

Section 4 : Normes et standards

FRANÇOISE BANAT-BERGER
CLAUDE HUC



version 1

22 novembre 2011

Table des matières

Section IV : Normes et standards pour l'archivage numérique	5
Chapitre 1 : Objectif de la session.....	5
Chapitre 2 : Distinction entre norme et standard.....	5
Chapitre 3 Les organisations de normalisation et de standardisation.....	6
Chapitre 4 : Du standard à la norme.....	9
Chapitre 5 Quelques considérations importantes sur l'usage des normes et des standards.....	10
Chapitre 6 : Les normes et standards utilisés pour l'archivage numérique.....	11
Questions : NORMES ET STANDARDS POUR L'ARCHIVAGE NUMÉRIQUE	19

Section IV : Normes et standards pour l'archivage numérique

A. Chapitre 1 : Objectif de la session

Les normes et les standards vont jouer un rôle capital dans l'archivage numérique. Nous avons commencé à voir que cet archivage impliquait la résolution d'une multiplicité de problèmes distincts. Pour résoudre au mieux chacun de ces problèmes, nous aurons à nous appuyer sur des normes et des standards spécifiques facilitant l'émergence d'une solution. Au dessus de ces normes et standards propres à chaque domaine particulier, nous aurons aussi des normes généralistes qui traitent de l'archivage numérique dans sa globalité.

L'objectif de cette partie est de présenter une vue synthétique de cet ensemble de normes et standards et de donner un certain nombre d'éléments sur la manière de les utiliser.

Cependant, avant de regarder ces normes et ces standards, il est nécessaire de clarifier le vocabulaire, de prendre connaissance des principaux acteurs de ce domaine et de disposer de quelques éléments sur la manière de les utiliser.

B. Chapitre 2 : Distinction entre norme et standard

Ces deux termes sont souvent utilisés l'un à la place de l'autre alors qu'ils relèvent d'instances fort différentes. Cette confusion est essentiellement liée au fait qu'en anglais, il n'existe qu'un seul mot, le terme "standard" pour désigner les deux concepts.

2.1 - Norme

Les normes sont des ensembles de règles approuvées par des **instances officielles** en charge de la normalisation. Elles offrent une certaine garantie de stabilité et de pérennité.

Complément

La définition proposée par l'Office Québécois de la langue française^{1 2} est assez proche de la définition de l'ISO : «Ensemble de règles fonctionnelles ou de prescriptions techniques relatives à des produits, à des activités ou à leurs résultats, établies par consensus de spécialistes et consignées dans un document produit par un organisme, national ou

1 - <http://www.oqlf.gouv.qc.ca/ressources/bdl.htm>

2 - <http://www.oqlf.gouv.qc.ca/>

international, reconnu dans le domaine de la normalisation».

Des informations complémentaires intéressantes sont proposées dans le texte de Ghislaine Chartron intitulé « *Standards-Normes-Documents numériques : Introduction générale*³ »

2.2 – Standard

Les standards  sont définis par des groupes qui n'ont pas de mandats officiels des gouvernements. Ces groupes peuvent être :

- industriel ou commerciaux : par exemple PostScript ou PDF sont des standards de fait qui sont définis par la société Adobe,
- collégiaux comme : comme
 - o le W3C (*World Wide Web Consortium*)⁴
 - o le consortium Unicode⁵,
 - o le consortium OASIS (*Organization for the Advancement of Structured Information Standards*)⁶,

Un standard est aussi un procédé ou un service qui est largement répandu. Un standard peut être ou non une norme.

Les anglophones sont souvent amenés à désigner les standards "standard de facto" et les normes par "standard de jure" pour différencier clairement la nature des documents concernés.

Attention

Il arrivera, dans la suite de ce cours, d'utiliser le terme « norme » pour désigner des normes et des standards, afin d'alléger le texte.

C. Chapitre 3 Les organisations de normalisation et de standardisation

Globalement, la situation est assez complexe

3.1 - Les organismes de normalisation

Les organismes de normalisation nationaux sont mandatés par les gouvernements.

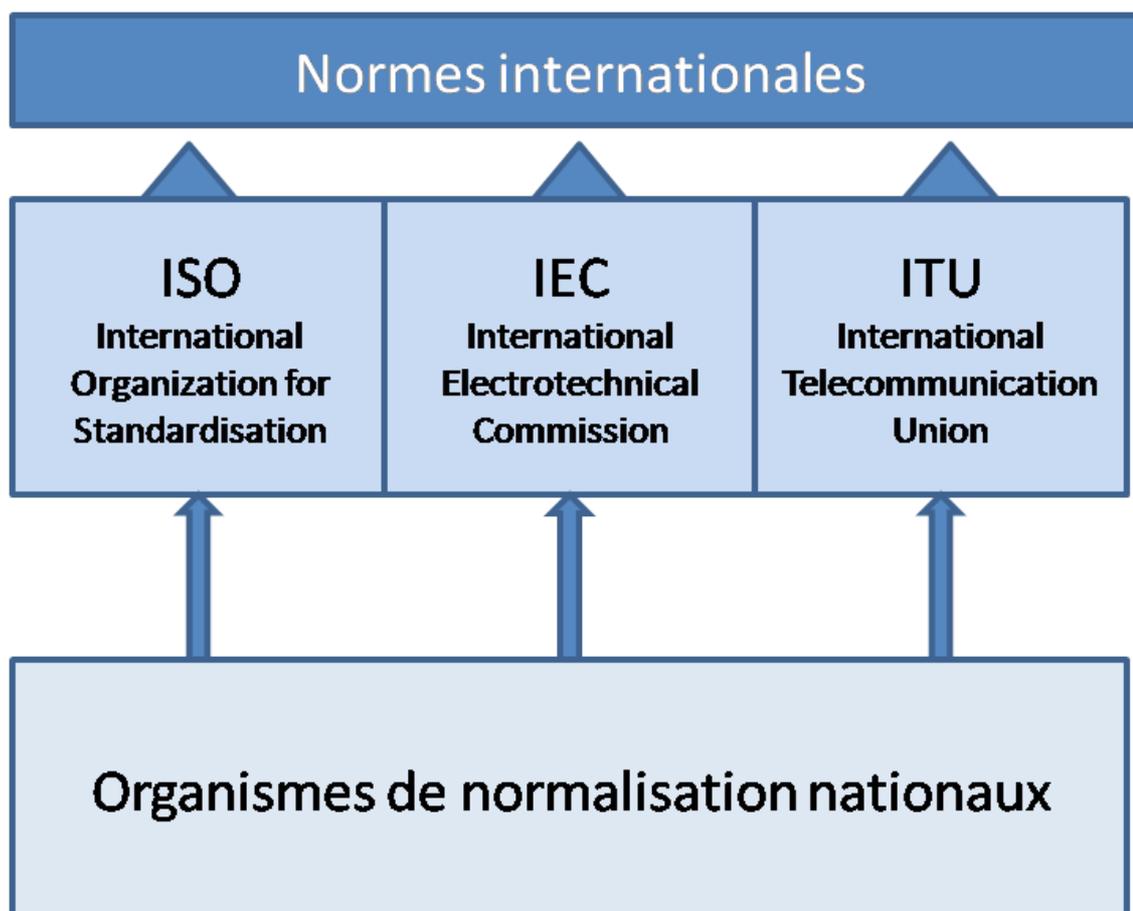
Ces organismes nationaux apportent leur contribution et dispose de droits de vote au sein des organismes internationaux.

3 - <http://urfist.enc.sorbonne.fr/anciensite/presse/standard/coursintro.htm>

4 - <http://www.w3.org/>

5 - <http://unicode.org/>

6 - <http://www.oasis-open.org/home/index.php>



Le schéma de principe du processus de normalisation internationale

Complément

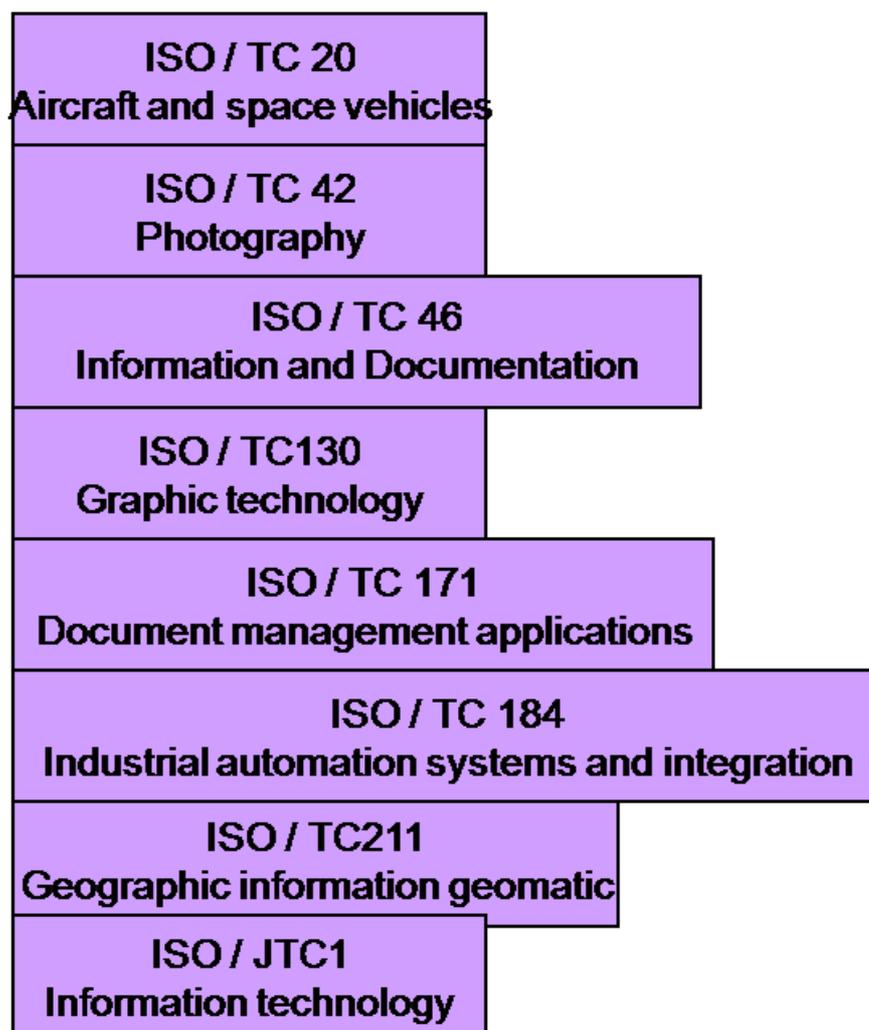
Les organismes nationaux peuvent **déléguer certaines de leurs responsabilités à des associations professionnelles**.

- En France, par exemple, le *BNAE, Bureau national de l'aéronautique et de l'espace*⁷ a reçu une délégation de l'AFNOR pour le développement et l'approbation des normes dans le domaine de l'aéronautique et de l'espace,
- Aux Etats-Unis, l'Institut of Electrical and Electronics Engineers (IEEE), fort de 360 000 membres de 175 pays a été accrédité par l'ANSI pour élaborer des normes dans son domaine de compétence. C'est en particulier IEEE qui a produit la norme aujourd'hui partout utilisée dans les ordinateurs, de représentation des nombres (ANSI/IEEE Std 854-1987 IEEE Standard for Radix-independent Floating-Point Arithmetic).

Il existe aussi des **organismes de normalisation au niveau continental**. C'est le cas du Comité européen de normalisation (CEN).

Les organismes de normalisation ont une **organisation hiérarchique, structurée par métier**. Cette organisation ne se prête pas bien à l'émergence de normes transverses et multiformes comme l'archivage numérique, chaque métier ayant tendance à trouver des solutions propres même si le problème à résoudre est le même que pour d'autres métiers. Ainsi l'ISO compte plus de 200 comités techniques (nommés TC pour Technical Committee), eux-mêmes subdivisés en un certain nombre de sous-comités. Un nombre important d'entre eux s'intéresse à des activités au cours desquelles des documents numériques, des images, des plans, des descriptions d'objet à deux ou trois dimensions..., à conserver durablement seront produits.

7 - <http://www.bnae.asso.fr>



Quelques uns des comités techniques de l'ISO, produisant des normes qui seront utilisées pour l'archivage numérique

3.2 - Les groupes de standardisation

Les groupes produisant des standards ont des structures très variées. Il existe :

- des groupes collégiaux (Unicode, W3C, OASIS, ECMA...)
Ils rassemblent, au plan international, des contributeurs provenant du secteur industriel, des universités, des institutions de recherche, etc.
- des groupes "métier"

C'est le cas du CCSDS (Comité consultatif pour les systèmes de données spatiales) qui est à l'origine de la norme OAIS (Open Archival Information System), fondamentale pour l'archivage numérique et dont nous parlerons en détail dans la partie 5.

D'autres groupes « métier » ont une vocation beaucoup large d'organisation, de coordination entre les membres, la standardisation ne constituant qu'une part plus ou moins grande de leur activité.

C'est naturellement le cas du Conseil international des archives (CIA) qui a publié les normes de description archivistique ISAD(G) (International Standard Archival Description-General), ISAAR(CPF) (International Standard Archival Authority Record for Corporate Bodies, Persons and Families), ISDF (International Standard for

Describing Functions) et ISDIAH (International Standard for Describing Institutions with Archival Holdings).

C'est également le cas de la société des archivistes américains (SAA) qui maintient la DTD EAD (Encoded Archival Description) pour la description des documents d'archives.

- des Institutions

Ainsi la library of congress (bibliothèque du Congrès des États-Unis) a publié et maintient le standard METS (Meta Data Encoding Transmission Standard) utilisé pour l'empaquetage des données et des métas-données.

(<http://www.loc.gov/ead/> , <http://www.loc.gov/standards/mets/>)

- des entreprises privées.

C'est le cas de Adobe qui est à l'origine du format PDF, mais il y en a beaucoup d'autres,

- des institutions ou des entreprises peuvent également produire des standards à vocation purement interne.

Cela de façon à homogénéiser certaines pratiques, développer et améliorer la qualité, les performances, la satisfaction des clients... Ces standards internes peuvent couvrir les domaines les plus divers, depuis les modèles à utiliser pour la création des documents jusqu'aux procédures à suivre en cas d'incident de production.

Mais il existe aussi des recouvrements :

Exemple

Le CCSDS par exemple est un groupe de standardisation qui joue également le rôle de sous-comité technique au sein de l'ISO : Comité technique 20 (Avions et véhicules spatiaux), sous-comité 13 (systèmes de transfert d'information et de données spatiales)

D. Chapitre 4 : Du standard à la norme

Il est assez fréquent qu'un standard, lorsqu'il est internationalement reconnu et utilisé devienne une norme.

Complément

- *HTML, Hypertext Markup Language pour les pages Web*
 - o **Standardisé par le consortium W3C (version 4.0 en 1997)**
 - o **Normalisé par l'ISO en 2000 (ISO 15445:2000)**
- *PNG, Portable network Graphics, pour les images*
 - o **Standardisé par le W3C en 1999**
 - o **Normalisé par l'ISO en 2004 (ISO 15948:2004)**
- *PDF – PDF/A*
 - o **PDF - format propriétaire (ADOBE)**
 - o **PDF/A est devenu la norme ISO 19005-1 en 2005**
 - o **PDF 1.7 est devenu la norme ISO 32000-1 en 2008**
- *ODF, Open Document Format, pour les documents bureautiques*
 - o **Standardisé par OASIS (mai 2005)**
 - o **Normalisé par l'ISO en 2006 (ISO 26300) en mai 2006**
- *OOXML, Office Open XML pour les documents bureautiques*

- o **Standardisé par ECMA en 2006**
- o **Normalisé par ISO en 2008 (ISO 29500)**
- *Le format de métadonnées Dublin Core*
 - o **Défini en 1995 par un groupe d'expert et d'utilisateurs des métadonnées de l'Internet,**
 - o **Normalisé par l'ISO en 2003 (ISO 15836)**

Soulignons enfin qu'au sein des groupes de standardisation, le W3C revêt une importance toute particulière :

- C'est un groupe promoteur des standards ouverts et libres de droit,
- Tous les standards du W3C sont disponibles en libre accès sur <http://www.w3.org/>
- Le W3C centre l'essentiel de son activité sur la structuration de l'information sous forme numérique,
- C'est lui qui est à l'origine de la standardisation de HTML, de XML (et de tous les langages spécialisés dérivés (par ex RDF, MathML, XHTML...), de PNG pour les images raster, de SVG (scalable vector graphics) pour les graphiques vectoriel

E. Chapitre 5 Quelques considérations importantes sur l'usage des normes et des standards

La norme peut être un instrument de pouvoir :

« Les normes, en donnant une capacité d'agir sur le comportement des autres acteurs et à les conformer à un modèle défini, constituent un instrument de pouvoir. Au sein des organismes qui les élaborent, le choix des normes (sujet et contenu) constitue donc un enjeu d'influence et de puissance. Les membres des organisations de normalisation cherchent à imposer leur vision du monde et à conforter leur propre stratégie de conquête et de protection des marchés. »

« À Armes Égales » Rapport au Premier Ministre par Bernard Carayon, 2006

Ce n'est pas parce qu'une norme existe qu'elle est utilisée :

Si la norme n'est que très peu utilisée ou pas utilisée du tout, il n'y aura pas ou peu d'outils logiciels facilitant son usage, sa qualité n'aura pas été éprouvée.

Exemple

Un exemple particulièrement frappant est celui de la norme ODA (Open Document Architecture). ODA a été défini comme *format d'échange* de documents. C'est un standard ECMA publié dès 1985, il a été encouragé par la Commission Européenne et il est devenu la norme ISO 8613 en 1989. Cette norme est homologuée par l'Afnor comme une norme française toujours en vigueur mais jamais utilisée !

Critères à prendre en compte pour le choix des normes et standards :

Les premiers critères seront sans doute

- l'adaptation aux besoins,
- la complexité et le coût de mise en œuvre dans le contexte de sa mise en application,
- les aspects politico-industriels ne doivent pas être ignorés.

Mais constituent également des éléments qui pèseront sur les choix finaux

- l'éventuelle obsolescence : si les normes constituent des règles dont la stabilité et la durée de vie est supérieure à celle des technologies purement propriétaires, il n'en

- reste pas moins qu'elles deviennent obsolètes au bout d'un certain temps, la dépendance par rapport à des technologies en passe de disparaître.

Attention

Prix des normes

Le plus souvent, le texte des standards sont en accès libre et gratuit.

Ceux des normes sont payants et chers ; ce prix est parfois un frein sérieux à leur diffusion et à leur utilisation.

Il est bon de savoir que dans certains cas, lorsqu'un standard a été normalisé, le même document peut être payant auprès de l'organisme de normalisation et gratuit auprès du groupe de standardisation qui l'a élaboré. L'organisme de normalisation se contentant de changer la couverture du document pour y apposer son nom et son logo. C'est en particulier le cas des standards du CCSDS, dont le Modèle de Référence OAIS, qui sont accessibles librement sur *le site du CCSDS*⁸

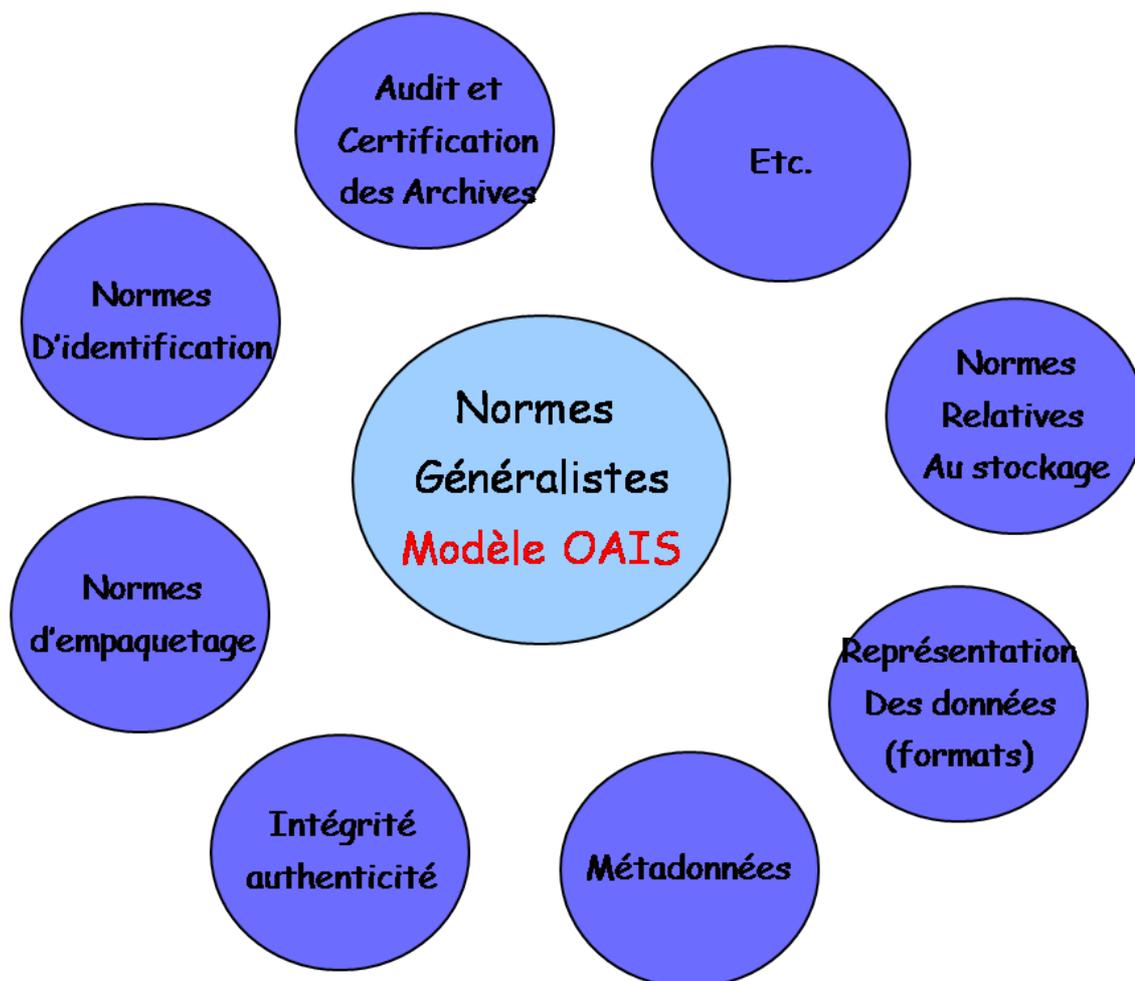
F. Chapitre 6 : Les normes et standards utilisés pour l'archivage numérique

Face à une technologie qui ne cesse d'aller d'innovations en innovations, face à ces changements constants et rapides, face aux besoins majeurs d'une certaine compatibilité technique entre les acteurs de l'archivage numérique, les normes et standards apportent un semblant de stabilité sur laquelle nous pouvons nous appuyer. C'est la raison pour laquelle, ces bases consensuelles vont jouer un rôle essentiel pour cet archivage.

6.1 - Première vue d'ensemble

Il ne s'agit pas ici de dresser un inventaire qui ne sera jamais exhaustif des normes et standards utilisables. Nous proposons une classification et une identification des principaux domaines normatifs. Ces principaux domaines sont représentés dans la rosace ci-après

8 - <http://public.ccsds.org/publications/default.aspx>



Les principaux domaines normatifs de l'archivage numérique

Les normes généralistes et plus spécifiquement le modèle de référence OAIS sont celles qui abordent la question de l'archivage numérique dans sa globalité. Elles seront l'objet de la section 5 de ce module.

Les autres domaines seront analysés plus en détail dans les sections 6 à 10, mais il est utile, à ce stade du cours, d'en donner un premier aperçu.

6.2 Principaux domaines normatifs à considérer

Ces domaines sont nombreux et il n'est pas imaginable ni imaginé qu'une même personne soit en mesure de connaître en profondeur tous ces domaines. Il sera cependant indispensable que tout responsable d'une Archive numérique dispose d'une vue d'ensemble de tous ces domaines et puisse s'appuyer sur des compétences spécialisées pour chaque d'eux quand cela sera nécessaire. Nous faisons référence, dans ce qui suit, à un certain nombre de normes sur lesquelles nous reviendrons dans les parties suivantes de ce cours.

Nous pouvons citer, dans ce panorama d'ensemble, un certain nombre de documents qui relèvent plus des bonnes pratiques ou des recommandations que des normes ou standards.

Stockage :

c'est tout ce qui concerne la préservation des bits.

On trouvera pour cela

- d'une part des guides et recommandations sur la stratégie de stockage, l'organisation d'un service de stockage
- d'autre part des normes sur les supports d'enregistrement

Exemple

- La norme ISO 9660 pour les CD-Rom
- Les normes ISO 13421 et 13962 pour les DLT (Digital Linear Tape)

Ces normes sur les supports d'enregistrement sont **utilement complétées par les évaluations de certains de ces supports.**

Exemple

Le Laboratoire national de métrologie et d'essais (LNE) en France a conduit un certain nombre d'études sur l'usage des CD et des DVD en vue de l'archivage et émis un ensemble de recommandations dans ce domaine. Ces informations et rapports sont disponibles sur le site de la *Direction des archives de France*⁹.

Formats de représentation de l'information :

Qu'il s'agisse de texte, d'image, d'images 3D, de graphiques vectoriels, de documents multimédia, d'audio, de vidéo, de données techniques, de données scientifiques, **toutes les formes d'information numérique se réfèrent à des formats.**

Il existe des formats généralistes et des formats métier :

Complément

- Dans les formats généralistes, on trouvera par exemple le PDF, les principaux formats bureautiques ODF et OOXML, le méta-langage XML, le format d'image PNG. Le développement spectaculaire de l'usage méta-langage XML a constitué une évolution majeure dans le domaine de l'accès, du partage et de l'archivage des données. PDF, ODF et OOXML reposent aujourd'hui largement sur ce méta-langage.
- Dans les formats 'métier', on pourra citer FITS (Flexible Image Transport System) pour l'astronomie ou encore les différentes applications du langage STEP (STandard for the Exchange of Product model data) pour l'industrie.* Dans les formats métier, les normes relatives aux données géographiques tiennent également une place importante.

Métadonnées :

Rien de très nouveau sur le principe et l'on n'a pas attendu le numérique pour inventer les métadonnées.

Depuis des siècles, les catalogues des bibliothèques répertorient des métadonnées propres à chaque ouvrage.

Cependant, avec le numérique, des **besoins nouveaux** sont apparus, par exemple :

- disposer de métadonnées techniques garantissant la capacité future à restituer une information compréhensible à partir d'un ensemble de séquences de bits,
- retrouver immédiatement les documents correspondant à un besoin donné, au sein de millions d'autres documents, éventuellement stockés de façon répartis sur des sites géographiquement distants

En fait, la généralisation du numérique a entraîné l'émergence de nombreux formats de métadonnées

- à caractère généraliste comme le **Dublin Core** (ISO 15836)
- ou orienté métier comme la **norme ISO 19115 de métadonnées géographiques**

9 - <http://www.archivesdefrance.culture.gouv.fr/gerer/archives-electroniques/stockage/>

qui propose plus de 400 rubriques descriptives.

La section 9 de ce cours est consacrée aux métadonnées.

Intégrité et l'authenticité :

Il s'agit ici des normes qui vont nous permettre d'apporter des preuves de l'intégrité et de l'authenticité des documents.

Ces normes vont couvrir :

- la cryptographie
- les algorithmes de calcul d'empreinte
- les protocoles de communication sécurisés
- les processus de signature électronique.

Un certain nombre de standards ont été proposés par le projet NESSIE (New European Schemes for Signatures, Integrity, and Encryption) de la Commission Européenne.

Le NIST (National Institute for Standards and Technology), américain propose également un ensemble de standards et en définitive, l'ISO a entériné les choix du NIST et de NESSIE et normalisé les principaux algorithmes de calcul d'empreinte numérique.

Ce sujet sera développé dans la section 10, couvrant l'intégrité, l'authenticité et les aspects juridiques....

Empaquetage :

Les standards d'empaquetage permettent d'associer, au sein d'une structure globale cohérente et portable des objets numériques pouvant être composés

- d'un ou d'un ensemble de fichiers,
- de métadonnées caractérisant ces objets et ces fichiers,
- des liens existant entre les fichiers eux-mêmes ou entre les composants de l'objet et les métadonnées

Ces standards permettent aussi de définir des mécanismes applicables aux objets numériques : ils peuvent exiger, par exemple que tel objet doit être ouvert en utilisant telle application, ou encore que les différents fichiers de l'objet doivent être ordonnancés et traités de telle façon pour pouvoir restituer leur contenu.

Nous parlons d'empaquetage à l'instar des paquets expédiés par la poste. En effet, un paquet postal permet de transporter toutes sortes d'objets, mais il sera toujours doté d'une adresse du destinataire libellée de façon normalisée, d'une adresse de l'expéditeur, d'un emballage répondant à un certain nombre de règles de base.

Les standards d'empaquetage vont permettre d'identifier, de décrire, de manipuler les composants numériques élémentaires d'un objet complexe

Les standards METS (Metadata Encoding and Transmission Standard) de la Bibliothèque du Congrès et XFDU (XML Formatted Data Unit) du CCSDS relèvent de ce domaine qui sera développé dans la partie 9 sur les métadonnées.

Identification :

Elle consiste à identifier un objet numérique de manière unique au sein d'un domaine au sein de l'archive ou plus souvent aujourd'hui, au sein d'un domaine beaucoup plus vaste, au sein d'une communauté.

Ce besoin n'est pas nouveau et a déjà été résolu par exemple avec l'usage de l'ISBN (International Serial Bibliographic Number) pour les ouvrages publiés ou de l'ISSN (International Standard Serial Number) pour les revues.

La question de l'identification des documents et plus généralement de tous les types de ressources numériques se pose de manière particulièrement aiguë dans le domaine numérique. Des standards d'identification (URI, Uniform Resource Identifier) et de nommage (URN, Uniform Resource Name) ont été définis par le consortium W3C, mais aujourd'hui, les navigateurs ne savent pas les interpréter. Les navigateurs ne savent interpréter que les URN (Uniform Resource Locator) qui sont des standards de localisation.

Chacun sait aujourd'hui, à quel point, les URL ne sont pas pérennes. Combien de fois avez-

vous constaté que l'adresse d'une ressource sur Internet, adresse que vous aviez mémorisé, n'était plus valide.

Il y a un besoin très important de disposer d'identifiants pérennes.

Différents services d'affectation d'identifiants pérennes comme ARK « Archival Resource Key » ou les DOI « Digital Object Identifier » ont vu le jour. Ils seront regardés plus en détail dans la partie 9 consacrée aux métadonnées.

La certification des archives numériques :

Le document numérique fait souvent un peu peur. Sa pérennisation pose toute une série de nouveaux problèmes. Jusqu'à quel point pouvons-nous avoir confiance dans les organismes en charge d'assurer cette pérennisation ?

D'une manière générale, lorsqu'une entreprise ou une institution veut démontrer qu'elle fait bien son travail, qu'elle est apte à fournir des produits ou des services satisfaisant aux exigences des clients et à la réglementation applicable, qu'elle vise en permanence à accroître ses performances et la satisfaction de ses clients, elle met en place un système de contrôle de la qualité et entreprend un processus de certification ISO 9001. Cette certification étant délivrée pour une durée définie par des auditeurs indépendants, eux-mêmes accrédités par un organisme agréé pour délivrer ces accréditations.

Dans le domaine de l'archivage numérique, plusieurs approches émergent en matière de certification :

- une approche basée sur un audit de conformité du service d'archive numérique par rapport à une spécification technique existante comme la norme Afnor NF Z 42-013 « Spécifications relatives à la conception et à l'exploitation de systèmes informatiques en vue d'assurer la conservation et l'intégrité des documents stockés dans ces systèmes, mars 2009 ».
- une approche plus ambitieuse et couvrant réellement la totalité de la problématique de l'archivage numérique, approche basée sur deux normes pour la certification des archives numériques en cours d'élaboration au CCSDS
 - Metrics for Digital Repository Audit and Certification
 - Requirements for bodies providing audit and certification of digital preservation management systems

Les questions liées à la certification des archives seront abordées dans la partie 5.

Autres domaines normatifs : parmi les autres domaines normatifs à prendre en compte dans l'archivage numérique, il convient de citer également

- tout ce qui concerne les systèmes de management de la sécurité de l'information (incluant la confidentialité, la protection des informations sensibles, les plans de continuité d'activité)
- ainsi que les questions relatives à l'externalisation de prestations d'archivage,

Complément

Le projet InterPARES 3 a publié un document contenant une liste impressionnante des normes et standards utiles à la préservation de l'information numérique *General Study: International Standards Relevant to the InterPARES 3 Project*.¹⁰

Ce document, qui référence près d'une centaine de normes et de standards potentiellement utiles, manque un peu de structuration. Le rangement par numéro ISO ne sera que d'une faible utilité pour le lecteur. Des normes importantes, comme la norme ISO 10646 définissant le jeu universel des caractères, n'y sont pas mentionnées.

6.3 – Les normes et standards sont comme des poupées russes

Il faut savoir qu'un certain nombre de normes se présentent comme des poupées russes,

10 - http://www.interpares.org/display_file.cfm?doc=ip3_gs04_international_standards_v1-4p.pdf

c'est-à-dire qu'elles s'appuient sur d'autres normes, ces dernières pouvant également s'appuyer sur d'autres normes et ainsi de suite.

Attention

- La mise en application d'une norme d'un certain niveau n'implique pas que les normes de niveau inférieur soient appliquées correctement.
- Dans bien des cas, le choix d'une norme ne suffit pas, il convient donc d'ajouter à ce choix, des conditions d'applications particulières obligatoires.

Complément

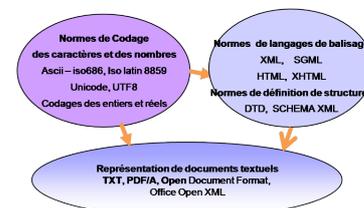
Prenons le cas de la norme ISO 26300, Open Document Format :

- Cette norme s'appuie en premier lieu sur un standard de structuration de document nommé Relax NG, standard qui a été développé par le consortium OASIS, puis normalisé par l'ISO ((ISO/IEC 19757-2),
- Relax NG est basé sur le langage XML, standard du W3C,
- XML utilise par défaut, le codage des caractères UTF-8. UTF-8 est une forme de codage directement dérivé du jeu de caractères universel défini par la norme ISO 10646 (connue également sous le nom d'Unicode). Mais XML peut aussi s'appuyer sur d'autres types de codage, le jeu de caractère utilisé étant spécifié en début du document XML.

Nous allons retrouver de telles situations dans les normes descriptives de formats de données, les normes de métadonnées, d'identification, d'emballage et d'autres encore.

La mise en application d'une norme d'un certain niveau n'implique pas que les normes de niveau inférieur soient appliquées correctement.

En ce qui concerne la représentation de l'information sous forme numérique, notre préoccupation devra toujours couvrir dans sa totalité, la chaîne logique qui conduit de la séquence de bits jusqu'à la restitution d'un document intelligible.



Nous avons dans le passé rencontré de multiples problèmes autour de ce sujet. Un exemple très simple est celui de l'enregistrement, sous Windows 2000, avec Word 97, d'un fichier au format HTML, en vue d'insérer ce fichier sur un site Internet.

Constat : comme l'illustre l'image ci-après, après transfert sur le site Internet, la visualisation de ce fichier apparaît correcte lorsqu'il est examiné à partir du système d'exploitation Windows et n'est pas correcte lorsqu'il est visualisé à partir d'un autre système d'exploitation comme Linux. Certains caractères spéciaux ayant été remplacés par des points d'interrogation.

- le Comité de Direction d'Aristote a officialisé, le 28 Septembre 2000, la Direction afin d'expliquer ce que sont les motivations et les objectifs de
- Le dépouillement des priorités affectées par chacun aux différents thème

Vu sous LINUX

Vu sous Windows

- le Comité de Direction d'Aristote a officialisé, le 28 Septembre 2000, la Direction afin d'expliquer ce que sont les motivations et les objectifs du
- Le dépouillement des priorités affectées par chacun aux différents thème

Mise en évidence d'anomalies dans l'affichage de certains caractères par le navigateur

Explication : Examen du fichier HTML à l'aide d'un éditeur de texte. Il convient de savoir qu'en début de chaque fichier HTML, un attribut, nommé « charset », permet de spécifier le jeu de caractères utilisé dans le fichier. Dans le cas présent, nous pouvons constater que le fichier a été créé, non pas avec un jeu de caractères normalisé, mais avec un jeu de caractères propriétaire, à savoir le jeu de caractères « Windows-1252 » propre au système d'exploitation Windows. Ce fichier reste néanmoins tout-à-fait conforme à la norme HTML qui n'impose pas un jeu de caractère particulier.

Fichier MS Word 97 - enregistré au format HTML sous Windows

```
<HTML>
```

```
<HEAD>
```

```
<META HTTP-EQUIV="Content-Type"
CONTENT="text/html; charset=Windows-1252">
```

```
<TITLE>Association Aristote – Groupe
Pérennisation des Informations
Numériques</TITLE>
```

Mise en évidence de l'usage d'un jeu de caractères propriétaire

Conclusion : Si nous voulons archiver à long terme, un fichier au format HTML, nous devons spécifier :

- Que le fichier doit se conformer strictement à la norme ISO 15445:2000 qui définit le format HTML en version 4.0, (une version 5.0 sera certainement normalisée ultérieurement),
- Mais aussi que le jeu de caractères utilisé doit être un jeu normalisé à l'ISO.

Cet exemple permet de comprendre, ce que nous disions plus haut, que dans bien des cas, **le choix d'une norme ne suffit pas mais qu'il convient d'ajouter à ce choix, des conditions d'applications particulières obligatoires.**

Questions : NORMES ET STANDARDS POUR L'ARCHIVAGE NUMÉRIQUE

Avez-vous compris tout ce qui vient de vous être enseigné ?
Si vous voulez le vérifier, faites les exercices proposés ci-dessous.

Si vous ne savez pas répondre, ne regardez pas trop vite le corrigé, travaillez à nouveau la (les) section(s) précédente(s) où vous découvrirez les solutions.

Bien sûr, si vous n'y arrivez vraiment pas, vous pouvez consulter les réponses. Ne les lisez pas avec précipitation mais avec une grande attention et surtout essayez de comprendre.

Remarque sur la limite de nos exercices

Nous vous proposons des exercices sur le numérique avec toutes les réserves que cela comporte en raison notamment des évolutions technologiques. Nous sommes en effet dans une discipline récente qui est en constante évolution. Les principes qui ont été posés dans ce module resteront sûrement durablement valides, par contre le nombre de solutions aujourd'hui valides ne le seront plus que partiellement demain ou plus du tout.

Par exemple : dans un contexte organisationnel donné, ce que nous pouvons recommander en matière de formats ou de moyens de stockage, pourra être remis en cause dans un ou deux ans.

Il est donc nécessaire que les partenaires d'un projet d'archivage numérique s'approprient les principes mais ne procèdent à l'analyse de la situation et des contraintes qu'au moment où le projet se met en marche et cela avec les partenaires concernés ; en effet, lorsqu'un archiviste peut se trouver seul au sein de son organisation, l'archivage numérique ne doit pas être une activité individuelle : ce sera toujours la mise en commun d'un ensemble de compétences complémentaires. Donc rien ne remplacera les exercices en vraie grandeur.

Nous rappelons que les exercices du PIAF sont à l'usage d'une auto-formation. Nous proposons à cet effet des types d'exercices de mémorisation, d'accompagnement, de positionnement afin de permettre à l'utilisateur de vérifier l'acquisition d'une culture minimale précise.

Les exercices ci-dessous porteront surtout sur les principes puisqu'il n'est pas possible d'aller trop loin en matière de solution.

Exercice 1

Les normes apportent une vraie solution durable et stable pour résoudre la problématique de l'archivage numérique

- Vrai
- Faux

Exercice 2

Citer les huit domaines normatifs importants pour l'archivage, tels qu'ils ont été identifiés

dans la section 4

- | | |
|--------------------------|--|
| <input type="checkbox"/> | Normes généralistes (OAIS) |
| <input type="checkbox"/> | Stockage |
| <input type="checkbox"/> | Représentation des données (formats) |
| <input type="checkbox"/> | Normes relatives aux documents photographiques |
| <input type="checkbox"/> | Normes relatives aux documents bureautiques |
| <input type="checkbox"/> | Métadonnées |
| <input type="checkbox"/> | Format des disques optiques numériques |
| <input type="checkbox"/> | Identification |
| <input type="checkbox"/> | Empaquetage |
| <input type="checkbox"/> | Format des images |

Exercice 3

suite des réponse possibles

- | | |
|--------------------------|--|
| <input type="checkbox"/> | Intégrité-Authenticité |
| <input type="checkbox"/> | Algorithmes de compression |
| <input type="checkbox"/> | Gestion documentaire |
| <input type="checkbox"/> | Audit et certification des archives numériques |
| <input type="checkbox"/> | Technologies de l'information |

Exercice 4

Le World Wide Web consortium (W3C), à l'origine des formats HTML, XML, PNG, etc. est un organisme de normalisation

- | | |
|-----------------------|------|
| <input type="radio"/> | Vrai |
| <input type="radio"/> | Faux |

Exercice 5

Le BNAE, l'AFNOR, l'ANSI, l'IEEE, le CEN sont tous des organismes qui élaborent des normes.

- | | |
|-----------------------|------|
| <input type="radio"/> | Vrai |
| <input type="radio"/> | Faux |

Glossaire

Norme

Document, établi par consensus et approuvé par un organisme reconnu, qui fournit, pour des usages communs et répétés, des règles, des lignes directrices ou des caractéristiques, pour des activités ou leurs résultats, garantissant un niveau d'ordre optimal dans un contexte donné (définition officielle de l'ISO/CEI).

Standards

Convention fondée sur un consensus plus restreint que pour la norme, généralement élaboré entre des industriels au sein de forums ou de consortiums (source Afnor).

Bibliographie

[Premier ouvrage de synthèse sur l'archivage numérique en langue française.]

- BANAT-BERGER F., HUC C., DUPLOUY L., L'Archivage numérique à long terme, les débuts de la maturité? Paris, La Documentation française, 2009.

[Norme de référence essentielle pour comprendre le problème posé par l'archivage numérique]

[http://public.ccsds.org/publications/archive/650x0b1\(F\).pdf](http://public.ccsds.org/publications/archive/650x0b1(F).pdf)