Huma-Num
la TGIR des humanités numériques

Les dossiers thématiques

#1

Les services de conservation de données proposés par Huma-Num

Dans un monde de la recherche de plus en plus numérique, la question de la conservation et de la préservation des données est devenue un enjeu majeur. Huma-Num propose aux équipes de recherche des solutions pour assurer la sauvegarde, la pérennité et le partage des données.

En savoir plus

TGIR Huma-Num
Pôle communication
190, avenue de France
75013 PARIS

huma-num.fr

INFORMER PARTAGER DIFFUSER

Ont collaboré à ce numéro

Gérald Foliot, CNRS - TGIR Huma-Num,
Olivier Rouchon, CINES,
Philippe Prat, CINES,
Michel Jacobson, Archives de France,
Françoise Banat-Berger, Archives de France,
et Nicolas Larrousse, CNRS - TGIR Huma-Num,
qui en a coordonné sa réalisation.

PRÉSENTATION

Nicolas Larrousse
Ingénieur de recherche
CNRS

Aujourd'hui, les données de la recherche sont produites nativement sous forme numérique ou proviennent de la numérisation de données analogiques. Le passage au numérique apporte un gain évident pour la transmission, la diffusion de ces informations et le travail collaboratif qui peut être effectué sur ces données. Mais paradoxalement, un objet numérique peut être plus fragile que son homologue du monde réel : en effet, une photo très abimée peut encore fournir de nombreuses informations, alors qu'un fichier informatique est totalement inutilisable à la moindre altération.

Pour éviter ces pertes, Huma-Num propose à la communauté des producteurs de données numériques en Sciences Humaines et Sociales, un ensemble de services pour le traitement, le stockage, la diffusion et la conservation de données numériques, regroupés sur « la Grille de services ». Les services de conservation se déclinent en stockage sécurisé, exposition de données et archivage à long terme.

Le premier volet de ce dossier présente le stockage sécurisé associé aux outils de traitement proposés par la « Grille de services ».

Le second volet présente Nakala, le nouveau service proposé pour exposer les données.

Enfin, le troisième volet est consacré au service d'archivage à long terme mis en place par la TGIR, qui s'appuie sur l'infrastructure et les compétences du Centre Informatique National de l'Enseignement Supérieur. Les problématiques soulevées par l'archivage à long terme sont ensuite mises en perspective par la vision des Archives de France sur le devenir des données de la recherche en format numérique.

Ce dossier thématique est une version complétée et mise à jour du guide initialement publié en mars 2013.

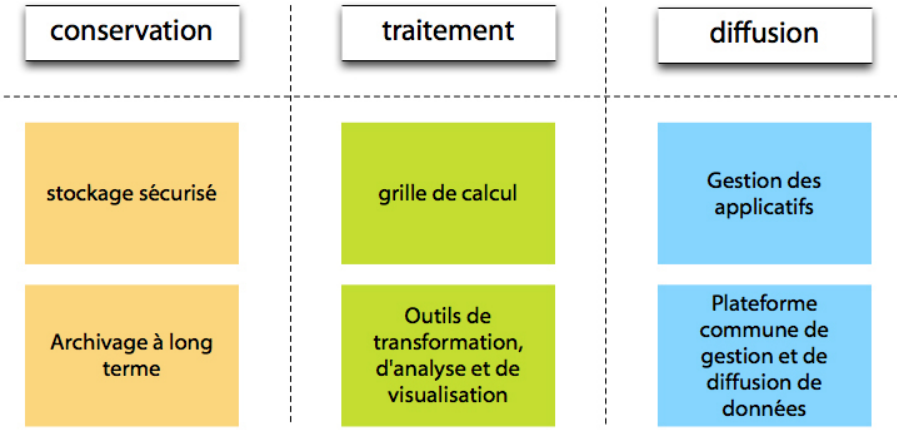
Juin 2014

COMPRENDRE LE STOCKAGE SÉCURISÉ

Gérald Foliot
TGIR Huma-Num

La grille de services de la TGIR Huma-Num s'organise selon deux dimensions : la première concerne le workflow (ou enchaînement de processus) de traitement des données : conservation/traitement/diffusion. La seconde est liée à la complexité de mise en œuvre. Toutes deux organisent les différents services de la grille représentés dans le schéma ci-contre.

Organisation de la grille par type de services et par niveaux



Dans cette présentation, nous voudrions nous arrêter plus particulièrement sur les problématiques de conservation et le sens opérationnel que nous donnons à ce terme. Si avant le numérique nous savions assez facilement conserver un document - il suffisait de poser un livre sur une étagère pour le garder - avec les documents numériques, les choses sont devenues plus éphémères. En effet, les supports sur lesquels se trouve l'information sont intrinsèquement fragiles, et les modalités d'accès et de restitution sont encore plus changeantes. La difficulté est de parvenir à avoir quelque chose de permanent dans un environnement qui ne l'est pas. Selon les usages que nous faisons d'un document ou d'une information, la conservation, c'est-à-dire l'action de garder quelque chose, révèle une diversité de pratiques. Celles-ci dépendent elles-mêmes des objectifs que l'on s'est fixés d'où une dimension de complexité qui dans l'ensemble des services proposés par Huma-Num va se traduire par trois niveaux : stocker, partager et archiver.

Le service de stockage sécurisé a pour fonction de garantir la conservation et l'intégrité de fichiers informatiques. Il ne permet pas de partager des fichiers, ni de les « pérenniser » de quelque manière que ce soit.

C'est un type de stockage que l'on dit sécurisé, car il met en œuvre deux protections particulières. La première consiste à procéder à une sauvegarde (ou backup) de fichiers informatiques d'un support à un autre (des bandes magnétiques dans la grande majorité des cas). On réalise également un déport de ces bandes de sauvegarde dans un autre bâtiment. La seconde consiste à mettre en oeuvre un mécanisme de vérification de l'intégrité de chaque copie (calcul de cheksum ou somme de tous les bits d'un message), et ce dans le but de détecter toute erreur de transmission.

Le service de stockage sécurisé proposé par Huma-Num s'inscrit dans l'infrastructure globale de stockage du Centre de Calcul de l'IN2P3. Celle-ci repose sur iRods pour le stockage et Tivoli Storage Manager (TSM) pour les robots de sauvegarde sur bandes magnétiques.

iRods est un outil de distribution de données permettant l'accès à des données réparties sur différents sites et sur des supports hétérogènes. C'est-à-dire qu'il permet, à partir d'un outil unique, de stocker des données sans se préoccuper de leurs emplacements réels sur tel ou tel équipement. C'est donc un système unifié d'accès à de larges surfaces de stockage (à l'heure actuelle cela représente 120 To de disponibles pour les SHS, facilement extensibles).

Une caractéristique peu connue – et particulièrement importante - de ce protocole d'échange d'information est l'intégration de la vérification des cheksums (somme de contrôle évoquée plus haut). En effet quand on transfère un fichier informatique via un réseau, rien ne permet de garantir que la communication réseau est correcte. Le transfert peut se terminer alors même qu'une partie des données transmises est erronée. Par exemple, il n'est pas rare de trouver sur internet des fichiers images ayant une bande blanche ou noire, car il y a eu une erreur dans le transfert du fichier. L'image reste lisible et valide informatiquement parlant, mais une partie de l'image a été transformée en bit de poids zéro.

La configuration du backup des serveurs iRods sur les robots de bandes dispose d'un temps de rétention fixé à un an. Durant un an, il est donc possible de retrouver sur les bandes un fichier qui aurait été effacé par erreur, ce qui représente donc une protection supplémentaire particulièrement importante.

Stockage sécurisé ne veut pas dire stockage crypté. En effet, les fichiers informatiques arrivent en clair sur les systèmes et sont conservés en clair sur les supports informatiques. Ce qui veut donc dire que si une personne peut intercepter une communication réseau ou s'il y a un vol de bandes magnétiques ou de serveurs au CC-IN2P3 (notons tout de même que les locaux sont gardiennés 24/24 et sous vidéo surveillance), il serait possible à un tiers d'accéder à ces fichiers. Pour se prémunir contre ce risque, il conviendrait de crypter les fichiers - ou à défaut de crypter la communication, via un tunnel VPN (virtual private network) par exemple. Le cryptage des fichiers implique cependant de pouvoir conserver de manière sûre la clé de cryptage elle-même, car si elle se est perdue, les fichiers ne seront plus lisibles par personne.

> La difficulté est de parvenir à avoir quelque chose de permanent dans un environnement qui ne l'est pas.

Le stockage sécurisé ne garantit pas non plus que les fichiers puissent être relus. Si le mécanisme de checksum permet de vérifier l'intégrité d'un fichier du point de vue informatique, cela ne veut pas dire qu'il existe encore un logiciel capable de relire ce format de fichier. Par exemple, il est tout à fait possible de stocker un fichier NouveauLivre.qxd ou DessinDeBatiment.max. Or, ces formats de fichier ne sont lisibles que par les logiciels Quark X'Press pour le premier et 3DS Max pour le second. Aucun autre logiciel n'est capable de lire ces formats particuliers. Rien ne garantit que dans 10 ans, le logiciel 3DS Max existera toujours ou qu'il pourra relire ses propres formats (par exemple la version actuelle de Microsoft Word est incapable de relire correctement un fichier Word antérieur à 1995). A ceci, il faudrait également ajouter la question du contenu même du fichier : pourra-t-on se souvenir dans 10 ans de ce que contient le fichier DessinDeBatiment ?

Afin de pouvoir répondre à ces problématiques particulières de préservation du contenu, il faut avoir recours à un niveau plus sophistiqué de conservation à travers l'archivage à long terme. Sa mise en œuvre est plus complexe et nécessite une description particulière. Cette question est l'objet d'un autre article dans ce dossier.

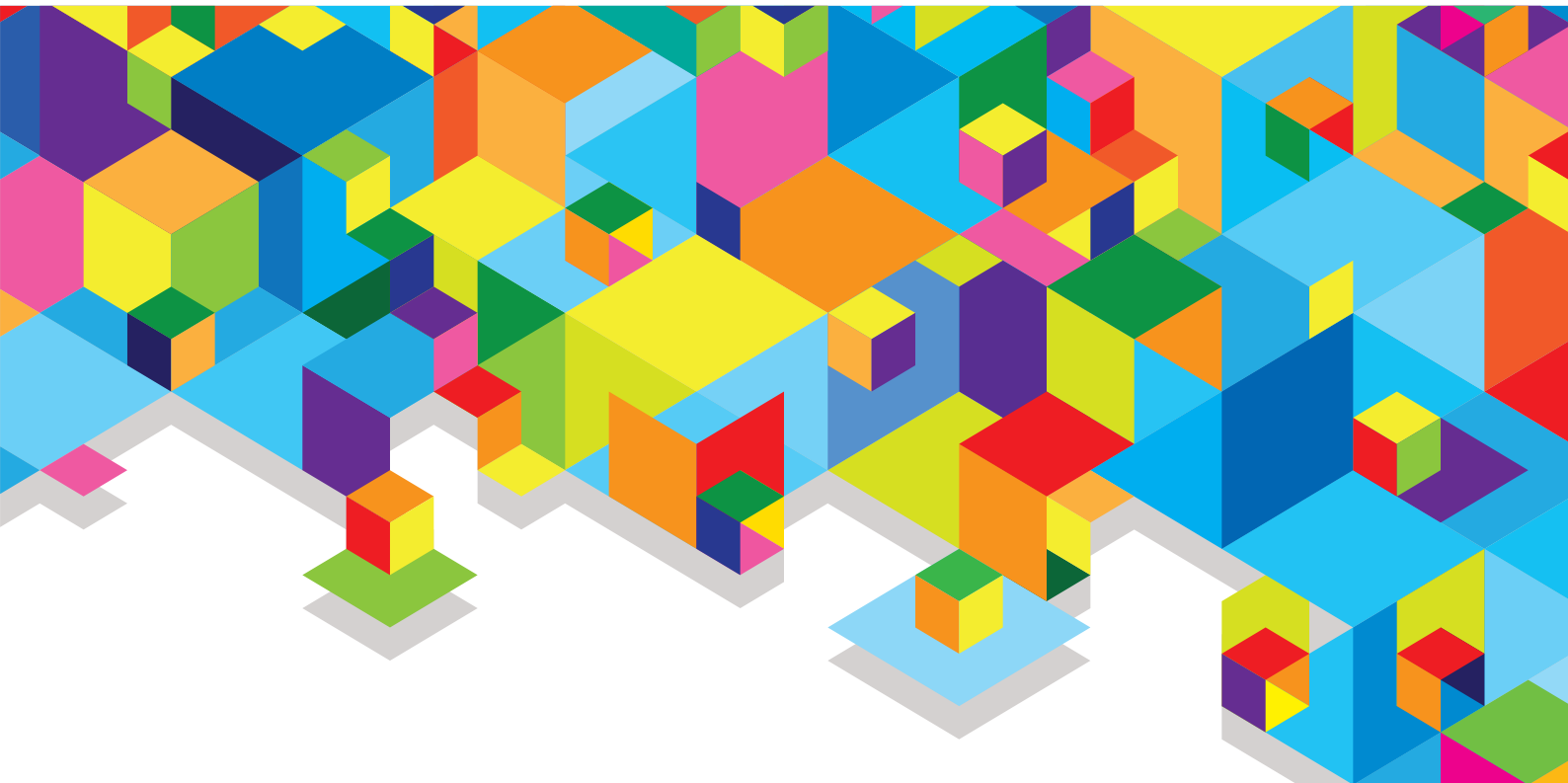
Sauvegarder le travail collaboratif : la solution de partage Sharedoc

S'il constitue une solution pour conserver une copie distante des données ou stocker de façon intermédiaire des fichiers avant archivage, le stockage sécurisé ne résout pas, dans un travail de recherche, le problème de l'exploitation des données à court et à moyen terme.

Il n'est pas rare en effet, à un moment donné de la recherche, d'avoir besoin de travailler à plusieurs sur des fichiers. C'est ce que l'on désigne généralement sous le terme de travail collaboratif, bien que cette dénomination réfère à des usages tout à fait différents pouvant aller d'un simple wiki, à l'édition simultanée et multi-utilisateurs d'un document.

Il semble difficile de proposer un outil capable de couvrir l'ensemble des pratiques de travail collaboratif. D'ailleurs, à notre connaissance, il n'existe aucun système couvrant tous les usages possibles dans ce domaine. Néanmoins, Huma-Num met à disposition un service de partage de fichiers qui utilise le stockage sécurisé : la plateforme Sharedoc. Le principe est de mettre à disposition des espaces de partage de fichiers accessibles via un simple navigateur ou via le protocole Webdav.

Si le protocole http est utilisé pour consulter des serveurs web, son extension WebDav permet, quant à elle, de procéder à l'opération inverse, c'est-à-dire de transférer, depuis son ordinateur, des fichiers vers un serveur distant afin de les partager. Ce protocole constitue une solution particulièrement efficace pour partager et manipuler des fichiers, comme on le ferait en local avec une simple clé USB. C'est une solution efficace pour bénéficier d'un stockage sécurisé des données.



NAKALA

Un service pour l'exposition de données

Nicolas Larrousse
TGIR Huma-Num

Après une phase de tests menée avec de futurs utilisateurs, la TGIR Huma-Num a lancé en juin 2014 NAKALA, nouveau service qui permet la mise à disposition des données et métadonnées de la recherche.

De quoi s'agit-il ?

Il s'agit d'un nouveau service qui s'inscrit dans un dispositif cohérent mis en place par Huma-Num autour des données numériques de la recherche. Sa mise en œuvre est liée au constat que de nombreux producteurs de données scientifiques ne disposent pas de l'infrastructure numérique nécessaire pour garantir un accès sécurisé, persistant et interopérable à leurs données.

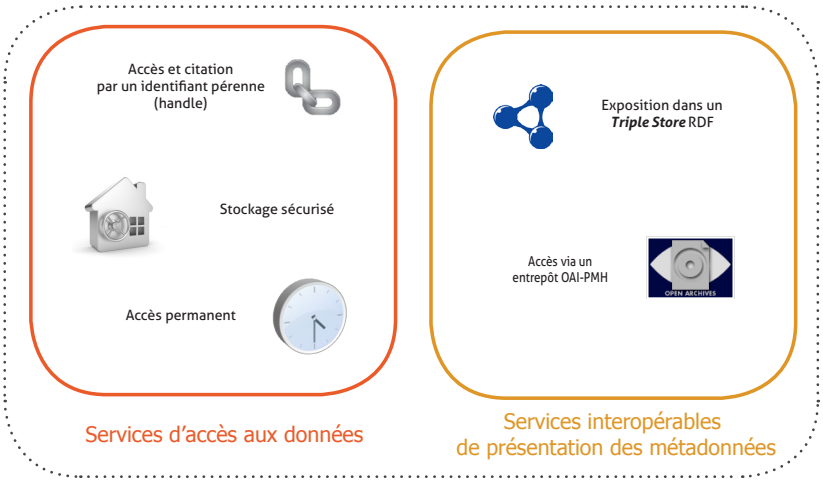
Nakala permet donc à des équipes de recherche, qui en font la demande, de déposer leurs données numériques (fichiers texte, son, images, vidéo) dans un entrepôt sécurisé qui assure à la fois l'accessibilité permanente aux données et leur citabilité dans le temps. Les technologies mises en oeuvre permettent de rendre interopérables les métadonnées, c'est-à-dire de pouvoir les connecter à d'autres entrepôts existants, et de les rendre moissonnables par des services spécialisés, ce qui permet d'en accroître la visibilité.

Avec ce nouveau service, les producteurs de données sont déchargés de la gestion purement technique et peuvent se consacrer à la valorisation scientifique de leurs données.

Dans le détail :

NAKALA propose deux grands types de services : des services d'accès aux données elles-mêmes et des services de présentation des métadonnées.

Les services proposés par NAKALA



Les services d'accès aux données permettent :

- d'associer un identifiant pérenne à chaque donnée

Les données peuvent ainsi être citées indépendamment de leur localisation réelle : la technologie proposée est basée sur l'attribution d'identifiants de type handle, qui possèdent un mécanisme d'adressage générique, sans lien avec l'institution qui porte le service. C'est ce qui rend pérenne l'accès à la donnée, même en cas de changement ou d'évolution de l'institution qui porte le service.

- d'assurer un accès permanent aux données et d'assurer leur sécurité

Les données sont stockées sur l'infrastructure gérée par la TGIR et sont ainsi sauvegardées régulièrement. La TGIR possède ses propres serveurs sécurisés au sein du centre de calcul de l'IN2P3-CNRS, partenaire des SHS depuis 2003.

Les services de présentation des métadonnées permettent :

- Une exposition basée sur les technologies du Web de données

NAKALA dispose d'un entrepôt RDF (Resource Description Framework) de type Triple Store, qui permet de partager les informations en utilisant les principes, méthodes et technologies du Web de données. L'utilisation de ces technologies standardisées permet de bâtir des applications de valorisation de ces données, par exemple des cartes interactives.

Il sera aussi envisageable de les connecter à d'autres entrepôts existants comme DBPedia (http://fr.dbpedia.org/), la version adaptée au Web de données de Wikipedia.

- Un accès interopérable utilisant le protocole OAI-PMH

NAKALA permet à chaque producteur de données de disposer de son propre entrepôt OAI-PMH, le standard d'interopérabilité des métadonnées utilisé depuis 1999 dans le domaine de la recherche et de l'enseignement supérieur.

Les données peuvent ainsi être « moissonnées » par des services spécialisés, comme par exemple ISIDORE, mais aussi Europeana, Gallica, etc., ce qui permettra d'en accroître la visibilité.

En pratique les métadonnées descriptives seront exprimées classiquement, en utilisant le format standard Dublin Core étendu (dcterms).

Ce que ne propose pas NAKALA...

NAKALA ne propose pas de moteur de recherche, ni de dispositif d'enrichissement de données, ni de site Web d'éditorialisation scientifique.

En revanche, il est prévu que les données puissent facilement être intégrées dans un outil d'éditorialisation (par exemple un CMS, un carnet d'hypothèse.org, une revue électronique, etc). De même, il est tout à fait possible d'utiliser ISIDORE pour enrichir ces données avec des référentiels scientifiques.

Comment fonctionne NAKALA ?

Les services proposés par NAKALA sont simples et bien identifiés, les outils de gestion associés le sont également.

L'interface de gestion permet d'avoir une vue d'ensemble de son espace et de gérer les données et les collections de manière simple et aisée. Cette interface permet également de déposer des données si ces dépôts sont de taille modeste.

Pour insérer d'importants volumes de données, NAKALA propose un dispositif de traitement par lots. Cet outil est développé en langage Java et peut ainsi fonctionner sur les principales plateformes existantes.

Comment utiliser NAKALA ?

Tout comme pour les autres services de la TGIR, il suffit de soumettre une demande à cogrid@huma-num.fr

http://www.nakala.fr

L'ARCHIVAGE À LONG TERME

Entretien avec Olivier Rouchon
CINES

Olivier Rouchon[1], vous travaillez au département archivage et diffusion au Centre Informatique National de l'Enseignement Supérieur, basé à Montpellier. Pouvez-vous nous définir en quelques mots ce qu'est l'archivage à long terme ?

Olivier Rouchon : L'archivage à long terme des documents électroniques consiste à conserver le fond et la forme d'un document et l'information qu'il contient – c'est-à-dire dans son aspect physique comme dans son aspect intellectuel – sur le très long terme, de manière à pouvoir le rendre accessible et compréhensible. Ce concept diffère de la sauvegarde sécurisée dans la mesure où il ne s'agit pas seulement de préserver le train de bits d'un fichier ou de migrer des fichiers pour faire de la place sur un disque, et qu'il ne constitue pas non plus l'ultime étape du stockage des données avant l'oubli ou la perte définitive.

Aujourd'hui, la plupart des fichiers informatiques de plus de dix ans sont quasiment illisibles, conséquence de plusieurs facteurs tels que l'obsolescence des formats de fichier, la détérioration des supports physiques, la disparition des logiciels ou des matériels de lecture, ou encore la perte d'information sur le contenu. Il est donc essentiel de procéder à des choix, de mettre en place des processus et des pratiques afin de réduire ces risques.

C'est à cette tâche que le CINES travaille depuis 2004 : afin de mettre en place un service pour l'archivage à long terme du patrimoine scientifique, qui permet à tout organisme produisant ou collectant en grande quantité des documents électroniques, dont le contenu possède une valeur patrimoniale avérée pour la communauté scientifique, d'initier un projet d'archives dans le respect du contexte législatif archivistique français.

Quels sont les types de données actuellement archivées au CINES?

Actuellement, plusieurs projets sont en exploitation sur la plateforme PAC du CINES. Le premier – le plus ancien dans la chronologie des projets déployés – concerne l'archivage à long terme de thèses collectées par l'ABES – Agence Bibliographique de l'Enseignement Supérieur – par l'intermédiaire de l'outil STAR, initié suite à l'arrêté du 7 août 2006 relatif aux modalités de dépôt, de signalement, de reproduction, de diffusion et de conservation des thèses ou des travaux présentés en soutenance en vue d'un doctorat. Cette disposition prévoit qu'un doctorant qui doit déposer son mémoire de thèse dans la bibliothèque de l'Université dans laquelle il va effectuer sa soutenance, peut le faire soit sous forme papier, soit sous forme électronique. Dans ce deuxième cas, la bibliothèque va déposer les documents numériques constituant le mémoire de thèse à l'ABES grâce à l'application STAR, puis ajouter des informations qualifiant la thèse, sous forme de métadonnées. Après plusieurs phases de validation, la thèse sera archivée au CINES. L'intégration de la plateforme PAC à STAR a été réalisée en 2007, et le projet monte en charge au rythme de l'implantation de l'application STAR dans les universités.

Les autres projets concernent pour la plupart l'archivage de documents scannés dans le cadre de programmes de numérisation de masse initiés par des bibliothèques. Parmi ceux-ci, on peut citer le projet Persée, qui vise à valoriser et à préserver des collections rétrospectives originales au format papier, dont certaines ont près de cent ans. Les revues sont numérisées et traitées de manière automatique, les collections décrites. Les revues ainsi traitées sont ensuite mise en ligne via le portail Persée. Cette chaîne de numérisation est ensuite interfacée avec la plateforme PAC, et les documents électroniques créés sont déposés au CINES pour

1 Olivier Rouchon est actuellement responsable du département Calcul intensif au CINES.

une préservation à long terme. Le principe de ce dispositif d'intégration a été repris pour d'autres grands programmes tels que la numérisation de fonds anciens en droit par la bibliothèque Cujas, la numérisation de collections thématiques de documents sur les neurosciences, la biologie, la géologie et la physique-chimie par la bibliothèque universitaire Pierre et Marie Curie, ou encore la numérisation des documents sur la médecine de la bibliothèque inter-universitaire de Santé.

Enfin, d'autres projets traitent de l'archivage des documents textuels versés dans la plateforme HAL ou de corpus oraux dans le cadre d'un projet pilote réalisé pour le TGE Adonis[2].

Quelle est la nature du partenariat proposé avec Huma-Num ?

Dans le cadre de ses missions de conservation des données numériques, Huma-Num s'appuie sur les infrastructures et les compétences du CINES (suite au rapport Olof Bärring, 2008). La convention entre Huma-Num et le CINES, récemment renouvelée pour une période de cinq ans, permet à la TGIR de proposer ce service à des conditions avantageuses. Dans ce cadre, le CINES réalise des copies d'archives en interne et les réplique au centre de calcul de l'IN2P3 ainsi qu'à la Bibliothèque nationale de France.

Le projet pilote réalisé initialement pour le TGE Adonis a permis de développer et de mettre en exploitation une solution mutualisée de pérennisation et d'accès aux données orales numériques. L'objectif principal était de valider le schéma retenu pour les données orales, tant sur le plan des fonctionnalités d'ensemble que sur le plan de la répartition des tâches et des responsabilités entre les acteurs. Ceci afin de pouvoir étendre ce schéma aux autres données créées, gérées et utilisées par différents producteurs de ressources numériques en Sciences Humaines et Sociales.

Cette infrastructure est désormais opérationnelle, et des discussions sont en cours pour intégrer de nouvelles thématiques au sein du dispositif.

Comment s'articule la collaboration entre Huma-Num et le CINES ? Quels sont les rôles des deux parties ?

Huma-Num est le maître d'ouvrage de la plateforme qui est proposée aux producteurs de données numériques. Elle s'appuie sur le CINES, pour l'expertise et les services de préservation à long terme, et sur le Centre de Calcul de l'IN2P3, Institut national de Physique Nucléaire et de Physique des Particules, pour la partie accès à l'information.

La TGIR Huma-Num est le point de contact unique entre la communauté productrice de documents numériques et le CINES. Elle recense et accompagne les services versants potentiellement intéressés par la mise en place d'un projet de conservation à long terme de documents numériques.

Comment se déroule la mise en place d'un projet d'archivage ?

Les services versants doivent remplir quelques conditions essentielles, parmi lesquelles :

- Appartenir à un organisme dont la mission est liée à l'Enseignement Supérieur ou la Recherche, collecter ou produire un ensemble

2 Le projet pilote a été réalisé par le TGE Adonis, de 2008 à 2009 Depuis mars 2013 le TGE Adonis et IR-Corpus ont fusionné. Huma-Num est l'unité issue de cette fusion.

d'objets numériques présentant un intérêt scientifique ou pédagogique au niveau national.

- Avoir une stratégie et un besoin pour un archivage patrimonial – à ne pas confondre avec une sauvegarde ou un stockage sécurisé de données – pour cette collection d'objets numériques, éventuellement un plan de classement.

- Vérifier le contexte légal de production des objets numériques : droits de propriété intellectuelle et droits d'auteur sur les documents à archiver.

- Avoir et/ou maintenir un référentiel d'informations décrivant la collection d'objets numériques, à partir duquel des métadonnées génériques décrivant chaque archive et les métadonnées métier spécifiques pourront être extraites, et où les informations transmises par le CINES pourront être stockées.

L'organisme souhaitant initier un tel projet effectue une demande qui est examinée par Huma-Num, avant son acceptation puis son lancement.

Le projet démarre par une phase dite préliminaire durant laquelle des informations générales sont échangées, telles que l'identification des informations à pérenniser, la liste des données qui seront transmises au CINES (format, taille, nombre…), accompagnées de leurs métadonnées, l'analyse de tous les aspects de faisabilité (sécurité, aspects légaux, coûts et risques…) ou encore l'évaluation de la volumétrie et des ressources requises.

Une seconde phase dite de définition permet ensuite de déterminer plus en détail les points qui feront l'objet d'une contractualisation des échanges, ce qui inclut la définition précise des objets à transférer, des précisions sur les termes et conditions du protocole de transfert (questions sur les restrictions d'accès, la communicabilité au public), la planification des transferts physiques ou encore la formation du personnel du service versant à l'utilisation du système d'archivage.

La TGIR Huma-Num coordonne ces deux phases avec l'appui du CINES pour son expertise archivistique et technique. Lorsqu'elles sont terminées, et les ajustements applicatifs éventuels effectués, les versements d'archives peuvent commencer.

Pour conclure, nous vous proposons de revenir sur un point qui semble essentiel et sur lequel nous devons insister : l'importance des métadonnées.

Les métadonnées sont les données qui permettent de préserver les informations décrivant les objets numériques. On parle alors d'informations de pérennisation et d'informations de représentation, qui sont des concepts définis dans le modèle OAIS. Il s'agit donc de porter un focus particulier sur leur niveau de qualité. Plusieurs types de contrôle peuvent être effectués, lesquels pourront être facilités par l'adoption de standards : en effet, les métadonnées normalisées sont décrites dans un référentiel qu'il est facile d'utiliser comme base de vérification pour s'assurer du niveau de qualité. L'éventail de tels jeux de métadonnées est très large, du plus générique pour la description de ressources numériques, tel que le Dublin Core, aux plus spécifiques à un domaine précis, comme les données géographiques, en passant par les métadonnées techniques de préservation comme PREMIS ou METS. Leur intérêt est donc primordial, car ce sont sur elles que s'appuient la recherche et la compréhension des objets numériques.

La certification Data Seal of Approval. Une certification pour les centres d'archivage ?

Le Data Seal of Approval est une accréditation attribuée aux centres d'archivage numérique à long terme, qui mettent en place des procédures d'assurance qualité afin de garantir l'accessibilité et l'intelligibilité des informations qui leur sont confiées. Cette accréditation a été initiée par quelques institutions engagées dans la problématique de l'archivage à long terme des données de la recherche. En attribuant ce sceau de confiance, elles ont non seulement souhaité garantir la durabilité des informations, mais aussi promouvoir les bonnes pratiques pour l'archivage numérique en général.

La certification est l'aboutissement de la consolidation d'un organisme ou d'un service, dont elle matérialise la reconnaissance de la qualité et du professionnalisme. Elle constitue donc un moyen d'instaurer des relations de confiance avec les communautés d'utilisateurs, et peut devenir, en outre, un levier pour obtenir des budgets auprès des organismes de tutelle. Le service d'archivage du CINES a depuis longtemps pris des initiatives autour de l'accréditation de ses activités de préservation numérique, lesquelles se sont matérialisées ces dernières années par l'obtention de l'agrément du Service Interministériel des Archives de France pour la conservation d'archives électroniques publiques intermédiaires, ou encore par l'obtention du Data Seal of Approval en 2010.

L'ARCHIVAGE NUMÉRIQUE : DE L'IMPORTANCE DES FORMATS

Philippe Prat
CINES

Le choix des formats d'archivage numérique est une question centrale dans la mise en place d'un processus d'archivage à long terme. En effet, de ces différents choix dépendent la pérennité et l'accessibilité aux contenus ainsi archivés.

Rencontre avec Philippe Prat.

TGIR Huma-Num : Vous assurez au CINES, la veille technologique autour des formats et vous avez mis en place en amont du processus d'archivage un processus de contrôle des fichiers numériques. Pouvez-vous nous en expliquer les raisons et les enjeux ?

Philippe Prat : Disons que l'archivage est à l'information ce que la sauvegarde est à la donnée. L'archivage doit donc se préoccuper, au-delà de la simple sauvegarde d'un train de bits, que l'information que l'on conserve soit exploitable dans le futur.

Nous savons d'expérience que les formats informatiques évoluent dans le temps et que la notion de compatibilité ascendante d'une « chaîne » de format montre parfois vite ses limites. La gestion du risque d'obsolescence des formats des fichiers que l'on archive est donc une préoccupation majeure. Se prémunir de ce risque reste encore un challenge dans la mesure où les méthodes et les contrôles à mettre en œuvre ne font pas partie d'un standard ou d'une norme mais découlent souvent d'un processus interne. En outre, chaque nouveau format à archiver peut induire son propre lot de complexités.

Le phénomène de l'obsolescence technologique fait que l'on ignore si un format sera toujours utilisé sur le très long terme. S'il ne l'est plus, les outils pour l'exploiter ont également de grandes chances de ne plus exister. Nous serons alors dans l'incapacité d'exploiter ce fichier et aurons échoué dans notre mission d'archivage. Par contre, si nous effectuons une veille technologique sur le format, nous serons capables d'observer que ce format commence à ne plus être utilisé et nous pourrons alors enclencher une procédure de migration vers un nouveau format.

Envisager cette migration consiste d'abord à identifier l'information que l'on souhaite préserver dans le fichier, à trouver un format cible capable de la contenir et à créer un nouveau fichier à partir de l'ancien. Pour se donner la garantie dès le départ de pouvoir migrer un format, il faut être certain que l'organisation des bits dans le fichier correspond bien à la spécification du format. Car si ce n'est pas le cas, la transformation du fichier source vers le fichier au nouveau format peut échouer. C'est principalement pour cette raison que l'on contrôle de manière forte le fichier.

Quelle méthodologie avez-vous mise en place pour contrôler l'éligibilité des fichiers à l'archivage à long terme ?

La spécification d'un format est un document souvent très complexe (pouvant aller jusqu'à des milliers de pages) qui décrit comment les données (les bits) doivent être organisées dans le fichier.

Garantir la validité stricte du format par rapport à sa spécification demanderait donc des moyens de développement gigantesques et

serait techniquement très difficile. C'est la raison pour laquelle nous avons mis en place au CINES des moyens pour contrôler au mieux le format et pour s'assurer qu'il soit « migrable ». Mais cela nécessite une stratégie, voire de l'imagination, pour chaque type de format, car comme je le disais, il n'existe pas un validateur officiel par format. Quand nous recevons, par exemple, une vidéo ou une image au CINES, nous effectuons virtuellement une transformation du fichier vers un autre format. Si cette transformation réussit, nous disons que nous avons un niveau de confiance suffisant pour garantir le succès d'une migration future.

Afin que chacun puisse vérifier le format des fichiers qu'il utilise, le CINES met à disposition sur son site internet un outil en ligne qui met en œuvre cette méthodologie (voir encadré sur l'outil FACILE).

Quels sont les critères retenus pour rendre éligibles certains formats ?

Le format doit être libre de droits et sa spécification accessible. Ce sont deux critères essentiels pour pouvoir – si cela s'avère nécessaire malgré toutes les précautions prises en amont – développer un outil pour exploiter ou transformer le format. Il faut également avoir une « confiance » suffisante à propos du format, et s'assurer que l'on a un grand nombre d'utilisateurs et d'outils qui gravitent autour. C'est pour cela que nous sélectionnons uniquement des formats largement utilisés.

Comment gérer-vous justement cette profusion de formats utilisés ?

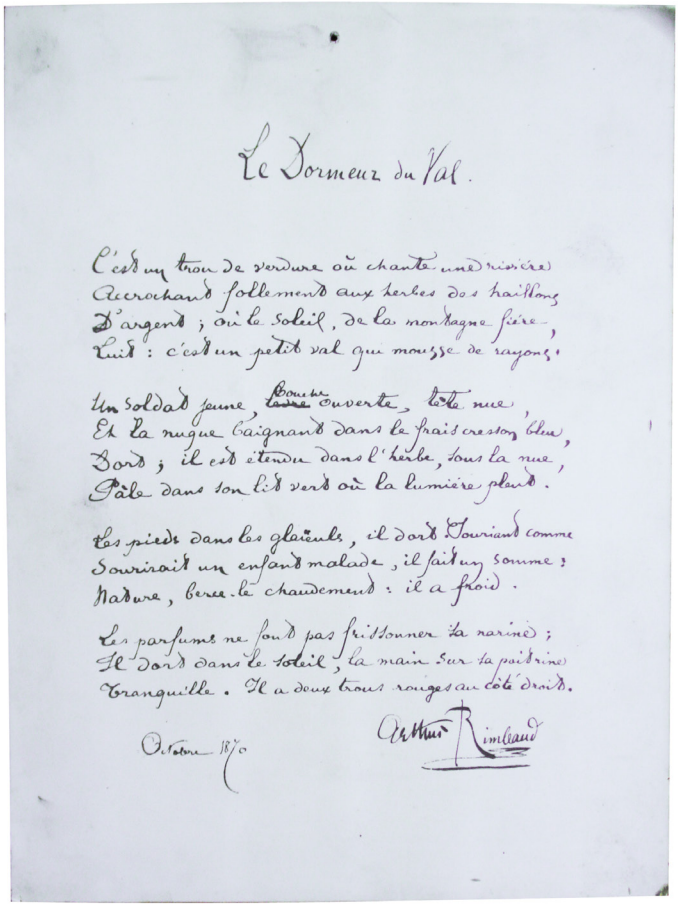
Nous les avons classés en plusieurs listes. Nous avons ainsi des formats « à l'étude », sur lesquels nous travaillons pour s'assurer qu'ils répondent à nos critères d'exigence. Nous disposons également d'une liste de formats « potentiellement archivables », c'est-à-dire qui répondent positivement à nos critères mais qui ne sont pas effectivement mis en œuvre sur notre plateforme, soit parce que la nécessité d'archiver ce format ne s'est pas imposée, soit pour des raisons évidentes de maintenance qui nous contraignent de limiter la liste.

Enfin, bien évidemment des formats « archivables » qui sont ceux effectivement acceptés par notre plateforme. Nous avons encore deux autres catégories : les formats en « voie d'obsolescence » qui sont les formats candidats à la migration vers un nouveau format. Et enfin des formats « obsolètes » qui ne sont plus exploitables.

Pour conclure, en dehors de ces questions techniques complexes, le choix d'un format ne dépend-il pas aussi de l'information que l'on souhaite transmettre ?

En effet. C'est d'ailleurs un point sur lequel je voulais revenir. Deux notions me tiennent à cœur : celle d'information et celle de donnée. Une information c'est l'interprétation d'une donnée. Une donnée devient donc une information lorsqu'on lui associe un sens. Une donnée n'a aucune signification en soi. Elle peut se présenter sous différentes formes, et peut contenir une multitude d'informations. C'est pour cela qu'il est important dans une perspective d'archivage d'identifier dans un document les informations que l'on souhaite transmettre.

Prenons un exemple. Imaginons que nous devons migrer le format de l'image d'un texte manuscrit. Nous pouvons retenir facilement deux formats cible : soit un fichier texte au format UTF-8, soit une autre image dans un nouveau format. Si nous ne savons pas quelle information il faut conserver, nous ne pourrons pas en toute sécurité choisir un format cible. En choisissant le format texte, nous pourrions être



Le Dormeur du Val.

C'est un trou de verdure où chante une rivière
Accrochant follement aux herbes des haillons
D'argent ; où le soleil, de la montagne fière,
Luit : c'est un petit val qui mousse de rayons.

Un soldat jeune, bouche ouverte, tête nue,
Et la nuque baignant dans le frais cresson bleu,
Dort ; il est étendu dans l'herbe, sous la nue,
Pâle dans son lit vert où la lumière pleut.

Les pieds dans les glaïeuls, il dort. Souriant comme
Sourirait un enfant malade, il fait un somme :
Nature, berce-le chaudement : il a froid.

Les parfums ne font pas frissonner sa narine ;
Il dort dans le soleil, la main sur sa poitrine
Tranquille. Il a deux trous rouges au côté droit.

Octobre 1870

Arthur Rimbaud

Photo manuscrit Le dormeur du Val, poème de Rimbaud (octobre 1870)

confrontés au mécontentement des personnes qui auraient souhaité conserver la forme de l'écriture originale. Nous aurons perdu cette information même si nous avons conservé la signification du texte qui, par ailleurs, ne peut avoir que peu d'intérêt. En revanche, si nous choisissons un format image en nuance de gris pour gagner de la place, nous pourrions faire face au mécontentement des personnes qui auraient préféré conserver la couleur du papier ou de l'encre, avec un spectre de couleur précis pour pouvoir en faire l'étude.

On peut donc dire pour conclure que la gestion des formats dans un contexte d'archivage reste assez complexe et doit être bien réfléchie. Improviser à ce niveau est impossible, car les conséquences sont irréversibles. Il demeure impératif de savoir identifier les informations à transmettre et de s'assurer qu'à tout moment elles soient compréhensibles et exploitables par les ayants droit. C'est une tâche difficile, je pense même que c'est un métier, peut-être celui d'archiviste, je ne sais pas ! Je constate seulement que cette compétence est trop rare, alors que les problématiques que nous soulevons devraient d'après moi être mieux prises en compte au niveau de la gestion documentaire ou du moins des solutions d'archivage.

Vérifier vos formats, c'est FACILE !

Le CINES a développé un logiciel pour contrôler les formats et notamment ceux qui sont éligibles à un archivage sur la plateforme PAC. Libre d'accès, simple d'utilisation cet outil vous permet de tester vos fichiers selon le même processus de validation que la plateforme.

Indiquez le fichier que vous souhaitez contrôler, déclarez son format. Les outils de contrôle feront le reste ! Une analyse de votre fichier sera effectuée et une réponse vous sera adressée.

Pour tester : http://facile.cines.fr
Plus d'information : http://www.cines.fr

QU'EST-CE QU'UNE ARCHIVE SCIENTIFIQUE ? LE POINT DE VUE DES ARCHIVES DE FRANCE.

Michel Jacobson
Archives de France

Les archives

Les archives ont une définition légale bien loin de l'image populaire véhiculée (papiers jaunis et cartons poussiéreux entassés dans les caves et les greniers) qui rapproche fortement ce terme des appellations « documents », « données » ou « informations » que l'on trouve par ailleurs. Les archives sont définies dans le Code du patrimoine comme « *[l]'ensemble des documents, quels que soient leur date, leur lieu de conservation, leur forme et leur support, produits ou reçus par toute personne physique ou morale et par tout service ou organisme public ou privé dans l'exercice de leur activité.* »[1]. Une telle définition signifie que la notion d'archives n'est pas liée à la nature du support utilisé. Ainsi, dans le Code, les archives numériques ne sont pas traitées différemment des archives papier. Il n'y a pas de distinction fonctionnelle entre un courrier sur papier en-tête avec une signature manuscrite et son équivalent numérique (fichier bureautique, courriel, etc.). Les deux supports peuvent porter le même contenu d'information, voire même la même valeur engageante pour celui qui l'a produite.

Une autre caractéristique des archives dans le Code est donnée par l'emploi du terme « document » que l'on pourrait définir comme un ensemble d'informations regroupées pour un usage particulier. Les archives renvoient donc à la fois à des informations et à des usages.

Les archives ne sont pas définies par leur domaine et les couvrent donc tous (juridique, administratif, scientifique, culturel, etc.). Elles peuvent contenir toutes sortes d'informations et revêtir de multiples formes (images, données numériques, textes, enregistrements sonores ou vidéo, plans, etc.). Par exemple, les données de l'état civil, les informations cadastrales, les minutes des jugements sont autant d'exemples d'informations conservées dans les services publics d'archives. Les factures, les bulletins de paye, les photos de famille sont également des archives (privées) que l'on trouve chez le particulier. Les données des chercheurs ou les données de la recherche constituent elles aussi des archives, qu'il s'agisse par exemple de cahiers de terrain, de relevés de mesures ou de documents contractuels.

Les archives publiques sont, toujours d'après le Code du patrimoine, « *Les documents qui procèdent de l'activité, dans le cadre de leur mission de service public, de l'État, des collectivités territoriales, des établissements publics et des autres personnes morales de droit public ou des personnes de droit privé chargées d'une telle mission...* »[2]. La production des agents de l'État, dans le cadre de leurs missions, relève donc typiquement, d'après le Code du patrimoine, du statut d'archives publiques. Que l'agent soit un juge, un officier d'état civil ou un chercheur, ils travaillent tous pour l'État et leur production ou les documents qu'ils reçoivent dans le cadre de leurs missions relèvent du domaine public mobilier. Les archives publiques sont imprescriptibles. Elles ne peuvent devenir privées, mais il est possible pour des archives privées d'entrer dans les collections publiques par des acquisitions, des dons, des legs, etc.

1 Code du patrimoine, livre II, article L. 211-1

2 Code du patrimoine, livre II, article L. 211-4

Le cycle de vie des archives

L'appellation « archives » est donnée aux documents dès leur création ou dès qu'ils sont reçus. La science archivistique distingue plusieurs âges dans le cycle de vie des archives. Les deux premiers âges dits « courant » puis « intermédiaire », commencent à la création ou la réception de l'archive chez le producteur et se termine lorsque le producteur n'en a plus l'usage. Cet usage couvre, à l'âge courant, principalement des besoins de fonctionnement puis, à l'âge intermédiaire, plutôt des besoins juridiques ou de bonne gestion. Durant ces premiers âges, le producteur est responsable des archives et les conserve de manière à pouvoir s'en servir si nécessaire. Le dernier âge, dit « définitif », n'est atteint que par une partie des archives, celle qui, au-delà de l'intérêt pour lequel elle a été constituée, revêt un intérêt historique, scientifique ou statistique. Ces archives (du moins celles qui sont considérées comme archives publiques), sont alors prises en charge dans des services publics d'archives pour conservation, traitement et mise à disposition suivant les délais de libre communicabilité fixés par le Code du patrimoine, sans plus de limite de temps. La responsabilité de leur conservation et communication et les coûts et obligations associés à celles-ci ne sont plus alors à la charge du producteur, mais à la charge du service d'archives.

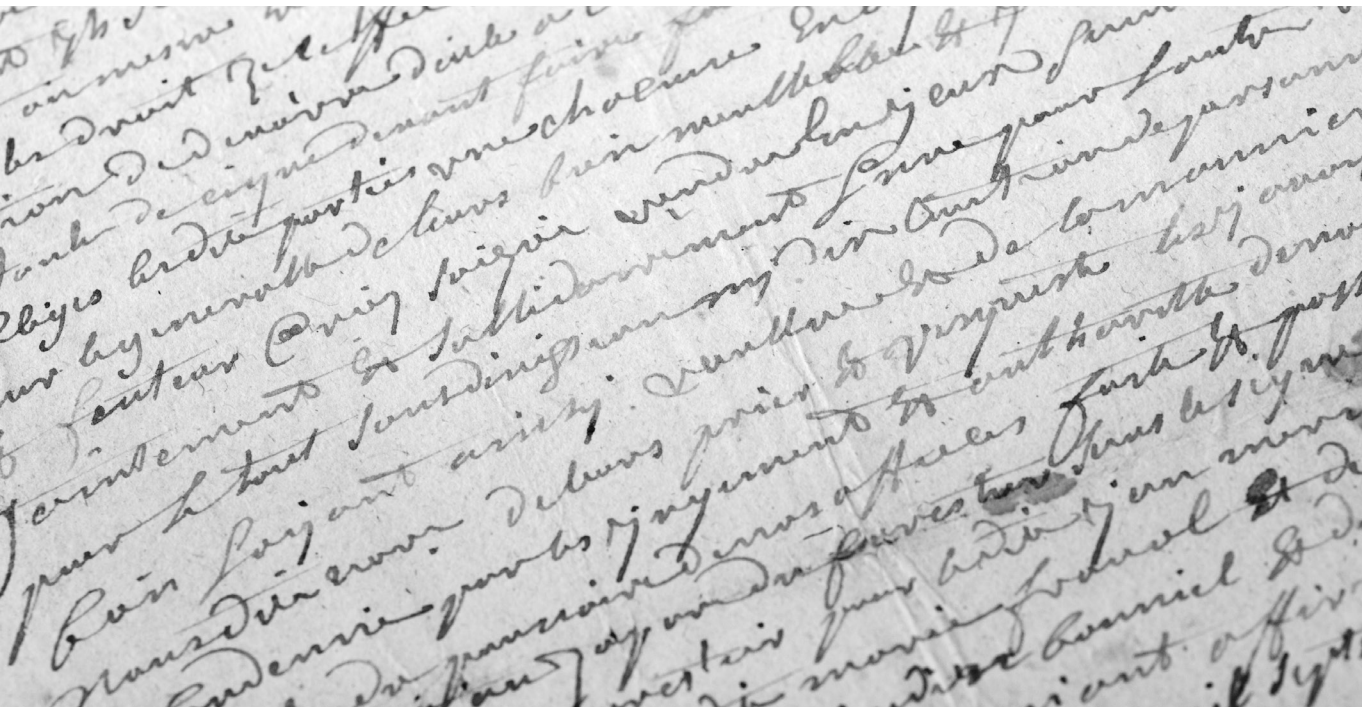
Cette théorie des trois âges peut sembler difficile à mettre en œuvre dans le cadre des données de la recherche, car le principe même de la démarche scientifique est qu'elle fonctionne par accumulation d'observations, de théories, de modèles et de critiques de ceux-ci. Il est donc parfois complexe de pouvoir déterminer une durée d'utilisation des données au-delà de laquelle l'organisation n'en aurait plus l'usage. La théorie des trois âges doit donc aussi être regardée de manière pragmatique et économique. Il est en effet rare qu'un producteur d'archives publiques ait la capacité de conserver ses propres archives sur le très long terme, car cela représente un coût non négligeable et demande des compétences qui ne sont normalement pas au cœur de son métier. Le versement dans les services d'archivage définitif représente donc souvent une opportunité pour les services et leur offre un moyen de lutter contre la perte, l'oubli et de s'ouvrir à d'autres communautés.

L'institution Archives en France

L'organisation institutionnelle de l'administration des archives de l'État date de la période révolutionnaire. En 1790, l'Assemblée constituante crée les Archives nationales en lui fixant comme missions la collecte, la conservation et l'organisation du libre accès des archives de la Nation auprès des citoyens. Ce dispositif sera complété peu de temps après par la création de services d'archives au sein de chaque département. On peut percevoir cette institution comme assurant la transparence des dépenses et plus globalement des activités de l'État, ou encore comme contribuant à une mémoire de cette activité et permettant de lutter contre l'oubli ou la réécriture des faits.

L'administration des Archives assure un double rôle :

- elle exerce le contrôle scientifique et technique (CST) sur les archives à l'âge courant et intermédiaire, alors qu'elles sont encore de la responsabilité du producteur. Ce CST est mis en œuvre à travers un certain nombre de dispositifs et d'actions dont la validation des tableaux de gestion (qui fixent par catégories de documents, les durées d'utilité administrative et leur sort final), la validation des plans de classement, les visas d'éliminations d'archives soumises par les services, l'instruction des demandes de dérogation pour la communication d'archives non encore librement accessibles, le contrôle des locaux de conservation, le contrôle sur les archives confiées à des prestataires externes agréés, etc. Il ne faut pas voir ce CST comme un instrument coercitif, même s'il peut prendre parfois cette forme pour des besoins de protection, mais comme une opportunité de mieux gérer son information en faisant appel à des professionnels. À ce titre, l'administration des Archives organise de la veille sur un certain nombre de points sensibles pour la pérennisation de l'information numérique comme les formats, les supports, les outils de gestion. Elle est active dans le domaine de la normalisation et de la réglementation ;
- le deuxième rôle des Archives est la prise en charge de la conservation sécurisée des archives définitives, leur traitement, leur accès et leur valorisation. Il s'agit cette fois d'une « compétence régalienne » et non plus d'un « service fonctionnel ». Un réseau d'opérateurs sur le territoire assure cette mission. Il s'agit des Archives nationales, des Archives départementales, régionales, municipales, etc. C'est ainsi que les archives définitives des établissements publics nationaux de type CNRS, et donc des chercheurs de ces établissements, sont versées, à l'issue de leur période d'utilité, au SCN (Service à Compétence Nationale) des Archives nationales.



Les archives et le droit

Les archives relèvent d'un régime particulier de communication, qui vise principalement à permettre aux citoyens d'y avoir accès. Il s'agit bien du citoyen et non pas uniquement de l'historien ou du chercheur, même si ces derniers font partie des usagers assidus des services d'archives. Le fait que certaines archives puissent également relever d'un régime de protection du droit d'auteur ne les empêchent nullement d'être des archives publiques. Dans le régime de communication particulier qui s'applique aux archives publiques, celles-ci sont réputées librement communicables de plein droit. Par dérogation à ce principe, certaines catégories d'archives, en fonction de la nature des informations qu'elles contiennent, peuvent n'être communicables qu'au terme d'un délai déterminé. Par exemple, lorsque la nature de l'information peut porter atteinte au secret médical, à la sécurité des personnes, ou à la vie privée. Cependant, quel que soit le type d'information, il est toujours possible de demander à l'administration des Archives une autorisation de consultation avant l'expiration du délai qui lui est associé. Cette autorisation pourra éventuellement être donnée par le directeur chargé des Archives de France au demandeur après avis du producteur de l'information. Pour ce qui est des archives privées, le service d'archives dépositaire se doit de respecter les conditions émises par les propriétaires.

Le Code du patrimoine a ouvert depuis 2009 la possibilité, pour un producteur, de confier ses archives publiques à un tiers agréé pour la période courante et intermédiaire, la période définitive restant de la compétence des services publics d'archives uniquement. Cet agrément[3] est donné par le ministère de la culture. Il permet à l'administration des archives de s'assurer que le tiers en question respecte un certain nombre de normes et de bonnes pratiques en matière d'archivage. Une distinction est établie entre archives numériques et archives sur supports traditionnels pour les normes à respecter (normes NF Z42-013 : 2009[4] et ISO 14721 : 2003[5] pour les archives numériques, et norme NF Z40-350[6] pour les autres). La durée pendant laquelle l'agrément est valable n'est pas la même dans les deux cas : cinq ans pour les archives sur supports traditionnels contre trois pour les archives numériques.

Les aspects examinés lors de l'instruction des demandes, s'agissant des archives numériques, sont pour partie des aspects relevant de la sécurité : sécurité des locaux (incendie, inondation, intrusion), sécurité des systèmes d'information, sécurité des données et pour une autre partie des aspects de compréhension et de mise en œuvre des règles métier (procédures de versement, de communication, de restitution, d'élimination, de modification, respect de contrôle scientifique et technique de l'Etat). Enfin pour une dernière partie, il s'agit de critères assez généraux de compétence, de procédure qualité, de formation, de documentation, etc. Les instructions se font sur dossier et sont complétées par des visites sur place.

C'est dans ce cadre que le CINES a demandé et obtenu cet agrément, ce qui autorise le CNRS à lui confier ses archives publiques numériques durant la période courante et intermédiaire. À l'âge définitif, les archives seront transférées aux Archives nationales qui sont l'opérateur compétent en la matière.

3 http://www.archivesdefrance.culture.gouv.fr/gerer/gestion-externalisee-des-archives

4 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.

5 Open Archival Information System (OAIS)

6 Prestations d'archivage et de gestion externalisée de documents - Service et mise en oeuvre des prestations

L'information numérique : quels défis pour les archives ?

La grande révolution du numérique pour les archives est principalement la dissociation de l'information de son support. Les deux étaient jusqu'alors plus intimement liés comme l'écriture manuscrite l'était avec son support papier. On parle alors de « dématérialisation de l'information ». L'information peut maintenant naître sous forme numérique et être échangée sous forme de flux dans des réseaux. On ne cherche plus à conserver les supports qui deviennent de simples vecteurs, dont le choix et le changement sont guidés par des préoccupations de marché (recherche d'un coût d'acquisition et d'entretien faible, d'un encombrement réduit, d'une consommation énergétique faible, d'une grande rapidité d'accès, d'une automatisation possible des traitements, etc.).

Gros avantage pour les archivistes, l'information numérique permet de dépasser les problèmes liés au maniement de supports originaux, qui était une des caractéristiques des archives (par opposition aux bibliothèques qui gèrent pour une large part des exemplaires imprimés). Le double numérique étant identique à l'original, un même document peut être conservé à plusieurs endroits à la fois ; il peut être consulté par un grand nombre d'utilisateurs en même temps. Ces utilisateurs peuvent être géographiquement très éloignés du document et n'ont plus nécessairement besoin de se déplacer pour le consulter. Un utilisateur peut consulter un document original sans risque de l'user, de le tâcher, de le falsifier ou de le subtiliser. En cela, l'information numérique porte en elle des qualités dont peuvent tirer parti les archivistes.

Mais l'information numérique présente aussi de nouvelles difficultés aux archivistes.

L'information numérique est codée. Pour parler de cette information, nous disposons du modèle et du vocabulaire associé proposés par l'OAIS. Cette norme conceptuelle précise les composants, les interactions, les fonctionnalités et les responsabilités à mettre en œuvre dans un système ou une organisation dont l'objectif est la préservation de l'information. Ce que l'on stocke (appelé dans le modèle OAIS «contenu de données»), ce sont des séquences d'unités binaires que l'on peut représenter de manière conventionnelle comme des 0 et des 1. La lecture de la séquence ne nous renseigne pas sur la nature de l'information qu'elle contient, sauf à connaître l'organisation (la syntaxe) de cette séquence. Cette syntaxe est décrite globalement (si on fait abstraction des couches basses) par les formats de représentation (XML, PDF, PNG, etc.). Cette information est nécessaire, en particulier pour utiliser des outils de visualisation ou de manipulation du contenu d'information (par exemple pour l'affichage d'une image ou pour la navigation dans une structure arborescente). Ces outils sont condamnés à évoluer pour suivre à la fois l'évolution des systèmes et celle des usages. Les systèmes eux aussi évoluent pour s'adapter aux nouveaux périphériques. Les

> Ce que l'on stocke, ce sont des séquences d'unités binaires que l'on peut représenter de manière conventionnelle comme des 0 et des 1.

périphériques et composants matériels évoluent en fonction des nouvelles technologies, etc. Les liens de dépendance entre ces briques du numérique forment une pile dont tous les étages sont soumis à des fréquences de changements relativement élevées et asynchrones.

À cela il faut souvent ajouter que l'information ne serait pas compréhensible sans des vocabulaires de données qui associent des définitions aux codes employés dans les données (par exemple la signification des balises utilisées dans un document XML). Si ces deux types d'informations (appelées dans le modèle OAIS respectivement «information de représentation syntaxique» et «information de représentation sémantique») ne sont pas disponibles, le contenu d'information ne sera pas compréhensible. Le modèle OAIS associe à ces informations de représentation, qui sont assez spécifiques au numérique, des informations de pérennisation, plus traditionnelles dans la pratique de la description archivistique (par exemple les informations de contexte et de provenance).

A l'heure où l'information est créée et manipulée sous forme numérique, le besoin de documentation est plus fort qu'il ne l'a jamais été. Le vieillissement des supports et des formats est beaucoup plus rapide qu'avant, et ce ne sont pas les préoccupations de conservation qui guident le marché. Conserver de l'information numérique dans ces conditions est une tâche qui demande de fortes capacités d'adaptation aux changements. Il n'est pas raisonnable de penser, pour un chercheur, qu'il va pouvoir se contenter de stocker de l'information sous cette forme pendant plusieurs décennies sans régulièrement dupliquer ses archives, les migrer d'une machine à une autre, d'un support à un autre, sans convertir des formats qui deviennent obsolètes, sans documenter sa production, sans utiliser un plan de classement, sans se préoccuper du nommage de ses fichiers. Le chercheur doit entrer le plus tôt possible dans un écosystème où l'information est gérée avec des moyens adaptés, où certaines tâches sont déléguées, automatisées éventuellement, voire rendues transparentes.

Les problématiques de codage, de formatage, de stockage, de gestion de l'intégrité, de gestion de la confidentialité, de gestion du cycle de vie des archives, de traçabilité doivent être envisagées le plus tôt possible, au risque de ne pas être capable à relativement court terme de préserver l'information.

Pour aller plus loin

À lire :

Françoise Banat-Berger, Laurent Duplouy, Claude Huc, *L'archivage numérique à long terme : les débuts de la maturité ?*, Paris, La Documentation française, 2009.

Claude Huc, *Préserver son patrimoine numérique*, Paris, Eyrolles, 2010.

Étude sur les formats PDF, http://www.huma-num.fr/ressources/guide-methodologique-le-format-de-fichiers-pdf/13-aout-2012

Journées OAIS, http://www.huma-num.fr/article/les-presentations-des-journees-oais-sont-en-ligne

Liens utiles :

ABES, http://www.abes.fr

Archives de France, http://www.archivesdefrance.culture.gouv.fr

ARK - identifiants, http://fr.wikipedia.org/wiki/Archival_Resource_Key

Centre de calcul de l'IN2P3, http://cc.in2p3.fr

CCSD, http://ccsd.cnrs.fr

CINES, http://www.cines.fr

Data Seal of Approval, http://datasealofapproval.org

HAL, http://hal.archives-ouvertes.fr

Modèle OAIS, http://fr.wikipedia.org/wiki/Open_Archival_Information_System

Persée, http://www.persee.fr

PIAF - archives, http://www.piaf-archives.org

PIN - Pérennisation des Informations Numériques, http://pin.association-aristote.fr/doku.php

Projet SPAR de la BnF, http://www.bnf.fr/fr/professionnels/conserver_spar.html

Les dossiers thématiques sont réalisés
par le pôle communication de la TGIR Huma-Num.

Retrouvez toute l'actualité d'Huma-Num sur :
http://www.huma-num.fr
http://humanum.hypotheses.org