

Le groupe PIN

Lieu d'échange, lieu de réflexion, lieu d'action

C. Huc, F. Banat-Berger, M.A. Chabin, D. Courtaud, J. Masanès, J. Poivre

1 LE CONTEXTE

La préservation à long terme des informations sous forme numérique constitue un enjeu capital, un problème grave et un défi commun dans le sens où il concerne aujourd'hui l'ensemble des secteurs d'activité.

1.1 UN ENJEU CAPITAL

L'enjeu est capital car d'une part, la quasi-totalité de l'information créée, gérée, transmise, stockée dans notre société l'est aujourd'hui sous forme numérique et d'autre part, la force probante reconnue par la loi à l'écrit électronique, va permettre de se passer entièrement du support papier.. Ne pas être en mesure de préserver convenablement et durablement cette information conduirait à une société sans mémoire. Mais surtout, les enjeux vont maintenant bien au-delà des questions culturelles et historiques, ces enjeux sont aussi politiques, économiques et sociaux. Que deviendrait une société incapable de justifier sur le moyen et long terme les droits des personnes, incapable d'assurer la protection de ses informations patrimoniales ? Tout le monde imagine que c'est impossible. Que deviendrait une société qui perdrait progressivement les avancées scientifiques et technologiques qui lui auront coûté tant de ressources humaines et financières ? C'est tout autant impossible mais peu nombreuses sont les entités à en avoir pris conscience aujourd'hui.

1.2 UN PROBLEME COMPLEXE

Le problème est également complexe. Il trouve sa première source dans l'obsolescence constante des technologies qui nous conduit à relever un défi paradoxal : comment préserver durablement de l'information avec une technologie qui n'a pas de pérennité et qui nous conduit à changer les systèmes d'exploitation, les logiciels, les ordinateurs tous les 5 ans et les médias de stockage tous les 7 ou 8 ans? Dans un tel contexte, le long terme ne s'applique pas seulement à la conservation définitive des documents. Le long terme sera défini comme une période suffisamment longue pour qu'on se soucie des conséquences que peuvent avoir sur l'information contenue dans un entrepôt numérique les changements technologiques, et notamment la gestion des nouveaux médias et formats de données d'une part, l'évolution des communautés d'utilisateurs d'autre part. Le vocabulaire et la navigation dans les connaissances archivées sont sujets à fluctuer au cours du temps et doivent être elles aussi prises en compte dans la pérennisation.

Cette complexité n'est pas seulement technique mais résulte de l'insuffisance de l'organisation généralement en place dans les organismes et de l'incapacité à penser l'information en terme de flux et de cycle de vie de l'information depuis son intégration dans un système d'information jusqu'à son archivage. Cette exigence en matière d'identification fine de l'information, de traçabilité des événements qui affectent le document tout au long de sa vie devient cruciale dans un environnement électronique, afin de prouver la fiabilité d'un système, l'intégrité et l'authenticité d'un document.

Enfin, cette complexité provient de l'insuffisance du réseau de normes, de législation et réglementation peu appropriées par rapport à ce contexte nouveau.

1.3 UN DEFI COMMUN

Toutes les institutions et toutes les entreprises sont concernées par la pérennisation de l'information numérique. On peut schématiquement classer les archives numériques en 3 grands domaines :

- Le domaine scientifique : ainsi, de nombreux phénomènes observés dans le domaine des sciences de l'univers sont totalement uniques et il est impératif de les conserver pour les confronter dans le futur, aux modèles élaborés pour expliquer et comprendre ces phénomènes. C'est le cas par exemple lorsque l'Agence Spatiale Européenne a lancé en 1985 une sonde spatiale pour observer la comète de Halley. Par ailleurs, l'analyse du comportement de notre planète et notre capacité à prédire son évolution suppose de disposer de données pendant des dizaines d'années, voire beaucoup plus. Ces archives prennent la forme de Centres de Données dont le financement est assuré par les organismes de recherches nationaux, par les agences spatiales. Ce domaine concerne aussi les archives des laboratoires de recherche médicaux, les archives de la météorologie, etc.
- Le domaine 'opérationnel' : il s'agit des archives en relation directe et permanente avec les activités industrielles, économiques, politiques, sociales, juridiques... de notre société. Ainsi dans le secteur industriel, les compagnies pétrolières disposent par exemple d'un gisement d'informations issues des multiples forages. Ces informations sont leur patrimoine. Dans l'aéronautique, l'énorme documentation des avions doit être préservée pendant 50 ans, et en permanence immédiatement accessible. Citons également le secteur de la santé ou des caisses de retraite car les dossiers des patients ou des salariés doivent être conservés pendant toute leur vie. Le domaine administratif ou juridique relève également de cette logique. Le plan stratégique de l'administration électronique (2004-2007) annoncé en février dernier par le Gouvernement vise, dans un objectif de simplifier et d'améliorer les services aux usagers, à dématérialiser des plans entiers de ces différents secteurs d'activité.
- Le domaine patrimonial représenté par les institutions en charge de gérer les archives de l'Etat et également de préserver le patrimoine culturel, artistique et historique : Les archives nationales, les grandes bibliothèques, les organismes en charge de préserver les enregistrements sonores ou vidéo....

Chacune de ces catégories dispose d'un ensemble de spécificités propres mais toutes ont en commun la majeure partie de la problématique de la pérennisation des informations numériques : collecte de l'information, stockage physique pérenne, gestion des migrations de média imposées par l'obsolescence des technologies, accès à cette information par les producteurs et plus largement les usagers, les chercheurs, les citoyens. Elles ont également besoin que les informations préservées soient parfaitement décrites, répertoriées, intelligibles et sécurisées.

2 LE GROUPE 'PERENNISATION DES INFORMATIONS NUMERIQUES' (PIN)

C'est en juin 2000 que le Centre national d'Etudes Spatiales (CNES) a organisé, conjointement avec la Bibliothèque nationale de France (BnF), une réunion interorganisme pour présenter et faire connaître en France le Modèle de Référence OAIS [1]. Ce Modèle constitue aujourd'hui une incontestable référence internationale et multidisciplinaire pour l'archivage numérique. A la suite de cette présentation, les différents participants appartenant pour la plupart à des institutions partenaires du CNES ou de la BnF ont pris la mesure des difficultés à surmonter pour résoudre le problème posé et de l'intérêt qu'il y avait à continuer à travailler ensemble. C'est ce qui a conduit à la création, en août 2000, d'un groupe de travail PIN au sein d'une association à but non lucratif existante, l'association Aristote. Aristote est une association qui regroupe la majorité des grands comptes publics français dans le but de partager leurs expériences et leurs problématiques communes dans les domaines des réseaux informatiques et des nouvelles technologies.

2.1 COMPOSITION DU GROUPE

A l'image de la typologie proposée ci-dessus, le groupe PIN rassemble des représentants expérimentés des trois familles d'archives que nous avons citées. La diversité des institutions représentées dans ce groupe,

les compétences et les points de vue croisés des participants en font un lieu unique pour l'émergence des solutions qui font aujourd'hui défaut. Son activité, comme son nom l'indique, est dédiée à l'ensemble des problèmes techniques, organisationnels, financiers, normatifs, méthodologiques, juridiques, etc. posés par la préservation à long terme de l'information sous forme numérique.

L'originalité et la force de PIN résident dans le caractère transverse de sa composition articulée sur plusieurs pôles :

- Des institutions patrimoniales comme la Direction des archives de France (DAF), le Centre des Archives Contemporaines de Fontainebleau (CAC), la Bibliothèque nationale de France (BnF) ou encore l'Institut National de l'Audiovisuel (INA). Plusieurs ministères comme la Justice, la Défense, l'Agriculture ainsi que les Archives de la ville de Paris sont également représentés ;
- Des organismes à caractère scientifique et technique comme le Centre national d'Etudes Spatiales (CNES), le Commissariat à l'Energie Atomique (CEA), l'Institut de Recherche en Informatique et en Automatique (INRIA) ou encore l'Institut Pasteur ;
- De grands groupes nationaux publics ou privés comme EdF-GdF, France Télécom ou Mederic (caisses de retraites).
- Des experts et chercheurs indépendants reconnus du domaine

D'autres comme le constructeur d'avion Airbus ou encore le Centre Européen de Recherche Nucléaire suivent les travaux du groupe au travers de son forum sans toutefois participer directement aux réunions de travail.

C'est parce la problématique de pérennisation est commune à tous qu'il a été possible de rassembler au sein de PIN, des représentants d'organismes qui n'avaient auparavant aucune relation particulière. Mais la diversité des organismes représentés n'est pas le seul facteur de richesse. La diversité des compétences et des expériences professionnelles des membres du groupe est tout à fait remarquable et caractéristique du croisement de compétences scientifiques, littéraires et juridiques :

- experts dans le domaine de l'archivistique,
- ingénieurs en sciences de l'information
- spécialistes de l'élaboration de normes du domaine,
- enseignants et formateurs de haut niveau
- gestionnaires de dépôt numérique
- juristes

2.2 FONCTIONNEMENT

Le groupe publie un site Web <http://pin.cnes.fr> sur lequel les comptes-rendus de réunion et les présentations des intervenants classées par thème, sont disponibles.

Le groupe dispose également d'un forum de discussion par messagerie. Il se réunit en séance plénière tous les trois mois à Paris. Les thèmes traités ont vocation à couvrir l'ensemble des facettes de la problématique :

- Le domaine normatif et méthodologique, la terminologie, les pratiques et procédures propres à chaque institution,
- L'organisation des archives numériques et leur coût,

- La pérennisation physique des données : le stockage pérenne,
- Les formats de données et de métadonnées,
- Les nouvelles technologies (XML, OAI...)
- etc.

Ces thèmes sont abordés de multiples façons : retours d'expérience, présentations de projets ou d'outils, comptes rendus de colloques et de conférences, synthèses de lectures. Certains sujets occasionnent la création d'un sous-groupe de travail qui se réunit indépendamment et rend compte de ses travaux lors des séances plénières.

Une animation collégiale du groupe reposant sur les représentants des institutions les plus actives a été mise en place. Enfin, PIN dispose du support et de l'expérience de l'association Aristote, notamment pour l'organisation d'activités tournées vers l'extérieur : séminaires, journées de formation... Des animateurs de PIN interviennent également comme organisateurs de séminaires Aristote (le Droit et les nouvelles technologies en 2002, Internet : la mémoire Courte en 2004) ou comme intervenants dans ces séminaires.

3 PROGRAMME D'ACTION

3.1 CREATION D'UNE CULTURE « PIN »

De par leurs responsabilités propres, les membres du groupe ont tous conscience du caractère stratégique du problème de la pérennisation. Nombreux sont ceux qui, dans leur domaine, au sein de leurs institutions respectives, ont pris l'initiative de conduire des actions de toute nature afin de progresser vers des solutions réalistes et fiables.

Le groupe PIN est un lieu de partage de tous ces savoirs et de toutes ces expériences. C'est un lieu de réflexion nourrie par la confrontation d'approches et de points de vues diversifiés. C'est aussi un lieu de partage dans un environnement non commercial et non concurrentiel. C'est enfin un lieu d'action d'où émergent peu à peu des esquisses de solution réalistes et applicables.

Cette expérience collective n'a pas de précédent en France. C'est pourquoi la première ambition du groupe PIN est d'être le creuset d'une formulation de la problématique et de ses solutions où se reconnaissent aussi bien les ingénieurs que les spécialistes des contenus, les décideurs que les opérationnels.

3.2 CONTRIBUTION A L'ELABORATION DE MANUELS PRATIQUES

Dans le cadre d'un sous-groupe nommé SAPIN (Sous-groupe des Archivistes de PIN), une contribution significative a été apportée à l'élaboration de deux manuels publiés par la documentation française sous les auspices de la direction des Archives de France :

- **Les Archives électroniques, manuel pratique** publié en 2002 par Catherine Dhérent [2],
- **L'archivage des documents bureautique, manuel pratique**, publié en 2004 par Joël Poivre (à paraître).

3.3 EVALUATION DES FORMATS DE DONNEES POUR LA PERENNISATION

Le choix de formats de données acceptables dans une perspective de préservation à long terme est une question cruciale. C'est sur ce choix que repose en bonne partie notre capacité future à restituer le contenu et/ou l'apparence des documents préservés sous la forme voulue. Cette question s'applique à toutes les catégories d'information numérique : texte, image, son, vidéo, documents multimédias, etc.

PIN a conduit dans ce domaine une réflexion méthodologique approfondie permettant de définir les différents critères qu'il faut prendre en compte pour évaluer au regard de leur aptitude à la pérennisation de l'information les formats informatiques. PIN a ensuite entrepris d'organiser une répartition du travail d'évaluation entre les différents organismes en mesure d'y contribuer : CNES, BnF, DAF... Il s'agit à ce stade de joindre nos forces et nos capacités respectives pour permettre à chacun de bénéficier d'un travail d'analyse qu'il n'aurait pas eu la possibilité de mener seul.

Enfin, connaissant le grand nombre de formats de représentation d'information existant sur le marché, sachant que pour chaque format, les versions se multiplient, il est évident que PIN ne peut pas maintenir seul une base de données d'évaluation des formats. Le besoin se situe au niveau international et doit être résolu à ce niveau. C'est dans cette perspective que le travail méthodologique de PIN a été traduit en anglais et que des discussions sont en cours, notamment avec la DPC (Digital Preservation Coalition) en Grande-Bretagne et le Global Digital Format Registry (GDFR) initiative internationale lancée aux Etats-Unis.

3.4 ACTIONS DE FORMATION

Dans un environnement où les concepts, les mots et les outils touchant à la pérennisation de l'information numérique fluctuent selon les acteurs et n'ont pas encore trouvé de normalisation stable, le groupe PIN entend diffuser son approche pluridisciplinaire et transverse de cette problématique.

C'est ainsi que dans un premier temps, différents membres du groupe notamment en raison des apports substantiels apportés par les travaux menés en son sein, ont été amenés, à la demande de différents organismes et écoles de formation, à coordonner, animer et conduire différentes actions de formation.

Dans un second temps, le projet de lancer et produire un module de formation en son nom propre s'est précisé : il s'agit de proposer aux responsables et acteurs d'archivage numérique à long terme, dans toutes les catégories d'institutions et d'entreprises évoquées plus haut, un discours organisationnel et technique commun, élaboré conjointement par des archivistes et des informaticiens et dispensés en commun à des archivistes, des informaticiens ainsi qu'aux producteurs d'informations.

4 CONCLUSION

PIN constitue un exemple unique en France de collaboration efficace entre archivistes et informaticiens qui illustre bien le fait que pérenniser une information numérique n'est pas l'affaire d'une seule personne ou d'un seul corps de métier mais nécessite la mise en commun de compétences complémentaires

5 REFERENCES

- [1] Huc Claude, Le Modèle de référence pour les systèmes ouverts d'archivage. L'étonnant parcours d'une norme unificatrice, Document Numérique, vol 4, n°3-4/2000 "l'archivage"
- [2] Catherine Dhérent, Les Archives électroniques, manuel pratique, La documentation française, 2002, disponible en ligne : <http://www.archivesdefrance.culture.gouv.fr/fr/archivistique/index.html>
- [3] Kenneth Thibodeau, Overview of technological approaches to digital preservation and challenges in coming years, in "The state of digital preservation : an international perspective", conference proceedings, council on library and information resources, Washington DC juillet 2002 <http://www.clir.org/pubs/reports/pub107/thibodeau.html>
- [4] Site Web du Groupe PIN ; <http://pin.cnes.fr>
- [5] Site Web d'Aristote : <http://www.aristote.asso.fr>
- [6] Site Web de la BNF ; <http://www.bnf.fr>

- [7] Site Web de la Direction des Archives de France ; <http://www.archivesdefrance.culture.gouv.fr/>
-