

# **LES COMPOSANTES D'UN DOCUMENT ÉLECTRONIQUE**

Rapport du Groupe de travail  
sur les métadonnées et les structures logiques

**Collection en ingénierie documentaire : 2**

**Janvier 1999**

Réalisé dans le cadre du Chantier en ingénierie documentaire  
Coordonné par : Richard Parent et Nicole Boulet

Conseil du trésor  
Sous-secrétariat à l'infrastructure gouvernementale  
et aux ressources informationnelles

# REMERCIEMENTS

Ce rapport est issu de la réflexion d'un groupe de travail qui s'est penché sur les composantes d'un document électronique.

Nous remercions de leur collaboration les personnes suivantes :

**Membres du groupe de travail :**

# TABLE DES MATIÈRES

<b>INTRODUCTION .....</b>	<b>1</b>
<b>CHAPITRE 1 .....</b>	<b>3</b>
<b>LES MÉTADONNÉES ET LES STRUCTURES LOGIQUES.....</b>	<b>3</b>
1.1 DÉFINITION ET CARACTÉRISTIQUES .....	3
1.2 RATTACHEMENT DES MÉTADONNÉES AU DOCUMENT.....	4
1.3 PRODUCTION DES MÉTADONNÉES.....	6
1.3.1 LE RÔLE DES AUTOMATES .....	6
1.3.2 LE RÔLE DES UTILISATEURS .....	7
1.4 TYPOLOGIE DES DOCUMENTS EN VUE DE L'ÉTABLISSEMENT DE PROFILS DE MÉTADONNÉES.....	7
1.4.1 TYPOLOGIE DES DOCUMENTS GOUVERNEMENTAUX.....	7
1.5 PROFILS ET SOUS-PROFILS DE MÉTADONNÉES POUR LES DOCUMENTS GOUVERNEMENTAUX.....	12
1.5.1 ÉTABLISSEMENT DE SOUS-PROFILS À PARTIR DE TYPES DE DOCUMENTS.....	12
1.5.2 PRÉSENTATION DE SOUS-PROFILS DE MÉTADONNÉES POUR LES DOCUMENTS DE RÉFÉRENCE, LES DOCUMENTS DE TRANSACTION ET LES MESSAGES DE COURRIER ÉLECTRONIQUE .....	13
1.5.3 ÉTABLISSEMENT DE PROFILS ET DE SOUS-PROFILS EN FONCTION DE NORMES.....	17
<b>CHAPITRE 2 .....</b>	<b>18</b>
<b>LES STRUCTURES LOGIQUES DES DOCUMENTS.....</b>	<b>18</b>
2.1 LE MODÈLE XML .....	18
2.1.1 UNE EXPÉRIENCE XML : LE CAS DES FORMULAIRES .....	20
2.1.2 BÉNÉFICES À TIRER DU FORMAT XML DANS LE CAS DES FORMULAIRES .....	21
2.2 UNE NORME POUR LES MÉTADONNÉES : LE FORMAT RDF .....	22
<b>CHAPITRE 3 .....</b>	<b>25</b>
<b>UN PORTRAIT COMPLET D'UN DOCUMENT ÉLECTRONIQUE.....</b>	<b>25</b>

# INTRODUCTION

Dans l'administration publique, les connaissances utilisées dans une organisation se trouvent en grande partie dans les écrits comme les textes de lois et de règlements, les directives et les manuels de procédure, d'interprétation ou d'application, les publications, les répertoires les plus divers, ainsi que dans les séries de centaines ou de milliers de notes, lettres, analyses, etc. Plus de 80 % de l'information gouvernementale existe sous forme de documents. Une **structuration** de cette information est possible aujourd'hui grâce à de nouveaux langages et outils informatiques qui permettent de décrire et de traiter les documents. Il importe d'**optimiser** l'utilisation du contenu informationnel qui est ainsi produit, emmagasiné, transmis, rendu accessible ou autrement utilisé dans le cadre de l'activité gouvernementale et plus particulièrement à l'ère de l'inforoute.

De plus en plus, les documents gouvernementaux sont non seulement produits sur support électronique, mais ils sont diffusés, publiés et conservés sur un tel support. Ainsi, le document électronique est au centre des préoccupations et prend une importance stratégique en devenant un outil de communication privilégié à l'intérieur du gouvernement et entre l'État et les citoyens.

En conséquence, pour faciliter la **communication**, et parce que l'information que contiennent les documents est souvent réutilisée durant leur cycle de vie, il faut tirer profit des nombreuses **régularités** de structures logiques présentes dans la masse de documents administratifs, juridiques, techniques et autres. À cette fin, il convient d'utiliser le potentiel de la technologie qui permet une meilleure gestion documentaire et, de ce fait, une plus grande accessibilité à l'information.

C'est dans cet esprit que le *Chantier en ingénierie documentaire* a abordé la problématique particulière reliée à l'utilisation des supports électroniques appropriés pour décrire de façon structurée les documents et leur contenu. À cette fin, il a mis sur pied un *Groupe de travail sur les métadonnées et les structures logiques* qui s'est mis à l'étude des composantes d'un document électronique.

Les objectifs du groupe de travail étaient, premièrement, de **catégoriser les types de documents** courants et de définir les pratiques concernant leurs métadonnées et leur structure logique ; deuxièmement, d'étudier les relations entre les éléments balisant la structure logique d'un document et les attributs et valeurs qui constituent les métadonnées associées à ce document. Ces travaux sur la **typologie et la définition des métadonnées** sont nécessaires pour répondre aux questions cruciales concernant les éléments qui permettent de repérer les documents électroniques, notamment les hyperliens, les adresses et les noms de ces documents. Ces travaux sont également nécessaires pour assurer la fonctionnalité du format des Pages vertes du Répertoire gouvernemental, lequel permet l'identification des documents de référence gouvernementaux de manière à les rendre accessibles et disponibles tant à l'intérieur du gouvernement qu'au sein de la population.

En outre, le besoin de définir les métadonnées d'un document électronique acquiert beaucoup d'importance pour l'automatisation des applications reliées au Web, car les métadonnées favorisent une meilleure utilisation de ces documents par des robots/agents, y compris les moteurs de recherche. De plus, il existe actuellement un consensus dans les milieux spécialisés sur la nécessité de définir des métadonnées en vue d'automatiser la gestion des documents en réseau qui se trouvent au cœur de l'informatique moderne. Ce consensus s'est imposé en raison du fait

qu'aujourd'hui les documents électroniques ne sont plus réalisés en fonction d'« instructions-commandes », mais en fonction d'« objets-documents ». Cette nouvelle vague informatique est souvent qualifiée de centrée sur les données (*data-centric*). Il ressort de ce point de vue que la notion de document électronique se trouve au cœur des préoccupations sur le Web et que la gestion documentaire devient par le fait même le support des mécanismes fondamentaux de l'architecture du Web. En d'autres termes, pour s'y retrouver dans la masse documentaire sur le Web ou les intranets gouvernementaux, il faut plus que jamais se préoccuper de gestion documentaire.

L'efficacité des communications électroniques repose sur l'adoption de normes relatives à la gestion des documents et de leurs métadonnées. Il importe de **modéliser les documents en tant que contenants structurés dans lesquels le contenu se trouve inséré**. C'est pourquoi ce rapport propose l'exploration concrète du langage de balisage de données XML et de la norme de représentation de métadonnées RDF qui permettent de structurer à la fois les métadonnées et le contenu des documents.

La première partie du rapport présente la manière de structurer l'information électronique sur le document électronique au moyen des métadonnées, alors que la deuxième partie porte sur les structures logiques, c'est-à-dire sur l'organisation des données du document électronique. Enfin, un troisième point fournit un portrait global des composantes internes et externes d'un document électronique.

# CHAPITRE 1

## LES MÉTADONNÉES ET LES STRUCTURES LOGIQUES

### 1.1 Définition et caractéristiques

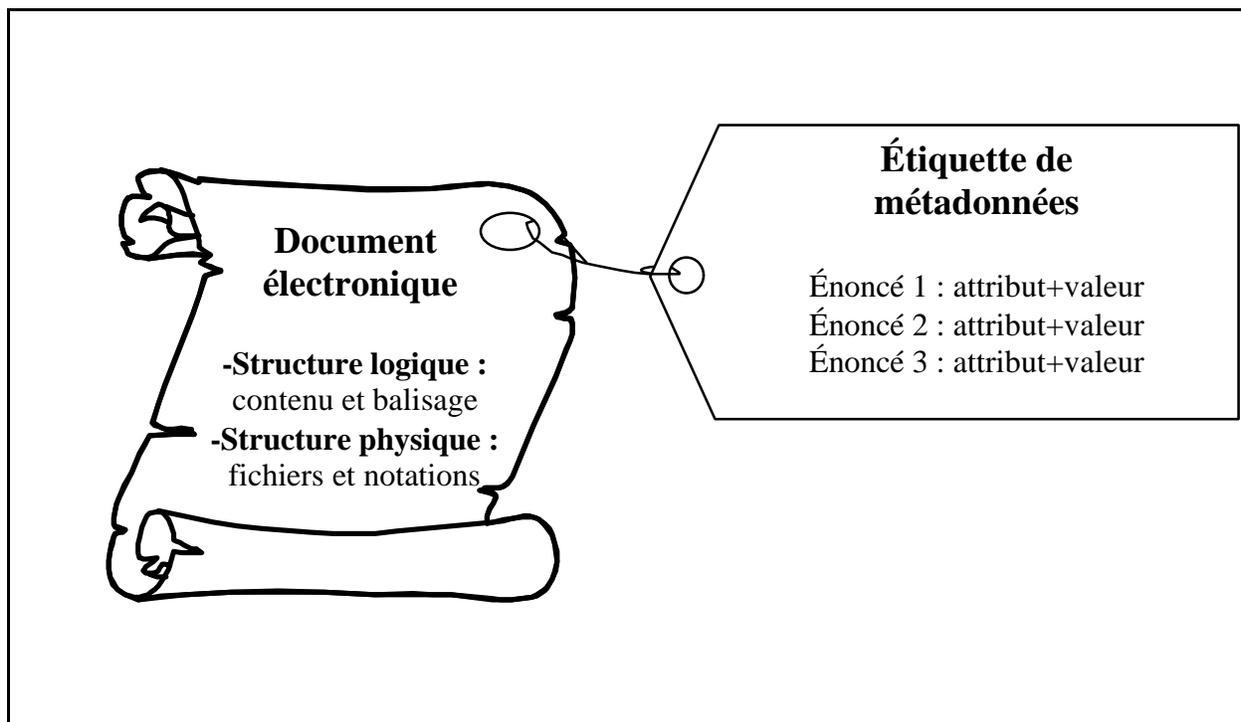
L'information stratégique qui est au cœur des développements en ingénierie documentaire est constituée de structures logiques et de métadonnées. Si l'on se réfère pense au modèle de document que représente un livre, ses métadonnées sont les éléments d'information extraits des premières pages et transcrites sur une fiche de référence ; quant à sa structure logique, elle se trouve représentée en partie par la table des matières.

En général, le terme métadonnées sert à désigner la description du document pris comme un tout, alors que la structure logique correspond au marquage des parties internes du contenu d'un document. Ces deux concepts, métadonnées et structures logiques, sont étroitement associés : ils ont en commun d'éclairer un contenu donné.

Les métadonnées d'un document peuvent être comparées à une étiquette collée sur une boîte de conserves ou attachée à un objet (voir figure 1). Elles identifient le document et décrivent certains aspects de son contenu, ce qui évite d'avoir à l'ouvrir uniquement pour en connaître le contenu.

Les métadonnées se présentent comme un ensemble d'énoncés à propos d'un document. Dans l'exemple « l'auteur est Jean Lévesque », l'énoncé comprend un nom d'attribut (« auteur ») qui prend la valeur « Jean Lévesque ». « La date de publication est le 27 novembre 1998 » constitue un autre énoncé possible pour ce même document. La liste des attributs possibles d'un document doit faire l'objet d'une convention sur un profil de métadonnées.

**Figure 1. Illustration du lien entre le document et l'étiquette de métadonnées**

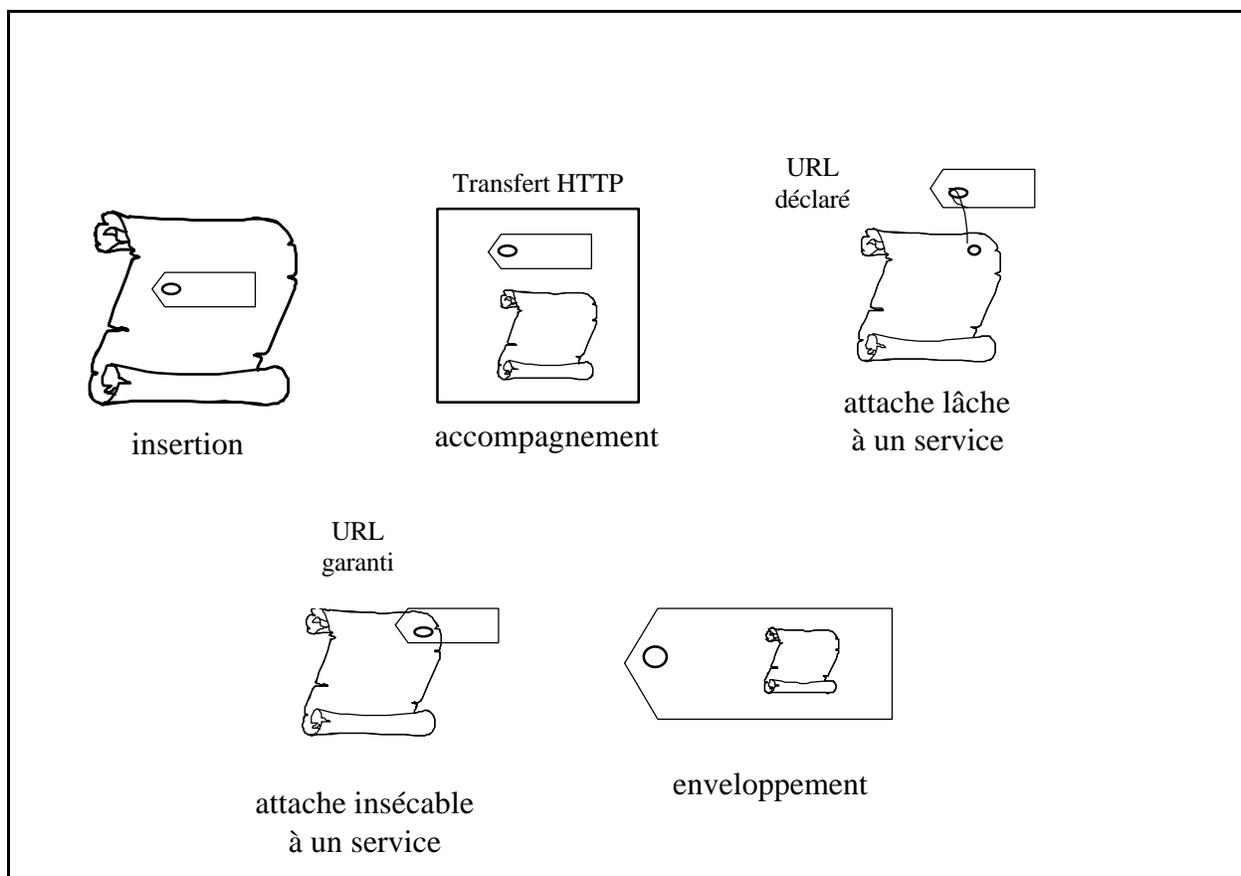


## 1.2 Rattachement des métadonnées au document

Comme les métadonnées sont une information à propos des données qui composent le document, il importe de pouvoir les rattacher au document ou de créer un lien entre elles. Il existe cinq façons d'effectuer un rattachement entre une « ressource » (le document proprement dit) et les métadonnées qui y sont associées (voir figure 2) :

- par insertion (*embedded*) : l'énoncé est contenu dans la ressource ;
- par accompagnement (*along with*) : l'énoncé est externe, mais véhiculé avec la ressource ;
- par attache lâche à un service (service bureau) : l'énoncé est externe, obtenu séparément, indiqué simplement avec un URL (*Uniform Resource Locator*) ;
- par attache insécable à un service (service bureau) : l'énoncé est externe, obtenu séparément, indiqué par un lien insécable grâce aux techniques cryptographiques ;
- par enveloppement (*wrapped*) : l'ensemble des énoncés contient la ressource.

**Figure 2. Les cinq modes de rattachement de l'étiquette de métadonnées au document**



La façon dont une étiquette est attachée à un document peut se limiter à indiquer l'adresse URL du document comme étant la valeur d'un attribut de localisation. Cependant, certains énoncés doivent avoir un rattachement garanti, comme dans le cas où l'attribut est la signature numérique d'un document ; le rattachement le mieux garanti fait appel aux techniques mathématiques de calcul d'une empreinte du document. Un algorithme donné d'une prise d'empreinte donnera toujours la même empreinte d'un document tant qu'il reste inchangé.

Le Web n'a pas été riche en métadonnées jusqu'à maintenant par rapport aux possibilités envisagées aujourd'hui, les principaux usages actuels étant les suivantes :

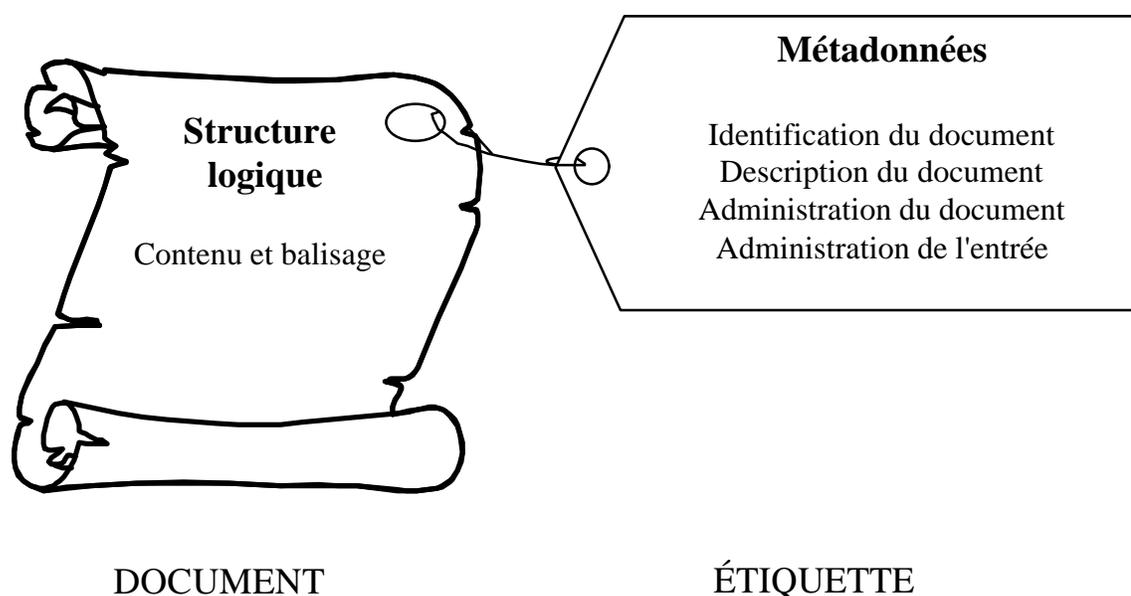
- en-tête des messages de courrier électronique sur Internet (protocole SMTP) ;
- en-tête HTTP : presque toujours présent pour indiquer le format (HTML, graphique...), utilisé parfois pour la taille de la page, rarement pour une date d'expiration, et pour l'encodage si le texte n'est pas en ASCII ;
- en-tête de page HTML : des métadonnées peuvent être indiquées par un élément « Title », un élément « Link » (employé pour le style surtout), ou par un élément « Meta », ce dernier étant sous-utilisé, sauf par des robots d'indexation comme Infoseek ;

- protocole d'exclusion des robots (REP) : métadonnées relatives au serveur pour indiquer aux robots d'indexation ce qu'il ne faut pas indexer.

### 1.3 Production des métadonnées

Les métadonnées sont produites par l'inscription des énoncés qui décrivent les attributs et les valeurs qui servent à l'identification, à la description, à l'administration du document et à l'administration de l'entrée relative au document. Les métadonnées sont produites et utilisées à la fois par les automates et par les personnes. Il importe de maximiser le rôle des automates dans le processus de production des métadonnées de manière que le travail humain ne soit sollicité que là où il est essentiel, c'est-à-dire quand un jugement de pertinence (par exemple, activité de classification) est requis.

**Figure 3. Illustration du rôle des métadonnées**



#### 1.3.1 Le rôle des automates

Les automates peuvent prendre en charge une grande partie de la création des métadonnées. Par exemple, certaines métadonnées sont destinées uniquement à la lecture par un ordinateur, dans un langage et un vocabulaire convenus, de manière que l'ordinateur puisse en quelque sorte interpréter la sémantique des valeurs rencontrées, au sens où des actions de traitement leur sont correctement associées ; par exemple, les contenus MIME (*Multimedia Internet Mail Extension*) vont déclencher un appel indiquant l'application requise pour afficher le texte, l'image, le tableau, produire le son, etc.

L'automatisation de cette saisie ne peut se réaliser que par la réutilisation d'éléments décrits antérieurement dans la chaîne documentaire. Si pour les documents papier tous les renseignements sur les éléments d'une référence dans les Pages vertes doivent être saisis par une personne, dans le cas des documents électroniques la plupart des éléments peuvent être saisis par des automates dans des conditions favorables.

Les automates peuvent aussi interagir avec les humains. Par exemple, certaines métadonnées sont exprimées de façon lisible pour les humains, mais elles sont également indexées par l'ordinateur, étant alors transformées en chaînes de caractères accessibles aux humains. Ces métadonnées peuvent toutefois ne pas être traitables automatiquement au même sens que les précédentes, c'est-à-dire que l'ordinateur ne peut pas accéder au niveau sémantique d'interprétation des humains pour les valeurs rencontrées, contrairement à ce qui se passe avec les contenus MIME. Les services rendus par les automates se limitent alors à véhiculer des chaînes de caractères placées dans les bonnes catégories et dont le sens est réservé à l'humain.

### 1.3.2 Le rôle des utilisateurs

Bien que la saisie d'éléments de métadonnées par les créateurs et les utilisateurs de documents doit être réduite au minimum, il demeure que l'intervention humaine est essentielle pour la production de certains éléments. Il convient que cette intervention soit réalisée le **plus près possible de la source** pour des raisons de coût et d'efficacité : en effet, l'indexation produite par le créateur du document s'appuie sur sa connaissance immédiate du contenu. Un tel mode de fonctionnement suppose que l'intervention demandée ne soit pas trop complexe et qu'une formation et des outils en permettent la réalisation.

Ainsi, l'opérationnalisation des métadonnées dans l'ensemble de l'appareil gouvernemental passe par **l'établissement d'un format commun** de métadonnées. L'adoption généralisée d'un tel format est un gage de rentabilité. Le format des Pages vertes du Répertoire gouvernemental fournit un profil de métadonnées pour les documents. Ce format comporte à la fois des éléments obligatoires et des éléments facultatifs. Le format prévoit aussi divers éléments généraux ou spécifiques à des types de documents associés à des séries ou à des collections.

## 1.4 Typologie des documents en vue de l'établissement de profils de métadonnées

Un profil de métadonnées est une définition du contenu de l'étiquette à attacher à un certain type de document. Dans les pages qui suivent, nous appliquerons ce concept aux documents gouvernementaux qui ont été catégorisés en trois types, à savoir les documents de référence, les documents de transaction et les messages de courriel.

### 1.4.1 Typologie des documents gouvernementaux

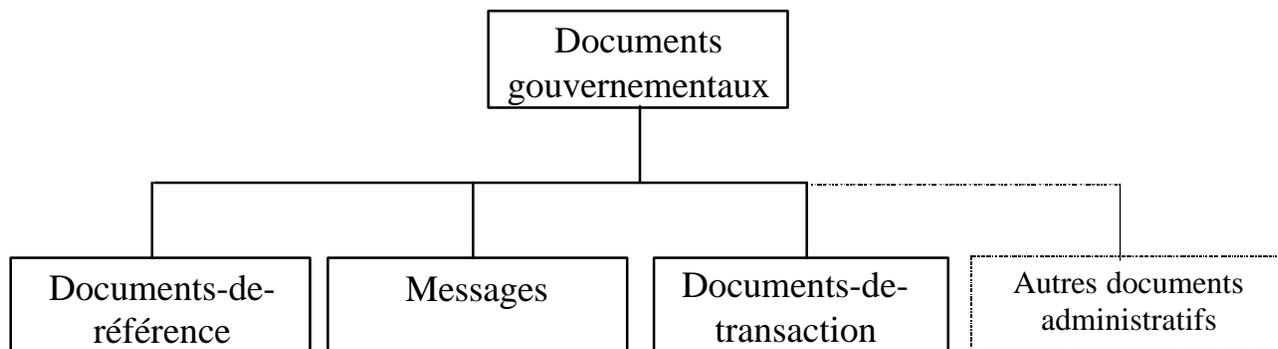
Avant de tracer un tel profil, il convient de proposer une certaine typologie des documents gouvernementaux. L'expression **documents gouvernementaux** s'applique à tous les documents produits, reçus et acquis par le gouvernement dans le cadre de ses activités. Une variété de

documents compose cet ensemble : du formulaire au projet de loi, des lettres aux messages électroniques, des notes aux rapports, etc. Plusieurs façons de subdiviser cet ensemble apportent un éclairage utile. Il y a notamment la distinction conventionnelle entre documents de gestion et documents de fonction. D'autres regroupements peuvent être basés sur le genre de contribution d'un document aux activités d'une organisation ; on distinguera par exemple les documents d'encadrement, d'opération, de liaison, de référence.

La description de documents aussi diversifiés que les documents gouvernementaux ne peut être effectuée avec un ensemble strict et obligatoire d'attributs, mais il est en principe souhaitable d'utiliser un maximum d'attributs communs. Cette idée a guidé la présente conception d'un ensemble d'attributs devant servir de format de description de tout type de document. Des métadonnées de portée générale doivent s'appliquer à tout document gouvernemental, quel qu'il soit, avant de distinguer les types et sous-types et d'étendre les sous-profils de métadonnées selon les besoins particuliers. C'est à ce travail **de typologie des documents et de définition des sous-profils de métadonnées** que plusieurs des participants au Chantier en ingénierie documentaire se sont consacrés au cours des derniers mois. Ce travail a permis de catégoriser les documents gouvernementaux en trois principaux types : les documents de référence, les documents de transaction et les messages de courrier électronique. Voici une description de la nature des documents que l'on retrouve dans chacune des trois catégories (voir figure 4) :

- documents de référence : documents utilisés afin de prendre connaissance d'une information générale pertinente à la réalisation d'un processus de travail (études, volumes, publications, informations publicitaires, communiqués, audiovisuel, formulaires vierges, spécifications techniques, tables, catalogues et répertoires). L'expression désigne les publications acquises (bibliothèques), mais aussi des documents produits dans l'organisation et diffusés sur les extranets et les intranets et qui se prêtent à une description semblable (c'est-à-dire avec sensiblement les mêmes attributs descriptifs ou métadonnées) ;
- documents de transaction : documents établissant certaines informations relatives à un échange entre deux parties et liant par signature une responsabilité de portée administrative, juridique ou financière entre les parties. Contrairement aux documents de référence, les documents de transaction s'adressent à un seul destinataire ou à un nombre limité de personnes. Il peut s'agir autant de documents produits que de documents reçus par une organisation ;
- messages de courrier électronique : ces documents ne sont pas des transactions et tous n'ont pas une valeur jugée suffisante pour être conservés. Leurs caractéristiques sont semblables à celles des documents de transaction.

**Figure 4. Sous-profils de métadonnées des types de documents gouvernementaux**



Les « Autres documents administratifs » sont un rassemblement hétérogène de documents qui ne sont inclus dans aucun des trois sous-profils qui ont été définis jusqu'à maintenant. Documents d'analyse, comptes rendus, définition de projets, il s'en trouve à foison. Il reste beaucoup à faire dans la définition de sous-profils pour mieux couvrir ces autres documents administratifs. Cependant, le travail déjà accompli sur les trois premiers sous-profils fournit une riche gamme d'attributs qui devrait permettre de couvrir la plupart des besoins généraux dans la définition de ces nouveaux sous-profils.

Le Groupe de travail sur les métadonnées et les structures logiques a privilégié ces trois sous-profils pour les raisons suivantes : 1) le travail effectué sur les attributs relatifs aux documents de référence offrait l'avantage de pouvoir puiser largement dans la tradition bibliothéconomique, ce sous-profil pouvant alors servir de base pour tous les autres sous-profils ; 2) le fait de regrouper ensuite les documents de transaction permet d'aborder l'incontournable défi de la signature numérique pour le commerce électronique ; 3) la prise en considération des messages rend compte du fait que l'échange de messages électroniques est une forme de communication en voie de se substituer à une grande partie de la correspondance sur papier et dont la conservation ne devrait pas être négligée trop longtemps encore.

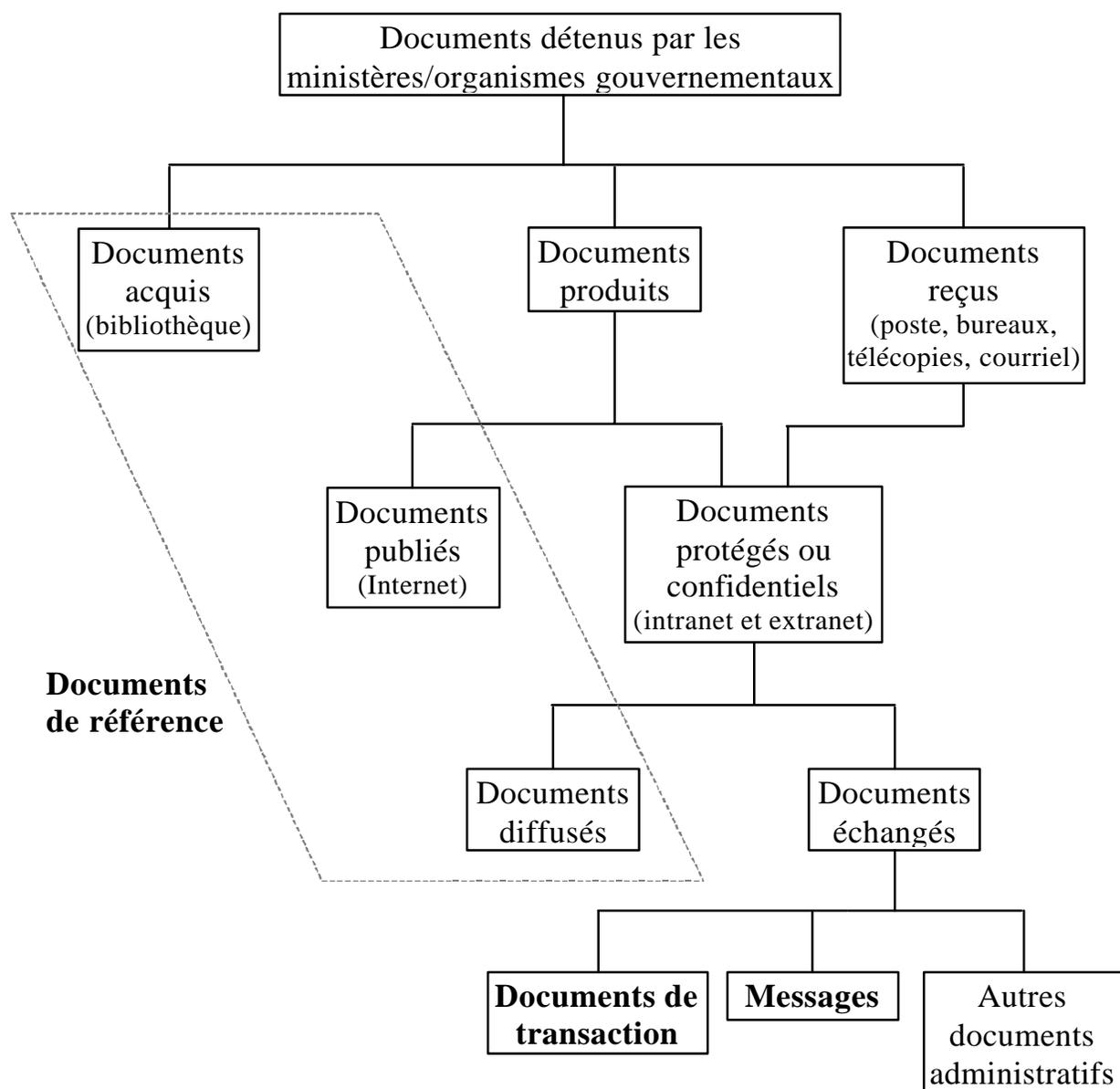
Cette typologie a été constituée après bien des tâtonnements quant aux critères à privilégier pour établir des groupes de documents homogènes dans leurs exigences de métadonnées. La figure 5 illustre les principales distinctions ayant ponctué le cheminement qui a permis cette façon de regrouper les documents, en précisant le nom de chacun des types de documents.

L'examen de la figure montre qu'une distinction importante, celle entre documents de gestion et documents de fonction, n'a pas été retenue parmi ces distinctions initiales. Ces deux types de

documents sont présents dans chacun des trois sous-profils définis. Toutefois, il est prévu d'utiliser cette distinction quand viendra le temps de définir des sous-profils spécifiques parmi les Autres documents administratifs. En particulier, il faudra miser sur le recours à la classification des documents de gestion établie par Louise Gagnon-Arguin (PUQ, 1998), qui établit plus d'une centaine de catégories soigneusement analysées et décrites selon une méthodologie innovatrice.

Tous les documents peuvent être protégés et échangés selon divers niveaux de confidentialité, tant les documents produits par l'organisation que ceux reçus de l'extérieur (clients, partenaires).

**Figure 5. Mise en perspective des trois profils de métadonnées (en gras) d'une classification plus générale des documents gouvernementaux**



## 1.5 Profils et sous-profils de métadonnées pour les documents gouvernementaux

Un profil général des métadonnées d'une classe de documents doit rester ouvert et extensible. Cependant, un vocabulaire commun désignant les attributs et les valeurs doit servir de socle pour assurer le partage des données et l'interfonctionnalité des applications et des services communs de gestion documentaire électronique. Les métadonnées sont le nom donné aux divers groupes d'attribut-valeur qui caractérisent les objets d'une classe. Lors de l'établissement d'un profil, il importe de se poser les questions suivantes : quels attributs doit-on prévoir ? et parmi eux, lesquels sont obligatoires, lesquels ont une valeur fixe, lesquels ont une valeur par défaut, etc. ?

### 1.5.1 Établissement de sous-profils à partir de types de documents

Le Chantier en ingénierie documentaire avait pour objet la création de profils pour une classe de documents, soit les documents gouvernementaux. Cette classe de documents a donné lieu à la création de sous-classes correspondant aux trois types de documents dégagés, à savoir les documents de référence, les documents de transaction et les messages de courriel, afin de dégager les sous-profils de métadonnées pertinents pour chacune de ces sous-classes de documents.

Outre le sous-profil des documents de référence, un sous-profil des documents de transaction représente le monde des documents administratifs, en particulier ceux qui sont signés et qui doivent être conservés. Un nombre croissant de ces documents de transaction circuleront par réseau, et il est donc urgent de proposer une façon de structurer cette information. Un troisième sous-profil représente les messages de courriel, car ces derniers prennent une importance accrue dans les communications menant aux prises de décision. Il n'est donc plus possible de les négliger et il faut prévoir dans les métadonnées des attributs permettant leur conservation, du moins la conservation de certains d'entre eux. Il faut souligner ici que la plupart des documents de transaction sont aussi des messages en raison de leur nature même. Cependant, beaucoup de messages ne sont pas des documents de transaction, et un profil distinct reconnaît leur importance accrue dans la création de l'information, dans sa communication et dans la prise de décision.

C'est ainsi que chaque sous-profil a donné lieu à la publication de lignes directrices destinées aux créateurs et aux gestionnaires de documents ainsi qu'aux informaticiens :

- *lignes directrices pour la description d'un document-de-référence enregistré à des fins de conservation ;*
- *lignes directrices pour la description d'un document-de-transaction enregistré à des fins de conservation ;*
- *lignes directrices pour la description d'un message de courriel enregistré à des fins de conservation.*

### **1.5.2 Présentation de sous-profils de métadonnées pour les documents de référence, les documents de transaction et les messages de courrier électronique**

Dans le tableau qui suit, le profil général des métadonnées des documents gouvernementaux est divisé en quatre sections :

- Identification du document
- Description du document
- Administration du document
- Administration de l'entrée des métadonnées

Ces sections regroupent 56 attributs. Dans l'axe horizontal, le tableau est subdivisé en trois sous-profils qui rendent possible la comparaison des attributs pour les trois types de documents.

## Comparaison des attributs pour les Documents-de-transaction et les Messages par rapport aux attributs du format des Documents-de-référence

La convention d'écriture pour les noms d'objets et d'attributs veut que ces noms s'écrivent avec traits d'union s'ils sont formés de plus d'un mot. De plus, la lettre initiale du nom est en majuscule s'il désigne une classe d'objet (ex. : Document-de-référence) et en minuscule s'il s'agit d'un attribut (ex. : identifiant-de-structure-logique).

### LÉGENDE :

Idem : même concept et même intitulé que pour le Document-de-référence  
 x : le concept désigné par un attribut ne s'applique pas  
 Autre nom : signifie que le concept désigné par un attribut s'applique, mais sous un autre intitulé

DOCUMENT-DE-RÉFÉRENCE	DOCUMENT-DE -TRANSACTION	MESSAGES
<i>Identification du document</i>		
Titre	Autre nom : Nom-administratif-courant	Autre nom : Sujet-du-message
Auteur	Autre nom : Créateur/signataire	Autre nom : Expéditeur
Autres-agents-de-production	x	x
Organisation-source	Idem	Idem
Date-de-publication	x	x
Codification	Idem	Idem
Langue-du-document	Idem	Idem
x	Catégorie-d'affaires	x
x	Destinataire-de-transaction	Destinataire-du-message

DOCUMENT-DE-RÉFÉRENCE	DOCUMENT-DE -TRANSACTION	MESSAGES
<i>Description du document</i>		
Résumé	x	x
Raison-d' être	x	x
Programme-gouvernemental	Autre nom : Cadre-réglementaire	x
Information-supplémentaire	x	x
Mots-clés	Idem	Idem
Période-couverte	x	x
Référence-géographique-par-nom-de-lieu	x	x
Référence-géographique-par-coordonnées	x	x
Source-des-données	x	x
Méthodologie	Idem	x
Accessibilité	x	x
- Distributeur	x	x
- Caractérisation	x	x
- Procédé-de-commande	x	x
- Support-technique	Autre nom : Fichier(s)-du-document Exécution-du-document	Autre nom : Pièces-jointes Exécution-de-pièces-jointes
- Lien-accessible	x	x
- Type-de-lien-accessible	x	x
- Durée-de-disponibilité	x	x
- Conditions-d' utilisation	x	x
- Point-de-contact	x	x
Renvoi	x	x
• titre-du-renvoi	x	x
• lien-du-renvoi	x	x
• type-du-renvoi	x	x
• relation-avec-l' entrée	x	x
x	Type-de-transaction	x

DOCUMENT-DE-RÉFÉRENCE	DOCUMENT-DE -TRANSACTION	MESSAGES
<i>Administration du document</i>		
Limites-d'accès	Idem	Idem
Règle-de-conservation	Idem	Idem
Document-institutionnel	Idem	x
Document-essentiel	Idem	x
Sous-classe-d'objet (document-de-référence)	Idem (autre valeur : document-de-transaction)	Idem (autre valeur : message)
x	Signature-numérique/Date de signature	Signature-numérique/Date de signature
Création-horodatée-du-document	Idem	Idem
Modification-horodatée-du-document	Idem	Idem
Identifiant-séquentiel-de-document	Idem	Idem
Nom-du-modificateur-de-document	Idem	Idem
x	Réception-horodatée	Réception-horodatée
x	Lien-avec-transactions-antérieures	Lien-avec-messages-antérieurs
x	Action-subséquent	x
x	Références-de-vérification	x

DOCUMENT-DE-RÉFÉRENCE	DOCUMENT-DE -TRANSACTION	MESSAGES
<i>Administration de l'entrée</i>		
Création-horodatée-de-l'entrée	Idem	Idem
Nom-du-créateur-de-l'entrée	Idem	Idem
Modification-horodatée-de-l'entrée	Journalisation : historique détaillé des opérations, y compris modifications	x
Nom-du-modificateur-de-l'entrée	(Journalisation)	x
Identifiant-séquentiel-d'entrée	Idem	Idem
Identifiant-séquentiel-d'origine	x	x
Date-de-révision	x	x
Langue-de-l'entrée	x	x

Le tableau précédent permet de constater que l'établissement d'un profil peut donner lieu à la production d'un nombre important de sous-profils et d'un nombre encore plus important d'énoncés de métadonnées. Pour s'y retrouver, il convient de normaliser le processus de création de ces profils par le recours à des langages normalisés pour leur expression. L'importance des normes XML (*eXtensible Markup Language*) et de la norme RDF (*Resource Description Format*) sera exposée un peu plus loin.

### **1.5.3 Établissement de profils et de sous-profils en fonction de normes**

L'établissement d'un cadre de métadonnées requiert une capacité de schématisation, c'est-à-dire de représenter des modèles comprenant un certain nombre de concepts et leurs interrelations. Cette schématisation devrait être basée sur des normes de balisage de données et des conventions permettant de faciliter la saisie des énoncés de métadonnées (attribut et valeur).

Une norme de balisage de données appropriée à cette fin se retrouve dans la syntaxe XML, et le support technique permettant d'établir le profil de métadonnées devra être établi par les fournisseurs de progiciels de gestion de documents, lesquels devront développer leurs produits en tenant compte de la réutilisation des métadonnées du document tout au cours de la vie de ce dernier, c'est-à-dire de sa création jusqu'à son archivage ou à sa destruction. Au moment de schématiser les profils de métadonnées, il faut accepter qu'ils ne pourront probablement pas faire en sorte que toutes les métadonnées soient complètement interprétables par la machine, et certains schémas resteront implicites parce que largement connus (ex. : ISBN). Il demeure cependant que les schémas doivent être traitables par la machine et avoir un tronc commun. L'inscription de ces schémas dans un référentiel, c'est-à-dire dans un mécanisme d'emmagasinage pour les modèles et des outils permettant de traiter les données modélisées, ou dans un répertoire, est essentielle pour leur partage et leur réutilisation. Finalement, les métadonnées doivent pouvoir être associées à des schémas capables de répondre et d'être adaptés à une variété de besoins.

## CHAPITRE 2

### LES STRUCTURES LOGIQUES DES DOCUMENTS

Alors que les métadonnées s'appliquent généralement au document pris comme un tout, la structure logique concerne les éléments de contenu qui sont rassemblés et ordonnés dans un document. Ainsi, les métadonnées servent à maîtriser les documents pris comme un ensemble, mais il est aussi important de maîtriser l'information qu'ils contiennent, afin de la rendre accessible électroniquement. Pour ce faire, il faut s'attaquer au vaste domaine des structures logiques des documents en vue de modéliser les documents pour en faire des contenants structurés dans lesquels le contenu informationnel se trouve inséré.

Une telle démarche de structuration devrait amener plusieurs avantages significatifs :

- moindre risque de désuétude des données représentées de façon indépendante des outils de traitement ;
- moindre coût de publication (papier), de mise à jour, de diffusion ;
- valeur ajoutée à l'information, rendant celle-ci beaucoup plus utilisable et réutilisable ;
- augmentation de la qualité des documents produits et imprimés ou affichés à l'écran ;
- contrôle des documents qui se trouvent au coeur des processus de travail (*circuit de production*) ;
- modélisation de l'information d'une façon englobante.

Le balisage des données est déjà rendu accessible par les récents développements du langage de balisage des données XML ; la syntaxe XML a aussi servi de base à l'élaboration d'une norme permettant de schématiser les métadonnées, à savoir la norme RDF.

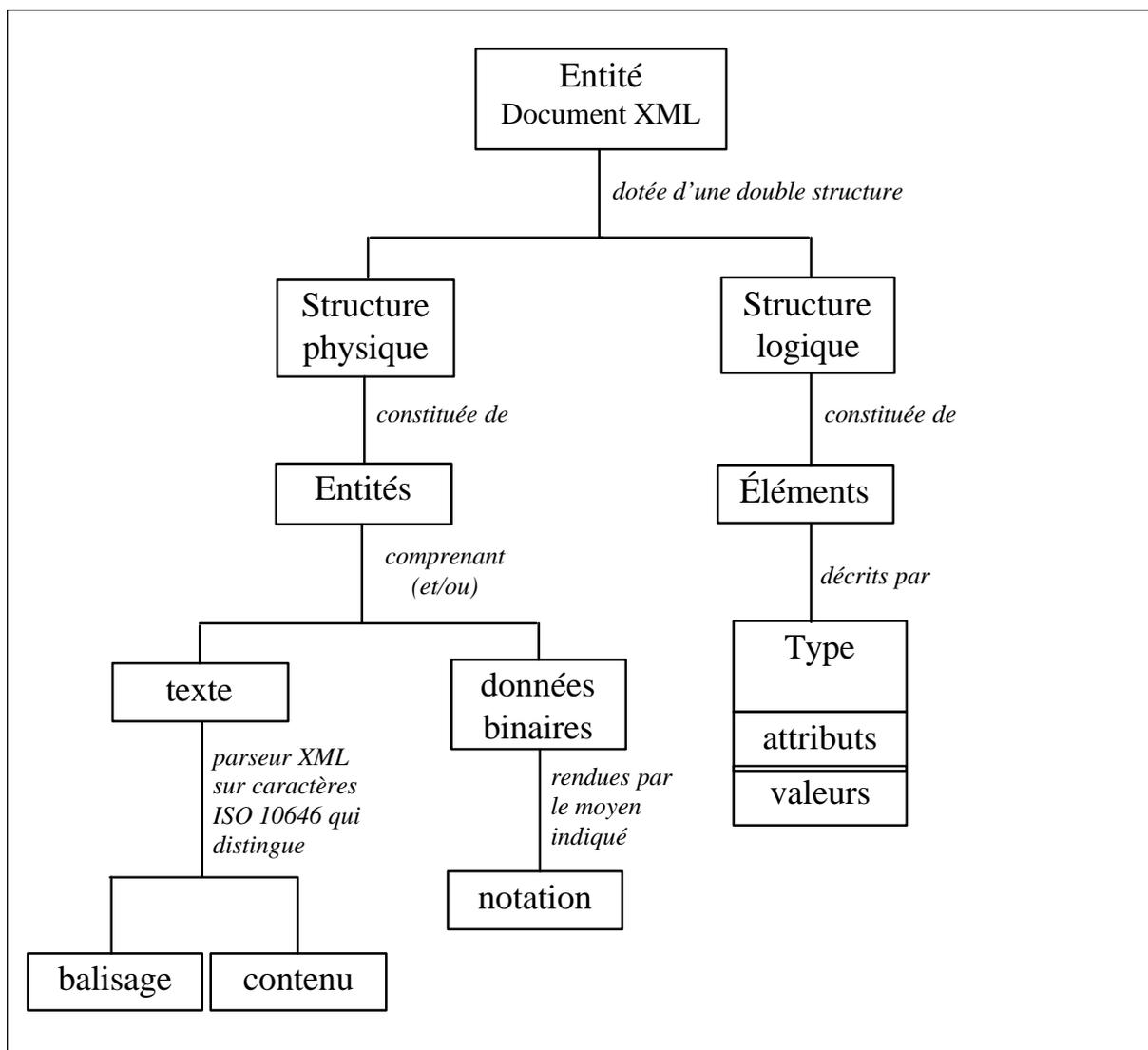
#### 2.1 Le modèle XML

Le modèle qui est en voie de s'imposer sur le Web est celui d'une syntaxe de représentation du document XML (héritier de SGML). Le sigle XML signifie « eXtensible Markup Language », ou langage extensible de balisage. Cette syntaxe représente un document comme une arborescence avec deux branches principales (voir figure 6) :

- Une **structure logique** déployée sur une séquence linéaire avec les balises et les noms d'éléments insérés dans la séquence du texte isolés par des délimiteurs. Le texte est constitué de caractères, dont une partie correspond aux données caractères (*character data* : lettres, chiffres, autres caractères) dans le document, et l'autre partie au balisage (*markup*). Le balisage encode une description de l'arrangement des données du document, de sa structure, et des paires arbitraires attribut-valeur qui, lorsqu'elles sont associées à la structure globale, constituent les métadonnées du document. Ainsi, la structure logique permet de désigner l'élément de contenu titre du document, lequel peut devenir une valeur d'attribut définie par renvoi (instruction *valueref*) à l'élément de contenu.

- Une **structure physique** constituée d'unités de stockage appelées entités et qui contiennent soit du texte, soit des données binaires. Ces données binaires ne sont pas traitées par le processeur XML, mais plutôt selon l'indication de la notation qui les accompagne et qui indique quelle unité de traitement ou quel programme externe doit être mis en œuvre pour les présenter.

**Figure 6. Illustration d'un document XML avec une branche physique et une branche logique**



### 2.1.1 Une expérience XML : le cas des formulaires

Au printemps 1998, une expérience de modélisation basée sur la syntaxe XML a été amorcée et a permis de démarrer la production d'un Guide pour la création de structures logiques pour les formulaires. Il s'agit d'un choix stratégique pour plusieurs raisons. Premièrement, ce type de document se prête particulièrement bien à la modélisation, car le formulaire vierge a justement pour fonction de rendre explicite une bonne partie de sa structure logique, laquelle permet de diriger la saisie des données à insérer. Deuxièmement, leur volume considérable dans la masse documentaire gouvernementale et en fait un élément de dépense important (leur traitement surtout !) du budget d'opération de l'administration. Troisièmement, le formulaire en tant que concept générique n'aura pas une existence réduite dans le cyberspace, bien au contraire. Enfin, la facilité de développer et de déployer aujourd'hui des formulaires XML (*ipso facto* HTML lorsque requis) sur le Web offre une occasion propice aux besoins de la réingénierie dans l'Administration.

Examinons maintenant la notion de formulaire et les fonctions de la syntaxe XML, de manière à faire ressortir la symétrie de ces deux structures logiques.

D'une part, le terme formulaire désigne de façon générique une catégorie de documents qui se particularisent par une délimitation bien définie du contenu informationnel ; les formulaires constituent un moyen stable d'emmagasinage d'information choisie et ils offrent une interface de saisie et d'affichage familière. Les formulaires constituent une classe de documents au même titre que les manuels, les rapports, les lois, les notes, les contrats, etc. La plupart de ces classes de documents se retrouvent dans plusieurs domaines ; on peut les considérer comme les catégories grammaticales sous-tendant une partie de la structuration de l'information.

D'autre part, la norme XML est un langage de description générique et réutilisable entre logiciels de même fonction ainsi qu'entre les fonctions de la chaîne documentaire. La norme XML permet en effet de séparer le contenu du processus s'y appliquant : ainsi, seules les marques de balisage appropriées pourront être retenues pour une fonction donnée. Il devient plus facile d'y définir des conventions d'application et l'accès est ouvert à tout fournisseur de fonction ou de processus pouvant être appliqué à un formulaire. Le format XML permet de s'organiser au niveau de la structure logique d'un document tout en couvrant l'aspect des fichiers physiques.

On peut alors comprendre qu'en XML (orienté objet) une classe comme les lois puisse être considérée comme une seule DTD (définition de types de documents) représentant un seul modèle logique. Par opposition, les formulaires, pour leur part, forment un arbre de DTD de taille imposante.

Les formulaires électroniques se présentent sous plusieurs formes : des systèmes d'impression sur demande, des images électroniques (PDF), des fichiers propriétaires produits par des logiciels de bureautique comme Word, WordPerfect, Excel, Lotus, etc., ou des formats HTML et XML du Web. Si toutes ces possibilités techniques doivent être couvertes dans le choix de solutions appropriées, il faut se concentrer sur les formulaires du Web, HTML et XML ; en outre, la norme XML est à privilégier pour la modélisation des données et des documents, tandis que HTML reste utile, voire nécessaire, pour les solutions de présentation (affichage et impression). C'est donc via des modèles XML que l'information sera enregistrée parce que ce langage permet la précision requise pour opérer dans un contexte administratif et satisfaire aux exigences en

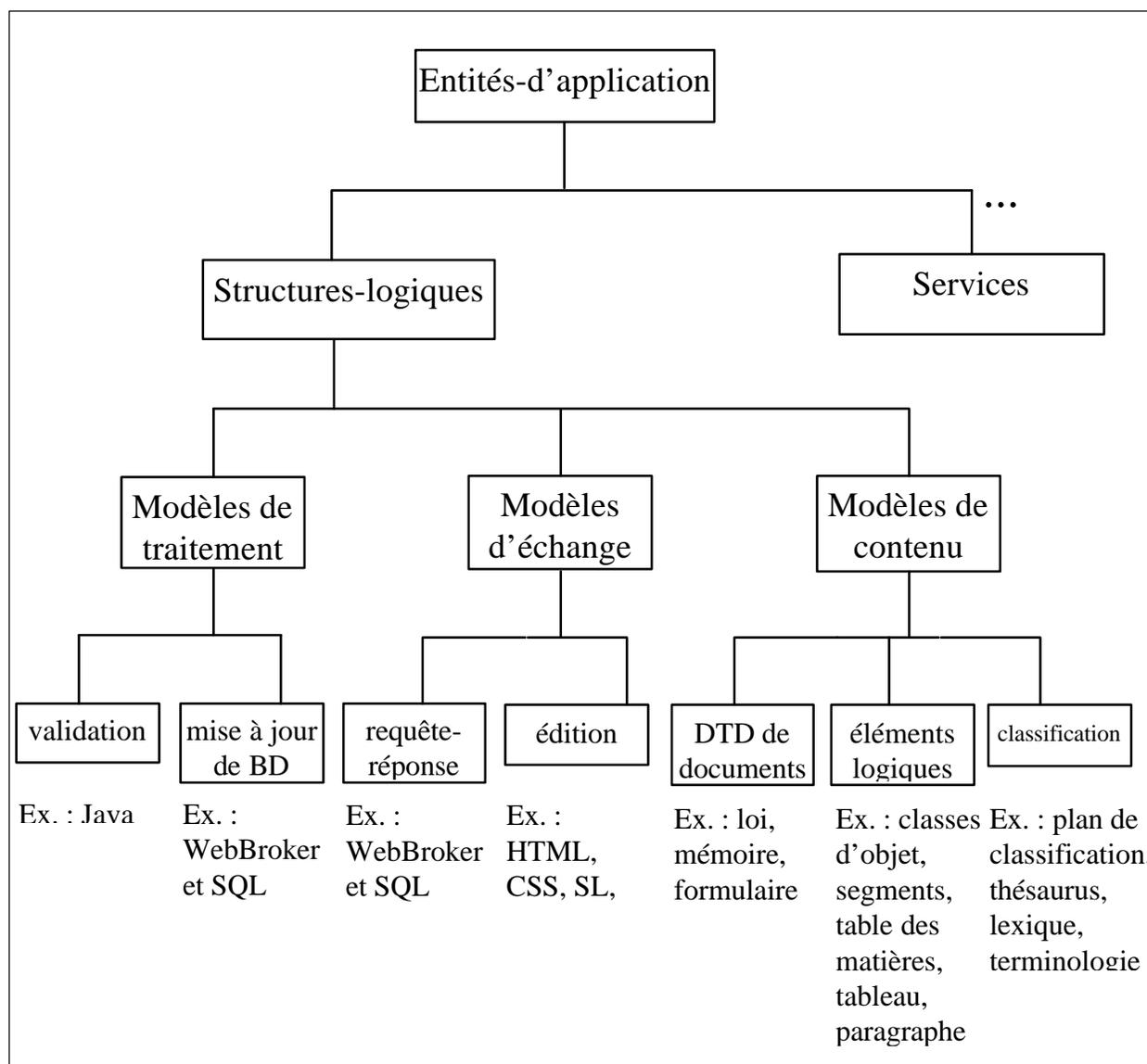
matière de collaboration, de circuit de production (*workflow*) et de contrôle, ainsi que pour transférer des données aux SGBD institutionnels hébergeant les systèmes administratifs conventionnels de paie, d'acquisition, etc.

### 2.1.2 Bénéfices à tirer du format XML dans le cas des formulaires

Le recours aux structures logiques offre trois paliers de bénéfices à obtenir du format XML :

- Le premier palier est constitué de la maîtrise des documents dans leur cycle de vie (même uniquement sur support électronique, les fichiers conservent tout de même une matérialité). La syntaxe XML est choisie comme le langage de base de la modélisation logique des documents. Les éléments logiques et leurs caractéristiques sont réutilisés par d'autres composantes fonctionnelles pour l'édition et l'impression ou l'affichage, la sélection du formulaire à remplir, la saisie assistée et la validation, l'acheminement des formulaires remplis pour traitement et emmagasinage. La structure logique décrite par la syntaxe XML est marquée sur les instances de document balisées, se prête à une interfonctionnalité forte entre logiciels et plates-formes et permet une réutilisation plus étendue dans chaque fonction et entre les fonctions.
- Le second palier est constitué de la maîtrise logique de l'information contenue dans les documents. La modélisation XML, comme les autres approches orientées objet, appréhende les objets informationnels sous l'angle sémantique du domaine ; c'est pourquoi elle s'intègre facilement aux bases de données qui peuvent aisément recueillir les renseignements transmis par formulaire. Ce second palier est porteur de beaucoup plus grands bénéfices, en termes qualitatifs du moins. Le format XML sera très présent dans un éventuel référentiel gouvernemental. Le référentiel serait une partie du Répertoire gouvernemental servant à l'emmagasinage de modèles de contenu. La figure 7 schématise le contenu éventuel qui pourrait y être rendu disponible pour le fonctionnement courant des échanges entre applications en réseau.
- Le troisième palier relève d'une dimension plus qualitative visant à favoriser l'unicité de la présentation visuelle de l'image gouvernementale, grâce à l'emploi d'un langage de style et de présentation (XSL).

**Figure 7. Illustration du contenu du référentiel avec une sous-classe Structures-logiques sous la classe Entités-d'application du Répertoire gouvernemental**



## 2.2 Une norme pour les métadonnées : le format RDF

Alors qu'au point précédent la syntaxe XML représentait la structure logique des données contenues dans le document, nous proposerons ici une norme pour représenter la structure logique des métadonnées. Le projet de norme principale pour les métadonnées est le format RDF (*Resource Description Framework*), qui tire sa force de sa capacité de représenter tous les

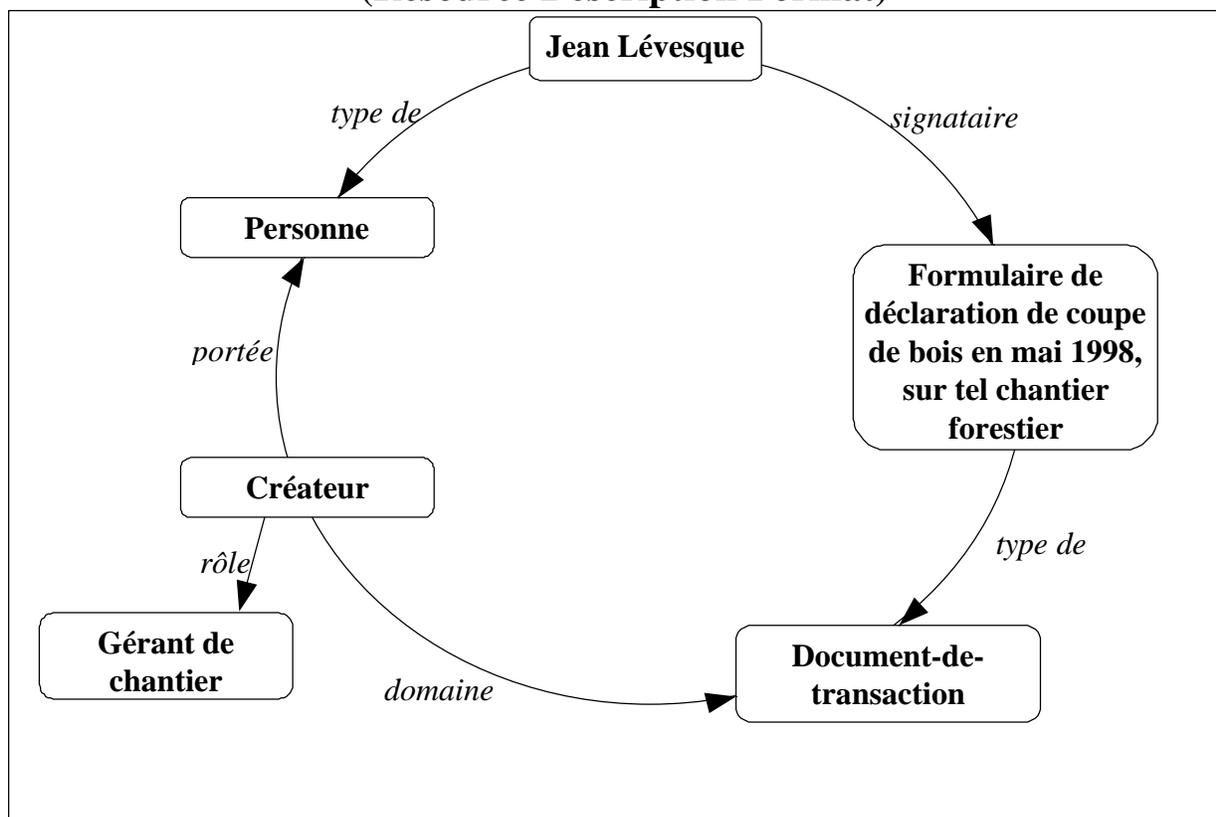
types de métadonnées et de fournir une base pour leur partage et leur traitement entre applications. Ainsi, alors que le format XML fournit la syntaxe générale des données du document, la norme RDF est associée à la description des métadonnées qui peuvent être attribuées au document. Ces dernières informant sur le sens des données, il devient alors possible de construire des voies d'accès plus directes, soit entre applications opérant sur des plates-formes distinctes, soit entre des éléments très précis d'information n'importe où sur le Web et dans toute autre application.

On pourrait se poser la question suivante : pourquoi utiliser la norme RDF ? Pourquoi ne pas se contenter de joindre les métadonnées à la structure logique du document XML ? En effet, une étiquette de métadonnées en elle-même ressemble étrangement à un document structuré. Mais, en réalité, l'étiquette n'est pas un document et le format RDF a été adapté à ses caractéristiques particulières :

- l'étiquette du document est emmagasinée de façon relationnelle (en tableau) avec une série de propriétés (paires attribut-valeur) qui n'ont pas d'ordre intrinsèque, tandis qu'un document (voir figure 9) est représenté comme une chaîne de caractères (linéaire) qui est découpée avec des balises portant des noms ;
- l'étiquette du document est emmagasinée comme une entrée dans un répertoire ou dans toute autre forme de SGBD ;
- l'étiquette est une forme parasitaire du document qui disparaît si son hôte, le document, est détruit (sauf dans certains cas pour des exigences particulières de la politique de conservation) ;
- le balisage et l'étiquetage peuvent partager des valeurs : un élément de document, le nom de l'auteur par exemple, est transféré comme valeur d'attribut sur l'étiquette de métadonnées.

En RDF, les métadonnées prennent la forme de PROPRIÉTÉS (incluant les attributs) qui servent à fournir de l'information sur les OBJETS (incluant les documents) auxquels elles sont rattachées. Les types de propriété ont la particularité d'être aussi des objets et peuvent donc être eux aussi décrits par des propriétés. La distinction entre données et métadonnées n'est pas tranchée ; au contraire, une donnée vue d'ici devient une métadonnée vue de là (figure 8).

**Figure 8. Illustration du mode de représentation  
des métadonnées en RDF  
(Resource Description Format)**



Le noyau du format RDF est la représentation de noms de propriété et de leurs valeurs. Les propriétés représentent :

- des attributs de ressources ;
- des relations entre ressources.

Dans ce modèle de données, à la fois les ressources décrites et les valeurs les décrivant sont des noeuds dans un graphe étiqueté orienté. Les arcs relient des paires de noeuds qui portent des noms de types de propriété.

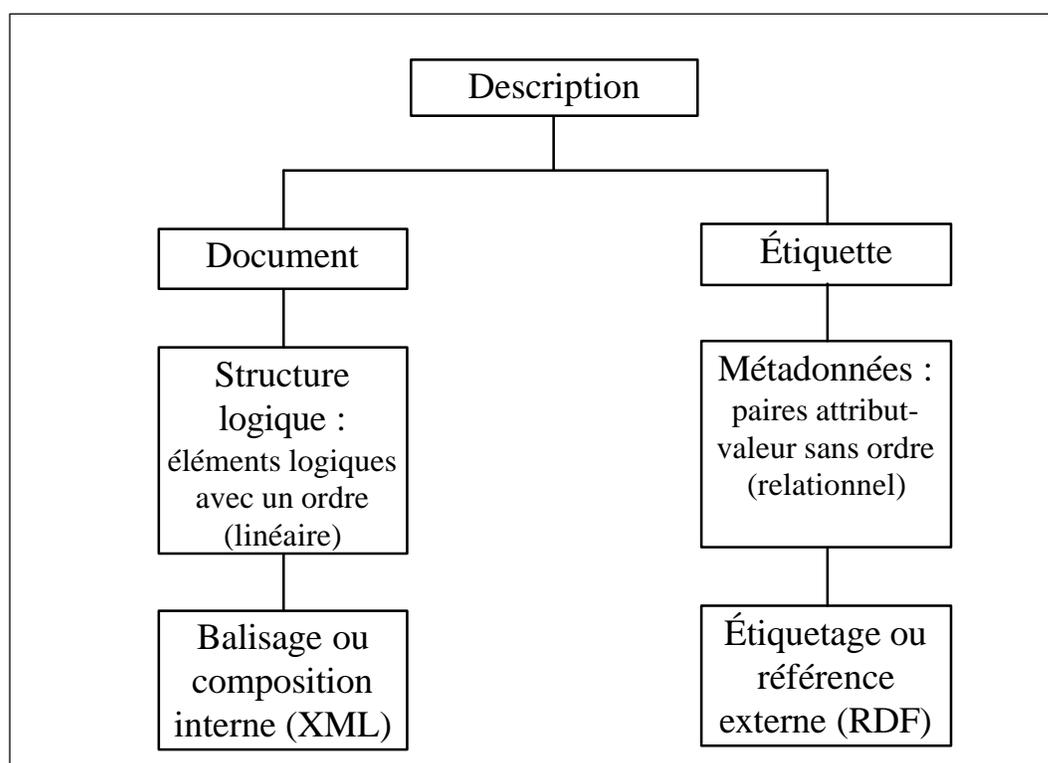
L'ensemble des propriétés d'un énoncé et les contraintes sur les valeurs sont définis dans un ou plusieurs schémas (incluant les profils de métadonnées) et identifiés par une adresse (URI ou URL).

## CHAPITRE 3

### UN PORTRAIT COMPLET D'UN DOCUMENT ÉLECTRONIQUE

Les normes XML et RDF permettent donc de tracer un portrait global d'un document électronique dont les composantes internes et externes peuvent être ainsi schématisées (voir figure 9) :

**Figure 9. Description interne et externe d'un document**



Le contexte électronique dans lequel le monde évolue présentement propose des défis importants à relever en matière d'organisation, de repérage et de transmission de l'information sous forme électronique. Pour assurer la gestion électronique des documents électroniques et pour que ceux-ci puissent traverser le temps en retenant les informations stratégiques qui éclairent le contexte de leur création, il est recommandé de s'appuyer sur les forces conjuguées des normes qui sont proposées dans ce rapport.