

> L'informatique a envahi notre société, modifié nos pratiques quotidiennes, transformé les méthodes de travail. Toute l'information produite, transmise, traitée, stockée dans ces nouvelles pratiques est sous forme numérique. Les obsolescences constantes de la technologie et l'absence d'organisations appropriées rendent cette information de plus en plus vulnérable au temps. La pérennisation de cette information pose des problèmes nouveaux à nombre d'organismes et d'entreprises. La recherche scientifique n'échappe pas à cette difficulté. Une norme fondamentale, le modèle de référence OAIS, nous permet de comprendre en profondeur le problème posé. Ce modèle constitue une base indispensable pour concevoir des solutions qui commencent à voir le jour. <

La pérennisation des informations sous forme numérique : risques, enjeux et éléments de solution

Claude Huc



3, chemin du Couderla,
31320 Vigoulet-Auzil, France.
claud.huc@free.fr

La vulnérabilité de l'information numérique face au temps

L'informatique, au travers des ordinateurs, des réseaux, des moyens de stockage, des nouvelles technologies logicielles, a profondément transformé le visage de notre société. Elle a modifié l'ensemble de nos pratiques tant professionnelles que personnelles. Du comptable au chercheur, de l'ingénieur au documentaliste, de l'architecte au bibliothécaire, on ne compte plus les métiers qui exigent que l'on passe une majeure partie de son temps devant un écran. L'informatique a généré des changements majeurs dans les méthodes de conception et de fabrication industrielles. La photographie argentique a pratiquement disparu. La radiographie suit le même mouvement. Le domaine de la création artistique n'est pas épargné par ces évolutions. Même les métiers relevant plus de l'artisanat que de l'industrie sont profondément touchés : dans quelques années par exemple, bon nombre d'artisans bijoutiers auront abandonné leurs pinces et leurs petits marteaux pour les remplacer par des outils de conception et de fabrication assistées sur ordinateur.

Cette évolution a pour corollaire que toute l'information créée, manipulée, traitée, stockée, échangée, diffusée dans notre société est une information sous forme numérique. Cette forme de représentation de l'information est incroyablement puissante. Le coût de son stockage est de plus en plus faible, il est possible de la préserver sans la moindre altération, nous disposons d'outils puissants pour créer des documents complexes et pour rechercher les informations utiles. Cette information peut revêtir les formes les plus diverses : documents textuels ou composites issus des outils de bureautique (ouvrages, publications...), images issues de la photographie numérique, images de synthèse, graphiques vectoriels, données et observations scientifiques, son, vidéo, documents multimédia, etc. Cette omniprésence de l'information sous forme numérique est accentuée par l'interpénétration des domaines que sont l'informatique, la photographie, la téléphonie et la télévision. La croissance de la production de contenus numériques est vertigineuse.

Et pourtant, en arrière plan des avantages immenses que l'information numérique nous procure se cache un risque majeur, celui d'une grande vulnérabilité au temps, explicable et avérée.



Lorsque nous disposons d'une information sous forme numérique (nous parlerons d'objet numérique), nous n'avons rien d'autre qu'une suite de 0 et de 1 tels que l'illustre la *Figure 1*. Il ne suffira généralement pas de se référer à un format standard du marché pour pouvoir restituer le contenu informationnel car ces standards existent en très grand nombre (supérieur au millier) et sont aussi en constante évolution. L'objectif recherché est à la fois de préserver l'intégrité de l'objet numérique mais aussi la capacité à restituer son contenu sous une forme intelligible et compréhensible par ceux qui l'utiliseront plus tard.

Les causes technologiques de cette vulnérabilité

Il suffit généralement d'ouvrir un livre pour pouvoir prendre connaissance de son contenu alors qu'un dispositif de lecture complexe impliquant du matériel et du logiciel est indispensable pour l'accès à une donnée numérique. La pérennité d'un tel dispositif de lecture dépend de celle de son fabricant ou de son bon vouloir, le bon vouloir de l'éditeur de logiciel pouvant être de vous garder sous sa dépendance. L'indisponibilité d'un seul élément au sein du dispositif de lecture peut compromettre définitivement un patrimoine de données.

Chacun sait à quelle vitesse évoluent les technologies numériques et à quel rythme les nouvelles éliminent les anciennes. Nous sommes confrontés ici à la faible pérennité des supports de données, à la disparition pure et simple des technologies de stockage les unes après les autres (*Figure 2*). Une telle disparition signifie en pratique que les matériels utilisés par cette technologie ne sont plus fabriqués et disparaissent du marché avec pour conséquence un accroissement exponentiel des coûts de maintenance des équipements existants. Nous sommes confrontés aussi au changement régulier des systèmes d'exploitation et des versions des logiciels que nous utilisons.

Le retour d'expérience de la conservation à long terme de documents et de données depuis plusieurs dizaines d'années dans le domaine de la recherche scientifique spatiale est éloquent à cet égard : média de stockage illisible, média en parfait état mais des équipements de lecture/écriture qui disparaissent du marché soit parce que le marché est trop limité, soit parce que la technologie est remplacée par une autre, moins coûteuse et plus performante, fichiers avec des structures propriétaires relevant d'un système d'exploitation qui a changé, documents textuels saisis avec un outil de traitement de texte propriétaire qui a disparu et qu'il faut saisir à nouveau, formulations mathématiques qu'il faut saisir à nouveau car l'éditeur d'équations n'est plus compatible avec le précédent, rupture des compatibilités ascendantes des versions successives de logiciels,...



Figure 1. Un objet numérique, c'est cela, et quand vous n'avez que cela, vous n'avez pas grand-chose...

On peut dire sans risque de se tromper qu'une information sous forme numérique devient vulnérable au-delà d'une période pouvant se situer entre 5 et 10 ans suivant les cas. Nombre de documents, de publications, de bases de données ont vocation à être préservées beaucoup plus longtemps. En d'autres termes, il s'agit de trouver le moyen de pérenniser des informations à l'aide d'une technologie qui n'a aucune pérennité.

Mais la technologie n'est pas seule en cause...

La pérennité de documents et de données sous forme numérique implique la mise en place d'une organisation dont c'est le mandat et la responsabilité, avec des compétences et des moyens adéquats. Cette nouvelle activité implique des ressources nouvelles alors même que les décideurs sont loin d'avoir tous conscience des enjeux et des risques dans ce domaine. Le problème des formats propriétaires fermés se heurte à la politique de certains éditeurs de logiciels. Certains d'entre eux disposent d'un monopole mondial sur leur spécialité. La fâcheuse habitude de traiter sans cesse les questions dans l'urgence nous fait remettre à plus tard tout ce qui peut relever du long terme, ce long terme n'étant en réalité que la durée de vie des technologies, c'est-à-dire 5 à 10 ans, parfois moins.

Un problème qui se pose à tous et une mutualisation indispensable

Les plans détaillés d'un avion de ligne doivent être conservés pendant 50 ans, ce sera une centaine d'an-



Les membres du groupe de travail PIN au sein de l'association Aristote [2]

Organismes à caractère scientifique et technique	Le Centre National d'Études Spatiales, le Commissariat à l'Énergie Atomique, le Centre National de la Recherche Scientifique, le Centre Informatique National de l'Enseignement Supérieur, l'Institut Géographique National, le Centre International pour la Recherche en Agronomie et le Développement...
Organismes et entreprises relevant du domaine de la santé	L'Institut Pasteur, l'Agence française de sécurité sanitaire des aliments, Sanofi Pasteur
Organismes en charge du patrimoine historique et culturel	L'UNESCO, La Bibliothèque nationale de France, la Direction des Archives de France, les Archives nationales, diverses archives régionales, départementales et municipales
Représentants des ministères	Agriculture, Défense, Justice (la gestion des dossiers judiciaires peut parfois durer 30 ans !), culture et communication, Écologie
Industriels ayant des besoins de pérennisation d'information	Airbus industrie
Ou encore...	Des opérateurs de télécommunication (France Télécom), des institutions du domaine artistique (Institut de Recherche et de Coordination Acoustique/Musique (IRCAM)...

Tableau 1. Le groupe PIN en France.

nées pour une centrale nucléaire. Les dossiers de retraite des salariés, les dossiers médicaux des patients dans un hôpital seront conservés pendant des dizaines d'années. Que dire encore des fichiers de l'État Civil, ou des données d'observations qui tracent l'évolution de notre planète, des publications et des données scientifiques, ou encore des documents historiques et patrimoniaux ?

Cet inventaire très partiel nous fait percevoir à la fois la gravité du problème posé et en même temps le très grand intérêt qu'il y a pour tous ces secteurs d'activité, à mutualiser leurs efforts, les questions essentielles à résoudre étant pratiquement les mêmes pour tous. Plusieurs initiatives vont dans ce sens. Nous pouvons ici citer le groupe PIN (Pérennisation des Informations Numériques) en France [1]. La diversité des membres du groupe, illustrée dans le *Tableau 1*, est un révélateur probant de la nécessité de partager les analyses, les retours d'expérience, les succès et les échecs, voire parfois les investissements et les infrastructures.

Les éléments de solutions

Avant de résoudre le problème, il convient de le comprendre. Une norme internationale fondamentale, publiée en 2003, issue du travail des centres de données scientifiques du domaine spatial, mais enrichie des contributions des grandes bibliothèques et des archives institutionnelles répond à cette première question. C'est le Modèle de référence OAIS (*open archival information system*) disponible en version anglaise [3] et en version française [4].

Ce modèle nous apporte l'ensemble des concepts nécessaires à cette compréhension. Il définit l'Archive comme une organisation chargée de conserver l'information pour permettre à une communauté d'utilisateurs d'y accéder et de l'utiliser. L'Archive couvre donc

fondamentalement la conservation de l'information, l'accès à cette information et l'intelligibilité de cette information pour ceux auxquels elle est potentiellement destinée. Le Modèle décrit en détail l'ensemble des fonctionnalités qui devront être prises en charge par toute entité responsable de la pérennisation d'objets information sous forme numérique : récupération des objets auprès des producteurs puis validation ou transformation de ces objets, stockage, gestion des données archivées, accès, administration. À ces fonctions de base s'ajoute une fonction essentielle de planification de la pérennisation visant à gérer tout ce



Figure 2. Silos de stockage à grande capacité. Contrairement aux idées reçues, la question du stockage pérenne ne joue qu'un rôle relatif dans la question de la pérennisation de l'information numérique.

qui change, tant du côté des technologies que du côté de la communauté des utilisateurs.

Le Modèle OAIS définit également une taxinomie des informations. À côté des objets numériques qui constituent l'objectif, la cible, de la pérennisation, il conviendra de conserver d'autres informations indispensables à la restitution et à l'usage de l'information cible : l'information dite de représentation est indispensable pour passer des *bits* à l'information intelligible. Cette information de représentation étant elle-même numérique, elle devra disposer de sa propre information de représentation. De ce point de vue, la vision est réursive. Il sera également nécessaire de conserver des informations sur la provenance de l'objet, sur les motifs de sa création. Une information d'identification devra lui être attribuée. L'information d'intégrité sera indispensable pour vérifier que l'objet n'a pas été modifié sans que cette modification ait été tracée. Elle sera également nécessaire dans un contexte de documents à valeur probante. Ce dernier point nous amène naturellement à la problématique juridique dans son ensemble et en particulier celle qui concerne la propriété intellectuelle. Il s'agit d'un sujet complexe qu'il conviendra de développer par ailleurs.

Bien qu'issu du domaine spatial, le Modèle OAIS a été adopté au plan national et au plan international par un grand nombre de bibliothèques, d'archives institutionnelles et d'entreprises industrielles. Un certain nombre de plates-formes d'archivage numériques ont été développées ou sont en cours de développement. Certaines d'entre elles ont spécifié que le Modèle OAIS était applicable dans son intégralité. C'est le cas pour le développement de SPAR (Système de Préservation et d'Archivage Réparti) à la BnF (Bibliothèque nationale de France).

Les retours d'expérience disponibles nous montrent à quel point la synergie et la complémentarité entre, d'une part, les compétences techniques - spécifiquement dans les domaines de l'informatique concernés : stockage, formats de données, formats de métadonnées, bases de données, Internet... - et, d'autre part, l'expertise archivistique sur les contenus et l'intelligibilité de ces contenus est indispensable. Croire que l'archivage numérique est une question purement technique est une erreur.

En complément de cette norme essentielle, il sera nécessaire de s'appuyer sur un grand nombre de normes ou de standards spécialisés sur la question des média de stockage, sur les formats de données et de métadonnées, sur les identifiants pérennes, sur l'intégrité et l'authenticité des documents, sur l'empaquetage des données, sur les interfaces entre les Archives et les producteurs de documents, sur le processus de certification des archives.

Bon nombre de ces normes vont être construites comme des poupées russes à partir de normes généralistes en informatique parmi lesquelles on pourra citer la norme Unicode pour la représentation universelle des caractères ou la norme XML pour la structuration des documents. Nous n'oublions pas non plus que notre capacité à pérenniser un objet numérique n'est pas indépendante de la manière dont ce document a été produit, géré, contrôlé avant son intégration dans une Archive numérique.

Les initiatives européennes et internationales

Les problèmes posés dans cet article se posent naturellement au plan international.

Il est intéressant ici de mentionner quelques-uns des plus importants projets de recherche européens relevant d'investissements de plusieurs dizaines de millions d'Euros du Programme cadre de Recherche et Développement (PCRD) consacré à ce domaine :

- CASPAR : *Cultural, Artistic and Scientific Knowledge for Preservation, Access and retrieval* [5].

CASPAR centre ses objectifs sur la pérennisation de l'information dans les domaines artistique, scientifique et culturel. Il présente l'intérêt de rechercher, mettre en œuvre et diffuser des solutions innovantes basées sur le Modèle de Référence OAIS.

- PLANETS : *Preservation and Long-term Access through Networked Services* [6].

L'objectif de PLANETS est de développer un réseau de services et des outils d'aide à la préservation de l'information numérique dans les domaines culturels et scientifiques. En pratique, il est fortement centré sur les besoins des bibliothèques et des archives institutionnelles.

- DPE : *Digital preservation Europe* [7].

DPE vise à contribuer au regroupement et à faciliter les interactions entre les différentes expertises existant en Europe dans le domaine de la recherche, dans le domaine culturel, dans l'administration publique et dans l'industrie, sur la question de la pérennisation de l'information numérique.

D'autres initiatives européennes sont en cours. Au plan international, les projets sont également nombreux.

- Citons à ce stade, le projet INTERPARES - *International Research on Permanent Authentic Records in Electronic Systems* [8], financé par la Canada depuis 1999.

Conclusions

Nous savons que le chemin à parcourir avant d'aboutir à des solutions totalement fiables et reconnues est encore long. Les besoins de la pérennisation de l'information numérique entrent parfois en contradiction avec les intérêts de certains éditeurs de logiciels. La situation d'ensemble est très hétérogène. Les agences spatiales ont été les premières à devoir faire face au problème. Certaines institutions en France comme la BnF sont en train d'opérer une véritable mutation vers l'archivage numérique avec une vraie compréhension de tout ce que cela implique. D'autres organismes n'ont pas encore fait ce pas. La perte de données ou de documents numériques de valeur sera à terme le facteur déclenchant la nécessaire prise de conscience du problème par les décideurs. ♦

SUMMARY

The long term preservation of information in digital form: risks, stake and outline of practical solutions

All information is now created, handled, stored and often used in digital form: we observe a formidable growth in the volume of data created, transmitted and archived. However, the obsolescence of technologies and the lack of appropriate organisation lead to a spectacular vulnerability to time after 5-10 years. Every sector of activity in today's society is concerned, public and private organizations, scientific research, etc. The ISO standard Reference Model for an Open Archival information System provides us with a high level of understanding of the problem to solve. This Model is essential for the emergence of practical solutions. ♦

RÉFÉRENCES

1. <http://www.aristote.asso.fr/PIN/index.html>
2. <http://www.aristote.asso.fr>
3. Organisation internationale de normalisation. ISO 14721-2003 Open Archival Information System - Reference Model, Genève : ISO 2003 : 164 p.
4. Organisation internationale de normalisation. ISO 14721-2003 Système ouvert d'archivage d'information - Modèle de Référence, Genève : ISO 2003 : 146 p.
5. <http://www.casparpreserves.eu/>
6. <http://www.planets-project.eu/>
7. <http://www.digitalpreservationeurope.eu/>
8. <http://www.interpares.org/>

TIRÉS À PART

C. Huc



Analyser

Viabilité cellulaire et Apoptose
Gènes rapporteurs
Instrumentation

Venez assister à l'atelier :

**« Application de la bioluminescence
à l'étude de la viabilité cellulaire et de l'ADMETox »**

Mardi 3 juin à 10 Heures
Forum Labo 2008, stand (H18-I19 Hall3), Paris.

www.promega.com/fr





> Grâce à *m/s*, vivez en direct les progrès des sciences biologiques et médicales

Chaque mois, avec les articles de référence de *M/S*

Chaque jour, sur www.medecinesciences.org



Médecine/Sciences

est indexé dans PubMed/Medline

Current Contents, série Life Sciences
EMBASE/Excerpta Medica
PASCAL
CABS
BIOSIS



Tarifs d'abonnement *M/S* - 2008 Mensuel - 10 numéros/an

Abonnez-vous à Médecine/Sciences

- > Des articles rédigés par des médecins et des chercheurs reconnus sur la scène internationale qui posent avec rigueur les bases des débats scientifiques.
- > Des synthèses, éditoriaux, dossiers techniques et analyses toujours replacés dans leur contexte pour que l'information soit la plus exacte, intelligible et objective.
- > La dimension humaine privilégiée, avec l'analyse des retombées diagnostiques, thérapeutiques, la prévention et l'éthique liées aux nouvelles avancées.

- > Un panorama clair et concis de l'actualité scientifique : des nouvelles, des brèves, des données chiffrées, des repères et perspectives pour qu'aucun fait significatif ne vous échappe.

Mon règlement :

Par mail edk@edk.fr

Uniquement pour les paiements par carte bancaire

Par fax en envoyant ce bulletin au 01 55 64 13 94

Uniquement pour les paiements par carte bancaire

N°

Date d'expiration Signature :

N° de contrôle au dos de la carte

Par chèque à l'ordre de Médecine/Sciences, en envoyant ce bulletin à :

Éditions EDK

2, rue Troyon
92316 Sèvres Cedex, France

Pour recevoir une facture, cochez cette case

Je souhaite m'abonner à *M/S* :

Nom : Prénom :

Adresse :

Code postal Ville :

Pays :

E-mail-obligatoire :

Je choisis l'abonnement :

	Particuliers		Institutions			Étudiants*			
	Papier + Électronique	Électronique seul	Papier	Papier + Électronique	Papier	Papier + Électronique	Électronique seul	Papier	
France	<input type="checkbox"/> 170 €	<input type="checkbox"/> 118 €	<input type="checkbox"/> 160 €	<input type="checkbox"/> 385 €	<input type="checkbox"/> 235 €	<input type="checkbox"/> 375 €	<input type="checkbox"/> 90 €	<input type="checkbox"/> 62 €	<input type="checkbox"/> 80 €
UE + autres	<input type="checkbox"/> 214 €	<input type="checkbox"/> 118 €	<input type="checkbox"/> 204 €	<input type="checkbox"/> 455 €	<input type="checkbox"/> 235 €	<input type="checkbox"/> 433 €	<input type="checkbox"/> 112 €	<input type="checkbox"/> 62 €	<input type="checkbox"/> 102 €

* Joindre un justificatif