

Section 5 - Modèle de référence OAIS et normes associées

version 1

FRANÇOISE BANAT-BERGER
CLAUDE HUC

22 novembre 2011

Table des matières

Section 5 : Modèle de référence OAIS et normes associées	3
Chapitre 1. Objet de la section.....	3
Chapitre 2. Premiers éléments.....	5
Chapitre 3. Le modèle d'information.....	6
Chapitre 4. Le modèle fonctionnel.....	13
Chapitre 5. Impact du modèle OAIS.....	15
Chapitre 6. Interface Producteur-Archive.....	16
Chapitre 7. Norme Afnor NF Z 42-013.....	19
Chapitre 8. Certification des Archives numériques.....	20
Chapitre 9. Conclusions.....	22
Questions : OAIS	24
Exercice : Révisez vos connaissances	27
Solution des exercices	29
Glossaire	32
Bibliographie	33
Webographie	34

Section 5 : Modèle de référence OAIS et normes associées

Chapitre 1. Objet de la section

Le modèle de référence pour un système ouvert d'archivage d'information (*Reference Model for an Open Archival Information System*) a été défini par le CCSDS. Il est devenu standard CCSDS en 2002 puis norme ISO 14721 en 2003.

Pour plus de simplicité, il est appelé « modèle OAIS ».

C'est un modèle abstrait qui nous permet de comprendre en profondeur toutes les spécificités de l'archivage numérique à long terme. Son caractère abstrait le rend parfois difficile à appréhender, mais il est essentiel.

Pourquoi est-il essentiel ?

Nous verrons à quel point ce modèle est essentiel pour l'émergence de solutions pratiques applicables dans le domaine de l'archivage numérique.

En effet :

- il définit les concepts indispensables à la compréhension et à l'analyse du problème,
- il propose une terminologie permettant aux diverses communautés concernées de dialoguer et de se comprendre indépendamment des vocabulaires spécifiques à leurs domaines respectifs,
- il contribue à l'élaboration de méthodologies.

Donc :

- il va donc constituer une base absolument incontournable pour la définition, l'élaboration et la mise en œuvre de solutions efficaces et fiables,

il sera enfin une référence pour le développement d'outils logiciels nécessaires à l'automatisation des processus.



Attention

- Le modèle OAIS ne définit pas une mise en œuvre.

Cette mise en œuvre sera d'ailleurs en partie dépendante des technologies et outils disponibles.

- Remarquons aussi que le modèle OAIS n'est absolument pas destiné de manière spécifique aux informaticiens, il doit permettre à toutes les personnes impliquées dans un processus d'archivage numérique, donc en premier lieu aux archivistes, de comprendre la logique d'ensemble de ce processus et par conséquent, de le piloter ou d'y participer de façon efficace.

Objectif de la section

L'objectif essentiel de cette section 5 est de présenter ce modèle aussi simplement que nous le pouvons.

Toutefois nous limitons notre présentation à ce qui est fondamental.

Il appartient au lecteur d'aller plus loin ensuite, en prenant directement connaissance de la norme.

Qu'allons-nous traiter plus spécifiquement ?

- Tous les concepts propres au modèle OAIS qui seront constamment réutilisés dans la suite de ce cours ; il est par conséquent essentiel de bien les comprendre.
- Plusieurs normes dérivées du modèle OAIS, qui traitent de l'interface entre le service d'archives numériques et les producteurs de documents ainsi que de la certification, seront également abordées dans cette section 5.
- Une autre norme généraliste d'une portée plus limitée, 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 » sera également analysée.
- La norme ISO 15489 consacrée au « Records management », autrement dit « gestion de l'archivage », doit également être considérée comme une norme généraliste, même si elle n'est pas spécifique au domaine numérique ni approfondie sur les questions de pérennisation (elle est abordée plus en détail dans le module 6 «Gestion des archives définitives»).

Le modèle OAIS existe en plusieurs versions

En accès libre et gratuit sous la forme de standard CCSDS :

- *Version 2002 en langue anglaise*¹
- *Version 2005 en langue française*² (traduction française de la version anglaise)

En accès payant (224 francs suisses) sous la forme de Norme ISO (14721: 2003) en langue anglaise ou française. Il s'agit strictement de la même norme que celle du CCSDS, mais dotée d'une couverture ISO).

La **version 2009** est en cours de normalisation :

Après une très large utilisation au niveau mondial et une réflexion approfondie, une nouvelle version du modèle OAIS est en cours de validation.

Tous les concepts essentiels ont été conservés, mais cette version comporte un certain nombre d'améliorations par rapport à la version de 2003 :

- elle aborde, plus en profondeur, la question des droits associés aux documents archivés
- et traite de la question de l'authenticité de ces documents.

Nous mettrons ce cours à jour dès que le document sera publié et normalisé.



Image 1 Version en langue française



Complément : Un processus de développement original

En 1996, le CCSDS a été saisi d'une demande de norme permettant de spécifier les conditions dans lesquelles il serait possible d'assurer l'archivage à long terme des données issues des expériences spatiales.

L'analyse préliminaire du CCSDS sur ce sujet a abouti à deux conclusions :

1. Il n'était pas possible de réaliser une telle norme car il n'existait aucun consensus sur le sujet. Il convenait d'élaborer en premier lieu un modèle permettant de définir les concepts et le vocabulaire indispensables à la caractérisation du problème à résoudre,
2. Le problème posé n'avait que très peu de spécificités propres au domaine spatial et il convenait donc de réfléchir à

1 - <http://public.ccsds.org/publications/archive/650x0b1.pdf>

2 - <http://public.ccsds.org/publications/NonEnglishVersions.aspx>

l'élaboration de ce modèle de la façon la plus large et la plus ouverte possible.

Le modèle OAIS a donc été réalisé par des représentants des grandes agences spatiales internationales, mais avec le concours et la contribution permanente de représentants d'archives nationales et de grandes bibliothèques.

La décision du CCSDS d'ouvrir le groupe de travail en charge du modèle, à des experts d'autres domaines, d'autres métiers, confrontés au même problème a réellement conduit à une réflexion transverse riche et fructueuse. C'est cette richesse qui est à l'origine du succès international de ce modèle.

Chapitre 2. Premiers éléments

Définir quelques termes essentiels afin d'éviter toute ambiguïté, identifier les acteurs et bien délimiter l'Archive numérique dans son environnement sont indispensables avant d'aller plus loin dans l'analyse. C'est le rôle de ce chapitre.

2.1. Vocabulaire

Nous avons, dans la section 3 de ce cours, proposé les définitions des termes « information » et « donnée ».

Il s'agissait de définitions proposées par le modèle OAIS.

Ajoutons deux définitions nouvelles :



Définition : Archive

Organisation chargée de conserver l'information pour permettre à une communauté d'utilisateurs cible d'y accéder et de l'utiliser.

Ce terme, équivalent à « service d'archives » ou « service d'archivage », couvre l'ensemble des activités allant de la prise en charge pour archivage jusqu'à la conservation et l'accès aux documents archivés.

L'Archive³ assure donc la pérennité des données, l'accès pérenne aux données, mais aussi la préservation, avec les données, de toutes les informations nécessaires à leur compréhension et utilisation.

Le modèle de référence s'applique à toutes les disciplines et institutions qui conservent des documents sous forme numérique. Les termes définis dans le modèle peuvent ne pas correspondre au vocabulaire d'une discipline particulière (par exemple : les archives traditionnelles, les bibliothèques numériques, les centres de données scientifiques).

Ce vocabulaire défini dans le modèle OAIS **n'a pas vocation à remplacer les terminologies existantes**. Il sera nécessaire, pour les disciplines et les institutions concernées, de trouver la correspondance entre certains de leurs termes propres les plus courants et les termes du modèle de référence OAIS. Par exemple : le terme souvent employé dans la littérature archivistique de « records » n'est pas utilisé dans le modèle de référence OAIS, mais peut être rattaché globalement à « Contenu d'information » à l'intérieur d'un « Paquet d'informations archivé ». Nous verrons ces concepts un peu plus loin.



Définition : Long terme (spécifique à ce module)

Période suffisamment longue pour qu'il soit nécessaire de prendre en compte les changements technologiques et, notamment, la gestion des nouveaux supports et formats de données ainsi que l'évolution de la communauté d'utilisateurs. Cette période n'est pas limitée dans le temps.

Cette définition fait apparaître une question nouvelle :

La distinction entre archives courantes³, archives intermédiaires³ et archives définitives³ conférerait un statut différent entre ces différentes catégories et permettrait éventuellement une gestion différenciée des actions.

Avec le numérique pour lequel le long terme peut, dans certains cas, être plus court que la durée de vie des archives courantes (par exemple : c'est le cas de dossiers de personnels qui sont ouverts tant que l'agent n'a pas quitté ses fonctions), il apparaît qu'un ensemble de dispositions communes doit impérativement être mis en œuvre dès lors qu'il s'agit de conserver des documents sur une durée supérieure à 10 ans, voire moins.

2.2. L'Archive dans son environnement

Cette première approche est simple. Elle permet d'identifier les acteurs principaux du problème et vise à délimiter clairement ce qui est de la responsabilité de l'Archive et ce qui ne l'est pas :

- Le producteur crée des objets numériques qui sont transférés à l'Archive. Le producteur tel qu'il est défini ici n'est pas nécessairement le créateur de l'information. Dans le domaine des Archives institutionnelles, on parle de « services versants » et une distinction est faite entre le service créateur – et donc propriétaire – de l'information et le service versant. Le producteur fournit les informations (sous forme de données numériques) à l'Archive dans le cadre d'un accord intitulé « Protocole de versement » qui définit les modalités de dépôt, le modèle de données à utiliser et les différents composants logiques à fournir.

- L'utilisateur recherche, sélectionne, commande et récupère les objets numériques qui l'intéressent.

- La direction/l'entité responsable (le « management ») représente les décideurs qui définissent la charte ou le mandat, le domaine de compétence, les priorités et les orientations générales de l'Archive, en cohérence avec la politique globale de l'institution ou de l'entreprise et en conformité avec les réglementations en vigueur. La direction

(humaines et matérielles), arbitre les conflits entre les acteurs et évalue les performances de l'Archive sous la forme d'audits.

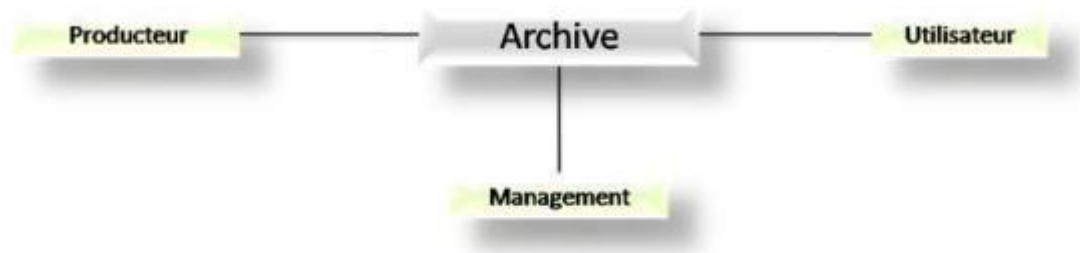


Image 2 Environnement de l'Archive

Dans cette approche, le modèle OAIS positionne l'Archive comme une entité dotée d'une certaine indépendance et de responsabilités propres bien identifiées. C'est un point de vue qui différencie ce modèle de la norme ISO 15489 consacrée au « **Records Management** ». En effet, dans la norme ISO 15489, on analyse le cycle de vie du document (numérique ou non-numérique) depuis le moment où il est fixé, stabilisé jusqu'à son archivage. La fonction d'archivage n'étant alors qu'une fonction parmi d'autres au sein du système d'information de l'entreprise.



Définition : Management (au sens de l'OAIS)

Décideurs de la politique globale de l'OAIS en cohérence avec la politique générale de l'organisme ou de l'entreprise.

- Il faut bien comprendre, dans la logique propre au modèle OAIS :
 - que c'est l'entité fonctionnelle "Administration" qui assure le pilotage du service au quotidien et qui prend toutes les décisions opérationnelles,
 - que c'est le "management" qui définit le mandat, la politique d'ensemble et qui, dans la plupart du temps, fournit également les ressources.

Chapitre 3. Le modèle d'information

Ce modèle d'information constitue l'une des grandes originalités de la norme OAIS. Il nous fait réellement percevoir toutes les conséquences et toute la spécificité qui sont induites par la nature même du document sous forme numérique.

3.1. Éléments de modélisation UML

Le modèle d'information de l'OAIS s'appuie sur le formalisme UML (*Unified Modeling Language*) de représentation **des objets** du domaine et plus spécifiquement sur ce qu'on appelle les diagrammes de classe.

Une classe décrit le comportement et le type d'un ensemble d'objets partageant des propriétés. Une classe est un concept abstrait représentant des objets concrets. Ces objets concrets sont appelés des instances de la classe. Pour désigner le concept d'automobile, on parlera donc de la « classe automobile » ou encore de l'« objet automobile ». L'« objet automobile blanche » constituera une sous-classe de la précédente et héritera de ses propriétés. L'automobile blanche qui est réellement dans votre garage sera une instance de cette sous-classe.

Nous présentons ci-après, les principaux éléments de ce formalisme que nous utiliserons :



Image 3 L'objet est représenté par un simple rectangle

On distinguera 3 catégories de relation :

- les associations matérialisées comme ci-dessus par un trait simple entre deux objets. On pourra, sous forme de texte, expliciter la nature de cette association,
- la spécialisation présentée ci-après,

- la composition présentée ci-après.

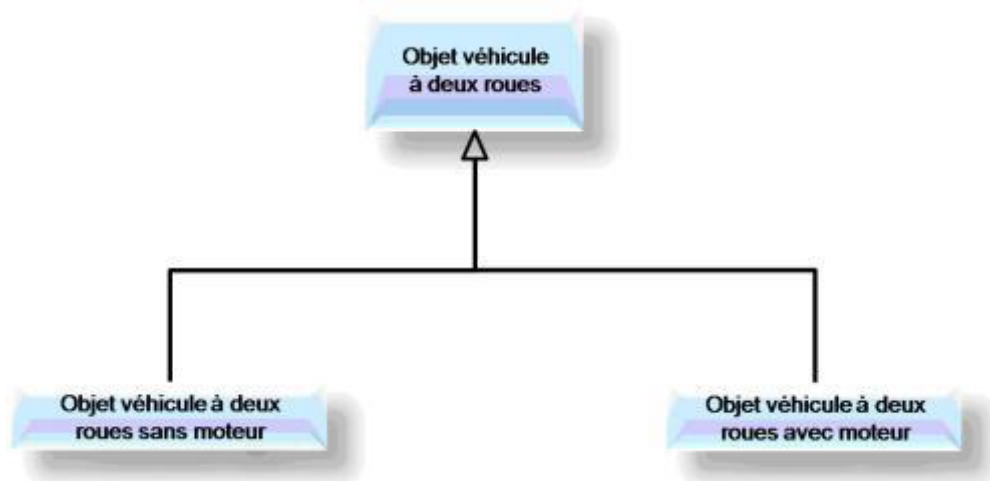


Image 4 Relation de spécialisation

Un objet (une classe) peut être spécialisé en plusieurs objets différents (sous-classes) qui possèdent toutes les propriétés de l'objet « père » auxquelles s'ajoutent des propriétés supplémentaires.

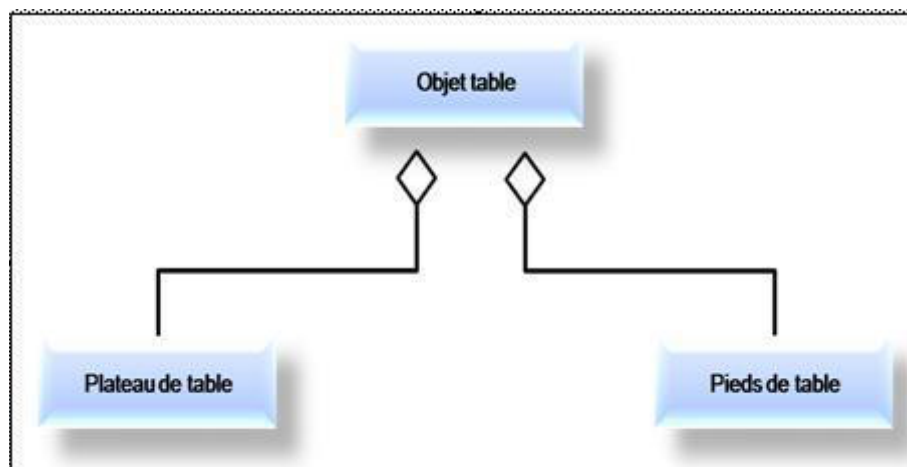


Image 5 La relation de composition

Légende :

Une classe peut être composée de plusieurs sous-classes

Enfin, le dernier élément à connaître concerne ce qu'on appelle la cardinalité :

Légende :

Ces règles de cardinalité permettent de définir le caractère optionnel ou obligatoire d'une relation, ainsi que son caractère unique ou multiple

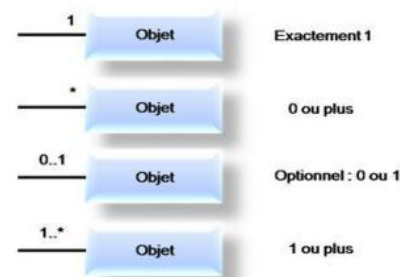


Image 6 Cardinalité

3.2. L'Information de représentation

Nous abordons ici le concept fondamental d'Information de représentation, concept que chacun peut pressentir mais qui est ici formalisé clairement.

Nous allons commencer par analyser un exemple.



Exemple

Supposons que nous ayons à conserver l'information contenue dans un objet numérique tout simple, un document numérique qui se présente sous la forme d'un unique fichier. Cette information qui constitue l'objectif essentiel de la pérennisation sera représentée sous la forme d'un objet appelé « **Contenu d'information** ».

Notre document se présente sous la forme d'une séquence de bits

```
00101110001100000010000001001101011000010110001101101
00101101110011101000110111101110011011010000000000000
11001000110000001100000011010100111010001100010011000
10011101000110010001100100010000000110001001101100011
10100011001100110010001110100011010000110100000000001
1111111101100011111111111000000000000000000000000000000
0010001100100100101000110000000000000000000000000000000
0000100000000010010000000000000000000000000000000000000
011111111111111110000000000000001100010000010111000001110
00001101100011001010100110101100001011100100110101100
00101011111111110110110000000010000100000000000000000101
1000010000000100000001010000010000000011100001011000
01010000010010000101011.....
```

Image 7 Si nous n'avons qu'une séquence de bits, nous n'avons pas grand-chose !

Nous pourrions penser qu'il suffit de connaître le format de notre fichier pour accéder à son contenu.

Légende :

En pratique, il existe aujourd'hui beaucoup plus de mille formats de fichier, chacun de ces formats pouvant exister en de multiples versions. Certains disparaissent, de nouveaux sont inventés.

3ds, aa, aac, abw, ace, ac3, ag, ai, aif, aiff, aiff, arc,
ark, art, atim, asf, asx, az, avi, blend, bmp, cbm, caf,
cad, cdf, cdg, cgm, cnt, cob, cpt, css, dat, csv, df,
DivX, diz, DjVu, doc, docm, docx, dot, dotm, docx,
dpa, ds, dts, dv, dvcam, dvcpro, dwg, dxf, eAAC+,
emf, eps, fits, flac, flash, flv, gif, hdi, hhtml, h264, ifc,
iges, iklap, jpg, lo, mhtml, mng, mkv, mpeg, mp2,
mp3, mp4, ppg, odp, ods, odt, ogg, ogm, Openexr,
pcrn, pdb, pdf, pdfa, pps, ppt, ppbx, prn, ps, psw, raw
rif, silk, sxw, stw, suw, svg, lgi, theora, flit, txt, uol,
vorbis, vqf, vrml, wav, wma, wrmf, wv, xcf, xhtml,
xls, xlsb, xlsx, xlt, xml, xps, XviD, x3d, zip.....etc etc des
centaines d'autres

Image 8 Mais combien de formats de fichiers ?

Dans notre exemple, nous pouvons ajouter que notre fichier contient des caractères codés conformément à la norme ISO 646 (identique à ce qu'on appelle aussi le codage ASCII). Cette norme définit le codage des lettres de l'alphabet non accentuées, des chiffres et de quelques caractères spéciaux. Pour que nous puissions dans l'avenir lire ce texte, il convient donc d'avoir la certitude que cette norme de codage soit préservée quelque part, ou alors de prévoir d'en assurer la conservation dans l'Archive.

Il est possible d'examiner le contenu de notre fichier avec un simple éditeur de texte qui transforme chaque octet du fichier en une représentation graphique correspondant au caractère codé dans cet octet.

Légende :

Nous pouvons visualiser le début de notre fichier avec le bloc-notes, mais nous sommes loin d'en percevoir le sens

120	58	1981-09-30	174.22	01225	3898.70	2352.45	-1.66
-80.00	-80.00	1.44	9	1010.76	16.17	55.84	58.61
3.98	120	58	1981-09-30	174.23	08152	4460.36	2347.36
3.35	-80.00	-80.00	1.44	1047.25	16.59	55.92	58.18
0.59	5.06	120	58	1981-09-30	174.24	20152	4338.25
-6.68	-9.09	-80.00	-80.00	1.33	9	1095.40	17.02
7.9	664.61	6.54	120	58	1981-09-30	174.25	273462
0.11	-80.00	-80.00	-14.88	-17.15	2	922	1135.18
5	-16.44	624.93	8.67	120	58	1981-09-30	174.26
33	-0.67	-12.14	48.75	120	58	1981-09-30	174.27
75.80	1.44	9	1010.76	16.17	55.84	58.61	1981-09-30

Image 9 Premier examen de notre fichier avec le bloc-notes

Ajoutons alors une information nouvelle consistant à préciser que notre fichier est constitué d'une répétition de séquences de longueur fixe égale à 187 octets. Chaque séquence se termine par le caractère « retour chariot » qui signifie qu'il faut passer à la ligne suivante. Toutes ces séquences sont structurées de manière identique et contiennent une succession de champs d'information d'une longueur définie, à savoir :

- 3 nombres entiers codés chacun sur 6 octets
- une séquence de 24 octets avec un codage interne spécifique,

- 3 nombres décimaux d'une longueur de 10 octets chacun,
- etc.

Muni de ces informations, nous pouvons regarder notre fichier avec un éditeur comme textedit qui prend en compte les sauts de ligne. En effet, sur les fichiers de texte, l'information "saut de ligne" présente une difficulté en ce sens qu'elle n'est pas codée de la même façon par tous les systèmes d'exploitation :

- sous Windows, le saut de ligne est codé sur deux caractères (le caractère "Carriage return" + le caractère "line feed",
- sous Linux, un seul de ces deux caractères suffit à déclencher un saut de ligne,
- si nous lisons, avec le bloc note de Windows, un fichier texte créé par un autre système, les sauts de ligne ne seront pas interprétés,
- par contre, ils le seront si nous utilisons un éditeur de texte compatible avec les deux systèmes.

Légende :

Cette représentation nous paraît déjà plus claire mais il nous manque encore des informations capitales, à savoir la signification de chacun des champs

```
120 58 5 1981-09-30T14:22:01.629Z 3898.70 2352.45 -1.66 -7.72 -18.61
120 58 5 1981-09-30T14:22:05.786Z 3894.64 2546.88 -1.49 -9.76 -18.96
120 58 5 1981-09-30T14:22:10.027Z 3925.05 2429.15 -1.49 -8.78 -18.27
.....
```

Image 10 Visualisation du fichier prenant en compte sa structure répétitive

Il est donc nécessaire de définir, un par un et de façon exhaustive, la signification de chacun des champs continus dans la structure répétitive.

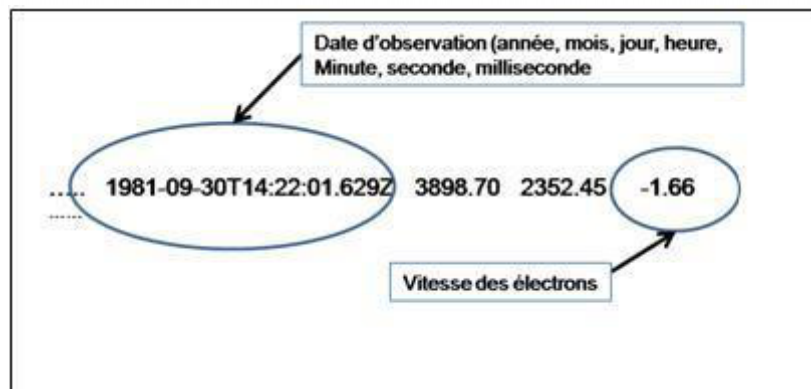


Image 11 Définir la signification de chacun des champs d'information

Nous voyons apparaître, au travers de cet exemple, une **première équation fondamentale** :

1. si nous voulons pérenniser l'information contenue dans un objet numérique, information qui sera modélisée par un objet « **Contenu d'information** », il n'est pas suffisant de conserver cet objet ;
2. il est indispensable de conserver avec cet objet, un ensemble d'informations, appelées « **Information de représentation** » qui nous permettront de passer des bits constituant l'objet numérique au contenu informationnel de cet objet.

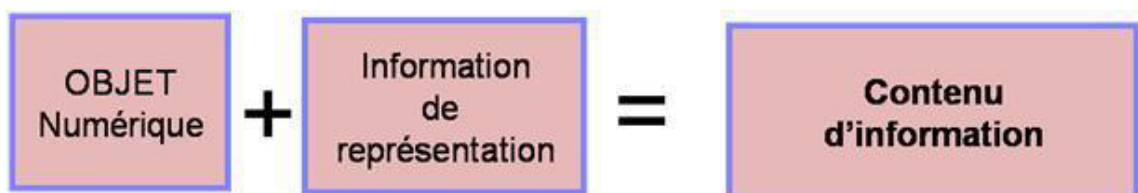


Image 12 Une équation fondamentale

Cette équation se traduit par le diagramme UML suivant :

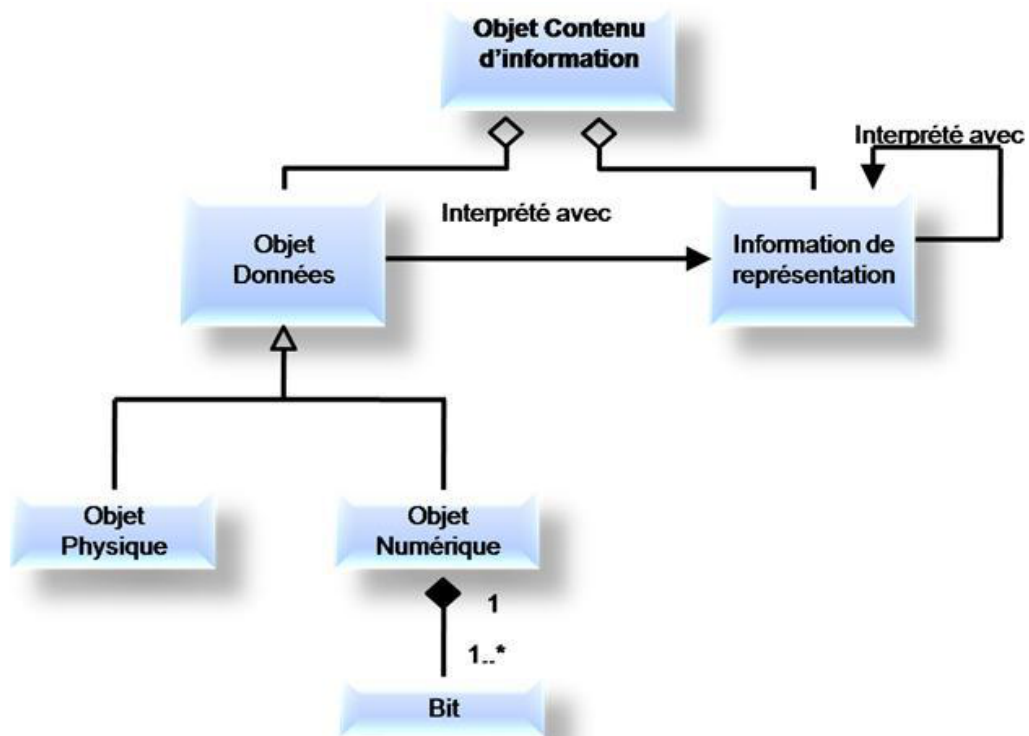


Image 13 L'objet Données, interprété avec l'Information de représentation, conduit au Contenu d'information

Ce diagramme appelle les commentaires suivants :

- L'Information de représentation étant basée sur des éléments existant également sous forme numérique (par exemple les tables de codage des caractères alphabétiques), elle est elle-même un objet numérique contenant de l'information à préserver, elle doit donc disposer de sa propre Information de représentation. C'est cela qu'exprime la flèche qui sort puis rentre dans le rectangle « Information de représentation ».
- L'objet Données peut aussi bien être numérique que physique (par exemple un ouvrage papier). Dans ce module de cours, nous nous intéresserons essentiellement aux documents numériques, mais **le modèle OAIS reste valide pour les documents non numériques**. C'est un point important car dans les années à venir, la plupart des services d'archives auront à gérer en même temps des documents non numériques et des documents numériques.

L'Information de représentation doit être en quelque sorte garante de l'intelligibilité de l'information que l'on veut conserver. Cette préoccupation n'est pas nouvelle. Lorsque l'archiviste constitue un dossier, il veille à ce que ce dossier soit cohérent et complet. Cette préoccupation prend cependant ici un caractère essentiel et systématique, lié aux multiples risques inhérents à l'information sous forme numérique mais aussi à l'instabilité grandissante du contexte dans lequel nous vivons.

L'information de représentation peut être concrétisée de la façon suivante dans le cas d'un document textuel rédigé par un service technique :

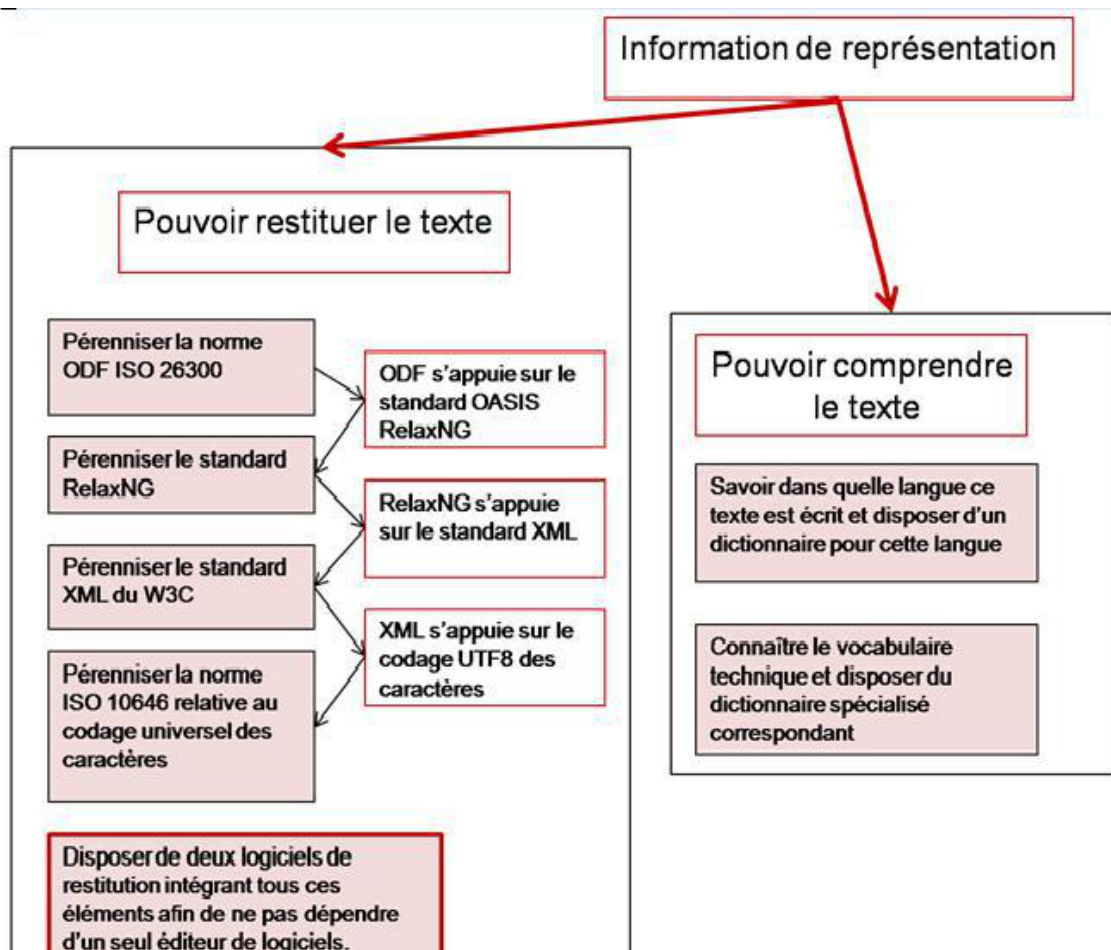


Image 14 Cas d'un document technique stocké dans un fichier bureautique au format ODF (Open Document Format)

3.3. Information de pérennisation

L'Information de représentation n'est pas suffisante pour que l'Archive assure l'ensemble de ses responsabilités. D'autres informations sont nécessaires, c'est ce qu'on appelle l'Information de pérennisation qui comprend :

L'Information de contexte qui décrit

- les liens entre le Contenu d'information archivé et son environnement,
- les raisons de la création de ce Contenu d'information,
- son rapport avec d'autres objets Contenu d'information. C'est l'information de contexte qui nous permettra de rattacher un objet à un ou à plusieurs ensembles (un fonds, un sous-fonds, un dossier...).

L'Information de provenance qui indique l'origine ou la source du Contenu d'information, qui trace tous les changements intervenus depuis sa création et qui identifie les intervenants de ces changements. Les changements majeurs peuvent être liés à une modification du format des données, modification qui sera dictée pour des raisons d'obsolescence de formats anciens, pour réduire des coûts d'utilisation de formats peu utilisés, ou encore pour répondre aux besoins des utilisateurs dont les besoins imposent un nouveau format.

L'Information d'identification permettant d'affecter un identifiant unique à chaque objet numérique. Cette identification qui existe pour les documents non numériques (ISBN...) prend cependant une importance toute particulière avec le numérique : les identifiants liés à une infrastructure d'ordinateur ou à l'architecture d'un système logiciel – comme c'est souvent le cas pour les adresses des ressources sur Internet – n'ont qu'une faible durée de vie en raison des changements réguliers de ces infrastructures et de ces architectures. L'exigence de disposer d'identifiants qui soient à la fois pérennes et compatibles (sans recouvrement) avec les identifiants utilisés par les institutions partenaires est renforcée par le développement des services « interopérables » permettant de rechercher des documents d'archives simultanément auprès de plusieurs institutions.

L'Information d'intégrité qui consigne les mécanismes de garantie d'intégrité des objets numériques et mémorise la ou les empreintes d'intégrité permettant d'assurer qu'aucun objet numérique archivé n'a subi de modification sans que celle-ci ait été tracée.

A ces catégories d'information, la nouvelle version du modèle OAIS ajoute **L'Information de droits d'accès** qui définit les régimes de droit attachés aux données (droits d'auteur,...), précise les restrictions et les contrôles d'accès aux objets archivés.

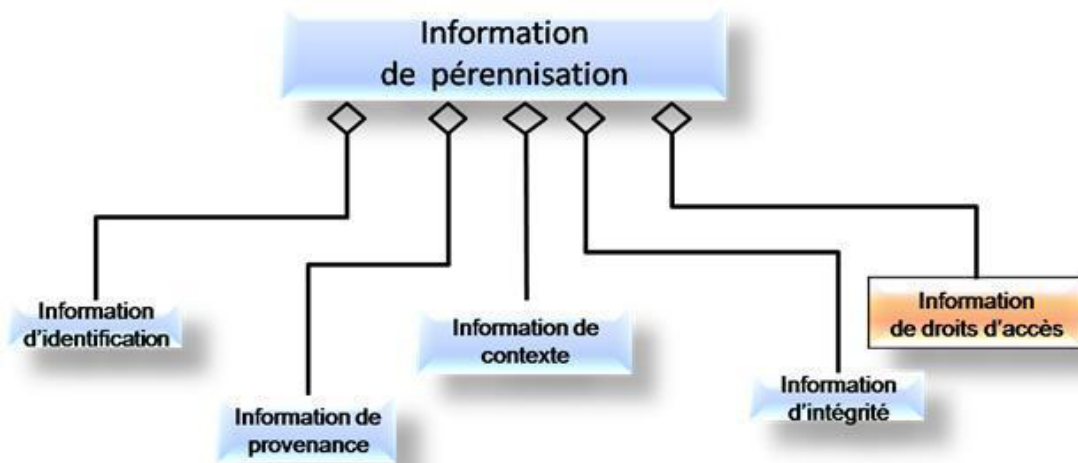


Image 15 L'Information de pérennisation et ses 5 sous-classes

L'Information de droits d'accès est marquée d'une couleur différente car elle fait partie de la nouvelle version du modèle, en cours de normalisation.

3.4. Paquets d'information

Un Paquet d'informations est un conteneur conceptuel au sein duquel le Contenu d'information et l'Information de pérennisation associée à ce contenu sont rassemblés. Le modèle définit trois types de paquets :

- le Paquet d'informations archivé (Archival Information Package, AIP) rassemble un Objet Données (un objet numérique) et l'ensemble des informations nécessaires à la pérennisation des informations contenues dans cet Objet Données, à savoir l'Information de représentation et les cinq composantes de l'Information de Pérennisation (contexte, provenance, identification, intégrité, droits d'accès). L'AIP apparaît ici comme l'objet fondamental manipulé par l'Archive.
- le Paquet d'informations à verser (Submission Information Package, SIP) est celui qui est fourni à l'Archive par le Producteur. Il peut parfois s'avérer nécessaire d'utiliser plusieurs SIP pour obtenir l'ensemble Contenu d'information et Information de Pérennisation, c'est-à-dire un AIP.
- le Paquet d'informations diffusé (Dissemination Information Package, DIP) est fourni par l'Archive aux consommateurs de données.

Cette notion de Paquet permet d'abord de donner une réalité conceptuelle à ce que l'Archive doit effectivement conserver. L'Archive devra en effet s'assurer que, quelle que soit la solution technique retenue, elle dispose effectivement des différentes catégories d'informations numériques constituant les AIP.

L'organisation et les formats retenus pour les Objets Contenu d'information d'un SIP peuvent être différents de ceux retenus pour les AIP. De même, il n'y a aucune raison qu'ils correspondent aux formats et à l'organisation que demandent des utilisateurs. Les Objets Contenu d'information sont donc amenés à être transformés pour répondre à différents besoins. Un DIP pourra donc être créé à partir de plusieurs AIP ou à partir d'une partie seulement d'un AIP.

Dans les mises en œuvre, les paquets peuvent avoir une réalité physique (mais le modèle OAIS ne l'exige pas car il se situe au plan conceptuel). Certaines institutions ont fait le choix de regrouper, au sein d'une même structure physique normalisée, l'ensemble des éléments des SIP, des AIP et parfois des DIP. Nous parlerons, dans la section 9 sur les métadonnées, des normes permettant la création et la gestion de ces Paquets d'information.

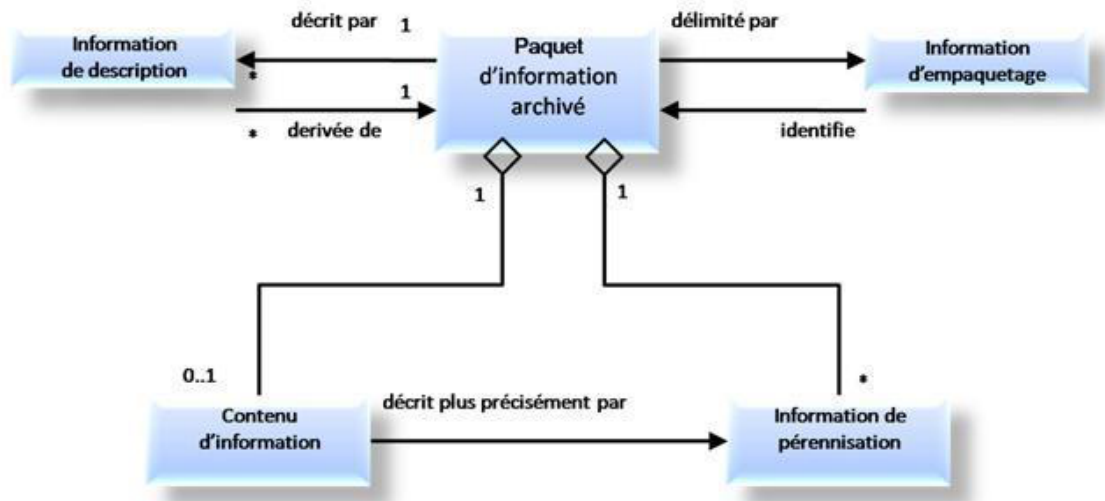


Image 16 La représentation modélisée du Paquet d'informations archivé (AIP)

Ce diagramme fait apparaître des objets déjà connus comme le Contenu d'information et l'Information de pérennisation, mais il fait apparaître également de nouvelles catégories d'information.

De quoi s'agit-il ?

- La description des liens réels ou logiques des différents composants d'un Paquet d'informations archivé constitue ce qu'on appelle **l'Information d'emballage**. En pratique, il s'agit de savoir comment nous avons organisé les différentes informations constituant un Paquet et nous devons conserver cette information aussi longtemps qu'elle sera utile : par exemple, la répartition des objets Contenu de données, Information de représentation, Information de pérennisation sur un ensemble de fichiers et de répertoires.
- Toute l'analyse qui précède est fondamentalement dictée par la préoccupation de pérennisation. Si l'on veut que les utilisateurs puissent rechercher, retrouver, évaluer, sélectionner, récupérer, analyser les objets Contenu d'information répondant à leurs besoins, il convient de mettre à disposition de ces utilisateurs une information dite **Information de description** qui permettra de réaliser les opérations en question. Cette Information de description n'est pas nouvelle, elle est extraite des informations précédemment décrites, elle en constitue le sous-ensemble utile aux processus d'interaction avec les utilisateurs. L'Information de description peut être reconstruite, régénérée, recalculée au fil du temps pour répondre à des besoins nouveaux des utilisateurs. Elle correspond par exemple à ce qui sera appelé « métadonnées descriptives ». Plus largement, toute information extraite de l'AIP et qui jouera un rôle dans le processus de recherche d'informations sera une information de description.

Chapitre 4. Le modèle fonctionnel

Le modèle fonctionnel de l'OAIS est un peu plus facile à comprendre que le modèle d'information car il est plus proche des tâches concrètes à mettre en œuvre.

Ce modèle est composé de six entités fonctionnelles.

- L'entité « Entrées » insère dans l'OAIS les objets numériques provenant du producteur
- L'entité « Accès » transmet les objets numériques aux utilisateurs
- L'entité « Gestion de données » organise la description du contenu des fonds de l'OAIS
- L'entité « Stockage » préserve les AIP sur le long terme
- L'entité « administration » assure le fonctionnement et la coordination d'ensemble
- L'entité « planification de la pérennisation » assure le travail de veille et de suivi afin de permettre à l'Archive de conserver les AIP dans un contexte en constant changement.

Le modèle fonctionnel identifie également :

- les interfaces entre les entités fonctionnelles
- et les interfaces avec les Producteurs et les Utilisateurs.

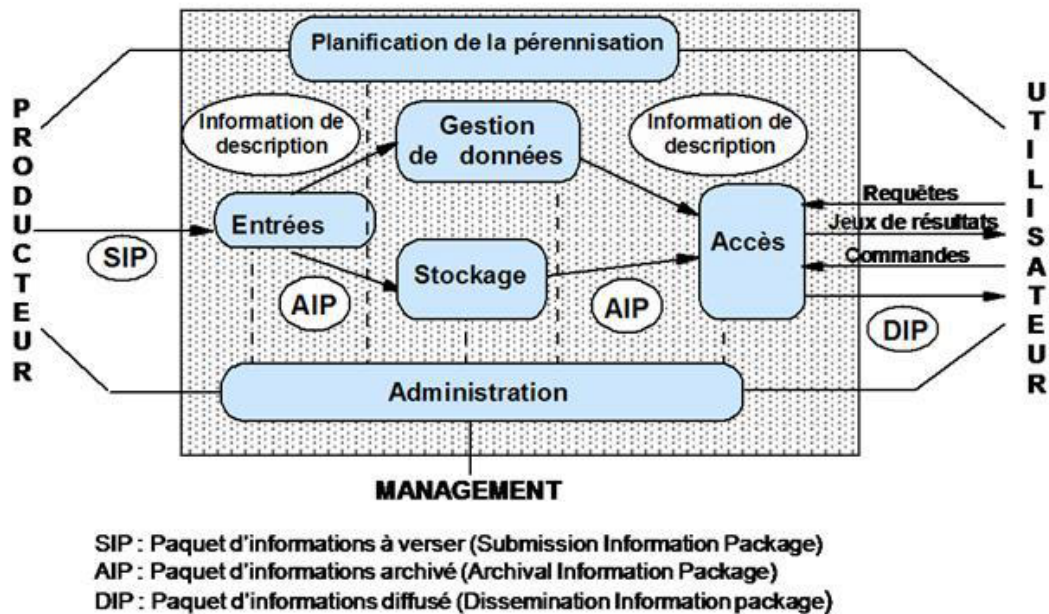


Image 17 Schéma d'ensemble du modèle fonctionnel

Le modèle tente de couvrir toutes les activités essentielles d'une Archive. Pour autant, le modèle ne spécifie ni une conception, ni une mise en œuvre particulière. Les implémentations réelles peuvent regrouper ou décomposer différemment les fonctions présentées. Toute implémentation, pour être conforme à la norme, devra être compatible avec le modèle fonctionnel et avec le modèle d'information.

C'est ce que nous allons préciser.

4.1. Entité « Entrées »

Cette entité récupère les objets numériques transférés par le producteur. Elle gère le mécanisme de dépôts (versement) des SIP, les contrôles d'accès associés et l'ensemble des interactions entre le producteur et l'Archive intervenant au cours du processus de versement.

Elle vérifie la conformité des SIP reçus par rapport à ce qui a été convenu entre l'Archive et le producteur, sous la forme d'un protocole de versement.

Elle assure si nécessaire un contrôle de l'intégrité des objets transférés.

Lorsque les SIP sont jugés valides, l'entité « entrées » va créer les Paquets d'information archivés (AIP) qui constitueront l'information de référence à conserver.

Au cours de cette tâche, elle sera amenée à compléter les métadonnées du Producteur par des métadonnées propres à l'Archive.

Les AIP seront confiés à l'entité « stockage ». L'entité « entrées » va également extraire, à partir des AIP, les **Informations de description** qui seront utilisées pour la recherche et la communication.

4.2. Entité « Stockage »

Cette entité est schématiquement chargée de conserver les bits.

- elle prend en charge les AIP fournis par l'entité « entrées »,
- et elle restitue ces AIP sur demande de l'entité « accès ».

L'entité « stockage » gère la hiérarchie de stockage (stockage en ligne, stockage avec accès différé...) pour répondre aux exigences de qualité de service des entités utilisatrices.

Elle assure les migrations des supports chaque fois que cela est nécessaire.

Elle procède aux contrôles réguliers et permanents de l'intégrité des informations qui lui ont été confiées.

Elle duplique les objets numériques et stocke la copie dans une installation physiquement séparée pour assurer leur restitution en cas d'incident ou de sinistre

4.3. Entité « Gestion de données »

Elle assure le maintien à jour et l'administration de la base de données qui contient les Informations de description. Cette base de données contient donc une description complète du fonds de l'Archive.

Cette entité reçoit des interrogations des utilisateurs via l'entité « accès » et renvoie les réponses. C'est donc à ce niveau que peuvent être mises en jeu les différentes techniques de recherche d'information : recherche par mots clés ou combinaisons logiques de mots clés, recherche en texte intégral, indexation.

Elle génère également des rapports sur l'état des fonds archivés et sur l'évolution et l'utilisation de ces fonds.

4.4. Entité « Accès »

Elle met en œuvre l'interface avec les utilisateurs.

Elle doit donc contrôler l'accès des utilisateurs au système d'archivage ainsi que les droits de ces utilisateurs à accéder ou non à tel ou tel objet numérique archivé.

Elle gère ensuite les différentes interactions entre le système et l'utilisateur : interrogations et réponses à ces interrogations, commandes d'objets archivés, génération des Paquets d'information diffusés (DIP) à partir des Paquets d'information archivé (AIP) reçus de l'entité « stockage » et transfert vers l'utilisateur de ces DIP.



Complément : Les services à valeur ajoutée

On peut remarquer ici que le DIP n'est pas nécessairement identique à l'AIP. L'Archive peut en effet offrir différents services répondant aux besoins des utilisateurs : fournir les documents d'archives dans un format souhaité par l'utilisateur et pas nécessairement dans le format dans lequel le document est archivé, extraire un sous-ensemble d'un objet correspondant à la demande de l'utilisateur. Cela est particulièrement intéressant dans le cas de données géographiques pour lesquelles on ne transmettra que la zone demandée par l'utilisateur.

4.5. Entité « Administration »

L'entité « administration » a une fonction globale. Elle assure le fonctionnement et la coordination d'ensemble, elle prend toutes les décisions relevant du fonctionnement de l'Archive et mène dans ce cadre une série d'activités parmi lesquelles on peut citer :

- le choix des règles et normes applicables en matière de versement (formats et documentation à fournir), de sécurité (contrôle d'accès), de migration (définition de retraitement des supports), etc.,
- les décisions de mise en œuvre des différentes migrations,
- la gestion du contrôle d'accès physique à l'Archive,
- l'administration du système informatique.

L'entité « administration » est en relation avec les acteurs externes :

- avec les Producteurs, elle négocie les protocoles de versement et veille au respect de ce protocole,
- avec le Management à qui elle rend compte des activités de l'Archive et dont elle reçoit éventuellement des directives,
- avec les utilisateurs pour régler les difficultés éventuelles d'accès au fonds.

4.6. Entité « Planification de la pérennisation »

Cette entité est particulière en ce qu'elle replace l'Archive dans le contexte de l'obsolescence des technologies et de l'évolution des besoins.

En effet elle a pour mission d'analyser les changements en cours et à venir et leurs impacts sur la capacité de l'Archive à conserver les informations archivées sous une forme utilisable et compréhensible :

- le suivi des ruptures technologiques implique un travail de veille, d'examen et de prototypage des technologies émergentes afin d'élaborer et de proposer des plans d'évolution du système informatique de l'Archive et de migration des données,
- l'analyse des évolutions de la communauté des utilisateurs peut conduire à proposer de faire évoluer les services offerts par l'Archive dans le sens voulu par cette communauté, par exemple la mise en place de nouvelles interfaces graphiques ; par exemple, à l'heure d'Internet, tout service qui en serait resté à un accès à l'information en mode ligne, sans interface graphique, serait voué à disparaître.

En dehors du modèle d'information et du modèle fonctionnel, l'OAIS aborde d'autres sujets : il définit une typologie des migrations et des modes de coopération entre les Archives numériques qui seront toutes deux développées dans la section 8 sur les stratégies de pérennisation

Chapitre 5. Impact du modèle OAIS

Un modèle mondialement utilisé

70 000 références sur Google en septembre 2009 (recherche stricte sur la chaîne de caractère « OAIS Reference Model »).

Des références en anglais, français, espagnol, grec, italien, russe, coréen, chinois, japonais, etc.

Le modèle est largement utilisé comme point de départ pour les activités de pérennisation :

- par les bibliothèques numériques (BnF...),
- par les archives institutionnelles (NARA, DAF...),
- par les centres de données scientifiques (NASA aux Etats-Unis, CNES en France, ESA pour l'Europe),
- par les entreprises industrielles (industries aéronautiques par exemple).



Complément : Une base conceptuelle pour définir des normes de mise en pratique

Le PAIMAS (Producer Archive Interface Methodology Abstract Standard) standardisé par le CCSDS en 2004 et norme

Le standard XFDU (*XML Formatted data Unit, structure and construction rules*), standardisé par le CCSDS en septembre 2008. Il traite de l'emballage de données et métadonnées pour les échanges et pour l'archivage.

Des normes en cours de rédaction :

- « *Producer Archive Interface Specification* » : proposition de mécanismes formels permettant la définition, le transfert et la validation des SIP,
- « *Metrics for Digital Repository Audit and Certification* » et « *Requirements for bodies providing audit and certification of digital preservation management systems* » : Audit et certification des archives numériques,



Complément : Des implémentations de plus en plus significatives

Le modèle OAIS ne spécifie pas une implémentation mais nombre d'implémentations sont basées sur les concepts de l'OAIS. Le projet LOTAR des industries aéronautiques européennes l'a pris en compte au plus haut niveau, les plateformes PIL@E (Plate-forme pilote d'archivage électronique) de la Direction des Archives de France ou PAC (Plate-forme d'archivage du CINES, Centre informatique national de l'enseignement supérieur) ont fait de même. Dans le cas du système SPAR (Système de préservation et d'archivage réparti) de la Bibliothèque nationale de France, le modèle OAIS constitue la référence fondamentale applicable.

Le Modèle OAIS est une base incontournable pour conduire la réflexion, l'analyse des besoins et des contraintes propres à chaque entreprise ou chaque institution en vue de la mise en œuvre d'une Archive numérique.

Naturellement, le caractère universel du modèle et son abstraction lui confèrent des limites. Sa mise en application concrète dans un domaine particulier implique que soient prises en compte les spécificités des objets numériques de ce domaine, les pratiques, la terminologie, les réglementations spécifiques. Aucune organisation visant à la création d'une Archive numérique ne pourra se dispenser de cette analyse.

Chapitre 6. Interface Producteur-Archive

Dans un processus d'archivage numérique, la phase de versement est l'une des plus critiques.

Lorsque l'Archive a reçu des documents numériques conformes à ce qu'elle attend et convenablement décrits, une part importante des obstacles a été franchie.

C'est au moment du transfert des objets numériques vers l'Archive que cette dernière prend véritablement la responsabilité de la pérennisation des informations contenues dans ces objets.

C'est aussi à ce moment que les difficultés peuvent s'accumuler :

- les objets numériques reçus ne sont pas conformes à ce que l'Archive attendait ;
- ce que le producteur devait verser n'a pas été défini avec suffisamment de précision ;
- découverte tardive d'anomalies dans les objets numériques versés (difficulté particulièrement difficile à résoudre).

Ces difficultés nuisent à la qualité de l'information archivée et augmentent le coût de l'opération.

Un certain nombre de normes interviennent dans la phase de versement. Les normes d'emballage sont décrites plus en détail dans la section 9 « métadonnées ».

Nous aborderons ici la **norme PAIMAS (*Producer archive interface methodology abstract standard, ISO 20652*)**, qui est **dérivée directement du modèle OAIS** dont on utilise le vocabulaire et les concepts.

La norme PAIMAS vise à définir pour l'Archive et le Producteur, une méthodologie de travail commune et structurée en quatre phases.

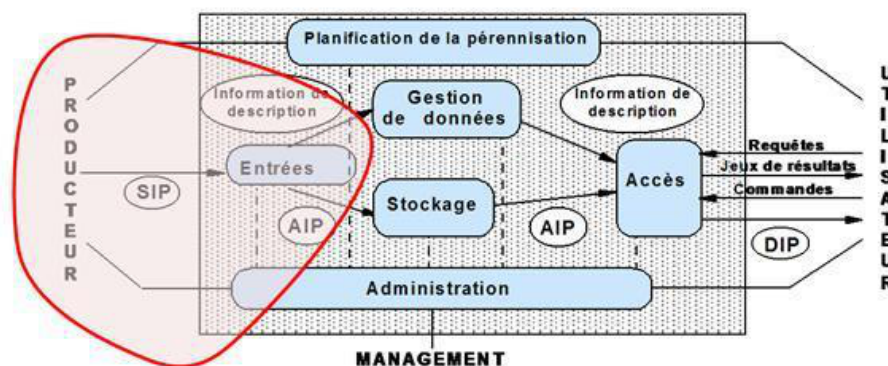


Image 18 Le domaine méthodologique couvert par le PAIMAS

Chaque phase est principalement caractérisée par une table des tâches à entreprendre ; chaque tâche est elle-même décomposée en une série d'actions à conduire par l'un ou l'autre ou par les deux partenaires du processus de versement. La norme PAIMAS se présente comme un guide générique qui pourra être spécialisé en fonction du cadre particulier de l'application et des spécificités des partenaires.

Les quatre phases du PAIMAS :

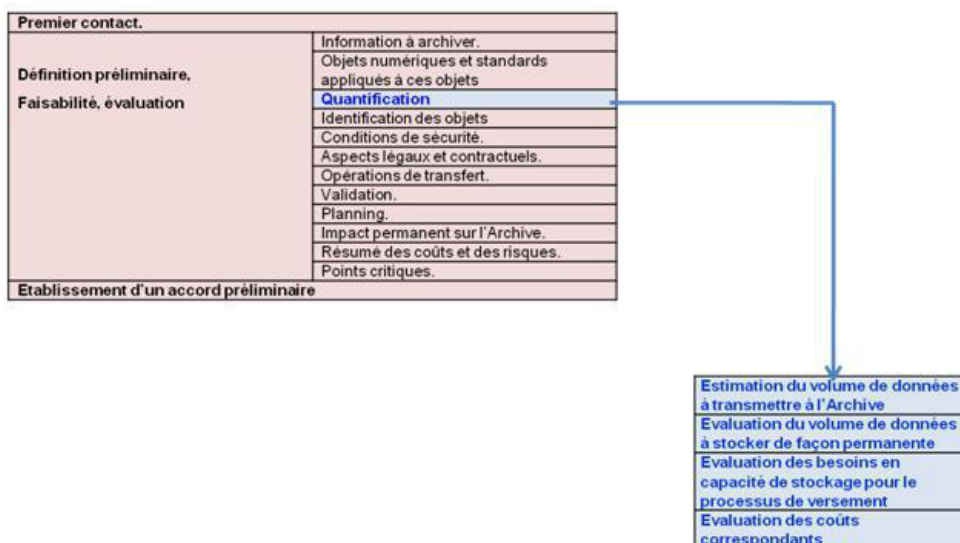


Image 19 PAIMAS, les tâches de la phase préliminaire et la décomposition de la tâche « quantification » en actions
Plus précisément, cette phase comporte une liste de tâches à conduire, et pour chaque tâche, une liste d'actions a été définie.

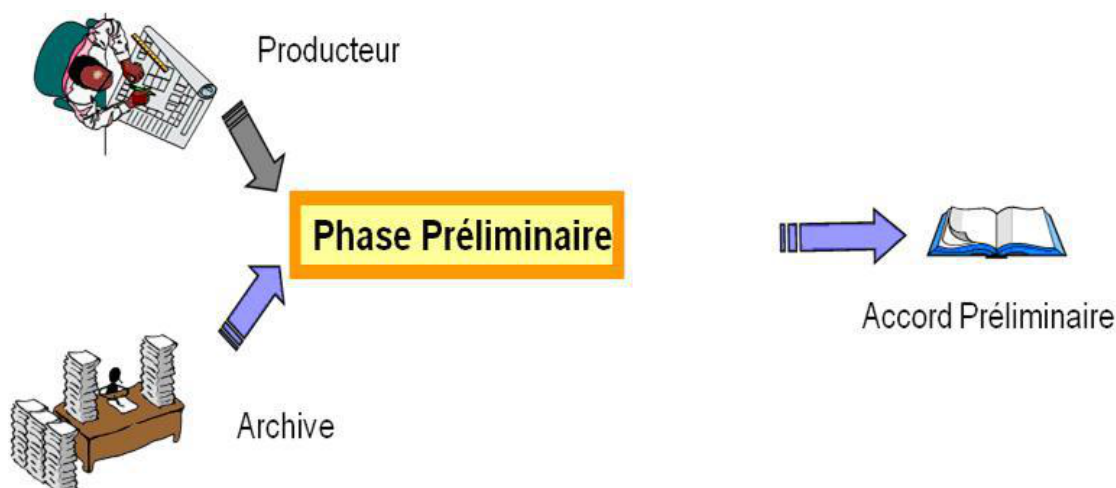


Image 20 PAIMAS, la phase préliminaire

Chaque action fait l'objet d'une description précise. Il est par ailleurs indiqué si l'action doit être conduite par l'Archive, par le Producteur ou par les deux.

Il en va de même pour les trois autres phases :

Légende :

La **phase de définition formelle** a pour objectif de négocier « le protocole de versement ». Ce protocole devra inclure une description complète, précise et formelle des objets numériques à livrer à l'Archive par le Producteur et un planning. Cette description doit faire l'objet d'un modèle formel, c'est-à-dire d'un modèle construit à l'aide d'un formalisme permettant à un individu ou une personne d'analyser et de comprendre sa cohérence et permettant également à la machine, à l'ordinateur, d'en faire un usage automatisé au cours du processus de transfert.. Ce formalisme peut lui-même s'appuyer sur un dictionnaire décrivant les différentes classes d'objets numériques qui seront transférées ainsi que leurs principales caractéristiques techniques et sémantiques.



Image 21 PAIMAS, la phase de définition formelle



Image 22 PAIMAS, la phase de transfert

Légende :

La **phase de transfert** correspond au transfert effectif des paquets d'information à verser. Elle inclut les tests du transfert puis sa mise en œuvre opérationnelle.

La

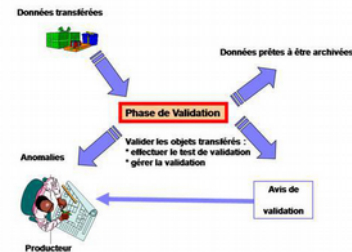


Image 23 PAIMAS, la phase de validation

phase de validation comprend la validation des paquets reçus par l'Archive et toutes les actions que cette validation peut impliquer. Différents types de validation des paquets sont envisagés : validation au fil de l'eau de chaque paquet, validation globale d'un ensemble de paquets constituant un tout logique, validation systématique, validation approfondie par échantillonnage.

Une vue synthétique de l'ensemble des phases peut être présentée ainsi :

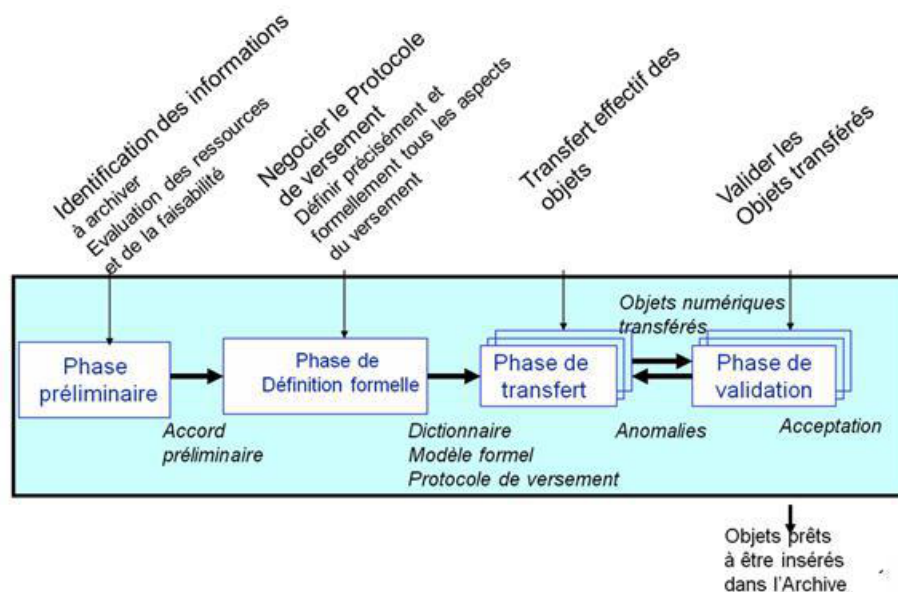


Image 24 PAIMAS, l'ensemble du processus de versement



Complément

Toutes tâches et toutes les actions définies dans la norme ne sont pas pertinentes dans tous les contextes, c'est pourquoi le PAIMAS a prévu un mécanisme permettant de spécialiser, d'adapter la norme pour un domaine ou une institution particulière. Dans une telle spécialisation, on pourra éliminer les tâches non pertinentes, ajouter d'autres tâches, prendre en compte les normes, les outils et le vocabulaire propre au domaine considéré.

Chapitre 7. Norme Afnor NF Z 42-013

Cette norme a pour titre : « 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 ». Elle a fait l'objet d'une importante mise à jour en mars 2009.

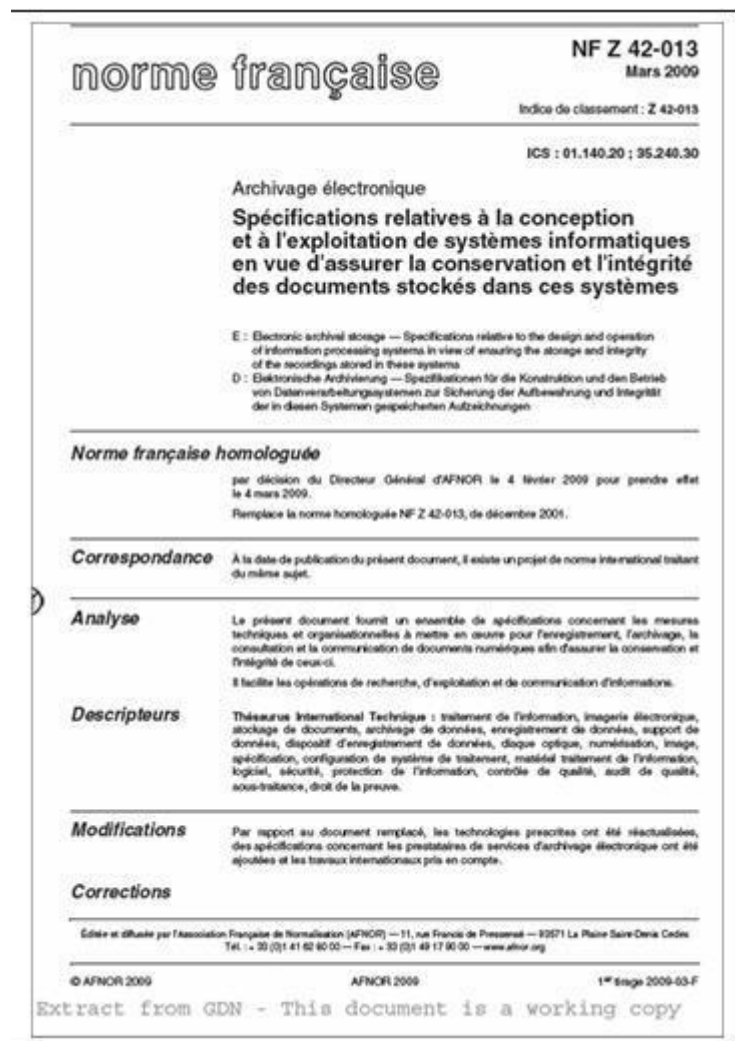


Image 25 La version publiée en mars 2009 de la norme Afnor NF Z 42-013

Elle se concentre sur les **caractéristiques du système informatique** sur lequel l'Archive numérique va appuyer ses activités. Elle se situe au niveau des spécifications techniques alors que le modèle OAIS se place au niveau conceptuel.

Elle inclut des spécifications relatives à la numérisation d'archives sur support papier avec toutes les exigences à respecter afin qu'on puisse être assuré que l'archive numérisée est bien conforme à l'original sur support papier.

Tous les aspects de l'archivage numérique sont abordés, et donc à ce titre, nous pouvons la ranger dans les normes généralistes. Certaines fonctions sont cependant à peine développées (c'est le cas notamment de la communication).

La norme met spécifiquement l'accent sur la question de l'intégrité des documents archivés et sur la capacité de l'Archive à apporter la preuve de cette intégrité.

Un point essentiel abordé et détaillé dans la 42-013 est celui de la nécessité de disposer d'une description détaillée de tous les constituants du système d'archivage et d'une traçabilité exhaustive de tous les changements au sein de l'Archive. Il doit être possible de démontrer la continuité de l'enregistrement de tous les événements tracés.

Chaque document archivé sera rattaché à un « profil d'archivage » défini comme un ensemble de règles applicables aux documents partageant les mêmes critères de confidentialité, de durée de conservation, de destruction et de droits d'accès.

Une gradation des niveaux d'exigence permet d'adapter les procédures au niveau du risque que l'Archive accepte de prendre. Un recours différencié à l'horodatage¹, au calcul d'empreintes, au cryptage et à la signature électronique sera utilisé en fonction des besoins.

Définition : Horodatage :

Fait de donner une heure certaine à une transaction, en se basant sur une source extérieure de temps incontestable et en signant les éléments de datation ainsi donnés (jetons d'horodatage).

On peut dire en conclusion que la nouvelle norme NF Z 42-013:2009, nourrie des retours d'expériences existants et des autres normes généralistes (OAIS, Records Management), apporte, au plan de la mise en œuvre, un grand nombre de principes, de règles et de solutions techniques de nature à donner au système d'archivage toutes les garanties de sécurité et d'intégrité qui peuvent être exigées pour les documents numériques.



Attention

Naturellement, plus le niveau de sécurité exigé est élevé, plus les coûts de développement et de fonctionnement de l'Archive seront élevés. Il est donc essentiel de savoir se situer au bon niveau au regard de ces exigences.

Chapitre 8. Certification des Archives numériques

Donner à une Archive numérique un label de qualité et de confiance, dire que nous disposons de garanties effectives suffisantes quant à la capacité de cette Archive à préserver des documents numériques à long terme, à s'assurer de leur authenticité, à maintenir leur intégrité tout en les gardant accessibles à la communauté des utilisateurs, tels sont les objectifs de la certification.

8.1. Qu'est ce que la certification des Archives numériques ?

Nous sommes ici sur une question de grande importance mais non encore stabilisée au plan des solutions envisageables.

Sur quoi pouvons-nous nous baser pour accorder une réelle confiance à une Archive numérique ?

Rien de ce qui se passe dans notre ordinateur n'est visible, sauf le résultat qui correspondra ou ne correspondra pas à nos attentes.

Qui n'a pas vu son PC bloqué sans raison apparente ? Et dans ces cas-là, il est bien difficile, sinon impossible, de connaître la cause exacte de ce qui s'est passé. Qui n'a pas perdu du temps parce que son logiciel de traitement de texte s'était figé avant même la sauvegarde de ce qu'on ne voulait pas perdre ? Combien de particuliers mais aussi d'entreprises ont perdu des informations à cause d'un virus ou en raison d'un piratage ?

Or on veut pouvoir s'appuyer sur cette technologie pour conserver des informations précieuses, vitales, et ce afin de pérenniser nos connaissances, de préserver notre mémoire ? Il y a de toute évidence besoin de trouver le moyen de créer cette confiance. Il est nécessaire qu'une Archive puisse démontrer qu'elle fait effectivement tout ce qu'il y a à faire pour conserver les informations dont elle a la responsabilité, qu'elle a les compétences pour le faire, qu'elle dispose des moyens nécessaires et qu'elle le fait bien. **Cette démonstration est ce qu'on appelle la certification des Archives.**



Complément

On peut définir ce processus de certification des Archives dans le même esprit que les processus de certification ISO 9000 (système de management de la qualité). En effet, le processus de certification ISO 9000, permet de montrer que l'organisme fait bien ce qu'il est supposé faire, qu'il le fait de manière satisfaisante pour les clients et qu'il a entrepris une démarche permanente d'amélioration de la qualité.

Une démarche voisine s'appliquant au domaine spécifique de la sécurité des systèmes d'information a été entreprise. Là encore, se pose la question de la confiance que l'on peut faire à un système d'information lorsqu'il gère des transactions bancaires, contient des secrets de fabrication, détient des informations relevant de la Défense Nationale... Cette démarche a donné lieu à la définition d'exigences applicables aux systèmes de gestion de la sécurité de l'information (ISO/IEC 27001:2005) et à l'énoncé d'un code de bonnes pratiques pour la gestion de la sécurité de l'information (ISO/IEC 27002:2005).

L'objectif est de **rentre** dans le concret de l'évaluation des mises en œuvre et des organisations.

La réflexion sur ce sujet pour les Archives numériques a été engagée dès 2003 - dans la foulée de la publication du modèle OAIS - par un groupe de travail international créé à l'initiative de la National Archives and Records Administration (NARA) et le Research Library group (RLG). Elle s'est ensuite poursuivie sous l'égide de l'OCLC (Online Computer Library Center).

L'objectif était d'élaborer des exigences de certification des archives dans une perspective de normalisation internationale. L'idée principale qui a été retenue et reprise dans tous les travaux ultérieurs a consisté à établir une liste de critères que l'Archive doit satisfaire et à classer ces critères par domaine. Ces critères doivent pouvoir donner des gages de confiance, de fiabilité et de pérennité de l'Archive

Ce travail a abouti à la publication en janvier 2007 du document *Trustworthy Repositories Audit & Certification : Criteria and Checklist*³, connu sous le nom de TRAC.

Par ailleurs, une collaboration entre le Digital Curation Center⁴ (DCC) en Grande-Bretagne et le programme européen Digital Preservation Europe⁵ e⁶ (DPE), a conduit à l'élaboration d'une méthode d'audit des archives numériques basée sur l'évaluation des risques. Elle a pour titre « Digital Repositories Audit Method Based on Risks Assessment (DRAMBORA) ». Cette méthode est essentiellement intéressante pour conduire des auto-évaluations.

3 - <http://www.crl.edu/content.asp?l1=13&l2=58&l3=162&l4=91>

4 - <http://www.dcc.ac.uk/>

5 - <http://www.digitalpreservationeurope.eu/>

6 - <http://www.digitalpreservationeurope.eu/>

Un processus de certification ne se réduit pas à l'examen de la conformité de l'Archive par rapport à un ensemble de critères. Ce processus pose d'autres questions qui doivent être parfaitement clarifiées :

- Qui peut effectuer un audit de certification ?
- Quelles sont les compétences nécessaires ?
- Quelles sont les garanties de neutralité et de confidentialité ?
- Qui peut délivrer cette certification ?
- Avec quelle accréditation ou légitimité ?

Toutes ces questions doivent également recevoir des réponses normalisées et incontestables.

8.2 Projet ISO-CCSDS en cours

Sur la base des travaux mentionnés ci-avant, un groupe de travail du CCSDS a été officiellement mandaté fin 2007, pour préparer l'élaboration d'une norme ISO dans ce domaine.

Le groupe est ouvert et les documents de travail sont accessibles sur le site <http://wiki.digitalrepositoryauditandcertification.org/bin/view/Main/WebHome>⁷

Un premier projet de norme, « *Metrics for Digital Repository Audit and Certification* », a repris et précisé la structure des critères de certification définie dans TRAC.

Domaine	Sous-domaine
Infrastructure organisationnelle	Gouvernance et viabilité organisationnelle
	Structure de l'organisation et personnels
	Cadre général de la politique de pérennisation et de la traçabilité
	Pérennité du financement
	Contrats, licences et engagements
Gestion des objets numériques	Acquisition des contenus
	Entrées : création de l'AIP
	Planification de la pérennisation
	Pérennisation des AIP
	Gestion de l'information
	Gestion de l'accès
Gestion des risques relatifs à l'infrastructure et à la sécurité	Gestion des risques relatifs à l'infrastructure technique
	Gestion des risques relatifs à la sécurité

Image 26 La structuration des critères de certification en domaines et sous-domaines

Pour chacun des sous-domaines, un ensemble de critères a été défini :

Sous-domaine	Critères de certification
Gouvernance et viabilité organisationnelle	L'Archive doit disposer d'un document explicitant sa mission de pérennisation, de gestion et d'accès à l'information numérique.
	L'Archive doit disposer d'un plan stratégique de pérennisation qui définit l'approche à long terme retenue pour remplir sa mission.
	L'Archive doit disposer d'un plan de réversibilité formel, de plans d'urgence, et/ou d'un dépôt de garantie dans le cas où cette Archive arrête son activité ou si l'institution assurant son management ou son financement modifie significativement son champ d'application.
	L'Archive assure une surveillance de son environnement organisationnel afin de déterminer à quel moment il devient nécessaire de mettre à exécution son plan de réversibilité, plan d'urgence et/ou utiliser son dépôt de garantie.
	L'Archive doit disposer d'une politique des collections ou de tout autre document qui définit les informations à pérenniser.

Image 27 Les critères de certification peuvent se référer aux aspects organisationnels et politiques

Sous-domaine	Critères de certification
Acquisition des contenus	L'Archive doit identifier le ou les contenus informationnels des objets numériques qu'elle devra préserver et leur propriétés.
	L'Archive doit spécifier clairement l'information qui doit être associée au Contenu d'information au moment de son dépôt (SIP).
	L'Archive doit disposer de spécifications permettant la reconnaissance et l'analyse syntaxique des SIP.
	L'Archive dispose des mécanismes pour authentifier le déposant pour chaque objet transmis.
	Le traitement d'ingestion de l'Archive doit vérifier que chaque objet numérique soumis est complet et conforme à ce qui a été spécifié.
	L'Archive doit disposer d'un contrôle physique suffisant sur les objets numériques afin de les préserver de façon efficace.
	L'Archive fournit au Producteur/déposant les informations adéquates aux points de contrôle prédéfinis dans les procédures de versement.
	L'Archive assure la traçabilité de toutes les actions et des processus d'administration nécessaires à l'acquisition des contenus.

Image 28 Ils couvrent aussi de nombreuses préoccupations sur les contenus archivés, un exemple avec le sous-domaine « Acquisition des contenus »

Chaque critère proposé fait l'objet :

- d'une formulation précise,
- d'un texte explicatif,
- d'un ensemble d'exemples sur les moyens permettant de justifier que le critère est satisfait.

Ce projet devrait aboutir à la publication d'une norme en 2010.

La connaissance des critères de certification sera aussi essentielle au moment de la conception et de la mise en place d'une Archive numérique puisqu'elle constituera un guide permettant de prendre à l'avance les dispositions visant à une certification ultérieure.

Un second projet de norme, intitulé « *Requirements for bodies providing audit and certification of digital preservation management systems* » devrait compléter les critères de certification sur l'ensemble des aspects relatifs à la procédure d'audit et les mécanismes d'attribution et de renouvellement de la certification.

8.3. Autres approches envisagées

L'hypothèse du CCSDS en matière de certification se réfère à une Archive prenant en charge l'ensemble des responsabilités telles qu'elles sont définies dans le modèle OAIS.

Il peut exister des contextes plus simples ou plus limités, par exemple :

- une partie des activités d'archivage est externalisée et le tiers archiveur souhaite être certifié sur le périmètre de responsabilité qui est le sien ;
- les documents à archiver sont des documents récurrents et les questions principales à résoudre se limitent à la conservation des bits et à la valeur probante des documents



Complément

Il existe des approches alternatives en matière de certification, qui sont en cours d'élaboration. Citons ici :

- construire un référentiel de certification en se basant sur la norme AFNOR Z 42-013 ; une étude est actuellement en cours chez AFNOR certification sur ce sujet ; ce référentiel pourrait répondre aux besoins des tiers archiveurs ;
- l'établissement d'une charte ou politique d'archivage dans les services publics d'archives constitue un référentiel de la sécurité de l'archivage électronique afin qu'il puisse être qualifié de « fiable » ; une grille d'audit peut être constituée à partir de ses différents chapitres et permet à un auditeur de contrôler la fiabilité d'un service d'archivage électronique ; cet audit peut aussi conduire, sous certaines conditions, à l'établissement d'une certification ;
- une autre approche qui vise à définir un référentiel de certification reposant sur la série de normes ISO 27000 (techniques de sécurité) ; l'idée de base réside dans le fait que pour des documents relativement simples (relevés de banque, factures, fiche de paie...), on peut considérer que 75% des exigences de l'archivage sont couvertes par les exigences de sécurité.

Chapitre 9. Conclusions

L'environnement normatif s'est considérablement enrichi en quelques années et constitue un apport irremplaçable pour l'archivage numérique.

Le cadre normatif reste cependant difficile à appréhender en raison de la diversité des points de vue et des périmètres couverts, mais aussi en raison des recouvrements partiels existant entre certaines normes. Cette situation complexe a pour origine essentielle la multiplicité des secteurs d'activités qui ont à faire face à des nécessités d'archivage numérique. Chacun de ces secteurs se retourne naturellement vers les organismes de normalisa

au sein de ces organismes, vers les comités techniques dont ils relèvent habituellement alors qu'une approche transverse serait probablement bénéfique pour tous.

Questions : OAIS

Objectifs

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

[Solution n°1 p 29]

Information disponible : le format du document est conforme à la norme ISO 19005-1 (Electronic document file format for long-term preservation PDF/A-1)

Proposition : Cette information relève de l'Information de représentation et doit impérativement être conservée.

☐

Vrai

☐

Faux

Exercice 2

[Solution n°2 p 29]

Information disponible : nous disposons d'une copie complète de la norme ISO 19005-1

Proposition : Il n'est pas indispensable de conserver ce document puisque l'ISO se chargera de le conserver.

☐ Vrai

☐ Faux

Exercice 3

[Solution n°3 p 29]

Information disponible : nous disposons de l’empreinte numérique du document archivé. Cette empreinte a été calculée avec l’algorithme SHA.

Proposition : Cette empreinte nous permet de connaître la provenance du document.

☐ Vrai

☐ Faux

Exercice 4

[Solution n°4 p 29]

Information disponible : Le document a fait, au moment de son archivage, l’objet d’une migration de format entre PDF 1.7 et PDF/A

Proposition : il s’agit d’une Information de provenance qui doit être conservée.

☐ Vrai

☐ Faux

Exercice 5

[Solution n°5 p 29]

Information disponible : Nous disposons d’un dictionnaire des mots techniques utilisé

Proposition : ce dictionnaire doit être conservé, il est l’un des constituants de l’information de représentation.

☐ Vrai

☐ Faux

Exercice 6

[Solution n°6 p 30]

Information disponible : Les différentes catégories d’information associées au document archivé, sont organisées au sein d’une structure de répertoires. Au plus haut niveau, nous avons un répertoire dont le nom est l’identifiant unique du document. Ce répertoire possède lui même plusieurs sous-répertoires permettant de ranger les différentes catégories d’information : un répertoire « Contenu » dans lequel sera rangé le document principal, un répertoire « Information-de-représentation » dédié à l’Information de représentation et enfin un répertoire « Information-de-pérennisation » permettant d’organiser les différentes catégories Information de provenance, de contexte, d’intégrité, de droits d’accès et d’identification.

Proposition : La description de cette structure de répertoire doit être conservée à long terme aussi longtemps que le document lui-même.

☐ Vrai

☐ Faux

Exercice 7

[Solution n°7 p 30]

Information disponible : Nous avons reçu un document signé, c'est-à-dire accompagné de sa signature électronique.

Proposition : Conformément au modèle OAIS, cette information relève de l'Information d'authenticité.

☐ Vrai

☐ Faux

Exercice : Révisez vos connaissances

[Solution n°8 p 30]

Exercice

L'information de pérennisation comprend 5 types d'information. Quels sont-ils?

☐	L'information de représentation
☐	L'information de contexte
☐	L'information de provenance
☐	L'information d'identification
☐	L'information d'intégrité
☐	L'information de droits d'accès
☐	L'information d'emballage
☐	L'information de description

Exercice

Dans le modèle OAIS, je contrôle l'accès des utilisateurs au système d'archivage ainsi que les droits de ces utilisateurs à accéder ou non à tel ou tel objet numérique archivés. Qui suis-je?

☐	L'entité « Gestion des données »
☐	L'entité « Entrées »
☐	L'entité « Accès »
☐	L'entité « Administration »

Exercice

Je suis une norme qui se concentre sur les spécifications techniques du système informatique sur lequel l'Archive numérique va appuyer ses activités. Qui suis-je?

☐	La norme ISO 14721
☐	Le modèle OAIS
☐	La norme AFNOR NF Z 42-013

Solution des exercices

> Solution n°1 (exercice p. 24)

- | | |
|----------------------------------|------|
| ☒ | Vrai |
| ☐ | Faux |
- Cette information sera indispensable dans le futur afin de toujours être capable de passer des bits du document numérique à son contenu intelligible.*

> Solution n°2 (exercice p. 24)

- | | |
|----------------------------------|------|
| ☐ | Vrai |
| ☒ | Faux |
- Les organismes de normalisation n'ont pas, à ce jour, pris des engagements internationaux explicites de conservation à long terme de tous les documents normatifs, y compris ceux qui sont périmés.

> Solution n°3 (exercice p. 25)

- | | |
|----------------------------------|------|
| ☐ | Vrai |
| ☒ | Faux |
- L'empreinte ne peut pas nous renseigner directement sur la provenance du document. Elle nous permet simplement de savoir si le document, tel qu'il est archivé, a été modifié (volontairement ou accidentellement) depuis qu'on a calculé son empreinte. L'empreinte fait partie de l'information d'intégrité. Il est vrai cependant que les signatures électroniques permettant de s'assurer de l'intégrité et de la provenance d'un document s'appuient aussi sur les calculs d'empreinte, mais ces empreintes ne constituent ici qu'un élément au sein d'un dispositif technique et organisationnel plus complexe.

> Solution n°4 (exercice p. 25)

- | | |
|----------------------------------|------|
| ☒ | Vrai |
| ☐ | Faux |
- L'information de provenance doit nous renseigner sur l'origine ou la source du document archivé, ainsi aussi sur son historique. Cet historique doit permettre de connaître et de tracer tous changements intervenus depuis la réception de ce document par l'Archive. Cette information de provenance doit donner aux utilisateurs futurs une certaine assurance quant à la fiabilité des contenus d'information.

> Solution n°5 (exercice p. 25)

- | | |
|----------------------------------|------|
| ☒ | Vrai |
| ☐ | Faux |

Nous savons que le vocabulaire évolue au cours de temps. Le vocabulaire scientifique et technique évolue de façon particulièrement rapide. L'objectif de l'archivage est de faire en sorte que le document soit conservé aussi longtemps qu'il soit intelligible et utilisable par les utilisateurs futurs sur le long terme. Ces utilisateurs devront pouvoir se reporter à la signification exacte de chacun des termes techniques, au moment de la création du document.

> Solution n°6 (exercice p. 25)

- | | |
|----------------------------------|------|
| ☐ | Vrai |
| ☒ | Faux |

Ce qui est indispensable, c'est que le Contenu ainsi que les Informations de représentation et de pérennisation soient conservées. L'Archive peut tout à fait décider à un moment ou un autre, d'organiser différemment le rangement de ces informations. La description de cette structure de répertoires relève de l'Information d'empaquetage

> Solution n°7 (exercice p. 26)

- | | |
|----------------------------------|------|
| ☐ | Vrai |
| ☒ | Faux |

Les éléments prouvant l'authenticité du document font partie de l'Information de provenance dans le modèle OAIS.

> Solution n°8 (exercice p. 27)

Exercice

- | | |
|-------------------------------------|---------------------------------|
| ☐ | L'information de représentation |
| ☒ | L'information de contexte |
| ☒ | L'information de provenance |
| ☒ | L'information d'identification |
| ☒ | L'information d'intégrité |
| ☒ | L'information de droits d'accès |
| ☐ | L'information d'empaquetage |
| ☐ | L'information de description |

Exercice

Solution des exercices

☐ L'entité « Gestion des données »

☐ L'entité « Entrées »

☒ L'entité « Accès »

☐ L'entité « Administration »

Exercice

☐ La norme ISO 14721

☐ Le modèle OAIS

☒ La norme AFNOR NF Z 42-013

Glossaire

Archive

Organisation chargée de conserver l'information pour permettre à une communauté d'utilisateurs cible d'y accéder et de l'utiliser (glossaire du modèle de référence OAIS).

Archives courantes

Documents et dossiers ouverts ou récemment clos gardés dans les bureaux pour le traitement des affaires.

Archives définitives

Documents qui, après évaluation, sont conservés sans limitation de durée.

Archives intermédiaires

Ensemble de documents qui, n'étant plus d'usage courant, doivent être conservés temporairement, pour des besoins administratifs ou juridiques (y compris les documents qui après tri seront conservés comme archives définitives)

Information

L'information est définie comme une connaissance pouvant être échangée. En pratique, l'information est donc un élément de connaissance susceptible d'être codé pour être conservé, traité ou communiqué.

Modèle de Référence

Cadre permettant la compréhension des relations essentielles entre entités d'un environnement donné, ainsi que le développement de normes ou de spécifications cohérentes liées à cet environnement. Un Modèle de référence s'appuie sur un nombre restreint de concepts unificateurs et peut servir de support à la formation et à la sensibilisation des non-spécialistes à ces normes (glossaire du modèle de référence OAIS).

Système ouvert d'archivage d'information (Open Archival Information System - OAIS)

Archive, avec son équipe et ses systèmes, officiellement chargée de conserver l'information et de la mettre à disposition d'une Communauté d'utilisateurs cible.

Le terme « ouvert » dans OAIS signifie que cette recommandation et les recommandations et normes ultérieures liées à l'OAIS sont élaborées dans des forums ouverts, et en aucun cas que l'accès à l'Archive est libre.

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.

Webographie

[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)