimante à l'arrivée. La résolution du scanner d'un télécopieur est de beaucoup plus basse qualité (entre 100 et 200 dpi).

#### Scanner

Il faut respecter un ensemble de règles pour garantir un résultat exploitable.

Il est bon de respecter un certain nombre de règles de base pour scanner des documents. La qualité du résultat peut en dépendre fortement.

Les textes ne doivent être conservés sous forme de fichier image que lorsqu'ils sont courts ou lorsque la reconnaissance optique de caractères n'est pas possible (signature...). Dans les autres cas, il est préférable d'envisager le codage des textes.



### Scanner un document

Pour choisir

- Tous les éléments nécessaires de l'enregistrement initial doivent être transférés sur le même support.
- Les relations entre l'enregistrement et le reste du système d'archives doivent être maintenus (par exemple, les autres enregistrements référencés).
- La qualité initiale du document (contraste, taille des caractères...) doit être suffisante pour permettre la meilleure impression possible après migration vers un format numérique.
- Avant scannage des documents, un test doit être conduit avec un échantillon.
- Même si le scannage est sous-traité, l'organisme demandeur doit contrôler la qualité et la complétude des documents numérisés.
- Les formulaires doivent être étudiés spécifiquement pour faciliter leur migration vers des supports numériques (taille des polices de caractère, emplacement des champs).

#### 4.3.4. De l'image scannée aux formats codés

#### Reconnaissance des caractères

L'OCR permet de générer des fichiers texte à partir d'images de textes.



### Reconnaissance des caractères

Notions de base

La reconnaissance optique des caractères (en anglais Optical Character Recognition, OCR) permet à un ordinateur de « relire » un texte.

Ce logiciel utilise un fichier image représentant le texte à lire (par exemple, un fichier généré par un scanner). Il analyse alors les dessins des caractères et permet de construire un fichier au format texte, éditable ensuite à l'aide de n'importe quel traitement de texte.



La reconnaissance n'est pas parfaite. Lorsque l'ordinateur hésite sur un caractère particulier, il le surligne pour qu'un opérateur puisse ensuite relire le texte et indiquer la lettre non reconnue. Plus rarement, l'ordinateur peut substituer une lettre à une autre. Les outils de correction automatique permettent d'optimiser la phase de correction. Dans le cas d'une reconnaissance optique de caractères, un contrôle doit toujours être effectué par un opérateur après la reconnaissance automatique. Ce contrôle peut être fait en utilisant des outils logiciels.

Le taux de reconnaissance d'une image propre d'un texte typographique s'établit à partir de 95 % (soit deux à trois erreurs par ligne de texte). Ce taux peut varier en fonction du langage de la source.

Après avoir scanné un document contenant du texte et des images, il est possible d'en extraire le texte grâce à un système de reconnaissance optique des caractères. Cela permet de retravailler le texte, d'en utiliser des parties ou de l'indexer pour faciliter la consultation.



Notions avancées

## Fichiers de télécopies

Il est souvent difficile d'utiliser des systèmes de reconnaissance optique des caractères avec les télécopies, à cause de leur faible résolution et de la faible qualité du résultat sur papier.

Une solution consiste à conserver les télécopies dans leur format numérique d'origine (ITU-T Groupe III ou ITU-T Groupe IV). Bien que ce format ne permette pas le retraitement des textes, il fournit une version compressée des images et des textes qui économise la mémoire nécessaire.

Les graphiques qui apparaissent sur le document de départ peuvent être vectorisés pour économiser de la place.

Dans certains cas, comme la numérisation de plans par exemple, la vectorisation apporte une véritable plus-value. Bien sûr, la vectorisation n'est efficace que sur les graphiques et les images constituées de contours.



Notions de base

## Vectorisation des images

Les graphiques sont constitués d'éléments simples (droites, courbes, rectangles...). Plutôt que de conserver les graphiques sous la forme d'une image constituée de points comme pour les photos, la vectorisation permet de retrouver les éléments de base du graphique.

Les avantages, comme pour la reconnaissance optique des caractères, sont de deux ordres:

- la place occupée en mémoire par le graphique est bien moins importante que pour une image (seuls les types d'éléments ainsi que les coordonnées de départ et d'arrivée sont conservés);
- il devient plus facile d'éditer le graphique ou d'en utiliser des sous-ensembles.

Lors de la phase de codage des données (reconnaissance des caractères ou vectorisation des graphiques), il est nécessaire de suivre quelques règles de base, comme celles proposées dans les encadrés de ce chapitre. Cela facilite le traitement de l'information numérique à tous les stades ultérieurs de son cycle de vie.

### ◆ Les télécopies

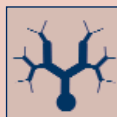
La reconnaissance des caractères est peu efficace sur les télécopies

### ◆ La vectorisation

Elle est efficace sur les graphiques, et économise de la mémoire.

### ◆ Codage

Quelques règles simples permettent d'obtenir un résultat de qualité.



## Codage des textes et graphiques

Pour choisir

- Les éléments d'un document qui ne sont pas codés doivent être conservés séparément dans un autre format adapté (par exemple, des photos ou des images).

### 4.3.5. D'un format numérique à l'autre

Il existe deux cas de figure où la transformation d'un enregistrement d'un format numérique dans un autre présente un intérêt:

- la migration d'un enregistrement dans un format plus pérenne pour faciliter sa relecture et sa consultation à long terme;
- l'ajout d'une structure à un «texte plat» (texte non structuré). Cela permet d'obtenir un document structuré ou une base de données plus aisément consultable.

#### Migration vers un format standard

La première utilité de changer le format d'une information numérique est d'assurer une plus grande pérennité.

Lorsqu'il s'agit de «migrer» l'information numérique d'un format vers un autre, il est important de prendre soin de ne pas perdre de données de façon involontaire. En effet, les fonctionnalités proposées dans les formats propriétaires et les formats standard ne correspondent pas toujours. Les deux exemples donnés ci-dessous illustrent cette difficulté.

Les formats standard pour la conservation et l'accès à l'information numérique sont présentés dans les chapitres 5 et 6.



## Pertes d'information

Exemple

### Exemple de perte d'information dans un document

Un document présente le texte suivant:

«Il est recommandé de clore ce projet (!)»

Une note de bas de page ajoute:

«(!) Sauf si les budgets initialement prévus sont finalement alloués.»

Le document initialement dans un format de traitement de texte propriétaire, est sauvegardé sans précautions dans un nouveau format standard. Si les notes de bas de page sont ignorées, le sens de la recommandation peut être grandement changé!

### Exemple de perte d'information dans une base de données

Une base de données dans un programme de gestion existe dans un format propriétaire propre à l'application. L'application étant très liée à un système d'exploitation et à une plateforme informatique propriétaire, il est décidé d'extraire les informations de la base pour les conserver dans un format indépendant. Si aucune précaution n'est prise, les relations entre les comptes avant et après modification du plan comptable peuvent être perdues.

Un deuxième cas de figure pousse à «migrer» d'un format numérique vers un autre. Lorsque l'on dispose d'informations non structurées, appelées souvent «texte plat», il peut être intéressant de leur ajouter une structure. C'est le cas souvent d'un texte scanné, auquel on souhaite ajouter des indications structurelles (table des matières, index...) afin de faciliter sa consultation. Il existe aujourd'hui des applications qui peuvent aider à redéfinir la structure d'un document en utilisant, par exemple, les types de polices de caractère utilisées pour les différents niveaux de titre.

Il est important, lors de la restructuration de documents scannés, de relier les éléments qui sont séparés par des sauts de page ou de colonne ou par l'insertion d'un graphique, d'une table, etc.

C'est également le cas de bases de données formatées sous la forme de textes plats (scannage d'annuaires ou de formulaires, anciennes bases de données conservées dans un format de bas niveau). Dans ce cas, la structure est souvent implicite, indiquée par des séparateurs (par exemple, des tabulations ou des points – virgules) ou par la place des champs sur la page scannée (colonne, place du champ d'un formulaire papier...).

Ce dernier cas de figure est d'autant plus important qu'il n'existe pas de norme officielle pour les formats de données structurées et qu'il est souvent nécessaire de conserver les données dans un format de bas niveau peu structuré (voir en particulier la section 5.2.4 sur les formats de données).

#### 4.3.6. Évaluation de l'information électronique

Conserver l'information n'a que peu d'intérêt s'il n'est plus possible de l'utiliser lorsque cela est nécessaire. Quand des informations doivent être saisies dans un enregistrement pour faire la preuve d'une activité, leur évaluation est particulièrement importante. Il faut évaluer avec soin leur authenticité, leur intégrité et leur valeur.

#### ◆ Structurer l'information

Il est possible d'ajouter une structure à un «texte plat» pour obtenir un document structuré ou une base de données.

#### ◆ Contrôler les enregistrements

La plupart des problèmes arrivent lors de la migration ou de la transmission de l'information

#### ◆ Éliminer des enregistrements

L'évaluation des enregistrements qui n'ont plus aucune utilité ou aucune valeur est une tâche importante.

#### ◆ Lignes directrices sur l'élimination

Définition de la meilleure façon d'éliminer des enregistrements par les services d'archives régionaux/nationaux.



### Élimination

Notions avancées

L'un des éléments les plus importants de l'évaluation est l'identification de l'information ou des enregistrements à détruire. Il n'est ni possible ni utile de tout conserver. Il faut détruire les enregistrements dès qu'ils n'ont plus aucune utilité ou valeur. Cela facilite l'accès aux enregistrements qui restent.

Cela vaut aussi pour certains éléments induits ou de peu de valeur à long terme qui se trouvent dans les bases de données. La suppression de ces éléments et le maintien de l'équilibre accroissent la vitesse d'accès et de traitement.

Il faut faire appel à la compétence des archivistes pour choisir les enregistrements à détruire.



### Lignes directrices sur l'élimination

Exemple

L'unité de stratégie de l'information (ISU) du cabinet du Premier ministre a mis au point, en association avec le bureau des archives de Tasmanie (Australie), des lignes directrices sur la procédure à suivre pour éliminer des enregistrements:

- identification des enregistrements qui ont une valeur persistante ou à long terme;
- les enregistrements temporaires doivent avoir une période de conservation;
- mise en place d'«autorités permanentes d'élimination» qui fournissent une autorisation légale continue pour la suppression d'enregistrements après un délai spécifié.

L'annexe n° 1 relative à l'élimination réglemente l'élimination des enregistrements administratifs communs à la plupart des organismes gouvernementaux.

La section 11 de l'annexe 1 autorise l'élimination de la documentation éphémère, de routine ou en double, qui n'est pas destinée aux archives officielles ou qui ne doit pas faire l'objet d'un enregistrement officiel, à savoir:

- les notes et les calculs qui sont utilisés uniquement pour préparer d'autres enregistrements comme la correspondance, des rapports, des documents budgétaires, des programmes et des tables,
- la documentation préliminaire qui n'est pas destinée à être conservée dans les archives de l'organisme gouvernemental et dont le contenu est intégré dans le système de tenue des archives ou enregistré comme un document commun. Il peut s'agir de brouillons et de premières moutures. Cependant, il peut être nécessaire, pour justifier et pour étayer l'évolution politique, de garder les brouillons des groupes de travail quand des commentaires sur ceux-ci ont entraîné de gros changements. Les petits changements rédactionnels ne sont pas considérés comme importants;
- les copies d'enregistrements communs ou d'enregistrements conservés à un autre endroit dans l'organisme.



## Contrôle de l'information

Pour choisir

Plusieurs contrôles peuvent être effectués sur l'information numérique:

- conformité aux normes et standards définis;
- protection en écriture;
- lisibilité des supports;
- comparaison avec les informations incluses dans la documentation (par exemple en essayant d'imprimer les premiers enregistrements);
- complétude du contenu.

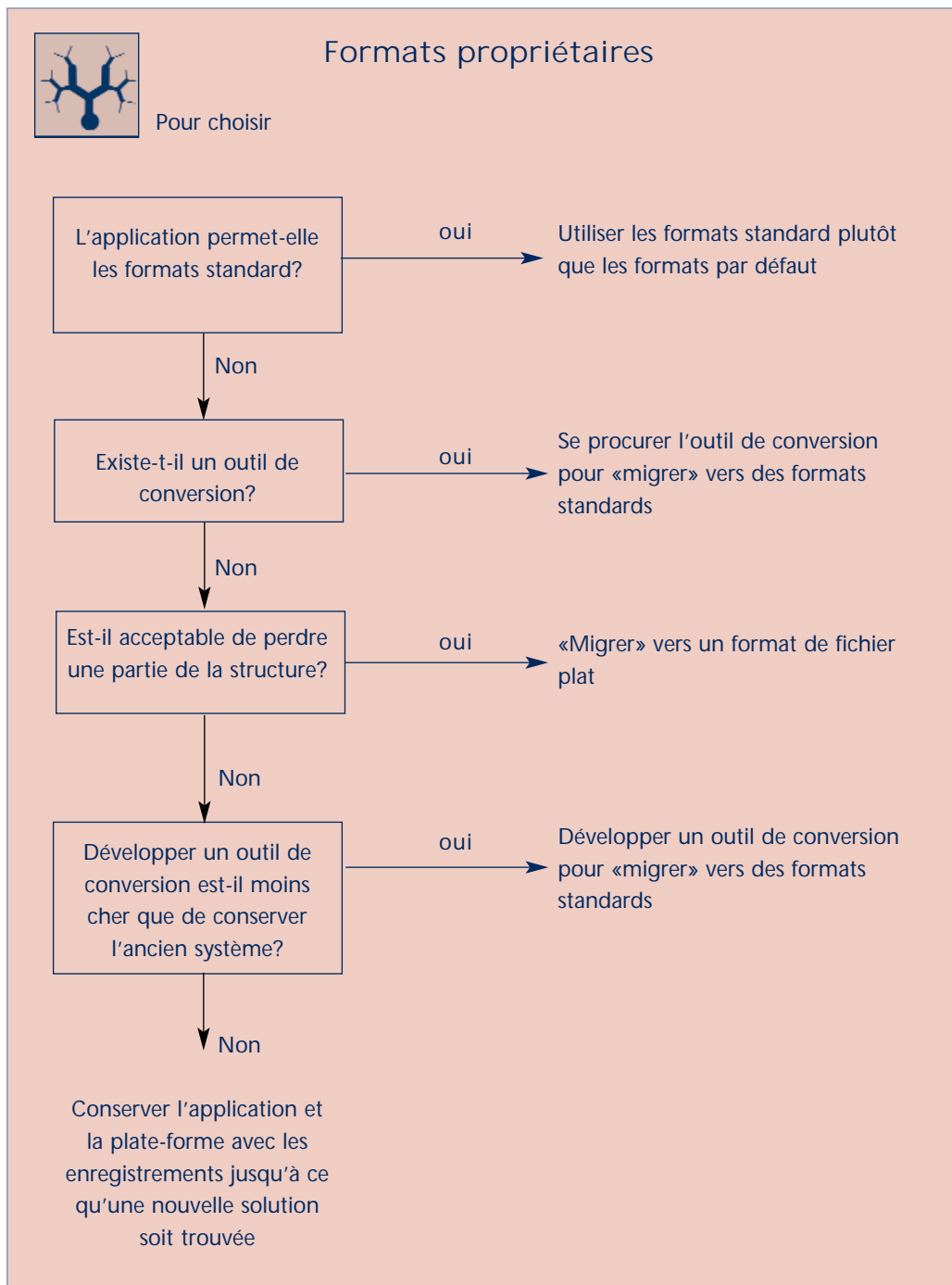
Cependant, seules certaines parties sont en général testées. Il arrive donc qu'un problème soit identifié lors d'un accès courant à un enregistrement (durant la phase d'accès et d'utilisation de son cycle de vie). Si l'information a été fournie par une autre organisation ou un autre service, celui-ci est prévenu afin qu'il essaye de reconstruire l'information initiale lorsque cela est possible. La plupart des problèmes arrivent lors des phases de migration (décrites dans cette section) ou lors de la transmission à un autre organisme (voir section 4.5), plutôt que lors de la conservation de l'information.

**Formats propriétaires** ▶  
Comment faire lorsqu'une application génère des données avec des formats propriétaires?

Lorsque l'application utilisée génère des données dans un format propriétaire, il peut être nécessaire de conserver le système complet pour pouvoir accéder à l'information. Cela inclut l'application elle-même, bien sûr, mais également la plate-forme informatique, la documentation et même le personnel ayant le savoir-faire pour faire tourner l'application et la plate-forme. Le coût d'une telle solution doit être étudié avec soin en fonction de la valeur intrinsèque de l'information.

Une alternative consiste à transférer l'information dans un format de bas niveau (un texte plat pour un document ou un fichier séquentiel plat pour une base de données). Dans ce cas, certaines informations risquent d'être perdues, en particulier des éléments de structure.

Une troisième solution consiste à «migrer» l'information dans un nouveau format plus standard. Le développement (ou l'achat) d'un outil de migration représente un certain coût. Lorsque les données doivent être conservées de manière à garder intégralement la structure, cette solution peut présenter un intérêt. Bien évidemment, le format cible dans lequel sont converties les données doit être le plus normalisé possible pour permettre une meilleure pérennité.



♦ **Choix d'une solution**  
 Il existe plusieurs solutions pour conserver à long terme des données. Le choix de la meilleure solution dépend de plusieurs critères

Figure 5 – Que faire des formats propriétaires ?

Il faut signaler que bien que le coût de migration de l'information numérique vers un nouveau format puisse être élevé, ne pas le faire peut avoir un coût encore plus important.

## 4.4. Gestion et classement de l'information numérique

### Gérer l'information numérique

- enregistrer la nouvelle information numérique,
- affecter un responsable,
  - coordonner les traitements,
  - classer l'information numérique,
  - décider du transfert.

Le but de la gestion de l'information numérique est de préserver la fiabilité, l'authenticité, l'intégrité et la testabilité de l'information dans le temps. Il faut pour cela que le contexte de l'information soit bien défini. Quand la teneur, le contexte et la structure suffisent pour constituer une preuve d'une activité, l'information devient un enregistrement.

Des tâches supplémentaires sont nécessaires pour gérer le cycle de vie de l'information numérique. La responsabilité de la gestion d'un groupe particulier d'informations numériques peut être transférée à un autre organisme ou à un autre service (voir section 4.5 sur le transfert).

La gestion de l'information numérique inclut les tâches suivantes:

- enregistrement de la prise en charge de l'information numérique. Cela nécessite la mise à jour de l'historique de l'information numérique;
- affectation. Lorsque plusieurs organismes ou services sont concernés, l'information numérique doit être affectée au bon organisme ou service et un gestionnaire de l'information numérique doit être identifié;
- suivi. Cette tâche consiste à coordonner les différentes phases de traitement (réception de l'information numérique, migration, conservation, utilisation, transfert à une autre organisation);
- classement. Il s'agit de faciliter la recherche d'une information numérique particulière (voir ci-dessous);
- décision de transfert à une autre organisation ou un autre service.

Un nombre croissant de groupes de personnes sont impliqués dans la gestion de l'information numérique. Les outils de flux des tâches (workflow) peuvent être d'une grande aide pour gérer les échanges au sein d'un groupe de coopération.

### Classification de l'information

Classer l'information représente l'une des tâches de gestion les plus importantes.

L'une des tâches les plus importantes, et peut-être la plus complexe, consiste à classer l'information. Le schéma de codification doit être suffisamment clair pour être compréhensible par d'autres organisations (en particulier si la responsabilité de l'information numérique doit être transférée à un autre service ou à une autre organisation).

En fait, la structure permet de retrouver une information particulière dans un document ou une base de données, alors que la classification permet de la retrouver dans l'ensemble des informations conservées.

Il faut utiliser le plan général de classement quel que soit le support de l'information. Des critères supplémentaires peuvent être utilisés pour indexer l'information afin de tenir compte de la spécificité de l'information numérique.



### Commission européenne

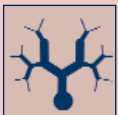
#### Exemple

Certains services de la Commission européenne utilisent un plan de classification de base qui fait référence au fonctionnement de cette institution, à l'administration, au personnel et au budget.

Plusieurs critères peuvent être utilisés pour classer et indexer l'information numérique. Ces critères sont particulièrement utiles dans le cas des messages électroniques.

- Type de document
- Dates (production, expiration)
- Auteur(s) (personnes, services)
- Signataire
- Destination (personnes, services)
- Copies (personnes, services)
- Numéro de l'information numérique, version
- Affectation (date, dossier, service...)
- Sujet
- Projet ou activité
- Mots clés
- Langage
- Nombre de pages
- Statut (officiel, non officiel), confidentialité
- Documents attachés, liens vers d'autres documents
- Autres catégories définies par l'utilisateur

◆ **Plan de classement**  
Plusieurs critères pour classer l'information numérique



## Description de l'archivage d'un enregistrement

Pour choisir

Le Conseil international des archives a rédigé des normes internationales pour la description de l'archivage (ISAD/G) dont il faut tenir compte pour définir une stratégie de classement.

L'identification de l'unité d'archivage de l'enregistrement est appelé «domaine de déclaration d'identité». Il est recommandé de ne pas le modifier pendant le cycle de vie de l'enregistrement.



## Courrier électronique

Notions avancées

Le courrier électronique peut être géré de différentes façons en fonction du contenu du courrier:

- il doit être traité comme une lettre «électronique» quand le contenu peut être archivé et considéré comme une déclaration «officielle»;
- il doit être traité comme une discussion informelle (par exemple, un appel téléphonique) quand le courrier n'a aucune utilité ultérieure comme c'est le cas pour une réponse automatique expliquant que la personne «n'est pas là et qu'elle lira son courrier dès son retour».

Comme le courrier électronique autorise aussi bien les lettres électroniques officielles que les discussions informelles, il est parfois difficile de décider comment les gérer. La compétence des archivistes en matière d'évaluation contribuera grandement à définir de nouvelles règles de gestion du courrier électronique.

Les archives australiennes donnent quelques exemples de documents éphémères et de routine dont l'enregistrement n'est pas nécessaire (voir les exemples dans l'encadré de la section 4.3.6).

◆ **Courrier électronique**  
Il doit être traité comme:  
• une lettre électronique  
ou  
• une discussion informelle  
en fonction de son contenu.

## 4.5. Transfert

À la fin de la partie active de leur cycle de vie, les enregistrements peuvent être transmis aux archives. Tous les enregistrements ne valent pas la peine d'être archivés. Les enregistrements sans utilité ou sans valeur doivent être éliminés avec l'aide des archivistes (voir l'élimination dans l'encadré de la section 4.3.6).

Deux types de transmission:

- la transmission physique,
- le transfert de responsabilité.

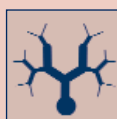
Il existe deux types principaux de transmission:

- la transmission physique d'une information numérique;
- le transfert de la seule responsabilité.

L'information est généralement transmise sous la forme d'un enregistrement (après avoir été saisi comme un enregistrement). Celui-ci peut être transmis à une autre organisation /un autre service ou transféré à des services d'archives.

Il est possible de séparer les fonctions de responsable de l'information numérique (responsable en particulier de sa gestion) de celles de conservateur de celle-ci. L'information numérique peut être conservée dans son service d'origine ou dans un organisme spécialisé. La facilité de copie ou de transfert par les moyens de communication électroniques ouvre la voie à la séparation des rôles.

Dans tous les cas, il est nécessaire d'effectuer les contrôles **AVANT** le transfert de l'information numérique.



### Réussir un transfert physique

Pour choisir

Les clés de la réussite d'un transfert physique reposent sur quelques règles simples:

- l'information doit être fournie dans son intégralité (y compris le contexte);
- les responsabilités doivent être clairement définies dans l'organisme fournisseur comme dans celui qui reçoit;
- lors de chaque transfert, les deux organisations (celle qui fournit l'information numérique et celle qui la reçoit) doivent s'assurer que l'information n'est pas altérée sans approbation préalable.



### Royaume-Uni

Exemple

Le *Public Record Office* au Royaume-Uni propose actuellement une nouvelle stratégie pour le transfert des enregistrements numériques.

Le transfert est effectué uniquement au travers d'un réseau électronique sécurisé. L'enregistrement est donc indépendant du support utilisé par le fournisseur.

L'organisme qui reçoit les enregistrements à conserver (dans ce cas le *Public Record Office*) va sélectionner un seul type de support pour archiver les enregistrements. Il conservera un contrôle sur cette technologie unique de façon à garantir la pérennité.



L'ensemble transféré comprend plusieurs éléments qui sont conservés conjointement:

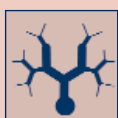
- l'enregistrement numérique lui-même;
- un audit;
- les métadonnées;
- un visualisateur fonctionnant sur un PC;
- un butineur fonctionnant également sur un PC.

Les micros PC sont considérés comme ayant une compatibilité ascendante suffisante pour garantir la relecture de l'enregistrement dans le futur. Chaque enregistrement est donc autonome et indépendant de tout problème de format.

Les contrôles effectués avant le transfert permettent de garantir la meilleure lisibilité des enregistrements.

#### ◆ Contrôler les enregistrements

Les tests effectués garantissent une bonne qualité du transfert.



### Contrôles lors du transfert

Pour choisir

La liste indiquée ci-dessous présente un exemple des étapes que l'organisation transmettrice doit suivre pour contrôler l'information avant son transfert. Cela doit se faire en étroite collaboration avec l'organisation réceptrice (service d'archives ou autre):

1. faire deux copies des données;
2. comparer les données avec la documentation fournie à l'aide d'un programme de statistique;
3. identifier et documenter les erreurs dans la documentation;
4. d'autres contrôles peuvent être effectués comme des vérifications inter-enregistrements et intervariables pour vérifier la consistance des données;
5. consulter le fournisseur de l'information numérique si l'identification des codes pose des problèmes ou s'il y a des erreurs et des inconsistances dans les données;
6. documenter les fichiers matériels en indiquant toutes les difficultés rencontrées.

## 5. Conservation à long et à court terme de l'information numérique

### Supports et formats de fichiers

Deux éléments clés pour la conservation à long terme.

La longévité physique des supports numériques est un point important. Il doit être complété par la longévité technologique. La conservation physique d'enregistrements numériques ne sert à rien si les matériels et logiciels existants ne peuvent plus traduire les données qu'ils contiennent pour les transformer en information utile.

Le support étant différent du contenu dans les enregistrements numériques, nous aborderons successivement les normes et standards pour les supports et les formats de fichiers. Nous nous concentrerons sur la longévité des supports et sur la maturité et la pérennité des standards.

### 5.1. Supports de données

#### Trois types de supports

Outre le papier:

- microfilm,
- support magnétique,
- support optique.

Il existe de nombreux supports utiles. Certains sont plus adaptés à une conservation à court terme, alors que d'autres sont plus orientés vers une conservation à long terme.

Outre le papier, on peut considérer qu'il existe trois grandes familles de supports (microfilms, supports magnétiques et supports optiques) et de nombreuses ramifications. D'autres types de supports moins connus ne seront pas décrits ici (bandes papier...).



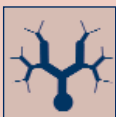
### Supports de conservation

#### Notions de base

- Les microfilms: ils sont largement répandus dans les archives, mais ne permettent pas aisément le retraitement ou la recherche à l'intérieur des enregistrements. Des normes existent et la longévité est excellente.
- Les supports magnétiques: assez anciens. Ils utilisent la polarisation de particules magnétiques dans un sens ou dans l'autre pour stocker chaque bit. Il s'agit en général de la sauvegarde de bandes à accès séquentiel.
- Les supports optiques: ce sont les supports les plus récents. Ils utilisent la déviation (ou non) d'un rayon lumineux par une microcuvette creusée dans le support pour indiquer la présence ou non d'un bit. Il s'agit en général de disques qui permettent un accès direct (accès plus rapide à l'information qu'avec une bande). Les densités obtenues sont importantes.

La bande magnétique et le microfilm sont très utilisés pour la conservation à long terme. Les supports optiques le sont de plus en plus, car ils sont particulièrement bien adaptés à cette conservation à long terme.

Les sauvegardes pour une conservation à plus court terme peuvent utiliser des supports plus variés, les aspects de pérennité et de longévité étant moins aigus.



## Supports magnétiques

Pour choisir

Les divers supports magnétiques sont plus ou moins normalisés et permettent des longévités variables. Les principaux supports sont décrits ci-dessous.

Disquette: la disquette 3' 1/2 est bien normalisée et accessible sur de nombreuses plates-formes (PC mais également Mac et Unix). Elle ne permet de conserver qu'une faible quantité d'informations (1,44 Mo en général) et sa faible longévité la réserve à la conservation à très court terme et à l'échange de fichiers.

Cartouche magnétique: la cartouche est très utilisée pour la sauvegarde sur moyen système. Il existe la cartouche quart de pouce et la cartouche demi pouce proposées par IBM.

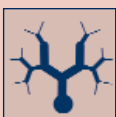
Bande magnétique: la bande 1600 bpi est lisible sur pratiquement tous les dérouleurs. Elle a été reconnue par X/Open comme format d'échange. La bande 6250, qui a une capacité de 112,5 Mo, est très utilisée pour les anciennes archives. Les bandes doivent être rebobinées tous les deux ans et réécrites tous les dix ou quinze ans sur un nouveau support (du même type ou d'un type différent).

Cassette vidéo 8mm: bien que normalisée, il n'existe pourtant qu'un seul fournisseur principal. La capacité actuelle typique est de 2,3 Go. La cassette doit être réécrite tous les deux ans.

Cassette DAT: normalisée, la cassette DAT (Digital Audio Tape) est principalement utilisée pour l'enregistrement audio. Elle est cependant également utilisée dans les technologies de l'information. La cassette DAT doit être réécrite tous les deux ans.

- Il existe plusieurs formats entièrement propriétaires pour le moment, notamment les disquettes à haute densité et les disques durs amovibles. De plus, la mémoire des cartes de crédit, bien qu'elle soit de capacité limitée, peut devenir un important support de conservation à l'avenir.

L'évolution rapide des supports optiques devrait leur permettre de s'imposer pour la préservation à long terme, alors que la supériorité actuelle des disques magnétiques en temps d'accès leur donne un avantage comme mémoire de masse.



## Supports optiques


Pour choisir


Les principaux supports optiques sont décrits ci-dessous.


CD-ROM: le CD-ROM est normalisé, c'est une bonne solution pour les nouvelles archives. Il existe une différence entre les CD-ROM pressés (plus adaptés à un grand nombre de copies mais plus fiables) et les CD-ROM gravés à l'unité (moins chers pour un faible nombre de copies mais moins fiables, sauf si une couche de vernis y est déposée ensuite). Les CD-ROM doivent être réécrits tous les dix ou vingt ans et ne nécessitent pas un environnement de stockage contrôlé, comme c'est le cas pour les bandes magnétiques.

◆ **Supports magnétiques**  
Il s'agit le plus souvent de bandes ou de cartouches.

◆ **Le disque optique**  
Il permet de grandes capacités et un accès direct contrairement à la bande.

 DVD (en anglais *Digital Versatile Disk* – disque numérique polyvalent): ce nouveau type de disque pourrait devenir un excellent support d'archivage à l'avenir. Il dispose d'une grande capacité (4,7 à 18 Go) et devrait bénéficier d'une large diffusion. Les lecteurs de DVD peuvent également lire les CD-ROM (seuls les lecteurs DVD plus récents sont en mesure de lire les CD-ROM gravés).

 Disque inscriptible WORM: (en anglais: *Write Once Read Many* – écriture unique plusieurs lectures) Il n'existe pas de standard et les WORM sont moins répandus. Ils doivent être réécrits tous les dix ou vingt ans.

 Disque optique réinscriptible: ces disques utilisent pour la plupart la technologie magnéto-optique qui combine les deux grandes technologies de stockage pour obtenir la rapidité, la densité et la possibilité de réécrire les données plusieurs fois. Il existe actuellement peu de standards. Ces disques doivent être réécrits tous les dix ou vingt ans.

Les logiciels propriétaires sont souvent utilisés pour avoir accès aux données stockées. Il est important de vérifier soigneusement qu'il est possible d'accéder facilement aux formats des fichiers et à la structure arborescente.

À cause de l'évolution rapide des technologies et de la longévité limitée des supports numériques, il est conseillé de réécrire périodiquement les archives numériques. Bien que cette réécriture ajoute des coûts supplémentaires, elle permet de résoudre beaucoup de problèmes posés par les formats non-standard et l'évolution des technologies. Cependant, la plupart des supports magnétiques et optiques ont des unités de détection et de correction qui permettent de réparer automatiquement les erreurs.

Lors de chaque réécriture, un choix doit être effectué pour décider:

- de garder les anciens fichiers tels quels;
- de «migrer» vers un support et/ou un format plus moderne (voir également la section 4.3.5: migration d'un format numérique à l'autre)

Les éléments dont il faut tenir compte sont non seulement les aspects financiers, mais également l'accessibilité, la lisibilité, la pérennité et la préservation de l'authenticité.

Conserver les enregistrements numériques  
Il est important de respecter quelques règles quant à l'environnement de conservation.



## Conservation à long terme

### Exemple

De nombreuses discussions ont lieu au sein de l'ISO, de l'ANSI et de l'ICA sur la meilleure façon de conserver les enregistrements numériques. Les quelques ordres de grandeur suivants fournissent des exemples pratiques:

- température moyenne de conservation:  $18^{\circ}\text{C} \pm 5^{\circ}$ ;
- humidité relative de conservation:  $40\% \pm 5\%$ ;
- fréquence de réinscription: 10 ans.

ISO: Organisation internationale de normalisation

ANSI: Institut national américain de normalisation

ICA: Conseil international des archives

Il existe de nombreux types de normes et standards en fonction de ce que l'on souhaite coder. La meilleure façon de procéder est de sélectionner un ensemble commun de standards dès le début pour faciliter la circulation de l'information. Il est préférable d'utiliser les mêmes formats pour la conservation à court et à long terme.



### Les différents types de formats

#### Notions de base

Les divers formats peuvent être regroupés en quelques grandes familles en fonction des objets qu'ils traitent :

- les images en mode point: elles sont constituées d'un ensemble de points. Le scannage d'un document, par exemple, fournit une image de ce type. Elles prennent une place significative et sont habituellement compressées avant d'être conservées. Elles peuvent servir de base à un codage ultérieur (pour retrouver un texte ou un graphique) ou conservées telles quelles (par exemple, pour une photographie). Le type de compression effectué définit le type de format du fichier de l'image. Les formats de télécopie constituent un cas particulier d'images compressées;
- les graphiques vectoriels: lorsque l'on utilise des graphiques ou des images constitués uniquement du dessin des contours, il est possible d'économiser beaucoup de place en utilisant un format vectoriel qui ne conserve que les coordonnées des vecteurs utilisés (segments de droite, arcs de cercle, etc.). Ce type de format n'est par contre pas adapté aux photographies;
- les textes: ils comprennent en général trois aspects séparés:
  - le texte plat lui-même qui rassemble des caractères codés,
  - la structure du texte (par exemple, le titre, les chapitres, les termes que l'on souhaite mettre en valeur, les listes, etc.),
  - les aspects de présentation (les termes mis en valeur seront en gras ou en rouge, etc.).
- les données: les possibilités de garder des données avec leurs fonctionnalités initiales sont très variables. Il n'existe actuellement aucun moyen économiquement réaliste de maintenir la capacité de calcul des tableurs. Pour les applications basées sur des modèles d'information et de traitement, les possibilités de conservation s'améliorent. L'extraction des bases de données, effectuée conformément aux normes et aux modèles exprimés dans les normes, est la bonne façon de procéder;
- les programmes: les programmes sont moins indépendants des plates-formes informatiques que les données. Les plates-formes elles-mêmes sont de plus en plus rapidement obsolètes;
- l'audio, la vidéo et tous les autres objets qui peuvent être inclus dans un enregistrement.

#### Plusieurs types de formats:

- images en mode point,
- graphiques vectoriels,
- textes,
- données,
- programmes,
- vidéo et audio,
- etc.



### Orientations futures

#### Notions avancées

Les documents et les bases de données devraient se rapprocher de plus en plus, dans le futur, de la notion de document composé et même de document orienté objet, c'est-à-dire un document regroupant plusieurs éléments autonomes (texte, images, son, vidéo) reliés entre eux.

En attendant, les standards en développement (OLE de Microsoft, OpenDoc d'IBM et APPLE, ou même le langage JAVA de Sun) ne sont pas suffisamment stables pour permettre aujourd'hui d'utiliser des enregistrements numériques qui intégreraient leur propre traitement (lecture, navigation, etc.).

### 5.2.1. Images en mode point

#### Compression des images

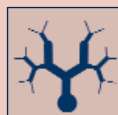
Elle peut être:

- sans perte,
- avec perte.

Deux types de compression sont utilisés pour les images:

- **la compression sans perte:** après avoir été compressée puis décompressée, une image est exactement identique à l'original. Cette spécificité impose des taux de compression bas d'environ 2;
- **la compression avec perte:** dans ce cas, les informations les moins utiles de l'image ne sont pas conservées. Notre œil, en effet, distingue nettement moins certains aspects des images que d'autres. Ce type de format permet des taux de compression plus élevés, qui dépendent du taux de dégradation acceptable des images.

Il existe de nombreux formats d'images. Il est nécessaire d'en choisir un mature et pérenne



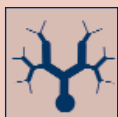
Pour choisir

## Formats d'images

Les principaux formats d'images (inclus les formats de télécopie et de vidéo) sont décrits ci-dessous.

- TIFF (*Tag Image File Format*): ce format est souvent utilisé pour les fichiers générés par les scanners. Il existe plusieurs possibilités suivant le nombre de couleurs choisies. Ce format n'engendre pas de perte mais ne permet qu'une faible compression. La dernière version 6.0 de TIFF permet d'avoir des possibilités multipages. Il faut éviter d'ajouter un codage supplémentaire en plus du format TIFF pour lui conserver une portabilité correcte (par exemple, les algorithmes de codage Packbits 32773 de l'ITU-T, le LZW ou encore le JPEG).
- GIF (*Graphics Interchange Format*): ce format vient de CompuServe et est largement répandu en particulier sur Internet. Il existe deux spécifications GIF 87A et GIF 89A. Les butineurs permettent souvent de lire les deux formats.
- JPEG (*Joint Photographic Experts Group*): cette norme internationale est de plus en plus souvent utilisée (également sur Internet). Il s'agit d'un format de compression avec perte qui permet un haut niveau de compression. Il s'agit certainement d'un bon choix en terme d'espace mémoire occupé et de pérennité.
- Télécopies: il existe deux formats de fichiers de télécopies suivant que l'on utilise une ligne téléphonique normale ou une ligne RNIS.
- Vidéo: il existe deux normes de formats vidéo:
  - MPEG-1 orienté vers les ordinateurs et le multimédia;
  - MPEG-2, plus récent, orienté vers la télévision numérique (inclus le son).
- Autres formats d'images: il peut être dangereux d'utiliser d'autres formats propriétaires pour les images (tels que BMP ou PCX) car leur pérennité n'est pas garantie à long terme. Le CD-Photo de Kodak n'est pas très utilisé pour les documents.

## 5.2.2. Graphiques vectoriels



### Formats graphiques

Pour choisir



CGM (*Computer Graphics Metafile*): il s'agit d'un format normalisé pour les graphiques vectoriels. Son utilisation est une bonne garantie de pérennité.



Formats spécifiques à des types d'applications: certaines applications nécessitent une approche spécifique avec des standards qui leur sont propres. C'est le cas, par exemple, des GIS (systèmes d'information géographiques) ou de la CAO (conception assistée par ordinateur).



Autres formats d'images: il vaut mieux éviter d'utiliser d'autres formats propriétaires (comme PICT sur Mac Intosh ou les *Windows Metafiles* de Microsoft, ainsi que de nombreux formats de logiciels particuliers), car ils n'apportent aucune garantie de pérennité à long terme.

### ◆ Les formats graphiques

Ils permettent de conserver la structure du graphique.

## 5.2.3 Textes

Il existe plusieurs types de texte, suivant que la structure et/ou la présentation sont conservés:

- un texte plat ne contenant qu'une suite de caractères est un fichier de bas niveau dans lequel il est difficile de naviguer à cause de la perte de la structure;
- un texte plat associé à une structure est très utile pour la navigation et fournit un fichier indépendant de la plate-forme matérielle;
- un texte complet (caractères, structure et présentation) n'est pas indépendant du matériel utilisé pour le lire (couleur ou non; texte prévu pour écran ou imprimante...).

### ◆ Un texte

Un texte incluant la structure mais non la présentation est indépendant de la plate-forme matérielle.



### Jeux de caractères codés

Notions avancées

Il existe trois grandes familles principales de formats codés de jeux de caractères

- l'ISO 646 équivalent (ou presque) à l'ASCII. Cette norme codée sur 7 bits ne permet pas de gérer les caractères spécifiques à l'Europe (par exemple, les accents);
- les jeux de caractères 8 bits. Il s'agit d'un surensemble de l'ISO 646. Il existe deux normes utiles pour l'Union européenne: ISO/IEC 8859-1 pour l'Europe de l'Ouest et ISO/IEC 8859-7 pour la Grèce. D'autres jeux existent pour l'arabe, l'hébreu et le cyrillique;
- le jeu de caractères universel (UCS) (*Universal Character Sets*). Ce jeu normalisé (ISO/IEC 10646) propose de coder la majorité des caractères et idéogrammes utilisés dans le monde à l'aide principalement de 2 octets (UCS-2 connu également sous le nom de Unicode) ou de 4 octets (UCS-4). Un codage supplémentaire à nombre d'octets variable (nommé UTF), facilite l'échange de caractères multioctets entre les machines car les caractères de contrôle 8 bits n'apparaissent jamais dans les séquences d'octets (par exemple, le code de contrôle de fin de chaîne est souvent 00. En UTF, aucun caractère autre que celui-ci ne comporte 00 parmi ses octets, alors que le caractère A, par exemple, est codé 00 64 en UCS-2).

### ◆ Des jeux de caractères normalisés

Il en existe pour répondre à des besoins et des cultures variés.

### La présentation ▶

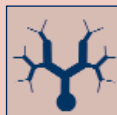
Il vaut mieux conserver la structure d'un texte plutôt que le style associé à chaque élément.

Beaucoup d'aspects de la présentation d'un texte sont dépendants de la plate-forme utilisée. Par exemple, il n'y a pas de sens à proposer un caractère clignotant sur une impression papier! En fait, la technologie évolue rapidement et les textes conservés peuvent être lus dans le futur avec des moyens qui ne sont pas encore conçus aujourd'hui.

Sauvegarder la présentation (par exemple les polices de caractères ou les macros d'un texte) pose des problèmes d'utilisation pour les textes conservés à long terme. La meilleure solution consiste à ne conserver que la structure du texte (ceci est un titre de premier niveau) plutôt que le style qui lui est associé (les titres de premier niveau sont en Arial/Helvetica 16 points). C'est au programme de visualisation qu'il revient de choisir un style pour les différentes parties du texte.

### Les textes structurés ▶

Il existe des normes et de nombreux formats propriétaires.



Pour choisir

## Les textes structurés

Il existe plusieurs normes et standards qui permettent de sauvegarder un texte structuré.

- SGML (*Structured General Mark-up Language*): Il ne s'agit pas d'une structure à proprement parler, mais plutôt d'un langage de balisage. Cette norme internationale est de plus en plus utilisée. Elle permet de conserver un texte et sa structure mais sans la présentation. Il existe des normes additionnelles telles que:
  - DSSSL: sémantique et spécifications des feuilles de style (*Document Style Semantics and Specification Language*);
  - SPDL: langage de description de page normalisé basé sur PostScript (*Standard Page Description Language*);
  - SDIF: format d'échanges de documents SGML (*SGML Document Interchange Formats*);
  - échange d'information sur les polices de caractères (*Font Information Interchange*);
  - DTD: définition de la structure des documents (il en existe plusieurs suivant les types de documents) (*Document Type Definition*).
- HyTime: il s'agit d'une extension de SGML qui permet à des documents composites d'inclure des objets multimédias.
- HTML: langage de balisage Hypertexte (*Hyper Text Mark-up Language*). Ce langage de balisage se présente comme une implémentation simplifiée de SGML. Il est particulièrement répandu sur les Internet sites. Cependant, il est encore peu stable et peu adapté aux documents longs. Il est préférable d'utiliser SGML pour la conservation à long terme des documents.
- ODA (*Office Document Architecture*): cette norme internationale permet de grouper dans un même fichier le texte, la structure et la présentation. ODA a été développé pour être utilisé dans la bureautique. Il n'est pas indépendant de la plate-forme utilisée.
- RTF (*Rich Text Format*): ce format est principalement utilisé par l'ensemble de logiciels Office de Microsoft. Il n'apporte aucune garantie de stabilité et de pérennité. Il est préférable d'utiliser les filtres d'exportation ou les outils de conversion vers des formats standard.
- PostScript: cette norme de description de page d'Adobe est largement utilisée pour envoyer ou imprimer des textes avec leur présentation. Il doit être aujourd'hui remplacé par des normes ouvertes.
- PDF: format de documents portable. Il permet la consultation uniquement de documents sur diverses plates-formes.

Beaucoup d'autres formats propriétaires existent.





## Applications spécifiques

Notions avancées

La structure des textes peut être spécifique avec des champs supplémentaires pour faciliter l'indexation. C'est le cas en particulier des messages électroniques.

Des normes ont émergé pour faciliter l'échange d'informations structurées, principalement pour les échanges commerciaux. Il s'agit de l'EDI (échange de données informatisées – *Electronic Data Interchange*) avec principalement la norme Edifact. En plus des définitions génériques d'EDI, des EDI spécifiques ont été créées dans divers secteurs (banque, automobile...). Bien que l'approche EDI soit tout à fait intéressante pour les échanges formels entre les organisations, la normalisation de l'EDI est lente et ne prend pas toujours en compte l'évolution rapide des technologies (comme les images, le son et la vidéo).

♦ EDI et messagerie électronique  
Deux applications particulières utilisant des structures particulières

### 5.2.4. Données et programmes

Il n'existe pas aujourd'hui de format normalisé de haut niveau pour les fichiers de données utilisés par les tableurs et les bases de données. Cela veut dire que pour être sûr de pouvoir lire les données après une longue période, il faut disposer d'un outil capable de lire l'ancien format ou conserver le logiciel lui-même.

En effet, il ne faut pas confondre l'interface entre le programme et les données (SQL par exemple, dans le cas d'une base de données) et le format du fichier où sont stockées les données.

Le problème est que les programmes sont moins indépendants des plates-formes matérielles que les données.

Lorsqu'il n'existe pas de format standard pour un type de données, la meilleure solution consiste:

- à utiliser un format propriétaire largement répandu qui puisse être relus par de nombreux outils ou
- à prévoir une stratégie de migration (ou de conservation du logiciel avec les données). La section 4.3 donne des éléments pour définir une telle stratégie.

Un problème similaire est posé par la conservation des programmes. Il est alors nécessaire de conserver le programme source ou de garder une plate-forme matérielle capable de faire tourner le programme.

♦ Les données et les programmes  
Ils posent un problème complexe car il n'existe pas de norme de format largement répandue pour eux.

## Different types de données

Ils conduisent à des choix différents:

- tableur,
- base de données,
- comptabilité / gestion,
- formulaire,
- signature numérique.



Pour choisir

## Fichiers de données

Il existe de nombreux types de fichiers de données qui peuvent être conservés. Les principaux sont détaillés ci-dessous.

- Tableur:** le format de fichier EXCEL peut être considéré comme un standard de fait propriétaire. Peut-être l'inclusion de tableaux dans SGML est-elle une première étape vers une norme ouverte.
- Base de données:** bien que de plus en plus de bases de données utilisent SQL comme langage d'interrogation, il n'existe pas de format de haut niveau pour les bases de données. La meilleure solution consiste à utiliser une base de données largement répandue ou à sauvegarder les données comme texte plat avec des séparateurs de champ (cela permet à une nouvelle application de recréer la base de données). Dans ce dernier cas, la structure de la base de données doit également être sauvegardée.
- Edifact** (*Electronic Data Interchange For Administration, Commerce and Transport* – échange électronique de données pour l'administration, le commerce et les transports): plusieurs types de message permettent d'échanger des fichiers de comptabilité et des factures.
- STEP** (*Standard for The Exchange of Product data* – norme pour l'échange de données sur les produits): ce groupe de normes facilite l'échange de données sur les produits industriels. Il comprend le langage EXPRESS pour la représentation et l'échange de données sur les produits. Ces normes sont largement utilisées.
- IDEF** (*Function and Information Modelling* – modélisation des fonctions et de l'information): IDEF0 est la méthodologie utilisée pour déterminer la portée de chaque protocole d'application dans la norme STEP. Elle est également utilisée pour la restructuration du déroulement des opérations. IDEF3 est consacré à l'amélioration du déroulement des opérations et fournit des formats graphiques standard pour le flux des tâches.
- Formulaire:** dans ce cas de figure, il est possible de ne conserver que les informations entrées et une copie simple d'un formulaire vierge. Il existe des travaux pour proposer un standard pour les formulaires (FIMS, HTML 3.0...).
- Signature numérique:** deux formats existent pour les signatures cryptées: DES et RSA. Elles peuvent être utilisées pour l'authentification des documents. La section 6.3.1 décrit les procédés de cryptage.

# 6. Accès et diffusion de l'information

## 6.1. Vers la société de l'information

La Commission européenne a adopté le concept de la société de l'information, une société dans laquelle l'information numérique jouera un rôle clé.

♦ **La société de l'information**  
Elle donne à l'information numérique un rôle de premier plan.



### Société de l'information

Notions avancées

Jacques Delors, à l'époque président de la Commission européenne, publie en janvier 1994 son livre blanc sur la croissance, sur la compétitivité et l'emploi.

Il charge alors Martin Bangemann, Commissaire européen en charge de l'industrie, de préparer un rapport avec un groupe de haut niveau constitué de dirigeants des industries des technologies de l'information, des télécommunications et des médias. Ce rapport, «L'Europe et la société de l'information planétaire», propose de prendre en compte, en plus des infrastructures, les services universels, les applications ainsi que les aspects sociétaux.

En février 1995, une première réunion du G7 est consacrée à la société de l'information et propose la mise en place de onze projets pilotes. Par exemple, le projet «gouvernement en ligne» propose de rendre le gouvernement plus proche des citoyens grâce aux nouvelles technologies.

La Commission continue à construire la société de l'information en Europe et elle a déjà contribué à sensibiliser le public et les décideurs. À cette fin, elle a mis sur pied différentes initiatives, dont son *plan d'action* qui présente une liste de toutes les actions importantes, notamment les mesures juridiques requises pour poursuivre la mise en œuvre de la société de l'information en Europe en remplissant quatre fonctions principales, à savoir:

- servir d'outil de navigation pour les États membres et autres institutions européennes (clarté et transparence);
- fournir des informations détaillées sur le développement du cadre réglementaire;
- fournir un service d'information ouvert à toutes les parties intéressées, notamment au public, sur la politique de l'UE, et permettre ainsi un dialogue constructif entre la Commission et tous les segments pertinents de la société, et
- fournir un outil de gestion interne pour la Commission.

Il devient fondamental que l'information numérique soit non seulement conservée mais également rendue accessible le plus aisément possible. Ainsi, l'information circule mieux à l'intérieur des organisations. En dehors des organismes publics, le citoyen peut accéder plus aisément aux informations publiques auxquelles il peut prétendre. Même après sa vie courante, l'information numérique peut être utilisée dans le cadre de recherches statistiques, scientifiques ou historiques.

### Les archivistes et la société d'information

Les archivistes sont les gardiens de la mémoire collective de la société de l'information.

### Vie privée et ouverture

Un dilemme tranché par les législations nationales et communautaires

La profession d'archiviste ainsi que les professions en rapport avec l'information font fondamentalement et activement partie de la société de l'information moderne. Il est fait appel à elles pour conserver la mémoire collective de la société de l'information – le DLM Forum sur les données lisibles par machine (Bruxelles, 18 au 20 décembre 1996) a été la première approche interdisciplinaire dans ce sens.

Une des difficultés, lors de la diffusion de l'information numérique, est d'identifier les bonnes informations à conserver ou à regrouper. En effet, l'information peut être traitée d'une façon qui n'a jamais été prévue par son auteur. Par exemple, les éléments d'une base de données peuvent être utilisés pour des recherches démographiques avec des critères de sélection particuliers.

Il existe un dilemme permanent entre la vie privée et l'ouverture. La législation fixe la frontière entre les deux. Il est important à cet égard de vérifier à la fois les directives européennes et la législation nationale.



## Information numérique et malvoyants

### Notions avancées

Les personnes ayant des problèmes de lecture ont recours à deux méthodes fondamentalement différentes pour lire l'information numérique:

1) L'information numérique n'est pas un problème pour les aveugles et les malvoyants car ceux-ci peuvent utiliser des programmes informatiques «standard» (traitements de texte, tableurs, bases de données...) en même temps qu'un autre programme qui tourne à l'arrière-plan et qui agrandit les écrans pour les malvoyants et propose des éléments supplémentaires (braille, voix synthétique) pour les aveugles.

Ces «lecteurs d'écran», qui agrandissent le contenu textuel d'un écran ou qui transforment en discours synthétique ou en braille temporaire (c'est-à-dire un dispositif avec une rangée de points tangibles qui peuvent reproduire les caractères braille sous le contrôle de l'ordinateur), existent aujourd'hui pour toutes les plates-formes informatiques.

La lecture de grands documents non structurés reste difficile car les malvoyants n'ont pas la capacité de «survoler» une page ou un document.

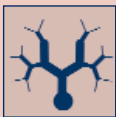
2) Les deux groupes peuvent lire beaucoup plus vite en utilisant des documents structurés et des programmes spéciaux de lecture (= butineurs 3D) qui envoient directement l'information audio, parlée et en braille à des dispositifs de sortie (synthétiseur ou ligne de lecture en braille). De plus, le texte sur l'écran est reformaté et présenté par ligne pour être facile à lire par les malvoyants. Ces programmes de lecture existent pour SGML et HTML.

## 6.2. Normes et standards pour l'accès aux données

### Diffusion de l'information

Il est nécessaire de sélectionner les normes et standards qui faciliteront l'accès aux données.

Certaines normes et certains standards sont plus adaptés pour rendre l'information accessible au plus grand nombre. Certains ont déjà été présentés dans le chapitre 5 sur les standards pour la conservation. La sélection proposée ci-dessous présente les mieux adaptés à la diffusion des informations. La conservation et la diffusion peuvent être effectuées si nécessaire dans des formats différents.



## Standards pour la diffusion

Pour choisir

Supports pour la distribution d'information:

- disquette 3 1/2 haute densité 1,44 Mo;
- CD-ROM standard;
- à terme le DVD (*Digital Versatile Disc*) peut constituer un support de choix.

Protocoles Internet pour l'échange et l'offre de documents sur réseau:

- TCP/IP (*Transfer Control Protocol, Internet Protocol*). Ces deux protocoles permettent des échanges par le biais d'un réseau. Ils sont largement utilisés dans le monde entier avec les réseaux Internet et Intranet;
- HTTP (*Hypertext Transport Protocol*) pour les serveurs Internet;
- FTP (*File Transfer protocol*) pour les serveurs de fichiers;
- SMTP (*Simple Mail Transfer Protocols*) pour les messageries électroniques;
- MIME (*Multipurpose Internet Multimedia Extensions*) autorise plusieurs formats dans le même message (par exemple, du texte avec différents jeux de caractères, des images...).

Formats pour les documents:

- SGML pour les documents portables. Le format HTML est plus adapté pour les documents courts tels que les pages d'accueil sur les Internet sites (*home page*). Les documents longs peuvent être placés directement au format SGML sur les serveurs Internet;
- formats propriétaires des traitements de texte (Microsoft Word, WordPerfect). Bien que ces formats ne soient pas recommandés pour une conservation à long terme, ils constituent des standards de fait largement utilisés dans l'environnement Windows des PC.

Interface d'accès aux bases de données

- SQL 2 (*Structured Query Language*) pour les bases de données relationnelles;
- ISAM (*Indexed Sequential Access Method*) pour les fichiers séquentiels indexés (interface de bas niveau).

## 6.3. La sécurité

---

### 6.3.1. Droits d'accès

Il existe de nombreux types de droits d'accès aux informations numériques:

- droit d'accès à la page de couverture;
- droit d'accès à tout ou partie d'un enregistrement;
- droit de visualisation et d'impression ou droit de visualisation seul;
- etc.

Rendre un fichier anonyme  
Une tâche plus complexe qu'il n'y paraît.

Les droits d'accès doivent être étudiés avec soin pour conserver le droit à la vie privée et l'anonymat de chacun. Il existe deux façons de rendre anonyme un fichier pour le mettre dans le domaine public:

- la suppression de certains champs (par exemple, les noms);
- l'agrégation d'informations pour fournir des statistiques.



## Anonyme ou non?

Exemple

Il faut prendre un certain nombre de précautions pour que la procédure utilisée pour rendre le fichier anonyme ne permette pas une sélection qui conduise à un faible nombre d'enregistrements.

Les noms sont, par exemple, supprimés d'une base de données. Une sélection permet de retrouver les personnes ayant telle profession dans tel village. Il est possible de faire des tris croisés pour trouver une personne unique et découvrir d'autres informations sur elle (comme sa tranche de salaire). Il peut, par exemple, y avoir peu de cardiologues dans un village de 1000 habitants...

### 6.3.2. Cryptage et authentification

Crypter les données:  
• pour assurer la confidentialité;  
• pour authentifier l'enregistrement.

Le cryptage des données peut avoir deux buts très différents:

- ne permettre la lecture d'un enregistrement qu'à des personnes sélectionnées;
- garantir que l'enregistrement a bien été produit par une personne particulière et n'a pas été modifié par quelqu'un d'autre.

Il est important, en matière de cryptage, de prendre en compte les particularités nationales. En effet, la législation en matière de cryptage est différente d'un pays à l'autre. Il peut y avoir des restrictions à l'exportation (comme aux USA) ou des restrictions à l'utilisation (comme en France).

Il ne s'agit pas simplement d'accéder à l'information ou de garantir un certain niveau de confidentialité. Il est également nécessaire de s'assurer que l'information à laquelle on accède est bien identifiée. L'authentification des enregistrements est un élément important pour obtenir une information qualifiée.



## Algorithmes de cryptage

Notions avancées

Il existe deux grandes familles d'algorithmes de cryptage:

- Les algorithmes à clé simple (symétriques):  
Type DES (*Data Encryption System*) (NIST FIPS 46-1, proche de l'ISO 8227 - DEA).  
Dans ce cas, la même clé de cryptage sert à coder et à décoder les données. L'émetteur et le récepteur d'un message codé disposent de la même clé.
- Les algorithmes à clé doubles (asymétriques):  
Type RSA (Rivest, Shamir & Adleman du nom de ses inventeurs)  
Dans ce cas, il existe deux clés pour chaque personne. Une clé que l'on garde secrète (la clé privée) et une autre que l'on diffuse largement (la clé publique). Un message codé avec une des clés ne peut être décodé qu'avec la deuxième clé.

Deux types d'algorithmes de cryptage:

- à clé simple, type DES;
- à clé double, type RSA.

Dans le cas d'un message confidentiel, l'émetteur envoie le message codé avec la clé publique du récepteur. Seul ce dernier pourra le relire à l'aide de sa clé privée.

Dans le cas d'un message authentifié, l'émetteur envoie le message codé avec sa propre clé privée. Toute personne ayant accès à sa clé publique pourra lire le message en ayant la garantie qu'il a bien été envoyé par l'émetteur et que personne n'a pu le modifier.

Le programme PGP (*Pretty Good Privacy*) de Philip Zimmermann est un logiciel de cryptage de type RSA mis dans le domaine public sur Internet. Cela malgré l'interdiction à l'exportation aux USA des algorithmes utilisant des clés suffisamment longues pour permettre une bonne sécurité.

Des serveurs d'authentification existent. Ils permettent d'obtenir la clé publique de quelqu'un avec une procédure garantissant son authenticité.

## 6.4. L'accès aux données

Il existe plusieurs façons de donner accès à l'information numérique:

- la salle de lecture électronique, c'est - à - dire une salle mise à la disposition du public, comprenant des équipements informatiques permettant de lire les enregistrements numériques;
- la copie de l'information numérique sans changement pour l'utiliser «chez soi»;
- la réalisation et la maintenance d'une «copie de consultation» de l'information numérique dans un nouveau format plus facile pour les utilisateurs (l'accès peut se faire par consultation sur place, par copie sur un support donné à l'utilisateur ou en ligne);
- utilisation d'un modèle d'accès général basé sur les métadonnées, pour générer automatiquement une vue de l'information numérique utile à l'utilisateur (l'accès peut se faire par consultation sur place, par copie sur un support donné à l'utilisateur ou en ligne).

Les deux derniers cas sont adaptés à une politique de diffusion sur Internet.

◆ Plusieurs façons de permettre aux utilisateurs d'accéder aux données



Exemple

### Allemagne

En Allemagne, pour les bases de données, il existe une différence entre les «copies de recherche», fournies dans un format adapté à la consultation par la communauté des chercheurs, et les «copies d'archivage», conservées dans un format de fichier plat pour éviter les problèmes de formats standard.

La diffusion de l'information nécessite également une politique d'information et de prise de conscience de la part des utilisateurs potentiels. Il existe deux stratégies qui peuvent être complémentaires:

- mettre l'information à disposition et laisser l'utilisateur la trouver grâce aux outils de navigation en ligne (diffusion passive);
- envoyer une information à un groupe d'utilisateurs ciblés (diffusion active).

Ces deux stratégies peuvent être utilement associées en donnant à un groupe cible des détails sur l'endroit où il peut trouver l'information.

La mise en place d'une politique de diffusion est indispensable pour permettre aux utilisateurs potentiels d'accéder à l'information.

◆ Diffuser l'information:

- mise à disposition de l'information (diffusion passive);
- envoi ciblé d'information (diffusion active).

## 7. Conclusions

Ce guide a présenté quelques exemples des solutions utilisées actuellement pour traiter l'information numérique ainsi que quelques suggestions pour définir une stratégie dans ce domaine.

Il n'existe pas une politique unique qui serait valable pour tous les pays et tous les cas de figure. Aussi est-il nécessaire de définir vous-même une stratégie pour l'information numérique, avec tous les acteurs impliqués.

Si ce guide vous apporte une aide dans cette tâche, il aura rempli son rôle.



## 8. Annexes

Annexe 1 — Terminologie

Annexe 2 — Questions en suspens

Annexe 3 — Pour comprendre la normalisation

Annexe 4 — Check-list pour définir une stratégie de l'information numérique

Annexe 5 — Prototype: quelles métadonnées faut-il créer?

Annexe 6 — Prototype: comment sélectionner les bonnes normes et les bons standards?

Annexe 7 — Index

Annexe 8 — Table des figures

Annexe 9 — Bibliographie

### 8.1. Terminologie

---

Ces définitions sont utilisables dans le cadre de ce guide. Il existe des définitions données dans les législations nationales qui doivent également être prises en compte.

Information

Élément de savoir qui peut être transmis.

Donnée

Représentation d'une information élémentaire dans un format qui permet de lui appliquer un traitement.

Données lisibles par machine (DLM)

Données dans un format adapté à la recherche, au traitement et à la transmission par un ordinateur numérique.

Document

Groupe consistant et cohérent de données, structuré de façon à présenter un raisonnement ou à servir de preuve de l'exécution d'une action.

Base de données

Groupe de données structuré de façon à faciliter la recherche et les traitements ultérieurs.

Enregistrement

Ensemble consistant de données enregistrées sur un support.

«Un morceau particulier d'information enregistrée générée, collectée ou reçue lors du commencement, de la conduite ou de l'achèvement d'une activité institutionnelle ou personnelle, et qui comprend une structure, un contenu ainsi qu'un contexte suffisants pour fournir une preuve ou une évidence de cette activité» (CIA).

Enregistrement numérique

«Un enregistrement dont l'information est enregistrée dans un format adapté à la recherche, au traitement et à la transmission par un ordinateur numérique» (CIA).

Support

Élément qui permet l'enregistrement, le stockage et la récupération des enregistrements.

## 8.2. Questions en suspens

---

1. Quelle est la valeur légale de l'information numérique?

La notion d'original utilisée avec les documents papier devient complexe avec l'information numérique. Il devient de plus en plus facile de dupliquer des enregistrements de telle façon qu'on ne puisse plus distinguer l'original de la copie. De plus, un enregistrement numérique peut être constitué uniquement de liens avec d'autres enregistrements numériques.

Une des solutions pour traiter ce problème est d'utiliser le cryptage avec des clés publiques et privées (voir la section 6.3.2 sur le cryptage des données). L'auteur crypte son enregistrement avec sa clé privée et toute personne peut récupérer sa clé publique sur un serveur d'authentification pour lire son enregistrement tout en ayant la garantie qu'il n'a pas été modifié depuis son cryptage.

2. La signification des termes peut varier d'un pays à l'autre

La traduction des différents termes n'est pas suffisante. Le problème du vocabulaire est bien plus complexe. Une solution consisterait à créer une liste de concepts avec les termes adaptés dans chaque langue.

## 8.3. Pour comprendre la normalisation

---

Trois étapes dans la maturité d'une technologie :

- standards de fait
- PAS
- norme

Pour permettre le jeu de la concurrence et garantir une bonne portabilité des données quel que soit le fournisseur, il est important de spécifier uniquement les interfaces et non les produits eux-mêmes. Ainsi, plusieurs fournisseurs peuvent proposer des produits compatibles. Les données ou les applications gagnent une meilleure pérennité.

Il existe plusieurs étapes qui définissent la maturité d'une technologie depuis les produits propriétaires jusqu'aux normes ouvertes:

- les standards de fait (*de facto standards*). Lorsqu'un produit prend une grande importance sur le marché, la compatibilité des applications et des données passe par le produit et l'évolution de la compatibilité dépend du fournisseur (par exemple, le traitement de texte Word de Microsoft);
- les spécifications disponibles publiquement (PAS – *Publicly Available Specifications*). Un consortium de sociétés représentatif du marché peut se réunir pour définir un standard d'interface. Dans ce cas, la définition des seules interfaces permet de développer plusieurs produits compatibles entre eux (par exemple, les spécifications d'X/Open ou de l'IETF);

- les normes (*de jure standards*). Des organismes officiels permettent de garantir un consensus sur une spécification. Dans ce cas, la spécification standard est appelée norme (par exemple, les normes de jeux de caractères de l'ISO).  
Une technologie, au fur et à mesure de sa maturité, permet l'émergence de standards de fait, de PAS puis de normes.

Il existe plusieurs organismes internationaux officiels:

- ISO: Organisation internationale de normalisation qui traite un très grand nombre de domaines;
- CEI ou IEC: Comité électrotechnique internationale.

Ces deux organismes ont créé un comité joint pour traiter des normes dans le domaine du traitement de l'information.

- UIT ou ITU: Union internationale des télécommunications. Son comité ITU-T traite particulièrement du domaine des télécommunications (il s'agit du nouveau nom du CCITT).

La procédure d'élaboration des normes comprend plusieurs étapes. Il est parfois possible de commencer à utiliser une norme dès l'avant-dernière étape nommée «projet de norme internationale» (*Draft International Standard – DIS*).

Au niveau européen, il existe des organismes équivalents:

- CEN: Comité européen de normalisation;
- Cenelec: Comité européen de normalisation électrotechnique;
- ETSI: Institut européen de normalisation des télécommunications.

Les deux premiers s'occupent en commun des technologies de l'information. Ils créent des normes européennes (EN) et des pré-normes européennes (ENV). Il s'agit souvent de reprises de normes internationales. Il existe cependant une différence au niveau légal: La décision 87-95 du Conseil des Communautés européennes, du 22 décembre 1986 rend obligatoire la référence aux normes européennes dans les marchés publics. Les normes internationales, pour leur part, reposent sur une base volontaire.

Ces organismes officiels développent ou entérinent des normes mais également des profils (ensemble de normes avec un choix d'options permettant leur bon interfonctionnement). Les profils internationaux développés par l'ISO/IEC sont appelés ISP (*International Standardised Profiles*).

D'autres organismes développent des spécifications:

- X/Open qui rassemble ses spécifications dans le X/Open Portability Guide (XPG);
- IETF (*Internet Engineering Task Force*) qui développe les spécifications Internet après avoir émis des «appels à commentaires» (*Request for Comments – RFC*);
- NIST du gouvernement américain (*National Institute for Standards and Technologies*) qui développe des profils appelés FIPS (*Federal Information Processing Standards*);
- et bien d'autres organismes encore qui travaillent sur les très nombreux aspects des standards pour le traitement de l'information.

Le tableau suivant présente les références de toutes les normes citées dans ce guide.

◆ Les organismes de normalisation internationaux dans le domaine informatique

- ISO/IEC
- ITU-T

◆ Les organismes européens de normalisation dans le domaine informatique

- CEN/Cenelec
- ETSI

◆ Plusieurs organismes développent des spécifications:

- X/Open
- IETF
- NIST
- etc.

Nom	Norme ou profil international	Norme ou profil européen	Autres specification	Remarques
<b>Supports de stockage (sections 5.1 et 6.2)</b>				
Disquette 3 1/2	ISO/IEC 9529-1 ISO/IEC 9529-2	EN 29529-1 EN 29529-2		
Cartouche 1/4'	ISO 8462-1 ISO 8462-1			
Cartouche 1/2'				
Bande 1600 bpi	ISO/IEC 3788:1976			
Bande 6250 bpi				
Cassette 8mm	ISO/IEC 11319 ISO/IEC 12246			
Cassette DAT				
CD-ROM	ISO 9660 ISO 10149			
WORM				
TMO				Développement en cours
DVD				
<b>Formats de fichiers d'images en mode point et de graphiques (section 5.2)</b>				
Images TIFF				
Images GIF				
Images JPEG				
Télécopie groupe III	UIT-T groupe III			anciennement
Télécopie groupe IV	UIT-T groupe IV			CCITT
Vidéo MPEG-1				anciennement
Vidéo MPEG-2				CCITT
Graphisme CGM	ISO/IEC 8632			
Graphisme CAO				
Graphisme GIS				
<b>Jeux de caractères (section 5.2.3)</b>				
7-bits	ISO 646			
8-bits west Europe	ISO/IEC 8859-1			
8-bis Greek	ISO/IEC 8859-7			
Multiocets	ISO/IEC 10646			
<b>Textes structurés (section 5.2.3)</b>				
SGML	ISO/IEC 8879	EN 28879		
DSSSL	DIS 10179			
SPDL	ISO/IEC 10180			
SDIF	ISO/IEC 9069			
Font information interchange	ISO/IEC9541			
DTD normalisée	ISO/IEC 12083			
HyTime	ISO/IEC 10744			
HTML			W3C HTML 3.0	
ODA/ODIFF	ISO 8613 FOD 26	EN 41509 EN 41515		
<b>Formats de données (section 5.2.4)</b>				
EDIFACT	ISO/IEC 9735	EN 29735		
STEP/ Express	ISO 10303			
IDEF0&3			IDEF	
FIMS				
<b>Protocoles d'échanges (section 6.2)</b>				
HTTP			IETF RFC	
FTP			IETF RFC	
<b>Interrogations de bases de données (section 6.2)</b>				
SQL	ISO/IEC 9075			version II
ISAM				
<b>Algorithmes de cryptographie (section 6.3.2)</b>				
DAS	ISO 8273			
DES			NIST FIPS 46-1	proche de DAS
RSA				

## 8.4. Check-list pour les DLM

---

Cette annexe présente les différentes questions que l'on peut se poser pour définir une stratégie pour l'information numérique. Les chiffres entre parenthèses renvoient aux sections du guide qui traitent ce sujet. Les prochaines révisions comprendront des vitrines avec des expériences et des clauses d'achat.

L'organisation n'aura pas obligatoirement une réponse sur chaque point, ou ne souhaitera pas forcément traiter tous les points (par exemple, la mise en place ou l'adoption d'un thesaurus). La liste des points non traités, non résolus, non fixés ou rejetés est également une information importante qui fait partie intégrante de la stratégie.

### I. Stratégie générale

- A. Identification des acteurs impliqués (4.1)  
Mise en place d'un groupe stratégique multidisciplinaire
- B. Liste de termes et concepts communs (8.1; 8.2)
- C. Besoins des utilisateurs: identification et suivi (4.1)
- D. Politique sur la valeur légale des enregistrements (8.2)
- E. Politique d'information et de formation des services (4.1)
- F. Technologie vieille (supports, documents objets...) (5.1; 5.2)

### II. Gestion de l'information numérique

- A. Identification des responsabilités (lors de chaque transfert) (3.2)
  - 1. Responsabilité de la gestion de l'information numérique
  - 2. Responsabilité de la conservation de l'information numérique
- B. Identification et enregistrement de l'information numérique importante (4.1; 4.2)  
(tailles et frontières avec d'autres informations numériques)
- C. Politique de documentation contextuelle de l'information numérique (2.1; 4.2)
- D. Définition d'un plan d'évaluation (4.3.6)
  - 1. Procédure pour l'information numérique de provenance douteuse (4.2)
  - 2. Procédure d'approbation pour la destruction ou la modification (4.2)
- E. Définition des règles pour le plan de classement (4.1; 4.4)
  - 1. Mise en place d'un classement par mots clés (thesaurus)
- F. Stratégie de cryptographie (4.3.2)
  - 1. Confidentialité
  - 2. Authentification
- G. Politique de transfert physique de l'information numérique (4.5)
  - 1. Liste des éléments à transférer
  - 2. Liste des contrôles à effectuer à chaque transfert
  - 3. Support utilisé pour les transferts physiques support d'origine, réseau sécurisé...

### III. Conservation (préservation) de l'information numérique

- A. Choix pour la conservation des «originaux» de l'information numérique (4.2)  
numérique seul ou copie papier
- B. Choix des supports de conservation (5.1)
  - 1. Conservation à court terme
  - 2. Conservation à long terme
  - 3. Environnement de conservation (température, humidité, fréquence de recopie...)
- C. Choix des formats de fichiers conservés (5.2)
  - 1. Fichiers images en mode point (5.2.1)
    - a) Type de compression utilisée (5.2.1)
    - b) Conservation des télécopies (4.3.4; 5.2.1)
  - 2. Fichiers graphiques (5.2.2)
    - a) Fichiers graphiques spécifiques (CAO, GIS)
  - 3. Fichiers audio, vidéo et multimédias (5.2.1)
  - 4. Fichiers textes (5.2.3)
    - a) Jeux de caractères (par défaut, acceptables)
    - b) Formats de textes structurés
    - c) Conservation ou non de la présentation
  - 5. Fichiers de données (5.2.4)
    - a) Bases de données (bas niveau ou propriétaire)
      - 1) Fichiers de comptabilité/gestion
    - b) Tableurs (bas niveau ou propriétaire)
    - c) Formulaires (5.2.4)
    - d) Documents composites (4.3)
  - 6. Programmes (5.2.4)
- D. Politique de conservation d'anciens systèmes et logiciels (4.3.6)  
(si nécessaire) Maintenance, documentation, savoir-faire...

### IV. Stratégie de récupération des documents papier

- A. Choix pour la conservation des documents papier (4.2)  
Identification des documents à scanner
- B. Charte qualité des documents prévus pour être scannés (4.3.3)
- C. Stratégie de reconnaissance des caractères (4.3.4)
  - 1. Choix d'utilisation de la reconnaissance des caractères
  - 2. Procédure de reconnaissance des caractères (4.3.4)
- D. Stratégie de vectorisation des graphiques (4.3.4)
- E. Procédure pour le regroupement des divers éléments d'une information numérique  
(textes, graphiques, images...) (4.3.4)

### V. Migration ou conservation du format de données (4.3.3)

- A. Politique de documentation des formats de données  
utilisés par les systèmes et les logiciels (2.4.2)
- B. Choix de migration ou de conservation des anciens formats (4.3.6)
- C. Choix d'ajouter une structure aux textes ou aux données récupérés (4.3.4)
- D. Choix de construction d'une base anonyme pour diffusion (6.3.1)
  - 1. Procédure de contrôle de la protection de l'anonymat
- E. Procédure d'analyse des pertes d'information dues à la migration (4.3.3)

## VI. Accès et utilisation des données

- A. Politique de droits d'accès (en visualisation et impression) (6.3.1)
- B. Mise à disposition des informations (diffusion passive) (6.4)
  - 1. Salle de lecture électronique (6.4)
    - a) Standards de consultation (6.2)
  - 2. Copie de l'information numérique sans changement pour diffusion (6.4)
    - a) Supports des copies (6.2)
  - 3. Copie de l'information numérique dans un format pour la diffusion (6.4)
    - a) Supports, protocoles, langages et formats (6.2)
    - b) Stratégie Internet (6.2)
  - 4. Modèle d'accès pour générer automatiquement un format adapté à l'utilisateur (6.4)
    - a) Supports, protocoles, langages et formats (6.2)
    - b) Stratégie Internet (6.2)
- C. Politique de promotion (diffusion active) (6.4)

### 8.5. Prototype: quelles métadonnées faut-il créer?

Cette annexe présente un exemple des métadonnées éventuelles qui se basent sur la proposition relative aux métadonnées faite par Dublin Core en décembre 1996. Vous trouverez d'autres informations à l'adresse Internet suivante: [http://www.purl.org/metadata/dublin\\_core](http://www.purl.org/metadata/dublin_core).

Ce prototype n'est pas la seule façon de procéder. Il ne fait que proposer des exemples qui pourront aider le lecteur à définir sa propre stratégie.

Les quinze éléments sont des options et la liste n'est pas limitative.

Le nom donné à la ressource par le Créateur ou l'Éditeur

TITRE

(Auteur ou créateur)

CREATEUR

La (les) personne(s) responsables au premier chef du contenu intellectuel de la ressource. Par exemple, les auteurs pour les documents écrits, les artistes, les photographes et les illustrateurs pour les ressources visuelles.

(Sujet et mots clés)

SUJET

Le sujet de la ressource, ou les mots clés ou les phrases qui décrivent le sujet ou contenu de la ressource. La définition de cet élément a pour but de promouvoir l'utilisation de vocabulaires et mots clés contrôlés. Cet élément pourrait inclure des données de classement selon un schéma normalisé (par exemple, les numéros de classement de la bibliothèque du Congrès ou des nombres décimaux Dewey) ou des vocabulaires contrôlés selon un schéma normalisé (comme les titres de sujets médicaux ou des descripteurs du thesaurus sur l'art et l'architecture).

Une description textuelle du contenu de la ressource, y compris des résumés pour ce qui ressemble à des documents ou des descriptions de contenu pour les ressources visuelles. Les collections futures de métadonnées pourraient comprendre une description informatique du contenu (analyse spectrale d'une ressource visuelle par exemple) qui peut ne pas être intégrable dans les systèmes de réseaux actuels. Dans ce cas, le champ pourrait contenir un lien avec cette description plutôt que la description même.

DESCRIPTION

L'entité responsable de la mise à disposition de la ressource dans sa forme actuelle, comme un éditeur, une faculté universitaire ou une entité institutionnelle. La spécification de ce champ a pour but d'identifier l'entité qui donne accès à la ressource.

EDITEUR

---

**COLLABORATEURS** La (les) personne(s) ou organisation(s) autres que celle(s) spécifiée(s) comme Créateur, qui ont apporté des contributions intellectuelles importantes à la ressource, mais dont la contribution est secondaire par rapport à celle des personnes ou entités définies comme le Créateur (par exemple, rédacteurs, transcripteurs, illustrateurs et organisateurs).

---

**DATE** La date à laquelle la ressource a été rendue disponible dans sa forme actuelle. Il est recommandé de la présenter comme un nombre de huit chiffres sous la forme YYYYMMDD tel que défini par ANSI X3.30-1985. Dans ce cas, la date du jour où cela est écrit serait 19961203, ou le 3 décembre 1996. Il existe d'autres possibilités mais, en cas d'utilisation, elle doivent être identifiées sans aucune ambiguïté possible.

---

**TYPE** La catégorie de la ressource, telle que page d'accueil, roman, poème, document de travail, épreuve, rapport technique, essai, dictionnaire. Le Type de ressource devrait être choisi dans une liste de types. Un groupe préliminaire de types est proposé à l'adresse (URL) suivante:  
<http://www.roads.lut.ac.uk/Metadata/DC-ObjectTypes.html>.

---

**FORMAT** La représentation des données de la ressource comme text/html, ASCII, fichier Postscript, application exécutable ou image JPEG. La définition de cet élément a pour but de fournir les informations nécessaires pour permettre aux personnes ou aux machines de prendre des décisions sur l'exploitabilité des données codées (quel matériel et quel logiciel sont nécessaires pour les afficher ou pour les exécuter par exemple).

Comme pour le Type de ressource, le format sera attribué à partir de listes énumérées comme les types de supports enregistrés sur Internet (types MIME). En principe, les formats peuvent comprendre des supports comme les livres, les feuillets et d'autres supports non - électroniques.

---

**IDENTIFICATEUR** Chaîne ou nombre utilisé uniquement pour identifier la ressource. Les URL et URN (le cas échéant) sont des exemples d'identificateur pour les ressources interconnectées. D'autres identificateurs mondiaux uniques, comme les numéros standard internationaux des livres (ISBN) ou d'autres noms officiels peuvent aussi être utilisés ici.

---

**SOURCE** L'œuvre, imprimée ou numérique, d'où cette ressource provient si c'est possible. Par exemple, le codage html d'un sonnet de Shakespeare pourrait identifier la version papier du sonnet qui a servi de base pour sa transcription en version numérique.

---

**LANGAGE** Langage(s) du contenu intellectuel de la ressource. Quand c'est possible, le contenu de ce champ devrait coïncider avec les trois codes de caractères Z39.53 pour les langues écrites.  
Voir: <http://www.sil.org/sgml/nisoLang3-1994.html>

---

**RELATION** Relation avec d'autres ressources. La définition de cet élément a pour but de fournir un moyen d'exprimer les relations entre des ressources qui ont des relations formelles avec d'autres, mais qui existent elles-mêmes comme ressources discrètes. Par exemple, des images dans un document, des chapitres dans un livre ou des pièces d'une collection. Une définition officielle de la Relation est actuellement en cours d'élaboration.

---

**COUVERTURE** Les caractéristiques spatiales et/ou temporelles de cette ressource. Une définition officielle de la Couverture est actuellement en cours d'élaboration. Les utilisateurs et les développeurs doivent comprendre que l'utilisation de cet élément en est encore au stade expérimental.

---

**DROITS** Un lien avec une note de droit d'auteur, une déclaration de gestion des droits ou un service qui donnerait des informations sur les conditions d'accès à la ressource. La spécification formelle des droits est actuellement en cours d'élaboration. Les utilisateurs et les développeurs doivent comprendre que l'utilisation de cet élément en est encore au stade expérimental.

---



## 8.6. Prototype: Comment choisir les bons standards?

Cette annexe présente le standard proposé pour chaque type de donnée. D'autres choix peuvent être faits pour chaque type en utilisant les informations données dans ces lignes directrices. Il est bon de définir un standard recommandé. Une autre façon de faire consiste à établir une liste des standards acceptables pour chaque type de fichier.

Ce prototype n'est pas la seule façon de procéder. Il ne fait que proposer des exemples qui pourront aider le lecteur à définir sa propre stratégie

Types de données	Standard recommandé	Commentaires
Jeux de caractères	ISO/IEC 8859-1	Une autre possibilité pour les pays d'Europe occidentale est Unicode (ISO/IEC 10646) si d'autres jeux de caractères sont nécessaires.
Texte structuré	SGML	
Images en mode point	JPEG	
Télécopies	ITU-T Groupe III	
Graphiques vectoriels	CGM	
Audio et vidéo	MPEG II	
CAO/PAO	STEP	
Comptabilité/facture	Edifact	
Autres fichiers de bases de données	fichier plat, virgule de séparation	Il n'existe aucun format standard de base de données. Le fichier plat permet la conservation à long terme si la structure de la base de données est bien documentée.
Fichiers cryptés	RSA	
Programmes	source ou version compatible avec PC	Il n'existe aucun standard pour les programmes compilés. Le code binaire Java, qui ne dépend pas de la plate-forme, peut contribuer à la conservation à long terme des applications Java.
<b>Supports pour la conservation à long terme</b>	DVD	Le DVD est relativement neuf mais il devrait être largement utilisé à l'avenir. Sa grande capacité et son utilisation facile peuvent aider à faire du DVD le support des archives. Plusieurs organisations ont choisi ce support. Il est recommandé de choisir un support ou un très petit groupe de supports pour faciliter leur utilisation future.

## 8.7. Index

<b>A</b>		<b>D</b>		<b>G</b>	
Accès aux données,	47	DES cryptage,	46	GIF,	38
Algorithmes de cryptage,	46	Destruction,	21	Graphiques vectoriels,	37
Anonymat,	46	Diffusion			
ASCII,	39	diffusion active,	47	<b>H</b>	
Authentification	46	diffusion passive,	47	HTML,	40
serveurs		DIS (projet de norme		HTTP,	45
d'authentification,	47	internationale),	51	HyTime,	40
		Disque optique			
<b>B</b>		réinscriptible,	36	<b>I</b>	
Bande magnétique,	35	Disquette,	35	IDEF,	42
Base de données,	13	DLM,	49	IDEFO,	42
fichier de données,	42	Document,	14	IDEF3,	42
		document orienté		IEC,	51
<b>C</b>		objet,	37	IETF,	51
Caractères,	39	Données,	11; 41	Images en mode point,	37
Cartouche magnétique,	35	Données lisibles		Internet,	47
Cassette DAT,	35	par machine,	16; 49	ISAM,	42
Cassette vidéo 8mm,	35	Droits d'accès,	45	ISO,	51
CCITT,	51	DSSSL,	40	ISP,	51
CD-ROM,	35	DTD,	40	ITU,	51
CEI,	51	<b>E</b>		ITU-T,	51
CEN,	51	EDI,	41		
Cenelec,	51	Edifact,	41; 42	<b>J</b>	
CIA,	7	EN (norme européenne),	51	Jeux de caractères,	39
Compression des images,	38	Enregistrement,	12	ASCII,	39
Comptabilité,	42	contenu,	13	UCS Universal	
Comptabilité/Gestion		contexte,	13	Character Sets,	39
(fichiers),	42	présentation,	13	JPEG,	38
Conseil International des		structure logique,	13		
Archives,	7	ENV (pré-norme		<b>L</b>	
Contenu,	13	européenne),	51	Langage de balisage,	40
Contexte,	13	ETSI,	51		
Courrier électronique,	31	Express	42	<b>M</b>	
Cryptage		<b>F</b>		Message	
DES,	46	Factures,	42	message	
PGP,	47	FIMS,	42	authentifié,	46
RSA,	45	FIPS,	51	message	
		Font Information		onfidentiel,	46
		Interchange,	40	Messages	
		Formats vidéo,	38	électroniques,	31; 41
		Formulaire,	42	classification,	31
		FTP,	45	Microfilm,	34

<hr/>		<hr/>		<hr/>	
N		R		T	
NIST,	51	Reconnaissance optique des		Tableur,	42
Normes,	51	caractères,	24	Technologie magnéto-	
normes		RFC,	51	optique,	36
européennes,	51	RSA cryptage,	46	Télécopie,	25; 38
O		RTF,	40	format de fichiers de	
		S		télécopies,	38
OCR (reconnaissance optique				Texte plat,	26
des caractères),	24	Scanneurs,	24	Textes,	37
ODA,	40	SDIF,	40	TIFF,	38
OLE,	37	SGML,	40		
OpenDoc,	37	Signature numérique,	42	U	
P		Société de l'information,	43		
		SPDL,	40	UCS (jeu de caractères	
PAS (spécifications disponibles		Spécifications disponibles		universel),	39
publiquement),	50	publiquement,	50	UIT,	51
PDF,	40	SQL,	42	V	
PGP,	47	Standards de fait,	50		
Phase de conception,	10	STEP,	42	Vectorisation,	25
Phase de création,	10	Structure logique,	13	Vidéo,	38
Phase de maintenance,	10	Supports		Vie privée,	44
Photo-CD,	38	supports			
Photographie,	37	magnétiques,	34	W	
Polices de caractères,	40	supports optiques,	34		
PostScript,	40			Workflow,	14; 42
Pré-normes européennes,	51			WORM,	36
Présentation,	13				
Profils,	51			X	
Programmes,	41				
Projet de norme				X/Open,	51
internationale,	51			XPG,	51

## 8.8. Table des figures

Figure 1 — Document ou base de données?	14
Figure 2 — Cycle de vie des données lisibles par machine	17
Figure 3 — Migration d'un document papier vers un document numérique	22
Figure 4 — Changement de format numérique	23
Figure 5 — Que faire des formats propriétaires?	29

## 8.9. Bibliographie

La présente bibliographie ne propose que quelques ouvrages de référence à la base du présent guide. Un grand nombre de documents internes aux organisations nationales et communautaires n'a pas été cité.

- **Les archives dans l'Union européenne.** *Rapport du groupe d'experts sur les problèmes de coordination en matière d'archives.* Commission européenne, secrétariat général. Luxembourg EUR-OP, Luxembourg, 1994, 104 p., ISBN 92-826-8234-X – Catalogue n° CM-83-94-741-FR-1
- **Actes du DLM-Forum sur les données lisibles par machine.** Bruxelles, 18 au 20 décembre 1996, dans: INSAR – Courrier européen des archives, supplément II, 1997, EUR-OP, Luxembourg, 376 p., ISBN 92-828-0112-8 – Catalogue n° CM-AC-97-SO1-FR-C (FR, DE, EN) sur Internet: <http://www.echo.lu/dlm/en/proc-index.html>
- **EPHOS – European Procurement Handbook for Open Systems.** Référence: EUR 14021
- **Guide pour la gestion des documents électroniques du point de vue archivistique.** Conseil international des archives, comité sur les documents électroniques, Paris, 1997. Sur Internet: <http://www.archives.ca/ica/p-er/francais.html>
- **B. Bauwens, F. Evenpoel et J. Engelen,** «Standardisation as a Prerequisite for Accessibility of Electronic Text Information for Persons who cannot use Printed Material», *IEEE Transactions on Rehabilitation Engineering*, vol 3, n° 1, p. 84-89, 1995.
- **Chr. Reeves et T. Wesley,** «Guidelines for Accessible Web page design», Brochure publiée par le Harmony Consortium (1997). Elle se trouve également sur Internet: <http://www.esat.kuleuven.ac.be/teo/harmony/guidelines>