

PAC

La plateforme d'archivage pérenne de documents électroniques au CINES

Olivier Rouchon

Centre Informatique National de l'Enseignement Supérieur
950 Rue de Saint Priest – 34097 Montpellier Cedex 5 – France
olivier.rouchon@cines.fr

Résumé

Pendant une période d'une quinzaine d'années, la question de la conservation à long terme de l'information sous forme numérique n'a été un sujet de préoccupation et de réflexion que pour de rares institutions scientifiques ou patrimoniales qui ont joué un rôle moteur dans la prise de conscience des risques encourus et dans l'émergence de normes de référence et de méthodologies dans ce domaine. La progression inexorable et exponentielle du numérique dans tous les domaines, les volumes à présent atteints et le caractère impératif de la préservation de l'information ont accéléré cette prise de conscience jusqu'au plus haut niveau de la hiérarchie des administrations et des entreprises.

Aussi, ces trois dernières années ont vu l'émergence d'une série de projets, dans les sphères publique ou privée, au niveau national, européen et mondial, visant à développer des infrastructures matérielles et logicielles dédiées à la conservation de l'information et à la mise en place de véritables services d'archivage à long terme des informations sous forme numérique. Nous allons essayer de présenter les défis, les choix et les processus mis en place au sein du CINES pour l'implémentation d'une telle infrastructure, qui est maintenant opérationnelle.

Mots clefs

CINES, PAC, archivage pérenne, documents électroniques, OAIS, formats de fichiers, métadonnées, DCMI, thèses, Persée ...

1 Introduction

Placé sous la tutelle de la DGRI (Direction Générale de la Recherche et de l'Innovation) du Ministère de l'Enseignement Supérieur et de la Recherche, le CINES (Centre Informatique National de l'Enseignement Supérieur) est un établissement public national à caractère administratif de création récente. Il a succédé au CNUSC (Centre National Universitaire Sud de Calcul) créé à Montpellier en 1980.

Selon ses statuts, le CINES est chargé de rendre des services informatiques aux établissements publics d'enseignement supérieur et aux organismes de recherche dans les domaines suivants :

- le calcul numérique intensif ;
- l'exploitation de bases de données notamment celles utilisées dans le secteur des bibliothèques ;
- l'expertise et la formation en matière de réseaux informatiques nationaux et internationaux et en matière de technologies associées (lien direct au réseau RENATER).

Le CINES a également, depuis 2004, une mission nationale d'archivage du patrimoine scientifique électronique et numérique.

Pour la remplir, le projet PAC (Plateforme d'Archivage du CINES) a été mis en place, qui vise à doter le CINES d'une plateforme et d'un service d'archivage numérique pérenne.

Une première présentation [1] des concepts de l'archivage pérenne de documents numériques et des débuts du projet PAC a été faite par Philippe Prat lors des Journées Réseaux 2005 à Marseille. Il s'agit aujourd'hui de présenter l'implémentation de ces concepts qui a été réalisée au CINES et qui est maintenant opérationnelle, de détailler les orientations et les décisions qui ont été prises face aux questionnements et obstacles, et éventuellement de les justifier.

1.1 Retour sur le contexte, le constat et la problématique de l'archivage pérenne

L'archivage pérenne des documents électroniques consiste à conserver le document et l'information qu'il contient dans son aspect physique comme dans son aspect intellectuel, sur le très long terme (trente ans et au-delà), et de manière à pouvoir le rendre accessible et compréhensible.

Or, la plupart des fichiers informatiques de plus de dix ans sont aujourd'hui illisibles, conséquence de plusieurs risques inéluctables :

- la connaissance perdue du contenu des fichiers ;
- le format de fichier inconnu ;
- le support physique détérioré ;

– le logiciel ou matériel de lecture disparu.

Il y a donc des processus et des pratiques de préservation à mettre en œuvre pour relever le défi de l'archivage pérenne.

2 Présentation de la Plateforme d'Archivage du CINES

Le projet PAC a démarré au printemps 2004 avec l'objectif de doter le CINES d'une plateforme et d'un service d'archivage numérique pérenne apte à répondre à la majorité des besoins, en privilégiant une approche généraliste, évitant ainsi le traitement des projets au cas par cas.

2.1 Les défis, orientations et choix retenus pour l'implémentation

Nous l'avons évoqué, la mise en place d'une infrastructure pour la préservation de documents électroniques implique un certain nombre de choix, de définitions de processus, pour réduire l'impact des risques dont la réalisation peut affecter la lisibilité de fichiers informatiques, et qui sont cités plus haut.

Tout d'abord, une équipe dédiée à la plateforme d'archivage a été constituée, chargée de couvrir les aspects organisationnels (définition et expertise des processus métiers et des méthodes), et culturels (renforcement des collaborations entre informaticiens, archivistes et bibliothécaires), en plus de l'aspect technique. A ce jour, l'équipe archivage du CINES est constituée d'un chef de projet, quatre ingénieurs (développement, système et réseaux, support) et une archiviste.

Ensuite, les aspects techniques ont été analysés, à la fois sur le plan théorique et sur le plan pratique, avec notamment les retours d'expérience présentés lors de réunions du groupe de travail PIN (Préservation de l'Information Numérique [2]). La future plateforme a commencé à se dessiner, en suivant des contours donnés par les normes internationales en vigueur.

Quelques exemples : pour éviter la perte d'informations relatives au document électronique et à son contenu, des métadonnées génériques (type DCMI – Dublin Core [3], METS, etc.) ont été utilisées. Elles permettent de décrire les attributs du document (auteur, titre, résumé, mots-clés, etc.) afin de le replacer dans son contexte et en préserver le sens. Ces métadonnées ont été développées et structurées suivant les normes internationales de description archivistique, ISAD(G) et ISAAR(CPF)¹.

¹ ISAD(G) : international standard for archival description, general.

ISAAR(CPF) : international standard archival authority record, for corporate bodies, persons and families

Il a également été décidé d'attribuer un identifiant unique et pérenne aux documents au moment de leur archivage pour permettre, notamment, de les retrouver et de les référencer. Plutôt que d'utiliser des identifiants persistants de type DOI (Digital Object Identifier [4]), le choix s'est porté sur une numérotation interne à l'application, car l'accès direct aux documents ou encore la possibilité de citation de ces documents depuis des systèmes d'information externes n'est pas une priorité. Cette numérotation respecte la définition et la structure d'une URI (Uniform Resource Identifier), en se réservant la possibilité de rejoindre ultérieurement le système DOI ou un équivalent (AOI, ARK).

Les formats de fichiers durables² ont été privilégiés afin d'éviter une obsolescence trop rapide des formats de fichiers acceptés par la plateforme. Tout document versé est vérifié pour s'assurer que les formats de fichiers qu'il contient sont conformes. En outre, des procédures de veille technologique et de migration logique ont été élaborées pour identifier les formats émergents ou obsolètes, et migrer les fichiers d'un format désuet vers un format pérenne.

Des outils de gestion du vieillissement des supports utilisés pour stocker les documents, ainsi que des procédures de migration physique ont été définis. Ceux-ci sont accompagnés d'un effort de veille et d'anticipation sur les technologies émergentes en termes de médias de stockage, permettant d'éviter la disparition du logiciel ou du matériel de lecture.

2.2 Les concepts et l'architecture logique

Le système PAC a été conçu comme un ensemble de trois serveurs logiques, s'inspirant donc du modèle proposé par la norme ISO 14721 (OAIS [5]).

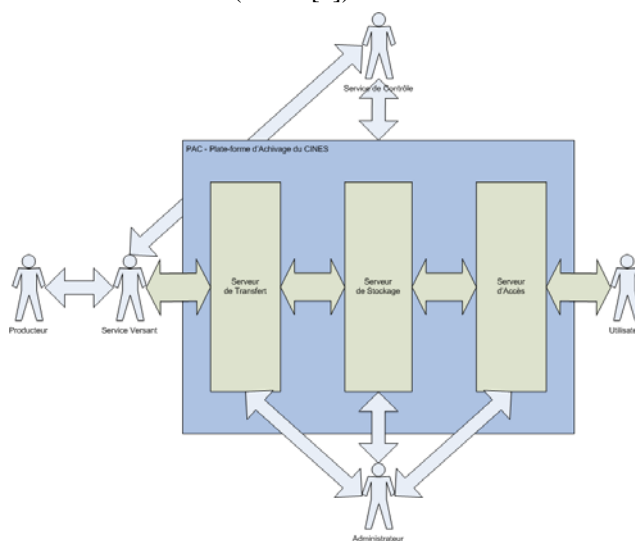


Figure 1 – Architecture logique de PAC

² Formats publiés, largement utilisés – ou promis à l'être – normalisés si possible : HTML, PDF, PDF/A, TXT, XML (texte), GIF, JPEG, PNG, TIFF (image).

- un serveur de transfert auquel le service versant les documents pourra transmettre ses archives ;
- un serveur de stockage où sont conservés les documents ;
- un serveur d'accès où le service versant et éventuellement les utilisateurs autorisés à consulter ses archives pourront rechercher et obtenir une copie des documents archivés.

2.3 Les acteurs

Pour décrire les principes de fonctionnement de la plateforme, il convient tout d'abord de décrire les rôles et responsabilités de chacun des acteurs qui seront amenés à interférer avec elle. Comme décrit sur la Figure 1, ils sont au nombre de cinq :

- le producteur, qui est une personne physique ou morale, publique ou privée, qui a produit, reçu et conservé des archives dans l'exercice de son activité ;
- le service versant, qui est une organisation qui transfère une archive à un service d'archives. Le service versant peut être également un producteur ;
- le service de contrôle, qui effectue le contrôle scientifique, juridique et technique des documents archivés, et éventuellement valide les demandes de communication d'archives ;
- le service d'archives (en l'occurrence le service PAC du CINES), qui reçoit le document à archiver transféré et est chargé de le conserver dans le temps pour permettre à une communauté d'utilisateurs d'y accéder ;
- l'utilisateur (ou service demandeur), qui est en relation avec le service d'archives pour trouver les informations archivées présentant un intérêt, et pour accéder au détail de ces informations, dans le respect de la législation applicable en matière de communication des archives.

3 Les échanges avec la Plateforme d'Archivage du CINES

Intéressons-nous à présent aux échanges de données entre le service PAC du CINES et les différents acteurs avec lesquels il sera en contact :

- le transfert d'archives, c'est-à-dire la transmission d'objets numériques à archiver par un service versant au CINES pour qu'il en assure la conservation. Le transfert peut éventuellement être précédé d'une demande de transfert ;
- la communication d'archives, c'est-à-dire la transmission de copies archives par le CINES à un

utilisateur ayant l'autorisation du service versant et/ou du service de contrôle compétent ;

- la modification des métadonnées et/ou des données pour en assurer leur préservation : information envoyée par le CINES au service versant pour le prévenir que ses archives versées au CINES vont être ou ont été modifiées afin d'en assurer une bonne conservation. Les modifications peuvent porter sur les données elles-mêmes (conversion de format) ou sur les métadonnées (ajout, correction, mise à jour) ;
- l'élimination des métadonnées et/ou des données, c'est-à-dire l'élimination d'archives soit à la demande du CINES, soit à la demande du service versant et/ou du service de contrôle compétent ;
- la restitution des archives au service versant ou au service producteur, ou transmission d'archives par le CINES à ces services en vue de leur en restituer la garde.

L'implémentation des processus d'échanges s'appuie (très) largement sur le Standard d'échanges de données pour l'archivage [6] défini par la DGME (Direction Générale de la Modernisation de l'Etat) et la DAF (Direction des Archives de France), qui est en cours de normalisation au niveau européen, et qui a déjà été utilisé pour l'implémentation du projet Pil@e des Archives de France.

3.1 L'initiation du projet d'archives

Un projet d'archives au CINES peut être initié à la demande de tout organisme produisant ou collectant en grande quantité des documents électroniques dont le contenu possède une valeur patrimoniale scientifique ou technique, et doté d'un système informatique pouvant être interfacé avec la plateforme PAC (sur laquelle il n'existe pas encore d'interface homme-machine pour un transfert interactif et manuel). La priorité du projet est alors définie, avec éventuellement la participation des organismes de tutelle, et un calendrier prévisionnel est mis en place.

Avant tout transfert, le responsable du transfert au service versant prend contact avec le service PAC du CINES (pac@cines.fr) afin de démarrer le projet d'archives, échanger des informations générales sur leur mode de fonctionnement respectif et élaborer un protocole de transfert où seront définis les composantes du projet d'archives.

Cette étape est divisée en deux phases, et doit aboutir à la création d'un document écrit fixant les responsabilités des deux parties co-contractantes.

Elle démarre par une phase dite préliminaire durant laquelle les points suivants sont abordés :

- l'identification des informations à pérenniser ;
- l'établissement d'une liste des données qui seront transmises au CINES (format, taille, nombre...),

accompagnées de leurs métadonnées obéissant au système PAC ;

- l'analyse de tous les aspects de faisabilité (sécurité, aspects légaux, coûts et risques...) ;
- l'évaluation de la volumétrie et des ressources requises.

Une phase dite de définition permet ensuite de déterminer plus en détail les points suivants :

- la définition précise des objets à transférer. Ce sont de préférence des objets dits primaires, c'est-à-dire issus d'un traitement (ex. numérisation de documents papier) ou d'un travail manuel (saisie de texte ou d'informations), et issus d'archives définitives – voir Figure 2. On évitera, par exemple, d'archiver des documents issus du retraitement d'objets primaires (ex. génération d'OCR) ;

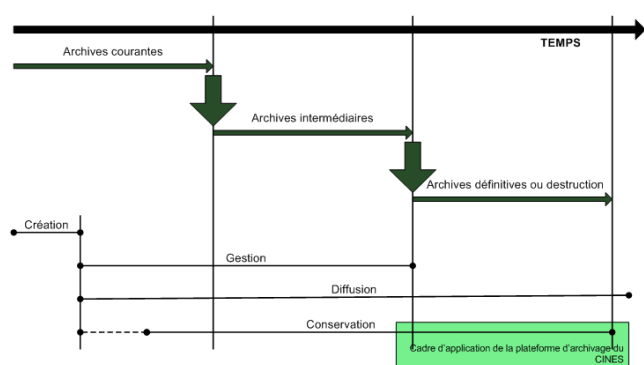


Figure 2 - Cycle de vie d'un document et cadre d'application

- les 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 ;
- la formation du personnel du service versant à l'utilisation du système PAC.

Une fois ces deux phases terminées, et les ajustements applicatifs éventuels effectués, les échanges peuvent commencer...

3.2 Le transfert des informations à archiver

Le service versant transmet par réseau via un protocole sécurisé, ou sur support amovible (suivant le protocole de transfert préalablement établi), les documents à verser au service PAC. Le transfert est repéré par l'application PAC, qui envoie par courriel un accusé de réception du transfert et procède alors à une série de contrôles de validité technique des données transférées, en s'assurant que les objets transférés respectent les conditions définies par le protocole de transfert : conformité des formats, structure de versement...

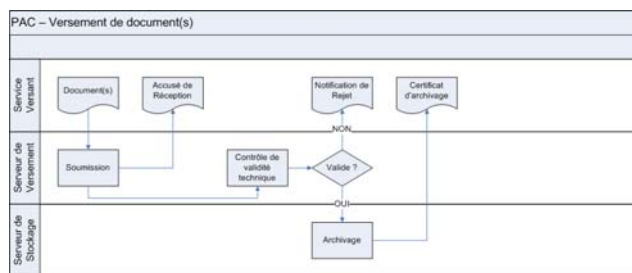


Figure 3 – Le transfert de documents

Si le résultat du contrôle est positif, le service PAC enverra un certificat d'archivage au service versant, signalant que l'objet à archiver est accepté et transféré sur le serveur de stockage. Cette notification comporte l'identifiant unique et pérenne ainsi que les métadonnées des archives transférées et la liste des contrôles effectués par le CINES sur le versement pour en tester la validité, afin que le service versant puisse garder trace et preuve des données transférées. L'identifiant de l'archive est un identifiant unique qui permettra au service versant ou à des utilisateurs de PAC de retrouver l'archive.

Si le résultat du contrôle est négatif, le service PAC enverra un avis de rejet, et le versement refusé sera déplacé dans un espace spécifique où le service versant pourra venir le récupérer afin de modifier les éléments non conformes. Un accusé de réception de l'avis d'anomalie devra être alors transféré par le service versant au service PAC.

3.3 Le stockage de documents

Le service producteur ou le service versant doivent tenir à jour les métadonnées des archives qui sont sous leur responsabilité. Pour cela, ils transmettent leur demande au service PAC qui réalisera la modification et transmettra un avis de modification.

Le service PAC s'engage à tenir à jour les métadonnées qui sont de sa responsabilité.

Le service PAC assure également la veille technologique concernant les évolutions d'environnement de stockage, les techniques et les technologies les plus optimales pour la conservation des archives.

De fait, il sera amené à réaliser certaines opérations d'exploitation courante (contrôle des supports, migrations de support à support, restaurations, contrôles de métadonnées) qui ne nécessiteront pas d'avertir le service producteur ou le service versant.

Enfin, le service PAC est chargé d'adresser annuellement au producteur la liste de ses documents archivés sur la plateforme de stockage, et doit fournir, à la demande du service producteur, ou du service versant, les résultats des tests de contrôle des supports.

Tout événement, action ou manipulation des données ou des métadonnées archivées est horodaté, historié et traçable.

Lors d'une modification de l'archive ou des ses métadonnées, le service PAC adresse au service versant un avis de modification, précisant les références des archives qui ont subi des modifications et la nature de celles-ci (ex. mise à jour des métadonnées, conversion de format). Le service versant doit en accuser réception.

L'élimination d'objets conservés au CINES est un processus exceptionnel qui ne se réalise qu'avec l'accord explicite du service versant, une première fois lorsqu'est établi le protocole de transfert, une deuxième fois en amont de la date à laquelle l'objet doit être détruit. Elle peut se produire selon deux scénarios³ :

- soit l'élimination est proposée par le service PAC ;
- soit l'élimination est demandée par le service versant.

Le service PAC envoie une demande d'élimination d'archives comprenant les métadonnées des documents concernés, au service versant, qui envoie en émission synchrone un accusé de réception de la demande d'élimination d'archives.

Le service versant fera suivre, après les vérifications nécessaires, un avis d'acceptation d'élimination ou un avis de refus d'élimination.

L'avis d'acceptation d'élimination doit comporter obligatoirement les métadonnées des archives concernées, afin de confirmer leur destruction. Une fois la destruction effectuée, le service PAC envoie au service versant une notification d'élimination d'archives, qui rappelle également l'identifiant des archives concernées par cette élimination.

En cas de refus d'élimination par le service versant, le service PAC doit notifier avoir bien reçu ce refus en envoyant un accusé de réception de refus.

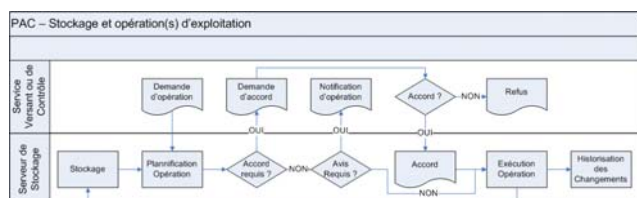


Figure 4 – Le stockage de documents

La restitution d'archives peut s'appliquer dans deux cas de figure :

- la restitution est à la demande du CINES ;
- la restitution est à la demande du service versant.

Le service PAC envoie au service versant⁴ une demande de restitution d'archives, dans laquelle figure les

³ Les procédures sont les mêmes dans les deux cas, seuls les interlocuteurs changent (ici, nous avons détaillé les procédures pour le cas où l'élimination est proposée par le service PAC du CINES

⁴ Le principe de la transaction est identique lorsqu'il s'agit d'une demande formulée par le service versant

métadonnées de la ou des archive(s) concernée(s). Le service versant répond par un accusé de réception, suivi d'un avis d'acceptation ou un avis de refus.

Après réception de l'avis d'acceptation, le service PAC envoie l'archive accompagnée d'un avis de restitution d'archives. Si l'archive parvient intégralement au service producteur, celui-ci envoie un accusé de réception des archives. Dans le cas contraire, le service producteur envoie un avis d'anomalie de restitution au service PAC.

3.4 L'accès aux documents archivés et leur communication

La communauté des utilisateurs ayant accès aux archives est définie par le service versant, en collaboration avec le service PAC, lors de la phase préliminaire du projet d'archives.

La demande de communication d'archive peut concerner autant les données, que ses métadonnées.

Avant de demander une communication d'archives, l'utilisateur se sera identifié auprès du service PAC, en utilisant l'outil en ligne mis à sa disposition sur le serveur d'accès, et aura pris connaissance des règles et spécifications en matière de communicabilité en vigueur au sein du service d'archives.

L'utilisateur une fois authentifié et autorisé par le service PAC pourra :

- consulter le sous-catalogue d'un projet d'archives particulier ;
- effectuer une recherche grâce à l'outil de recherche du serveur d'accès, qui interroge les métadonnées du système de gestion de données ;
- faire une demande de communication d'archives auprès du service PAC ;
- choisir éventuellement le mode de restitution : consultation en ligne, support amovible, transmission réseau.

La demande de communication est effectuée par l'utilisateur. Le service PAC envoie en retour un accusé de réception de la demande.

Deux cas de figure sont à envisager :

- soit l'archive est librement communicable, critère défini préalablement dans le protocole de transfert : en ce cas, le service PAC communique l'archive à l'utilisateur sans en référer au service versant. L'utilisateur retourne ensuite au service PAC un accusé de réception de communication si l'archive lui est parvenue avec succès, ou un avis d'anomalie de communication dans le cas contraire ;
- soit l'archive n'est pas communicable, auquel cas le service PAC réalise auprès du service versant une demande d'autorisation de communication. Si le service

versant agréé la demande, alors il transmet au service PAC une autorisation de communication, et le service PAC communique l'archive à l'utilisateur, qui retourne ensuite au service PAC un accusé de réception de communication si l'archive leur est parvenue avec succès, ou un avis d'anomalie de communication dans le cas contraire. Si le service versant refuse cette demande, il transmet au service PAC un refus d'autorisation de communication auquel le service PAC répondra par un avis de réception du rejet. Le service PAC transmettra ensuite à l'utilisateur un avis de rejet de la demande de communication.

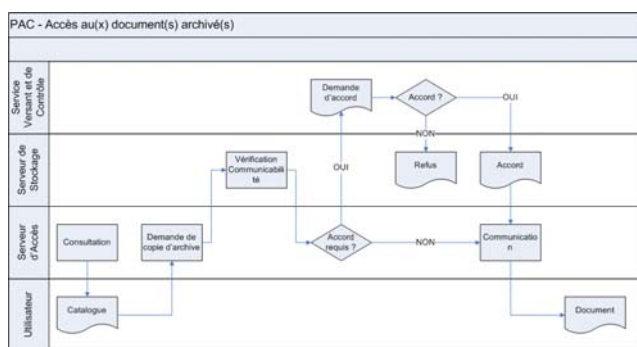


Figure 5 – La communication de documents

Le CINES est un établissement public, et à ce titre respecte les délais prévus par le Code du patrimoine en matière de communicabilité des archives et se conforme à la législation relative au droit intellectuel, au droit des auteurs, au droit de reproduction et au droit de représentation des œuvres de l'esprit.

Nous pouvons constater que ces échanges sont extrêmement codifiés, mais cette contrainte est le prix à payer pour garantir la fiabilité et le déroulement optimal et sécurisé des échanges entre chaque acteur.

4 Les principes de fonctionnement de la plateforme

Pour implémenter les procédures pour les échanges de données décrites ci-dessus, la plateforme PAC a été scindée en trois serveurs logiques (Transfert, Stockage, Accès). Ils seront responsables des trois types d'échanges que nous venons d'évoquer, et nous allons maintenant résumer les grandes lignes de leurs fonctions.

4.1 Le serveur de Transfert

Réception des paquets d'informations à verser (SIP – Submission Information Package)

- détection d'un nouveau transfert ;
- envoi d'un accusé de réception (courriel au format XML [7]) au service versant.

Contrôle des SIP

- vérification de la structure informatique des documents transférés ;

- conformité des métadonnées fournies au format XML, par rapport au schéma de référence [8] ;
- correspondance biunivoque entre la description des fichiers du SIP et les fichiers qui composent effectivement le document ;
- contrôle et validation du format des fichiers ;
- calcul de l'empreinte numérique de chaque fichier composant le document.

Création des paquets d'information archivés (AIP – Archival Information Package)

- création de l'identifiant unique et pérenne du document archivé ;
- mise à jour des métadonnées, ajout des métadonnées de préservation ;
- génération de l'AIP à partir du SIP et transfert vers le serveur de stockage.

4.2 Le serveur de stockage

Archivage des AIP

- copie multiple de l'AIP sur les différents médias ou supports qui composent le serveur de stockage ;
- envoi du certificat d'archivage au service versant ;
- vérification périodique de l'intégrité des AIP archivés par nouveau calcul des empreintes numériques ;
- migration technologique (logique – conversion de formats de fichier, ou physique – migration de support) ;
- fourniture d'états et de statistiques (interne au service PAC, ou destinés au service versant ou service producteur).

4.3 Le serveur d'accès

Contrôle de l'authentification de l'utilisateur

- outil en ligne.

Consultation du catalogue des AIP archivés

- module de recherche multi-critères.

Communication d'une copie d'un document archivé

- génération du paquet d'information diffusé (DIP – Dissemination Information Package) à partir de l'AIP ;
- mise à disposition de l'utilisateur.

5 Etat des lieux

Le projet a été initié avec le développement interne d'une première plate-forme pour valider les services attendus sur les deux projets pilotes alors à l'étude. La première version (v1.0) de la plateforme PAC est maintenant opérationnelle ; les technologies utilisées pour son développement viennent du domaine du logiciel libre – pour en faciliter le portage d'un environnement à un autre en cas de migration, et sont les suivantes :

Serveur primaire	Dell PowerEdge 1950 Biprocasseur Intel Xeon CPU 3.00GHz 1Go RAM 300Go disque interne 300Go disque externe (sur serveur secondaire via NFS)
Serveur secondaire	Dell PowerEdge 1850 Biprocasseur Intel Xeon CPU 2.80GHz 1Go RAM 300Go disque interne (accessible par serveur primaire via NFS)
Plateforme	Linux RedHat Enterprise server 4
Langage de programmation	Composants développés avec Java Development Kit 1.6
Base de Données	PostgreSQL
Analyse de format de fichiers	Utilisation de bibliothèques Open Source : Jhove, DROID, ImageMagick
Calcul des empreintes numériques	JDK 1.6 (hashing MD5 et SHA-256, etc.)
Envoi de courriels	Javamail
Exécution des modules	Déclenchement des routines (réception, vérification, archivage) à intervalles réguliers – traitement par lots
Protocole transfert	Soumission de SIP sur serveur de transfert via SFTP ou SCP
Interface utilisateur	Administration en mode web

5.1 Les projets d'archives en cours

Actuellement, deux projets sont en cours.

Le premier concerne l'archivage pérenne de thèses⁵ collectées par l'ABES – Agence Bibliographique de l'Enseignement Supérieur – par l'intermédiaire de l'outil STAR (Signalement des Thèses, Archivage et Recherche [9]), 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.

L'intégration de la plateforme PAC à l'outil STAR a été réalisée, et le projet est maintenant dans la phase finale des

tests de cette interface, avant une mise en production imminente du service.

Le deuxième projet, actuellement dans sa phase préliminaire, concerne l'archivage de revues en Sciences Humaines et Sociales (SHS), numérisées dans le cadre du projet Persée [10].

La volumétrie estimée est la suivante :

	Documents (cumul)	Volume (cumul)
Projet Thèses – ABES		
Fin 2007	1000	4Go
Fin 2008	2000	8Go
Fin 2009	3000	12Go
Projet Persée⁶ – UL2		
Fin 2008	3000	10To
Fin 2009	6000	20To
Fin 2010	12000	40To

Pour information, la production annuelle à venir de thèses et de fascicules Persée numérisés est estimée à 10000 et 3000 respectivement. C'est potentiellement le volume annuel à traiter après les trois premières années d'exploitation.

5.2 Les défis à venir

D'autres projets d'archives sont en cours de discussion, parmi lesquels l'archivage des documents versés dans les Archives Ouvertes (HAL [11]).

La plateforme actuelle ne permettant pas de gérer le volume de données prévu pour le projet Persée, et en prévision de projets à venir d'importance, il a été décidé de procéder à un appel d'offres pour l'acquisition au printemps 2008 d'une infrastructure capable de gérer de larges volumes (plus de 40 To).

Bibliographie

- [1] Michel Auffret, L'archivage pérenne des documents numériques. *Journées Réseaux JRES 2005*, <http://2005.jres.org/paper/47.pdf>, 6 Décembre 2005.
- [2] <http://vds.cnes.fr/pin/>
- [3] <http://dublincore.org/>
- [4] <http://www.doi.org/>
- [5] http://vds.cnes.fr/pin/documents/projet_norme_oais_version_francaise.pdf
- [6] https://www.ateliers.modernisation.gouv.fr/ministeres/projets_adele/a103_archivage_elect/public/standard

⁵ Documents nativement au format électronique

⁶ Documents issus de la numérisation de revues en Sciences Humaines et Sociales

[d.echange.archives_echanges_v0/downloadFile/file/archives_echanges_v0-1_description_standard_v1-0.pdf?nocache=1141748589.19](http://d.echange.archives-ouvertes.fr/archives_echanges_v0/downloadFile/file/archives_echanges_v0-1_description_standard_v1-0.pdf?nocache=1141748589.19)

- [7] <http://www.cines.fr/pac/avis.xsd>
- [8] <http://www.cines.fr/pac/sip.xsd>
- [9] <http://www.abes.fr/abes/page,428,star.html>
- [10] <http://www.persee.fr/>
- [11] <http://hal.archives-ouvertes.fr/>