

Diplôme de conservateur de bibliothèque

Mémoire d'étude / janvier 2016

## **Archivage pérenne en bibliothèque universitaire : bilan et perspectives**

**Elsa Ferracci**

Sous la direction de **Marie-Madeleine Géroutet**  
Responsable du service Bibliothèque numérique - SCD de l'université de Lille 1



## **Remerciements**

*Toute ma gratitude va aux nombreux professionnels sollicités dans le cadre de mon enquête qui ont eu l'obligeance de répondre à mes questions et de me consacrer du temps :*

*Catherine Renard (Bibliothèque inter-universitaire Cujas) ; Jean-François Vincent et Claire Ménard (Bibliothèque inter-universitaire de santé Paris Descartes) ; Olesea Dubois (Bibliothèque de Science Po) ; Rémi Gaillard (SCD de l'université Pierre et Marie Curie) ; Romain Le Nezet (SCD de l'université Université Sorbonne Nouvelle - Paris 3) ; Pauline Rivière (Bibliothèque Sainte-Geneviève) ; Anne-Sophie Traineau-Durozoy (SCD de l'université de Poitiers) ; Adeline Rege (SCD de l'université de Strasbourg) ; Frédéric Blin (BNUS) ; Stéphanie Bouvier (SCD de l'université d'Angers) ; Daniel Eymard (SCD de l'université de Toulon) ; Marie-Madeleine Géroutet (SCD de l'université de Lille 1) ; Anne Flahaut-Steiner et Solenn Bihan (SCD de l'université de Lille 2) ; Isabelle Westeel et Cécile Malleret (SCD de l'université de Lille 3) ; François-Xavier Boffy (SCD de l'université de Lyon 1) ; Maximilien Savoye (SCD de l'université de Lorraine) ; Olivier Legendre (SCD de l'université Clermont-Université) ; Jean-Marie Le Béhec (SCD de l'INPT, Université de Toulouse) ; Yvan Leclerc (Centre Flaubert, Université de Rouen) ; Stéphane Reecht et Christine Génin (BnF) ; Raphaël Tournoy (CCSD) ; Joachim Schöpfel (ANRT) ; Stéphanie Roussel (SIAF) ; Dominique Armand (CNDP) ; Geneviève Cron (COMUE SPC) ; Hélène Chaudoreille (COMUE PSL).*

*Je remercie en particulier Lorène Béchard (CINES), Laurent Duploux (BnF) et Julie Ladant (MENESR) dont l'aide a été décisive.*

*Sans toutes leurs participations, ce mémoire n'aurait pu voir le jour.*

*Qu'il me soit enfin permis de remercier vivement ma directrice de mémoire, Marie-Madeleine Géroutet, qui a proposé ce sujet et suivi avec attention la rédaction de ce mémoire et dont les remarques et conseils, toujours précis et constructifs, m'ont aidée à orienter au mieux mon travail.*

*Résumé :*

*Au vu des risques que crée l'obsolescence technologique, l'archivage pérenne des contenus numériques s'avère désormais incontournable et constitue un enjeu pour l'Enseignement supérieur et la Recherche. La préservation à très long terme nécessite cependant des compétences et des techniques spécifiques et implique des coûts, humains et financiers. Pour la bibliothèque universitaire se pose alors la question du rôle qui doit être le sien au regard de l'archivage pérenne des contenus numériques qu'elle est amenée à stocker, à diffuser ou à produire.*

*L'objectif de ce mémoire est de dresser un panorama de l'archivage pérenne dans l'Enseignement supérieur et la Recherche, de présenter et d'analyser un ensemble de retours d'expérience de bibliothèques universitaires ayant mené à bien, ou mettant actuellement en œuvre, ou encore envisageant un projet d'archivage pérenne, et d'en déduire les difficultés et les obstacles qui s'opposent à une réelle avancée de l'archivage pérenne en bibliothèque universitaire. Le mémoire envisagera la mutualisation, à divers niveaux, comme une réponse à ces difficultés.*

*Descripteurs :*

*Information électronique – Conservation*

*Archivage électronique*

*Bibliothèques universitaires*

*Patrimoine scientifique numérique*

*Numérisation*

*Abstract :*

*As a result of technological obsolescence, nowadays long-term archiving proves to be unavoidable and represents an issue in the field of Higher Education and Research. Nevertheless, long-term preservation requires specific skills and technics, as it implies human and financial costs. Academic libraries have then to define their role regarding long-term archiving of the digital contents they usually store, broadcast or produce.*

*The goal of this report is to give an updated overview of long-term archiving in the field of Higher Education and Research, to present and analyse a set of feedbacks from academic libraries having completed successfully or being considering a long-term archival project, and to deduce the difficulties and obstacles which are opposed to a significant progress in long-term archiving in academic libraries. The report will consider different levels of resource sharing as a response to these difficulties.*

*Keywords :*

*Digital preservation*

*Digital archiving*

*Academic libraries*

*Digital and scientific heritage*

*Digitisation*

## *Droits d'auteurs*



Cette création est mise à disposition selon le Contrat :  
« **Paternité-Pas d'Utilisation Commerciale-Pas de Modification 4.0 France** »  
disponible en ligne <http://creativecommons.org/licenses/by-nc-nd/4.0/deed.fr> ou  
par courrier postal à Creative Commons, 171 Second Street, Suite 300, San  
Francisco, California 94105, USA.



# Sommaire

<b>SIGLES ET ABRÉVIATIONS .....</b>	<b>9</b>
<b>INTRODUCTION.....</b>	<b>11</b>
<b>I. STRATÉGIES ET ENJEUX DE L'ARCHIVAGE PÉRENNE DANS L'ESR</b> <b>.....</b>	<b>14</b>
1. <i>Un nouvel enjeu patrimonial ? .....</i>	14
2. <i>Le paysage institutionnel et les stratégies nationales et</i> <i>internationales .....</i>	24
3. <i>Coûts et exigences techniques et réglementaires de l'archivage</i> <i>pérenne .....</i>	32
<b>II. PROJETS D'ARCHIVAGE PÉRENNE EN BIBLIOTHÈQUES</b> <b>UNIVERSITAIRES : RETOURS D'EXPÉRIENCE .....</b>	<b>41</b>
1. <i>Méthodologie de l'enquête et bilan général .....</i>	41
2. <i>Solutions adoptées et acteurs des projets .....</i>	42
3. <i>Projets d'archivage pérenne en bibliothèque universitaires : retours</i> <i>d'expérience.....</i>	49
4. <i>Les freins et les causes des abandons des projets.....</i>	56
<b>III. SITUATION DES BIBLIOTHÈQUES UNIVERSITAIRES : VERS UNE</b> <b>NÉCESSAIRE MUTUALISATION .....</b>	<b>60</b>
1. <i>Situation des SCD dans l'université et dans l'ESR .....</i>	60
2. <i>De la mutualisation à la réflexion sur les périmètres institutionnels</i>	75
3. <i>De nouvelles missions et de nouvelles compétences ?.....</i>	83
<b>CONCLUSION .....</b>	<b>91</b>
<b>SOURCES.....</b>	<b>93</b>
<b>BIBLIOGRAPHIE.....</b>	<b>97</b>
<b>ANNEXES.....</b>	<b>103</b>
<b>GLOSSAIRE.....</b>	<b>117</b>
<b>TABLE DES MATIÈRES.....</b>	<b>123</b>



## *Sigles et abréviations*

ABES : agence bibliographique de l'enseignement supérieur  
AIP : Archival Information Package (certificat d'archivage)  
ANR : Agence Nationale de la Recherche  
ANRT : Agence Nationale de Reproduction des Thèses  
AO : Archive Ouverte  
AOI : Archive Ouverte Institutionnelle  
BSN : Bibliothèque Scientifique Numérique  
BIU : bibliothèque inter-universitaire  
BU : bibliothèque(s) universitaire(s)  
CCSD : Centre pour la Communication Scientifique Directe  
CINES : Centre Informatique National de L'enseignement Supérieur  
CNRS : Centre National de la Recherche Scientifique  
CollEx : Collections d'Excellence  
COMUE : Communauté d'Universités et d'Établissements  
CRI : Centre de Ressources Informatiques  
CTLes : Centre Technique du Livre de l'Enseignement Supérieur  
DSI : Direction des Systèmes d'information  
DUA : Durée d'utilité Administrative  
EPRIST : association des responsables de l'information scientifique et technique des organismes de recherche français publics ou d'utilité publique  
ESR : Enseignement Supérieur et Recherche  
HAL : Hyper Article en Ligne  
HDR : Habilitation à Diriger des Recherches  
IGB : Inspection Générale des Bibliothèques  
IN2P3 : institut national de physique nucléaire et de physique des particules (CNRS)  
ISTEX : Information scientifique et technique d'excellence  
JISC : Joint Information Systems Committee  
LRU : Libertés et Responsabilités des Universités  
IRHT : Institut de Recherche et d'Histoire des Textes  
MCC : Ministère de la Culture et de la Communication  
MENESR : Ministère de l'éducation nationale, de l'enseignement supérieur et de la recherche  
METS : Metadata Encoding and Transmission Standard  
MODS : Metadata Object Description Schema

NERC : Natural Environment Research Council  
OAIS : Reference Model for an Open Archival Information System  
PAC : Plateforme d'Archivage au CINES  
PREMIS : PREservation Metadata Implementation Strategies  
SAE : Système d'Archivage Électronique  
SCD : Service Commun de la Documentation  
SHS : Sciences Humaines et Sociales  
SIAF : Service Interministériel des Archives de France  
SIP : Submission Information Package  
SPAR: Système De Préservation et d'Archivage Réparti  
STAR : Signalement des Thèses Électroniques, Archivage et  
Recherche  
TGIR : Très Grande Infrastructure de Recherche  
TO : téraoctet  
UMR : Unité Mixte de Recherche  
UMS : Unité Mixte de Service

# INTRODUCTION

---

La prise de conscience des risques encourus par les productions numériques du fait de l'obsolescence technologique s'est faite au long des années 2000 parmi les professionnels des bibliothèques universitaires, pour lesquels la sécurisation des ressources apparaît désormais comme une nécessité incontestable. Pour autant, si le stockage sécurisé assure aujourd'hui une conservation de premier niveau devenue la norme, la préservation à très long terme implique un changement d'échelle : les processus, l'organisation, le management, le niveau de technicité, les besoins d'anticipation, la réflexion sur les collections et enfin les coûts de l'archivage pérenne sont sans commune mesure avec les pratiques courantes de préservation numérique.

Les expressions désignant la préservation du numérique sont nombreuses - archivage numérique, électronique ou pérenne, préservation numérique, conservation numérique, stockage sécurisé, ou encore gestion de collection numérique-, mais ne renvoient pas aux mêmes procédés. La gestion de collection numérique par exemple comprend des processus visant à améliorer les données, mais il n'est pas encore question de pérennisation. Le stockage sécurisé (via des coffres forts numériques) suppose une garantie d'intégrité des fichiers et de leur sécurité (gestion des droits d'accès par exemple, de mots de passe, accès par connexion avec HTTPS, données chiffrées, etc.). Mais il n'est pas non plus question de gestion à long terme car aucun travail sur les formats n'est prévu.

Archivage n'est donc pas stockage – les deux notions sont à distinguer : le stockage implique une sauvegarde de contenus (papier, format électronique) entreposés en un lieu sûr dans l'attente d'un traitement ou d'une utilisation ultérieurs. De la même façon, la sauvegarde est une opération technique qui permet d'assurer « la continuité de l'exploitation d'un système informatique en cas d'incident »<sup>1</sup>. Ces deux processus ont un objectif strictement sécuritaire et de court ou moyen terme. L'archivage pérenne demeure quant à lui un processus d'archivage au sens plein du terme, qui ne s'écarte pas des exigences du processus traditionnel d'archivage, lequel implique diverses étapes dans le traitement du document, de sa sélection à sa valorisation. Il est donc plus qu'un circuit technique : il nécessite également des moyens administratifs et humains.

L'expression d' « archivage pérenne » est quant à elle relativement récente et peu fixée. On trouve aussi les termes de préservation et de conservation, qui tous deux évoquent une continuité avec les activités traditionnelles de conservation des bibliothèques ou des archives. Quand « préserver » et « conserver » évoquent une résistance contre l'usure du temps, le verbe « pérenniser », moins statique, suppose une approche dynamique et proactive visant à garantir l'authenticité sur le long terme des documents<sup>2</sup>. Quant à l'adjectif « pérenne », l'étymologie aussi bien que la définition du dictionnaire restent vagues ; la littérature spécialisée propose comme objectif des durées de tantôt 30 ans, tantôt 50 (hors DUA réglementaires des documents administratifs). Aujourd'hui, la tendance est de ne pas se prononcer sur une durée par nature irréaliste et de définir la pérennité de l'archivage en

---

<sup>1</sup> BÉCHARD Lorène, FUENTES HASHIMOTO Lourdes, VASSEUR Édouard, *Les archives électroniques*, Paris : Association des archivistes français, 2014, p. 8.

<sup>2</sup> Le mot « pérennisation » est celui que l'on emploie pour traduire l'anglais « preservation » du modèle OAIS, présenté en I 3.

fonction des besoins : il s'agira donc de conserver les documents aussi longtemps que jugé nécessaire par ceux qui en ont la responsabilité.

L'archivage pérenne des documents numériques consiste à « conserver le document et l'information qu'il contient dans son aspect physique comme dans son aspect intellectuel, de manière à pouvoir le rendre accessible à long terme et à pouvoir demander une remontée d'archive »<sup>3</sup>. Il exige de « mettre en place une organisation et un ensemble de moyens et de procédures visant à ce que les informations pérennisées restent accessibles, compréhensibles et exploitables au cours du temps »<sup>4</sup>.

Afin d'assurer une conservation sur le long terme, la procédure d'archivage pérenne consiste à :

- assurer la traçabilité, l'intégrité, la sécurité, la pérennité des données ;
- répondre aux exigences légales et réglementaires de conservation et de communication ;
- lutter contre l'obsolescence technologique ;
- faciliter l'accès à l'information.

L'archivage pérenne devra donc non seulement préserver, mais également permettre de reconstituer les informations conservées et de les communiquer à nouveau : comme tout système d'archivage, il se doit de maintenir le lien entre le document et sa communauté d'utilisateurs potentiels.

Dès lors, se pose la question du rôle de la BU face à l'archivage pérenne, dont l'envergure excède les missions traditionnelles de l'institution : il mobilise en effet aussi bien une réflexion de nature archivistique et juridique que des compétences et des moyens informatiques. De plus, un projet d'archivage pérenne repose par définition sur une typologie d'acteurs large : la bibliothèque universitaire, mais aussi la communauté scientifique en tant que productrice et gestionnaire de contenus, l'université, qui peut impulser une stratégie générale d'établissement ou de site et soutenir les projets, les institutions de service public, scientifiques et universitaires, qui ont mission de valoriser et pérenniser la production de la recherche, ou encore les instances de l'État, à l'origine de la politique nationale de l'IST. Il s'agit donc de déterminer des périmètres d'action et l'échelle à laquelle doit être assumée la responsabilité de conservation de telle ou telle ressource – celle de la bibliothèque, de l'université ou de la collectivité.

Au sein de ce paysage complexe, c'est plus largement la responsabilité patrimoniale de la BU qui est en question ; or le corpus réglementaire et légal fixe peu d'orientations et d'obligations en la matière. Dans un cadre encore peu fixé, que peut et que doit assumer la BU ? Si elle ne prend pas en charge certains contenus, qui en a alors la responsabilité ? Les BU sont d'ores et déjà fortement sensibilisées à la notion de patrimoine scientifique ; actrices pour la plupart d'entre elles de circuits nationaux tel STAR, elles sont également très impliquées dans le

---

<sup>3</sup> Définition extraite du site de l'ABES, « STAR-Archivage pérenne », [En ligne] (page consultée le 29/12/2015) : <http://documentation.abes.fr/aide/star/accueil/exports.html>. Voir également la définition de l'archivage pérenne selon le CINES : <https://www.cines.fr/archivage/un-concept-des-problematiques/le-concept-darchivage-numerique-perenne/> [En ligne] (page consultée le 21/12/2010).

<sup>4</sup> BANAT-BERGER Françoise, DUPLOUY Laurent et HUC Claude, *L'archivage numérique à long terme : Les débuts de la maturité ?*, La Documentation française : Paris, 2009, p. 21.

mouvement de l'Open Access. Or la valorisation et la diffusion de l'IST constituent des problématiques qui ne peuvent être séparées de celle de l'archivage pérenne, apte à garantir sur le long terme une accessibilité qui est la raison d'être des archives ouvertes. Plus largement donc, on voit qu'un projet d'archivage pérenne, au-delà de ses aspects techniques, supposera une réflexion de nature politique sur l'accès à la production scientifique, sur les missions de service public d'un établissement (qui peuvent excéder celles de la seule BU), sur la visibilité à long terme de l'institution, mais aussi sur les coûts de maintenance des archives et sur l'efficacité de l'utilisation des fonds publics attribués aux projets de recherche ou de numérisation patrimoniale. Aux questions « par qui » et « comment » s'ajoutera donc un « pourquoi », qui pourra permettre de légitimer des projets d'utilité commune mais jugés coûteux pour les universités.

Face à la complexité d'un projet d'archivage pérenne, des questions stratégiques se posent tout d'abord : quels documents ou données doit-on archiver de façon pérenne ? Pour quelles raisons et selon quels critères ? Quelles sont les obligations légales ou réglementaires ? Quel est l'état de la science dans le domaine, et quelles sont les stratégies nationales et internationales dont il faut tenir compte ? Dans un second temps, un abord pragmatique s'avère nécessaire. Nous proposerons ainsi un bilan sur les retours d'expérience de bibliothèques universitaires qui mènent ou ont mené un projet d'archivage pérenne. Enfin, nous reviendrons sur la place et le rôle de la BU au sein de l'université et plus largement de l'ESR et sur les évolutions qui se dessinent, à moyen ou long terme, en particulier en termes de mutualisation. Qu'implique ce nouvel enjeu en termes de compétences et de formation ? Suppose-t-il une évolution du métier et des missions des BU ?

# I. STRATÉGIES ET ENJEUX DE L'ARCHIVAGE PÉRENNE DANS L'ESR

---

## 1. Un nouvel enjeu patrimonial ?

### A. Un enjeu émergent, de nouveaux risques

Dans le cadre spécifique des bibliothèques, la Charte de la conservation dans les bibliothèques définit ainsi la préservation du numérique : « La préservation des documents numériques se définit comme l'ensemble des principes d'organisation, des programmes informatiques et des infrastructures mis en œuvre en vue du contrôle et de la régénération des informations numériques et numérisées ainsi que des métadonnées qui leur sont associées. (Note : Le terme de préservation est préféré à celui de conservation dans le domaine numérique en raison du caractère anticipatif de ce mode de conservation) »<sup>5</sup>. Sur le plan des processus, l'archivage pérenne du numérique ne diffère pas en effet fondamentalement d'un archivage des ressources papier, et requiert également une « organisation » et une « infrastructure » ; au-delà des aspects techniques et de la veille à exercer sur les formats, les supports et les technologies, c'est une démarche archivistique, qui implique une organisation, la définition de critères de sélection et une politique d'archivage.

L'échelle de temps joue cependant ici un rôle majeur. A court ou moyen terme, nombre d'établissements considèrent qu'assurer les diverses tâches en lien avec la conservation du numérique est une mission qui peut revenir à l'établissement – aux professionnels du SCD, ou aux informaticiens – et qui peut donc assurée en interne. Un stockage informatique sécurisé de bon niveau peut en effet garantir contre la perte accidentelle du document. Mais à un horizon plus éloigné (plus de vingt ans), les risques sont réels que l'information ne soit plus accessible, plus lisible, ou plus compréhensible du fait de l'obsolescence matérielle, logicielle, du format de fichier ou par perte de la signification du contenu<sup>6</sup>. L'archivage pérenne vise à anticiper ces risques. Plus profondément, la sécurité des archives va au-delà de la préservation des données et documents contre la perte ou le vol : c'est aussi une question éthique, celle de la responsabilité envers une forme de mémoire commune, qui pose la question de la fonction patrimoniale des institutions. Car la préservation numérique constitue un enjeu de taille pour les bibliothèques universitaires, confrontées au défi de la constante augmentation de leur « cyberpatrimoine ».

Si la documentation papier est loin de disparaître en bibliothèques universitaires, et tendrait d'ailleurs paradoxalement à s'accroître<sup>7</sup>, la tendance actuelle est en effet à la dématérialisation. Passage progressif du support argentinique vers la numérisation, évolution très rapide des techniques liées au

---

<sup>5</sup> « Charte de la conservation », Direction générale des médias et des industries culturelles, Ministère de la culture et de la communication [En ligne] (page consultée le 29/12/2015) <http://www.patrimoineecrit.culture.gouv.fr/conservation-restauration/charte.php>

<sup>6</sup> Sur la notion d'obsolescence, voir le site du CINES : <https://www.cines.fr/archivage/un-concept-des-problematiques/le-concept-darchivage-numerique-perenne/> (page consultée le 29/12/2015).

<sup>7</sup> Voir le constat du rapport de Benoît Lecoq et Charles Micol, *Le stockage des collections imprimées à l'heure du numérique*, Rapport n° 2014-029, IGB, octobre 2014.

numériques, disparition progressive des moyens de productions de microformes, numérisation suite à l'obsolescence voire la disparition des équipements de restitution : les évolutions technologiques ont un impact crucial sur les supports. Au patrimoine numérique déjà acquis au fil des nombreuses entreprises de numérisation menées par les bibliothèques universitaires<sup>8</sup> s'ajoute un véritable « cyberpatrimoine »<sup>9</sup> qui se développe de façon exponentielle : données brutes de la recherche, articles scientifiques déposés dans les AOI en format natif numérique, publications numériques onéreuses acquises par les SCD (livre numérique natif (ebooks), revues en ligne), littérature grise numérique dont thèses numériques et numérisées, cours en ligne, documentation pédagogique, etc., sans compter les courriels, ou encore le *big data*, dont il ne sera pas question ici. La conservation, pérenne ou non, de ce patrimoine immatériel, pose autant de problèmes de stockage que le papier en termes de coûts et de besoins d'organisation. Elle nécessite également une anticipation, du fait de l'obsolescence rapide des formats d'une part, de la potentielle disparition des documents d'autre part lorsque la conservation est opérée uniquement sur des serveurs en interne.

La croissance exponentielle de ce cyberpatrimoine - selon certaines estimations<sup>10</sup> les besoins en matière de stockage de données numériques augmentent de 127 % par année - représente un enjeu majeur pour les établissements. Les réponses immédiates concernent d'abord le stockage – constitution de centres de ressources informatiques (CRI) ou de Data Center (encore peu nombreux, mais plusieurs projets sont en cours). L'archivage pérenne est quant à lui encore peu répandu : car il implique une réflexion à très long terme ; il n'est pas toujours ressenti comme une nécessité, surtout en regard d'autres enjeux plus actuels – valorisation, diffusion et accessibilité des collections principalement ; ou encore en raison de la technicité et des coûts d'une telle démarche. Mais pour autant, une planification anticipée de la préservation numérique est indispensable, sous peine de faire peser des risques aux collections.

Le risque qui pèse sur le document numérique consiste tout d'abord en la perte de données – par « crash » d'un serveur, risque non négligeable<sup>11</sup> ; par impossibilité de restituer l'information dans son intelligibilité et son intégrité (obsolescence des formats et des supports). Dans le cas de données ou de fichiers nativement numériques (données de la recherche, thèses archivées au format numérique entre autres), ou de fichiers perdus, détruits ou voués à une perte prochaine (cas de documents menacés de destruction par la fragilité des papiers utilisés par exemple), c'est la valeur des données elles-mêmes qui peut être engagée, certaines n'étant pas reproductibles (enregistrements audio de sources désormais indisponibles, fichiers d'origine perdus ou illisibles, etc.).

Dans le cas des collections patrimoniales, pour lesquelles on conserve l'exemplaire papier, c'est la valeur ajoutée par le travail intellectuel des équipes qui est en jeu, telle la cohérence du fonds numérique constitué : par exemple, dans la bibliothèque numérique IRIS (SCD de Lille 1) les collections numérisées ne

---

<sup>8</sup> Pour un panorama du patrimoine numérique des bibliothèques universitaires, voir le catalogue Patrimoine numérique du MCC : <http://www.numerique.culture.fr/pub-fr/index.html>

<sup>9</sup> Expression de Frédéric Blin, in CAVALIER François et POULAIN Martine (éd.), *Bibliothèques universitaires: nouveaux horizons*, Paris, France: Éditions du Cercle de la librairie 2015, p. 85.

<sup>10</sup> Voir MILNER John, « A UK research data service (UKRDS) : the way forward for research data management ? », *Serials : The Journal for the Serials Community* (22) 1 : 83-85, 2009.

<sup>11</sup> Plusieurs cas de « crash » des disques durs des services informatiques ont été recensés durant les entretiens menés pour ce mémoire, le risque n'est donc pas inexistant.

proviennent pas toutes de fonds papier conservés à Lille 1 et résultent des choix de constitution opérés par les bibliothécaires et les chercheurs ; il en est de même pour Medic@, où des documents appartenant à d'autres bibliothèques sont numérisés, s'ils permettent de compléter les collections initiales ; sans compter les ajouts de documents numérisés à la demande, par exemple des particuliers. La perte totale ou partielle des fichiers, des données ou des métadonnées a aussi un impact en terme de coût pour la collectivité : coût humain (il faut compter 25 à 30 heures de travail par ouvrage numérisé en moyenne, production des métadonnées comprise<sup>12</sup>), éventuellement coût de la sous-traitance si la numérisation est effectuée par un prestataire, mais aussi prix du matériel mobilisé, ou des développements informatiques requis. Nombre de bibliothèques universitaires ont en effet lancé depuis les années 2000 des projets de valorisation numérique en profitant des progrès technologiques pour mettre à disposition des publics leurs ressources patrimoniales et scientifiques, par le biais entre autres de bibliothèques numériques<sup>13</sup>. Le contenu de ces bibliothèques numériques représente un coût d'investissement élevé, qu'il s'agit de préserver.

Les risques peuvent être d'origines très diverses<sup>14</sup> : environnementaux ; organisationnels ; liés aux supports d'enregistrement ; technologiques (en lien avec l'environnement matériel et logiciel concernant la lecture d'un document : obsolescence des formats et de leur environnement, etc.) ; ou concernant l'accessibilité technique. Pour se prémunir de ces risques, et à l'instar des plans de conservation que les SCD mettent en place pour leurs ressources papier, un projet préventif de préservation des documents numériques peut être envisagée. Un archivage pérenne envisage pour chaque risque une réponse technique ou une action liée au management (plan d'urgence, duplication hors site, plan de sécurité des systèmes d'information, veille, contrôles, alertes, formation, accompagnement au changement, mixité des technologies utilisées, documentation des environnements de consultation, choix de formats à risques limités, collecte d'informations (métadonnées, mots de passe), retrait des mécanismes de protection, collecte d'informations (métadonnées, documentation associée, etc.). Si, comme pour toute activité, il ne peut être question de supprimer tout risque, il demeure possible de déterminer un niveau de risque acceptable, qui sera à définir selon les besoins des utilisateurs et les moyens des « producteurs » des données et des établissements. La gestion de risques que propose l'archivage pérenne, et qui s'appuie sur le modèle OAIS (voir infra), est un ensemble de bonnes pratiques qui peut s'appliquer à toute collection et toute institution, le modèle OAIS étant un modèle abstrait, qui ne spécifie ni l'échelle, ni la nature des « objets » concernés. Certains outils permettent aux professionnels d'évaluer les risques de leur collection numérique, comme DRAMBORA, « Digital Repository Audit Method Based on Risk Assessment »<sup>15</sup>.

Les tiers archiveurs proposant un archivage pérenne intègrent bien sûr une gestion proactive des risques dans leurs processus, tel le CINES, qui propose cette définition de la gestion de risques :

---

<sup>12</sup> Source : Alexandra Gottely, responsable des collections patrimoniales et numérisation, bibliothèque Cujas.

<sup>13</sup> Cf. ANDRO Mathieu, ASSELIN Emmanuelle, MAISONNEUVE Marc, *Bibliothèques numériques : logiciels et plateformes*, Tosca consultants, Paris : ADBS Éditions, 2012.

<sup>14</sup> Sur ce point, voir BANAT-BERGER Françoise, DUPLOUY Laurent, HUC Claude, *L'archivage numérique à long terme : les débuts de la maturité ?*, op. cit., p. 147-160.

<sup>15</sup> DRAMBORA, [En ligne] (page consultée le 30/12/2015), <http://www.repositoryaudit.eu/>

« Ensemble des dispositions prises pour anticiper et prévenir les problèmes avant qu'ils ne surviennent effectivement. Au sein du service PAC, cette gestion prend la forme d'un plan de gestion des risques, composé d'une liste récapitulative et structurée des risques, d'un tableau de description, d'évaluation et de mitigation des risques (tableau de suivi), ainsi que d'un document de spécifications du plan de gestion des risques. »<sup>16</sup>

### B. Un prolongement de la vision patrimoniale ?

La Charte de l'UNESCO sur la préservation du patrimoine numérique en 2003 alertait déjà les producteurs scientifiques, les établissements culturels et les décideurs sur la fragilité du patrimoine documentaire numérique, et entérinait la nécessité de porter un nouveau regard patrimonial sur les objets numériques. Une conférence tenue en 2012 a réuni des professionnels des secteurs du patrimoine, des représentants gouvernementaux (ou décideurs politiques), des représentants de l'industrie informatique, pour « proposer des recommandations pratiques visant à assurer un accès permanent au patrimoine documentaire numérique ».

« Bien qu'aujourd'hui le savoir soit partout créé et rendu accessible en format numérique, il risque de disparaître et sa perte pourrait entraîner l'appauvrissement de l'humanité. Malgré l'adoption de la Charte de l'UNESCO sur la préservation du patrimoine numérique en 2003, les risques de disparition du patrimoine numérique restent encore méconnus. »<sup>17</sup>

La Recommandation de la Commission européenne sur la numérisation et l'accessibilité en ligne du matériel culturel et la conservation numérique de 2006 met quant à elle l'accent sur la « sauvegarde » du patrimoine européen et son accessibilité future<sup>18</sup>.

Les solutions d'archivage se sont depuis développées, et se sont emparées de cette problématique. « To help ensure that current and future generations have access to our collective digital heritage », « our organizational byline : Committed to our digital future »<sup>19</sup> : les arguments de « vente » ou d'*advocacy* des solutions d'archivage numérique (ci-dessus, ceux de Duraspace) reposent ainsi, au-delà des aspects techniques, sur la notion de patrimoine immatériel de l'humanité. Une bibliothèque numérique comme HathiTrust met également en avant la notion de bien commun, d'héritage et de conservation des biens publics, évoquant la transmission des collections numériques aux générations futures : « The HathiTrust Digital Library brings together the immense collections of partner institutions in digital form, preserving them securely to be accessed and used today, and in future generations »<sup>20</sup>.

Dans le cadre de l'ESR, l'archivage pérenne s'envisage selon BSN 6 sur la très longue durée, pour les formats aussi bien numériques que papier et pour des

<sup>16</sup> CINES, « Une plateforme en production » [En ligne] (page consultée le 29/12/2015) <https://www.cines.fr/archivage/une-plateforme-en-production/notre-glossaire/>

<sup>17</sup> UNESCO, « La Mémoire du monde à l'ère du numérique : numérisation et conservation », 26-28 septembre 2012, Vancouver, Colombie-Britannique, Canada, [En ligne] (page consultée le 30/12/2015) <http://www.unesco.org/new/fr/communication-and-information/events/calendar-of-events/events-websites/the-memory-of-the-world-in-the-digital-age-digitization-and-preservation/>

<sup>18</sup> [En ligne] (page consultée le 30/12/2015) <http://eur-lex.europa.eu/LexUriServ/LexUriServ.do?uri=OJ:L:2011:283:0039:0045:FR:PDF>

<sup>19</sup> DURASPACE.ORG [En ligne] (page consultée le 30/12/2015) <http://www.duraspace.org/about>

<sup>20</sup> HATHITRUST Digital Library, « HathiTrust Certified as Trustworthy Repository » [En ligne] (page consultée le 30/12/2015) <https://www.hathitrust.org/hathitrust-certified-as-trustworthy-repository>

documents ou données à la valeur très forte et à vocation patrimoniale. L'archivage pérenne peut alors être envisagé comme la réponse à une nouvelle vision patrimoniale à long terme des données, images et documents produits par les établissements de l'ESR (numérisations, produits pédagogiques, thèses, écrits scientifiques, données de natures diverses). Les établissements eux-mêmes affichent pour objectif de sauvegarder la mémoire scientifique de l'établissement - qui n'est plus seulement physique mais aussi immatériel- et sont conscients des enjeux de la préservation du numérique. La valeur symbolique d'un archivage sur le long terme du patrimoine intellectuel de l'université peut participer à motiver le lancement d'un projet. Le patrimoine n'est pas absent des obligations réglementaires des SCD, puisqu'il est bien mentionné dans le cadre régissant les obligations des conservateurs, qui ont la responsabilité des fonds « anciens, rares et précieux » selon le Code général de la propriété des personnes publiques<sup>21</sup>. L'archivage pérenne peut ainsi être inclus au sein des obligations découlant de cette mission, selon la manière dont on l'interprète.

### *C. Qu'archiver de façon pérenne ? Critères de sélection et typologie des documents à pérenniser*

Selon la nature des contenus, les critères de sélection et les décisions de conservation ne seront pas identiques.

#### **Une analyse préalable des collections**

L'archivage pérenne n'est pas en effet la réponse adéquate à tous les types de documents. Il est ainsi pertinent de définir des critères aptes à départager un archivage intermédiaire d'un archivage pérenne : certaines données « brutes » de la recherche par exemple n'ont pas vocation à être archivées de façon pérenne, et pourront être conservées par un archivage intermédiaire (plateforme ISAAC par exemple<sup>22</sup>). Un projet d'archivage pérenne devra ainsi déterminer les contenus numériques qui sont à préserver en priorité, ce qui suppose de recenser l'ensemble des ressources numériques produites ou reçues par la bibliothèque, de connaître les droits de diffusion et d'accès à ces ressources, d'évaluer la nécessité de les conserver, et de définir la durée de cette conservation. Cette démarche est globalement proche de la méthode archivistique, qui définit trois « âges » dans le cycle de vie des archives<sup>23</sup>. Déjà bien en amont du projet d'archivage pérenne donc, les projets de numérisation ou de constitution de collections numériques doivent procéder à une approche raisonnée, qui inclut une démarche d'audit des contenus, d'évaluation, de tri et d'inventaire, et qui devra répondre à un certain nombre de questions sur les collections : comment décrire et classer les contenus concernés, sachant que la logique de collections selon les concepts bibliothéconomiques et la logique d'objet supposée par l'archivage peuvent ne pas coïncider ? Qui est responsable de leur devenir ? Doit-on les conserver, et si oui,

---

<sup>21</sup> Code général de la propriété des personnes publiques, article L2112-1, LEGIFRANCE [En ligne] (page consultée le 29/12/2015)

<http://www.legifrance.gouv.fr/affichCodeArticle.do?idArticle=LEGIARTI000006361198&cidTexte=LEGITEXT000006070299>

<sup>22</sup> Sur ISAAC, voir III 1. B

<sup>23</sup> Archives courantes (phase de création), intermédiaire (phase de gestion et de conservation), puis destruction ou préservation sous forme d'archives définitives.

pendant combien de temps ? Il s'agit donc de proposer une définition des objets qui auront une vocation patrimoniale selon l'institution.<sup>24</sup>

Dans un premier temps, posons donc qu'une archive pérenne est constituée de préférence :

- d'objets issus d'archives définitives : on archivera uniquement des documents dont la valeur intrinsèque est reconnue suffisante pour occasionner les investissements financiers et humains liés à la procédure d'archivage pérenne ;
- des objets dits « primaires » (documents originaux). Ainsi, on n'archive généralement pas les PDF issus d'une opération de numérisation, et qui constituent le format de communication des documents sur le Web par exemple, mais les fichiers dits « masters »<sup>25</sup> : il est toujours possible de reproduire un PDF, quand le fichier « master » est unique ;
- enfin, d'objets dans un format identifié et vérifiable : format publié, normalisé...<sup>26</sup>

En amont, il importe que la numérisation respecte déjà des critères de mutualisation et de partage, l'archivage pérenne suivant alors naturellement. Il est ainsi conseillé de numériser surtout les documents rares, uniques ou à forte valeur ajoutée, c'est-à-dire à valeur patrimoniale ; et afin d'éviter les redondances, d'user des multiples possibilités techniques (moissonnage par le protocole OAI-PMH, entre autres) pour renvoyer l'utilisateur vers les numérisations d'ores et déjà effectuées par d'autres établissements. Cette approche sélective et de mutualisation permet également de réfléchir en amont à la question des coûts d'archivage, toutes les ressources universitaires ne pouvant être archivées pour des raisons de volumétrie évidentes. La Commission Bibliothèques numériques, dans son rapport 2012-2013, insiste sur ce point : « Il est également nécessaire de s'interroger sur ce qu'il convient d'archiver de façon pérenne, et de mettre davantage l'accent sur la valeur et la nécessité d'une sélection. Le caractère de rareté des documents doit être pris davantage en compte dans la décision de procéder à un archivage numérique pérenne »<sup>27</sup>. La valeur patrimoniale des données à archiver semble le critère déterminant la sélection des documents et des données. Cet abord des contenus par le critère de la valeur patrimoniale peut cependant s'avérer problématique. Plusieurs SCD enquêtés évoquent en effet l'intérêt relatif que peut présenter la préservation de telle ressource pour la communauté, ce qui revient à poser la question de la valeur du matériau envisagé, qui peut parfois paraître trop faible. Or la question, si elle peut sembler évidente à court terme, ne l'est pas sur une durée plus longue, un contenu pouvant acquérir une valeur historique ou

---

<sup>24</sup> Sur cette définition, et pour une première approche de la méthodologie de la préservation numérique, cf. WEBB Colin, PEARSON David, KOERBIN Paul, « Oh, you wanted us to preserve that?! Statements of Preservation Intent for the National Library of Australia's Digital Collections », *D-Lib Magazine*, volume 19, numéro 1/2, janv.-fév. 2013, National Library of Australia [En ligne] (page consultée le 30/12/2015). Doi:10.1045/january2013-webb

<sup>25</sup> Le terme « fichier master » vient de l'époque de l'analogique. A la fin de la procédure d'enregistrement, on disposait d'un fichier de référence à partir duquel se faisaient les copies qui à leur tour permettaient la reproduction. L'objectif était de disposer d'un enregistrement de référence sur un support aussi peu dégradé que possible. Aujourd'hui c'est le fichier obtenu en fin de processus d'enregistrement qui se nomme ainsi et qui devient la référence validée par le responsable. En effet l'enregistrement non plus analogique mais numérique ne se modifie pas car il est indépendant du support, ce qui n'est pas le cas de l'analogique. La notion d'original dans le monde numérique n'existe pas.

<sup>26</sup> Classification inspirée par les analyses de l'ABES, cf. « STAR-Archivage pérenne », [En ligne] (*op. cit.*).

<sup>27</sup> Commission Bibliothèques numériques, Rapport 2012-2013, p. 6 [En ligne] (page consultée le 30/12/2015) <http://www.enseignementsup-recherche.gouv.fr/cid75370/commission-bibliotheques-numeriques.html>

patrimoniale bien après le temps de sa réception. La réflexion à mener sur les collections est donc par nature complexe, et ne bénéficie pour l'instant d'aucun référentiel national d'aide à la décision. Pour d'autres types de production, les personnels des SCD peuvent également être amenés à se poser la question des droits : si les droits de diffusion ne sont pas acquis, l'archive conservera des documents dont il sera impossible de faire usage.

Pour des raisons comparables, il est également essentiel de connaître le statut légal des archives concernées. Le Code du Patrimoine prévoit précisément la destination des archives publiques. Par conséquent, il s'agit de considérer ce qui est prévu en fonction de la nature des archives. Si les archives sont considérées comme des archives publiques, l'archivage définitif devra être effectué au sein d'un service d'archives, selon la législation en vigueur édictée par le Code du Patrimoine<sup>28</sup>. L'archivage pérenne effectué au CINES ne constituera donc qu'une étape intermédiaire. Si les archives sont au contraire de nature privée, l'archivage pérenne au CINES peut représenter un mode d'archivage définitif. Cette différence de traitement explique que les archives de HAL, privées, puissent demeurer au CINES en archivage définitif, quand les documents versés dans la plateforme Huma-Num de la TGIR Adonis par le CNRS, opérateur de l'État, soient versés au CINES pour l'archivage intermédiaire, mais aux Archives nationales pour l'archivage définitif (à l'exception des données patrimoniales, des publications, ou des photothèques pour lesquelles Huma-Num est fournisseur d'archivage et qui ne constituent pas des archives publiques). Ce versement est rendu obligatoire par la loi. Il est donc nécessaire de faire nettement la partition entre archives publiques et privées, et de connaître le statut législatif des archives qu'on s'apprête à sauvegarder. Ainsi, les publications des chercheurs, soumises au droit d'auteur, bénéficient du statut du dépôt légal, et constituent des archives privées (la recherche relevant de la propriété intellectuelle) ; mais les brouillons des mêmes chercheurs, à supposer qu'on souhaite les archiver, constituent pour leur part des archives publiques. Les frontières sont parfois floues, et le statut ambigu de certains documents peut se trouver défini au cas par cas. Difficulté à laquelle s'ajoute celle de définir les DUA des données scientifiques.

### **Critères et typologies**

Des critères généraux de sélection peuvent être proposés selon les contenus, afin d'établir les conditions de l'éligibilité à l'archivage pérenne des objets numériques, en envisageant la valeur patrimoniale, intellectuelle et économique des documents et collections. Pour les bibliothèques numériques, on se demandera ainsi :

- le document est-il nativement numérique ? Ou est-il le produit d'un transfert de support ? Le problème se pose alors de façon différente, dans la mesure où il existe ou pas un support de substitution assurant une solution de pérennité.
- la numérisation est-elle une numérisation de sauvegarde ? (par exemple, cas de la numérisation d'exemplaires de presse se dégradant rapidement, ou sauvegarde d'*unica*).

---

<sup>28</sup> Les établissements de l'ESR dont les BU peuvent envisager un archivage pérenne au CINES pour leurs fichiers, que cet archivage soit intermédiaire ou définitif, le SIAF ayant délivré son agrément à l'opérateur dans les deux cas.

- quelle est la valeur de la bibliothèque numérique constituée ? Repose-t-elle sur sa complétude, sur sa valeur intellectuelle? Cette valeur nécessite-t-elle alors une démarche d'archivage pérenne ?

Pour les données de la recherche, l'archivage pérenne peut être envisagé comme solution par exemple pour les données de « longue traîne » ; pour les données orphelines<sup>29</sup> ; pour des données non reproductibles (données climatiques), car capturées en temps réel et donc irremplaçables ; au coût de production extrêmement élevé (données astronomiques) ; ou qui pourraient être reproduites, par exemple dans le cas des données expérimentales produites en laboratoires, mais avec un coût très important. Les données de référence déjà conservées ailleurs, ou certaines données issues de simulation, présentent un intérêt variable pour l'archivage pérenne.

Pour le fonds ancien, les critères établis traditionnellement pour les documents physiques peuvent quant à eux être repris comme critères d'éligibilité. Quant aux productions universitaires, pourront être retenus pour un archivage pérenne de préférence les documents validés par les scolarités de l'établissement, avec des critères qui seront définis en fonction des souhaits des UFR (note minimale ou pas, etc.).

Le NERC<sup>30</sup> propose également, suite à sa liste de critères de sélection des données dans le cadre d'un Data management plan, une liste de critères visant à guider la sélection de documents et de données pour un archivage pérenne. On relève ainsi les critères suivants, qui recoupent en partie les critères mentionnés ci-dessus :

- « Unicité (donnée ou document unique et non reproductible)
- Valeur historique et/ou scientifique (utilité pour l'avancée de la science ; données en appui de projets en cours ou de tendances scientifiques ; utilité pour les orientations futures de la communauté scientifique)
- Nouveauté (données qui créent un précédent)
- Intégrité et complétude (données et documents fournis avec leurs jeux de métadonnées)
- Potentiel de réutilisation et de re-distribution.
- Potentiel de citabilité et référencement (données qui pourront être citées ou référencées dans une publication)
- Intérêt économique : le coût de l'archivage, conçu non simplement comme stockage mais aussi dans ses aspects de management, partage, accès, sauvegarde et maintenance à long terme des données, suppose que la

---

<sup>29</sup> La LERU propose des critères de décision pour les données de la recherche : « Not all data can or should be kept for the long term. A selection and appraisal process for data to be retained has to take place, involving the definition of curatorial terms and conditions. (...) For data collections and raw data, the researcher should make a decision based on a number of factors, including costs in archiving data, the importance of making this type of research data available to the wider research community, the availability of suitable data curation infrastructure and services to undertake this activity ». In « LERU-Roadmap for Research Data –FOSTER », Advice Paper, n. 14, novembre 2013 [En ligne] (page consultée le 30/12/2015) <https://www.yumpu.com/en/document/view/36353062/1kdrob/27>

<sup>30</sup> Le NERC, Natural Environment Research Council, est la plus importante fondation (« funder ») anglaise pour la science environnementale, la formation et l'innovation, qui s'appuie sur les universités et les centres de recherche. <http://www.nerc.ac.uk/>

donnée ou le document présentent un intérêt économique ou que leur perte potentielle constitue un enjeu financier ». <sup>31</sup>

Plus concrètement, dans le cadre de l'ESR, BSN 6 propose de définir « plusieurs types d'objets numériques [qui] ont besoin d'être conservés :

- la production scientifique imprimée et non imprimée des établissements, notamment les thèses, les documents pédagogiques, la littérature grise, les archives scientifiques, et les archives administratives [on pourrait détailler cette production en mentionnant les articles, les conférences ou « Workshop », les livres ou chapitres de livres, les HDR, les mémoires, les brevets, les articles scientifiques de vulgarisation, les rapports...] ;

- les données issues d'opérations de numérisation : documents analogiques reproduits sous forme numérique, collections patrimoniales comme sources de la recherche, thèses numérisées ;

- l'édition électronique et les ressources acquises auprès d'autres producteurs ;

- les données brutes de la recherche » <sup>32</sup>.

Au-delà de ce périmètre, on notera que la diversité des données et des documents produits, en particulier en SHS, est très importante <sup>33</sup>. Pour autant, comme nous le verrons, le SCD n'a pas nécessairement la responsabilité de tous les contenus, en particulier concernant les deux derniers types de ressources de la liste élaborée par la BSN 6.

### **Et l'archivage du papier ?**

Si la problématique de l'archivage pérenne semble se concentrer en premier lieu sur les documents numérisés ou nativement numériques, il demeure indispensable de mener en parallèle une réflexion sur l'archivage pérenne des documents papier et sur un mode de conservation multi-support, tant du fait de la nécessité de la conservation et de l'accessibilité des collections en elles-mêmes que dans une optique de gestion de risque des documents numériques.

Les derniers états des lieux font état d'un accroissement continu des collections papier dans les universités, malgré l'essor du numérique <sup>34</sup>. Les établissements se trouvent confrontés au manque de place de stockage et au coût induit. En réponse, des solutions collaboratives ont vu le jour, via les plans de conservation partagée (PCPP), la création du CTLes, chargé, sous la tutelle du MENESR, de regrouper les collections des bibliothèques universitaires parisiennes puis de l'Ile-de-France, et la création d'un groupe de travail au sein des nouveaux CollEx, dans le cadre duquel la conservation partagée est appelée à tenir une place majeure dans la mesure où elle est conçue comme une amélioration des services à

---

<sup>31</sup> NERC, « NERC Data Value Checklist » [En ligne] (page consultée le 29/12/2015) <http://www.nerc.ac.uk/research/sites/data/policy/data-value-checklist>

<sup>32</sup> Olivier Rouchon, « RENATIS. Journée de présentation Bibliothèque Scientifique Numérique (BSN) », ENS Cachan, 4 Juillet 2012.

<sup>33</sup> Voir sur ce point JACOBSON Michel, LARROUSSE Nicolas, MASSOL Marion, « La question de l'archivage des données de la recherche en SHS (Sciences Humaines et Sociales) », *Archives et données de la recherche* (ICA/SUV 2014), juillet 2014, p. 4.

<sup>34</sup> Cf. LECOQ Benoît, MICOL Charles, « Le stockage des collections imprimées à l'heure du numérique », *op. cit.*, et F. BLIN Frédéric et POIROT Alain, « Le patrimoine des bibliothèques universitaires », in CAVALIER François, POULAIN Martine (éd.), *Bibliothèques universitaires: nouveaux horizons, op. cit.*

la recherche. Mais bien qu'engagée à divers niveaux, régionaux ou nationaux, la conservation partagée des collections physiques se heurte entre autres à l'abandon des projets locaux de silos de conservation (Lyon et Alsace entre autres) du fait des « arbitrages opérés sur les investissements publics »<sup>35</sup>. La réflexion se poursuit, avec notamment l'un des « segments » de l'infrastructure CollEx qui envisage des réserves mutualisées<sup>36</sup>, et semble vouée à passer de façon de plus en plus systématique par la mutualisation.

La réflexion sur la conservation du papier affecte la prise de décision en particulier dans les cas où il s'agit de substituer définitivement et entièrement une version numérique à une version papier. C'est alors que la question de la conservation pérenne se pose pour l'un ou l'autre support. Sur le plan de la gestion de risque, passer au « tout numérique » sans archivage pérenne des copies numériques effectuées constitue une prise de risque importante. Ainsi des universités qui souhaitent mettre en place un dépôt numérique des thèses d'exercice, non déposées dans STAR, se heurtent à la nécessité d'un archivage pérenne, ou dans l'impossibilité d'un tel projet, à celle de la conservation d'un exemplaire papier en cas de déficience du stockage sécurisé. Dans ce cas précis, la mutualisation répond en partie au problème, puisque la BIUS reçoit un exemplaire des thèses papier d'exercice. Aucune obligation légale ne régit cependant ce dépôt, ce qui pose problème quand une université passe au dépôt électronique : des liens renvoient vers les documents, la BIUS n'étant plus alors le dépositaire. Dans le cas d'autres documents, un archivage pérenne constitue la solution à la fois à la sécurisation des données et au stockage des exemplaires physiques qui ne sont plus considérés comme nécessaires.

Mais le « tout numérique » pose des questions en matière de gestion de risques. Outre que les supports de la microfiche ou de l'argentique sont également des supports de très bonne qualité, stables et pérennes, la multiplicité des supports offre une garantie de sauvegarde par répartition des risques. Ainsi les grands projets d'archivage s'appliquent-ils à envisager la conservation multi-supports dans le cadre d'une approche stratégique de l'archivage pérenne : le JISC<sup>37</sup> par exemple le préconise ; la bibliothèque du Congrès quant à elle demande aux universités américaines de continuer à recevoir les thèses sur microfiches<sup>38</sup>. Ithaka S+R— l'un des services de l'organisation Ithaka, qui comprend aussi PORTICO et JSTOR—conseille aux bibliothèques de conserver des copies papier dans certaines circonstances<sup>39</sup>.

---

<sup>35</sup> CAVALIER François et POULAIN Martine, *op. cit.*, p. 80-81.

<sup>36</sup> « Reconnu et partagé depuis près de trente ans par les responsables des politiques documentaires et universitaires de tous les grands pays, l'objectif visant à construire des réserves documentaires suppose que des réflexions, et des moyens adaptés soient mis en place dans chacun des domaines concernés : localisation, bâtiment et équipements spécifiques, ressources humaines, structures de gestion (...) une réserve mutualisée est conçue pour traiter efficacement des documents, des ensembles hétérogènes provenant des bibliothèques (des universités, des grands établissements, des collectivités territoriales, des centres de recherche...) ; elle est conçue pour regrouper les collections spécialisées et les rendre disponibles à distance (quitte à traiter l'impact des droits d'auteur), ou sur place pour des chercheurs. Elle accueille prioritairement les ensembles documentaires constitués dans le cadre des pôles scientifiques ou des centres d'excellence préexistants ». Cf. « Le stockage des collections imprimées à l'heure du numérique », *op. cit.*, p. 59-60.

<sup>37</sup> Anciennement Joint Information Systems Committee, organisation non gouvernementale de Grande-Bretagne.

<sup>38</sup> Library of Congress, « Digital preservation », [En ligne] (page consultée le 19/10/2015) <http://www.digitalpreservation.gov/>

<sup>39</sup> Ithaka S+R a publié ses recommandations dans « What to Withdraw? Print Collections Management in the Wake of Digitization », <http://www.sr.ithaka.org/publications/what-to-withdraw-print-collections-management-in-the-wake-of-digitization/>. Les cas dans lesquels le papier devrait être conservé sont les suivants, selon Daniela Bone et Peter Burns : « - Very rare or unique print copies that represent historic or aesthetic value (...) - Backup copies to fix scanning errors (...) - Initial scanning standards are inadequate or technology improves. (...) - Loss of digitized material requiring

La question se pose en particulier lorsque l'original est susceptible d'avoir une valeur patrimoniale <sup>40</sup> ainsi que pour les thèses. Depuis 2006, les thèses sont déposées au format numérique ; certains établissements n'ont pas encore adopté le dépôt numérique, mais l'ensemble des thèses sera déposé en numérique d'ici 2019, et la microfiche comme support de diffusion, abandonné. L'ANRT, chargé du microfichage des thèses (plus de 200 000 thèses microfichées par l'organisme à ce jour), devra alors changer d'activité. Plusieurs options ont été proposées pour réorienter son activité : soit numériser les thèses sur microfiches, les fichiers du CINES constituant alors un backup des fichiers de l'ANRT ; ou bien opérer une sauvegarde sur support argentique. Le groupe de travail de BSN 6, que l'ANRT a rejoint, n'a pas retenu ces options, mais prend en compte le multi-support dans sa réflexion. Les préconisations conjointes de l'ADBU et de l'EPRIST (2014) suggèrent par ailleurs de « mettre en place un schéma global d'archivage pérenne des thèses, qu'il soit mixte (analogique/électronique) ou exclusif en termes de support » et « partant, clarifier et repositionner les missions des acteurs concernés (ANRT, CINES, ABES) ». <sup>41</sup>

## 2. Le paysage institutionnel et les stratégies nationales et internationales

Face à une problématique encore peu développée dans les établissements, et à un paysage scientifique et technique mouvant, les recommandations et les solutions susceptibles de soutenir les professionnels dans leurs décisions et leurs pratiques émanent de l'État, des opérateurs mandatés par l'État, des grandes infrastructures publiques de recherche, des initiatives européennes, du groupe PIN, ainsi que de nombreux gisements d'informations.

### A. Les impulsions et orientations stratégiques de l'État

Pour la politique d'archivage pérenne de l'ESR, l'État « joue un rôle d'accompagnement stratégique des établissements et définit les grandes orientations des politiques nationales pour obtenir les meilleurs résultats au meilleur coût » <sup>42</sup>. Son rôle est de soutenir les opérateurs (CINES, BnF) et les programmes de portée nationale, d'organiser la coordination des acteurs de l'ESR

---

the print originals to be scanned again. (...) - Reliability of access, including such terms of access as licensing and pricing. - Scholarly needs, particularly regarding high-quality images. - Local faculty needs and campus politics. University libraries should take into account the needs of their institution's faculty, who may be reluctant to dispose of print entirely » (cf. BONE Daniela and BURNS Peter, « An Overview of Content Archiving Services in Scholarly Publishing », Allen Press, Inc., [En ligne] (page consultée le 30/12/2015) <http://bit.ly/1kfvfWB>).

<sup>40</sup> Voir sur ce point Frédéric Blin : « Une erreur fondamentale serait de croire que la préservation de la version physique originale d'un document numérisé serait inutile. D'une part, l'original a une valeur patrimoniale en soi et, d'autre part, la durée de la pérennité des données numériques n'est aujourd'hui de l'ordre que d'une trentaine d'années. Numériser pour faire de la place sur les rayonnages est pour l'heure un danger réel pour la transmission du patrimoine et de la mémoire. Les exemples de pertes des données conservées sur des CD-ROM sont légion, sans parler de la qualité technique des numérisations réalisées il y a quelques années, insuffisante par rapport aux besoins d'aujourd'hui », in « La numérisation des bibliothèques », *Pour la science*, novembre 2013, n. 433, p. 48- 51.

<sup>41</sup> ADBU, « Politiques de sites et IST : les préconisations ADBU / EPRIST » [En ligne] (page consultée le 18/09/2015) <http://adbu.fr/politiques-de-sites-et-ist-les-preconisations-adbu-eprist/>

<sup>42</sup> LADANT Julie, « La conservation pérenne des collections de l'ESR : enjeux, actions et perspectives au niveau national », Journées professionnelles du CTLes, vendredi 21 juin 2013 [En ligne] (page consultée le 18/09/2015) <https://www.ctles.fr/sites/default/files/ctles/documents/JPro2013Ladant.pdf>

pour éviter les redondances et garantir la cohérence, d'inciter à la mutualisation (dans le cadre des contrats de site notamment), de mesurer et évaluer l'impact des politiques publiques (indicateurs, statistiques, enquêtes, etc.) et d'organiser la coopération avec d'autres initiatives ou politiques publiques (autres ministères, projets européens, etc.) Globalement, la politique du MENESR définit :

- un opérateur national pour l'ESR, le CINES – ce qui exclut a priori la BnF comme interlocuteur des SCD ;
- un cadre de concertation et de réflexion, le segment 6 de la BSN ;
- un cadre de concertation interministériel pour le numérique, la Commission Bibliothèques numériques présidée par Pierre Carbone, qui n'est plus active.

Dans le contexte de l'autonomie des universités (lois dites LRU (2007) et ESR (2013)), le rôle de l'État ne peut être que d'incitation et de coordination. Pour autant, il est nécessaire d'établir une politique globale afin de sauvegarder à long terme le patrimoine scientifique numérique, mission qui incombe à BSN 6.

Le segment 6 de la BSN, Bibliothèque Scientifique Numérique<sup>43</sup>, est consacré à l'archivage pérenne. Notons que le travail sur l'archivage pérenne est toujours en cours au sein du groupe de travail de BSN 6 et que les stratégies sont encore matière à discussion, d'autant plus que la question des archives scientifiques et la problématique émergente des données de la recherche nativement numériques et de leur masse en croissance exponentielle reposent la question de l'archivage pérenne sous un angle nouveau. Des recommandations générales sont émises entre 2010 et 2012 afin d'impulser une première dynamique dans le domaine de l'archivage pérenne : on voit que la problématique est récente (moins de 5 ans de regard rétrospectif à porter). Plusieurs « époques » peuvent être dégagées, dénotant plusieurs approches stratégiques :

- dans un premier temps (2010-2014), un appel à projet a été lancé afin d'inciter les établissements à initier une démarche d'archivage pérenne. Cet appel à projet instaurait des financements de numérisation dans le cadre de BSN 5. La reconnaissance par BSN 5, par le biais d'un label, permettait ensuite d'être éligible à des financements pour l'archivage pérenne dans le cadre de BSN 6 : le label ainsi imaginé ouvrait droit à un archivage par le CINES au vu d'un certain niveau de qualité de la numérisation. Notons qu'à l'époque, l'archivage pérenne est intrinsèquement lié à la notion de bibliothèque numérique. Pour autant, cette politique est jugée au sein de BSN complexe à mettre en œuvre, et les priorités de cet archivage pérenne, trop peu claires. Le problème du financement s'est également posé.
- dans un second temps (2014-2015), la problématique de l'archivage est ré-envisagée. Il s'agit alors de définir précisément l'archivage pérenne et ses usages : sa différence par rapport au stockage sécurisé courant, ou la question des archives intermédiaires (pour les essais cliniques par exemple, produits par les universités de santé et de médecine, l'obligation légale de conservation est de 50 ans : quelle stratégie adopter à l'égard de ce type de documents?).

---

<sup>43</sup> BSN [En ligne] (page consultée le 30/12/2015) <http://www.bibliothequescientifiquenumerique.fr/>

Dans le cadre de la réflexion menée au niveau ministériel sur la coordination des divers acteurs et la mutualisation des moyens, la Commission interministérielle Bibliothèques numériques consacre l'un de ses trois axes de travail à l'archivage pérenne. La Commission est créée en 2011 pour poursuivre la réflexion initiée en 2009 dans le cadre du Schéma numérique des bibliothèques<sup>44</sup>. Présidée par Pierre Carbone, elle est composée à parité de membres désignés par le MCC et le MENESR et inclut dans ses principaux axes de travail l'archivage pérenne des documents numériques. Différents groupes de travail avaient été mis en place au sein du MENESR à la suite de ce rapport. La dimension interministérielle de cette Commission était en accord avec une vision large et mutualisée de l'archivage pérenne. Entre autre, la Commission envisageait de confier un rôle clef au groupe PIN.

Le groupe PIN<sup>45</sup>, « Pérennisation des informations numériques », a été mis en place en 2000 au sein de l'association Aristote pour toutes les catégories d'archives - scientifiques, patrimoniales, opérationnelles. Il est composé d'ingénieurs en sciences de l'information, d'experts en archivistique, de spécialistes des normes, de formateurs, de gestionnaires de dépôt numérique, de juristes ... Groupe de travail qui a pour vocation de proposer un cadre d'échanges et de retours d'expérience, il ne propose en son nom propre aucune orientation spécifique. Les principaux acteurs français de l'archivage participent au groupe PIN, tels le CINES, la BnF, le CNES ou le SIAF. Seule grande instance nationale de réflexion et de partage sur la problématique, le groupe PIN a pour équivalent NESTOR<sup>46</sup> en Allemagne, ou l'IIPC<sup>47</sup> et la NDSA<sup>48</sup> aux États-Unis.

Le groupe de travail sur la préservation numérique de la Commission suggérait que les actions soient menées au niveau de l'État et arguait de la nécessité de mettre au point une politique concertée des grandes institutions (MENESR, MCC). Pierre Carbone, Inspecteur des bibliothèques, est alors missionné et préconise de s'appuyer sur le groupe PIN et son réseau d'experts, afin entre autres de produire des études. La commission Bibliothèque Numérique proposait également dans son rapport de « confier des études au groupe PIN en vue de constituer un centre de compétences partagées, et en premier une étude sur les modèles économiques de tiers-archivage dans le monde ». Ces préconisations ne seront pas mise en œuvre, de même que l'étude initialement prévue par BSN 6 sur un panorama des systèmes d'archivage pérenne en France. Le groupe PIN poursuit donc son activité, mais sans être mandaté par l'État, et émet des préconisations normatives (procédures, processus, normes, bonnes pratiques) quant au record management et au patrimoine, en toute neutralité. Le groupe PIN demeure

---

<sup>44</sup> Rapport de Bruno Racine, président de la Bibliothèque nationale de France, remis le 22 mars 2010 à M. Frédéric Mitterrand, Ministre de la culture et de la communication élaboré dans le cadre du Conseil du Livre. Le rapport formule 11 recommandations principales visant à accélérer le développement numérique des bibliothèques françaises. Rapport téléchargeable en ligne : <http://www.Enssib.fr/bibliotheque-numerique/notices/48219-schema-numerique-des-bibliotheques> (page consultée le 30/12/2015).

<sup>45</sup> Voir la présentation du groupe sur le site de l'association Aristote par Laurent Duploux (BnF) : <http://pin.association-aristote.fr/lib/exe/fetch.php/public/presentations/2008/pin20080519-laurent-duploux-mets.pdf>

<sup>46</sup> NESTOR [En ligne] (page consultée le 30/12/2015) : « Nestor is a cooperation association including partners from different fields, but all connected in some way with the subject of digital preservation ». [http://www.langzeitarchivierung.de/Subsites/nestor/EN/Home/home\\_node.html](http://www.langzeitarchivierung.de/Subsites/nestor/EN/Home/home_node.html).

<sup>47</sup> IIPC Net preserve.org [En ligne] (page consultée le 30/12/2015) <http://www.netpreserve.org/>

<sup>48</sup> Library of Congress, « National Digital Stewardship Alliance » [En ligne] (page consultée le 30/12/2015) <http://www.digitalpreservation.gov/ndsai/index.html>

fondamental pour l'archivage pérenne de l'ESR, fournissant le principal cadre d'échanges des professionnels sur la question.

### ***B. Les opérateurs de l'État***

Au sein des institutions culturelles en France, la préservation numérique est assurée :

- pour les musées, les bibliothèques municipales et les archives : par la BnF ;
- pour les centres de recherche (CNES, CNRS, etc.) : par le CINES.

D'autres acteurs publics pour l'archivage patrimonial existent, mais sans concurrence entre eux : l'INA, le CNES, la DGFIP. Les bibliothèques universitaires, faisant partie du MENESR, n'ont pas d'opérateur dédié, mais ont pour interlocuteur naturel le CINES, orientation encouragée par BSN 6. La BnF et le CINES sont agréés tous deux par le SIAF et procèdent au même type d'archivage pérenne en matière de normes, de contraintes et d'objectifs. Les deux institutions se livrent à des échanges sur leurs pratiques au sein des groupes de discussion que constituent le groupe PIN et le groupe de travail de BSN 6, proposent des orientations stratégiques au sein des divers groupes de travail auxquels ils appartiennent et travaillent avec les institutions sur les choix de formats de fichiers et de métadonnées.

Leurs stratégies sont donc concertées ; pour autant, la question des périmètres des deux opérateurs reste encore à clarifier.

#### **Le CINES**

Le CINES<sup>49</sup> est l'organisme pilote concernant l'archivage pérenne au sein de l'ESR ; il dépend conjointement de la DGRI et de la DGESIP. Ses deux missions sont le calcul intensif et l'archivage pérenne. Le périmètre d'archivage pérenne du CINES englobe théoriquement les données de l'ensemble des établissements de l'ESR<sup>50</sup>, dont les thèses et la plateforme de revues Persée. Les services du CINES s'assimilent à du tiers-archivage : les établissements établissent un contrat avec le CINES sous forme d'une convention précisant les rôles et responsabilités des parties et une contribution financière. Deux types d'archives sont concernées : les archives publiques intermédiaires<sup>51</sup> (durée de conservation limitée dans le temps (DUA entre 5 et 90 ans) et les archives publiques définitives : thèses, archives numérisées, données scientifiques.

Depuis 2006, date de la parution de l'arrêté sur le dépôt électronique des thèses, une responsabilité réglementaire de conservation est établie, et le CINES a été choisi comme opérateur mandaté pour une conservation mutualisée. Depuis, le décret a été modifié pour inscrire l'archivage pérenne des thèses dans les missions de l'établissement. Le CINES, qui n'archivait en 2009 parmi les établissements du MENESR que les thèses par l'intermédiaire de l'ABES, a donné immédiatement

---

<sup>49</sup> CINES [En ligne] (pages consultées le 30/12/2015) <https://www.cines.fr/> ; <https://www.cines.fr/presentation/missions/>

<sup>50</sup> Liste des fonds conservés au CINES : <https://www.cines.fr/archivage/nos-references-nos-partenariats/les-fonds-conserves-au-cines/> [En ligne] (page consultée le 30/12/2015)

<sup>51</sup> Offre rendue possible par l'obtention d'un agrément du Service Interministériel des Archives de France le 14 Décembre 2010.

son accord en 2009 pour accompagner la bibliothèque Cujas dans son projet d'archivage pérenne. La BSG a obtenu à la même époque un accompagnement comparable, mais a fait le choix d'une sous-traitance par un prestataire. Ces deux établissements ont été les premiers à s'engager dans une démarche d'archivage pérenne. Les bibliothèques universitaires constituent ainsi historiquement les premiers utilisateurs du CINES pour sa mission d'archivage pérenne : les relations avec l'opérateur sont donc établies de longue date. Depuis, si le CINES tend à favoriser les projets à grosse volumétrie, ainsi que l'archivage pérenne des données scientifique, il n'oppose jamais de refus à un projet d'archivage pérenne d'un établissement de l'ESR même pour de très petites volumétries, comme c'est le cas par exemple à Lille 2 ou à l'université de Lorraine (où il est question de quelques centaines de gigas), en particulier quand il est question de sauvegarde des thèses et mémoires par exemple, documents qui sont en cohérence avec sa mission nationale d'archivage pérenne des thèses de doctorat. Les bibliothèques universitaires représentent environ un quart des services versants au CINES.

Les infrastructures de stockage reposent sur une convention entre le CINES et l'IN2P3 (CNRS) pour la réplication de données. L'organisme, dont l'archivage est effectué sur la plateforme PAC<sup>52</sup> (16 millions de fichiers archivés en 2015<sup>53</sup>), assure des études et une veille technologique, et fournit un haut niveau de service et d'accompagnement aux services versants : « disponibilité des équipements 24x7x365, supervision, sécurité, etc. ; accompagnement de l'établissement pendant la phase de mise en place du projet : participation d'un archiviste et d'experts ; assistance de niveau 1 et 2 (incidents, support, suivi) pendant l'exploitation du projet. »<sup>54</sup>

L'équipe dédiée à l'archivage pérenne est composée de 10 personnes. Le pool de référents projets a récemment été étendu. Une gestion de projet implique 5 à 6 personnes, qui assurent dans le même temps d'autres fonctions au sein de l'équipe. Durant le projet, un interlocuteur unique de profil informaticien assure le suivi des tests, des connexions, etc., tandis qu'un archiviste intervient sur chaque projet au début pour en définir les modalités, et soutenir l'élaboration du PPD, des paquets d'archivage, ou du mapping de métadonnées. Les coûts de personnels sont ainsi estimés :

« Coût des ressources humaines pour la mise en place d'un projet : 0,2 ETP / an ;

• Coût des ressources humaines pour l'exploitation d'un projet : 3,5 ETP ;

• Coûts des ressources humaines pour l'assistance aux utilisateurs : quelques jours / an »<sup>55</sup>.

---

<sup>52</sup> Sur PAC, voir ROUCHON Olivier, « PAC, la plateforme d'archivage pérenne des documents électroniques au CINES », in LEBLOND Corinne (dir.), *Archivage et stockage pérenne*, Lavoisier, 2009, p. 21-40 ; MASSOL Marion, BECHARD Lorène, « Le service d'archivage électronique à long terme du CINES et la plateforme PAC », Dossier « Archivage numérique pérenne », *La Gazette du CINES*, février 2013, p. 10 et suiv. ; CINES, « La plateforme PAC et ses évolutions fonctionnelles et techniques », 3ème journée des utilisateurs de l'archivage, 9 juin 2015, [En ligne] (page consultée le 29/12/2015) <https://www.cines.fr/wp-content/uploads/2015/06/PAC-Journ%C3%A9eUtilos-2015-06-09aprem.pdf> ; et sur le site du CINES, « Une plateforme en production » [En ligne] (page consultée le 30/12/2015) <https://www.cines.fr/archivage/une-plateforme-en-production/>

<sup>53</sup> <https://www.cines.fr/la-500-000eme-archive-au-cines/>

<sup>54</sup> ROUCHON Olivier, *Le modèle d'évaluation des coûts de l'archivage numérique*, CINES, Réunion plénière, 14 Mars 2013, [En ligne] (page consultée le 30/12/2015) [http://pin.association-aristote.fr/doku.php/public/reunion\\_pleniere/cr/cr\\_20130412](http://pin.association-aristote.fr/doku.php/public/reunion_pleniere/cr/cr_20130412)

<sup>55</sup> Ibid.

On voit que l'un des atouts de l'opérateur repose sur les personnels très qualifiés consacrés au suivi des projets, pour la conduite desquels aucune limitation de durée n'est prévue.

### La BnF et les périmètres CINES-BnF

La plateforme d'archivage de la BnF, qui repose comme PAC sur le modèle OAIS, est SPAR (Système de Préservation et d'Archivage Réparti)<sup>56</sup>. La masse de données y est plus importante que dans le monde académique : la plateforme archive ainsi 2 pétaoctets en octobre 2015, soit plus de 4 millions de paquets, ou encore 500 millions de fichiers. La BnF, qui y archive ses propres ressources, a également pour mission d'ouvrir ce système à d'autres établissements : ainsi, l'une des filières, ou « collections » de SPAR, est consacrée au tiers-archivage, la BnF assurant ainsi le même type de mission que le CINES.

Afin d'éviter la concurrence entre établissements publics, l'offre de service étant la même par ailleurs, une répartition a été définie par BSN 6 comme mentionné supra : les établissements dépendant de la tutelle du MCC s'adressent à la BnF, et ceux dépendant de la tutelle du MENESR, au CINES ; par ailleurs, une harmonisation des tarifs est également en discussion. Mais les établissements qui possèdent un patrimoine numérique cumulent à la fois la problématique de l'archivage pérenne et celle de la diffusion : deux projets distincts doivent ainsi être menés dans les SCD, l'un pour la mise en ligne et la valorisation, puis un second pour l'archivage pérenne, ce qui de plus fait courir le risque d'une absence de continuité technique ou temporelle entre les deux projets. Le « modèle » Gallica offrirait donc une offre « intégrée » séduisante pour les SCD, qui ne disposent pas au sein de l'ESR d'un service comparable assurant à la fois une diffusion et un archivage pérenne associé gratuit - gratuité qui s'inscrit dans le cadre de projets de coopération. Plusieurs SCD se sont d'ailleurs adressés à la BnF en espérant bénéficier de ce type de service. Mais pour les raisons de périmètres évoquées, les établissements de l'ESR ne peuvent y prétendre, non tant pour des raisons juridiques que du fait d'un « gentleman agreement »<sup>57</sup> respecté de part et d'autre. Ainsi, si la convention établie entre la BNU et le BnF, suite à un appel à partenariat lancé par la BNU auquel a répondu la BnF, a permis la mise en ligne de Numistral<sup>58</sup>, la bibliothèque numérique patrimoniale de la BNU, dans le cadre du lancement de Gallica Marque Blanche<sup>59</sup>, elle ne mentionne volontairement pas l'archivage pérenne de Numistral qui doit être assuré par une autre solution – projet encore en cours à la BNU.

---

<sup>56</sup> BnF, « SPAR, le système de préservation numérique de la BnF » [En ligne] (page consultée le 30/12/2015) [http://www.bnf.fr/fr/professionnels/spar\\_système\\_préservation\\_numérique.html](http://www.bnf.fr/fr/professionnels/spar_système_préservation_numérique.html). Voir sur SPAR : DERROT Sophie, FAUDUET Louise OURY Clément, PEYRARD Sébastien, « Preservation is knowledge: acommunity-driven preservation approach », *Proceedings of the 9th International Conference on Preservation of Digital Objects (iPRES)*, Toronto (Canada), octobre 2012, p. 1 à 8 [En ligne] (page consultée le 30/12/2015) <https://ipres.ischool.utoronto.ca/sites/ipres.ischool.utoronto.ca/files/iPres%202012%20Conference%20Proceedings%20Final.pdf>

<sup>57</sup> Source : Laurent Duploux, chef du service Numérisation, BnF.

<sup>58</sup> Bibliothèque Nationale Universitaire, « Numistral » [En ligne] (page consultée le 30/12/2015) <http://www.numistral.fr/> ; <http://www.bnu.fr/numistral/presentation-de-numistral>

<sup>59</sup> La BnF a développé une solution d'hébergement et de mise en ligne ouverte aux autres bibliothèques avec Gallica marque blanche. Le rapport d'activité 201369 de la BnF la définit ainsi : « [Il s'agit de] permettre la création de bibliothèques numériques développées et hébergées par la BnF, mais ne mentionnant ni Gallica ni la BnF. Il s'agit donc d'un service « dans les nuages », entièrement personnalisé aux couleurs du partenaire et disposant de toutes les fonctionnalités de Gallica existantes et à venir ». À ce jour, la principale réalisation de bibliothèque numérique passant par ce dispositif est Numistral de la BNU de Strasbourg. On peut y consulter les collections numérisées depuis 10 ans par la BNU et aujourd'hui intégrées dans Gallica. Les standards de description et de structuration des données doivent correspondre à ceux employés par la BnF, dont le DublinCore.

La question des périmètres entre établissements, d'ordre politique, est complexe, et appelle probablement, comme le préconisait le rapport de la Commission Bibliothèques Numériques, « une plus grande mutualisation entre les deux secteurs (Culture ; Enseignement supérieur et recherche) », qui propose également de procéder à des répliquions croisées « qui pourraient constituer une solution intéressante, particulièrement pour les numérisations les plus lourdes comme les fonds d'images » afin d'optimiser les coûts<sup>60</sup>.

### ***C. Les infrastructures de recherche du CNRS : IRHT, CCSD et TGIR-Huma-Num***

Des sections ou unités mixtes de recherche proposent des services ou des outils susceptibles de concerner l'archivage pérenne d'une partie des productions des établissements de l'ESR.

L'IRHT, l'une des sections du CNRS, assure ainsi l'archivage pérenne de Liber Floridus, devenu récemment le catalogue « Initiale »<sup>61</sup>. Fruit d'un partenariat entre le MENESR, l'IRHT et le CINES, Liber Floridus était une base qui, à terme, visait à proposer à la consultation l'ensemble des enluminures des manuscrits médiévaux conservés dans les bibliothèques de l'enseignement supérieur. En cours d'enrichissement, le catalogue Initiale recense actuellement les manuscrits enluminés issus des fonds des bibliothèques Mazarine et Sainte-Geneviève, soit un ensemble de près de 1700 manuscrits et 33 000 images. C'est l'IRHT qui en gère l'archivage pérenne auprès du CINES, et non les établissements eux-mêmes. La base Liber Floridus était hébergée par le CINES. L'hébergement de ce site remonte à l'époque où le CINES offrait des services de diffusion mais pas encore de service d'archivage. Depuis lors, les missions ont été réorientées et les services de diffusion du CINES ont quasiment disparu.

Le CCSD<sup>62</sup> est quant à lui une unité mixte de service placée sous la triple tutelle du CNRS, de l'INRIA et de l'université de Lyon, qui propose plusieurs « produits » aussi bien aux organismes de recherche qu'aux universités, depuis la Convention de partenariat en faveur des archives ouvertes et de la plateforme mutualisée HAL signée par 26 acteurs de l'ESR en 2013. Par cette convention, cosignée par l'Agence de Mutualisation des Universités et Établissements (AMUE), la Conférence des Présidents d'Université (CPU), la Conférence des Grandes Écoles (CGE) et 22 établissements de recherche, les signataires « souhaitent mettre en œuvre une politique scientifique et technique commune en consolidant les moyens permettant aux chercheurs de partager les résultats de leurs travaux sous forme numérique, et permettre leur réutilisation ; de les diffuser

---

<sup>60</sup> Rapport 2012-2013 de la Commission Bibliothèques numériques, *op. cit.*, p. 6. Voir également Pierre Carbone: « S'il existe un partage clair des compétences d'archivage pour les institutions relevant de chaque ministère (MESR et MCC), il est toutefois souhaitable d'aller plus loin dans la coopération nationale en favorisant des politiques partagées de sauvegarde entre la BnF et le CINES, ce qui garantirait une plus grande sécurité à chacun des acteurs et permettrait des économies d'échelle », in « Numérique et archivage pérenne : regard du président de la Commission Bibliothèques Numériques », *Bulletin des bibliothèques de France* [en ligne], n. 5, 2013 [consulté le 11 octobre 2015] <http://bbf.Enssib.fr/consulter/bbf-2013-05-0043-010>

<sup>61</sup> IRHT/CNRS, « Initiale » [En ligne] (page consultée le 30/12/2015) <http://initiale.irht.cnrs.fr/accueil/index.php>

<sup>62</sup> Centre pour la Communication Scientifique Directe (UMS3668). [En ligne] (page consultée le 30/12/2015) <https://www.ccsd.cnrs.fr/>

immédiatement et gratuitement à la communauté scientifique internationale, ainsi qu'au grand public ; d'assurer leur conservation à long terme. »<sup>63</sup>

HAL, « Hyper article en ligne », archive nationale centrale française, est l'un des produits proposés par le CCSD à la communauté scientifique qui vise à la mise en œuvre de cette politique scientifique et technique commune, et à assurer de manière transparente pour les services versants une conservation à long terme de cette production. Les partenaires de la convention sus-citée s'engagent ainsi à inciter leurs chercheurs à déposer dans les AO et à concilier les archives institutionnelles et l'archive nationale HAL. Les partenaires, universitaires entre autres, qui ont une archive ouverte institutionnelle locale, peuvent dupliquer son contenu dans l'archive HAL, sous réserve d'éligibilité à la plate-forme. Les établissements peuvent également faire la demande d'un portail spécialisé dans HAL, créé et personnalisé par le CCSD. L'investissement, le développement, la gestion de l'infrastructure sont supportés par une unité du CCSD. Les documents nativement numériques, mais aussi certains produits de numérisation peuvent être déposés dans HAL. Outre la visibilité internationale de HAL, ces dispositions permettent aux archives institutionnelles locales de bénéficier de l'archivage pérenne de HAL mis en œuvre avec le CINES sans surcoût<sup>64</sup>.

DUMAS est un autre produit du CCSD<sup>65</sup>, qui a pour objectif de « valoriser les activités pédagogiques et de préparation à la recherche des établissements d'enseignement supérieur ». Mise en production en 2008, cette plateforme permet le dépôt des mémoires, des thèses d'exercice, de pharmacie, etc., documents pour lesquels il n'existe pas de solution nationale comme STAR. Le dépôt se fait sous la responsabilité des professionnels de la documentation (dans les UFR, les SCD, les laboratoires de recherche) ou des enseignants chercheurs.

Enfin, le TGE-Adonis, devenue la TGIR (Très Grande Infrastructure de Recherche) Huma-Num à la suite d'une fusion d'unités, est une Unité Mixte de Services (3598) associant le CNRS, l'Université d'Aix-Marseille et le Campus Condorcet qui vise « à faciliter le tournant numérique de la recherche en sciences humaines et sociales »<sup>66</sup>. Lancé en 2004 par le CNRS afin de répondre à « la croissance de la production de données », à « leur diversité » et à « la fragilité des dispositifs supports » dans le secteur des SHS, la TGIR Huma-Num développe, par le biais de consortiums regroupant des acteurs des communautés scientifiques, une production raisonnée des données et propose des recommandations scientifiques et des bonnes pratiques. Ses objectifs sont le traitement, la conservation, l'accès et l'interopérabilité des données de la recherche. Les services de la TGIR sont proposés sous la forme d'une grille de services dédiés, d'une plateforme d'accès unifié (ISIDORE) et d'une procédure d'archivage à long terme au CINES : la TGIR Huma-Num accompagne la phase de dépôt et finance cet archivage<sup>67</sup>. La communauté des producteurs de données, dans le périmètre des SHS, peut donc

---

<sup>63</sup> Ministère de l'éducation nationale, de l'enseignement supérieur et de la recherche, « Convention de partenariat en faveur des archives ouvertes et de la plateforme mutualisée HAL » [En ligne] (page consultée le 30/12/2015) <http://www.enseignementsup-recherche.gouv.fr/cid71277/parteneriat-en-faveur-des-archives-ouvertes-plateforme-mutualisee-hal.html>

<sup>64</sup> Les premiers documents de l'archive ouverte HAL gérée par le CNRS-CCSD ont été archivés dans PAC le 29 mars 2010. Il s'agit d'articles scientifiques pluridisciplinaires de niveau recherche ainsi que de thèses françaises ou étrangères disponibles en accès libre sur le portail HAL.

<sup>65</sup> CCSD, « Dumas », [En ligne] (page consultée le 30/12/2015) <http://dumas.ccsd.cnrs.fr/>

<sup>66</sup> Huma-Num, « A propos de Huma-Num » [En ligne] (page consultée le 30/12/2015) <http://www.huma-num.fr/>

<sup>67</sup> Huma-Num, « Archivage à long terme » [En ligne] (page consultée le 30/12/2015) <http://www.huma-num.fr/services-et-outils/archiver>.

bénéficier par l'intermédiaire d'Huma-Num des compétences du CINES pour l'archivage (pérenne ou intermédiaire) de ses données, quel que soit le service choisi (ISIDORE, NAKALA (pour l'exposition de données), etc.<sup>68</sup>

#### *D. Les initiatives européennes*

De nombreuses initiatives européennes – programmes, projets, alliances ou ligues- se sont emparées de la question de l'archivage des données (projets Eudat et PERICLES, organisation European Alliance for Permanent Access, l'Open Preservation Foundation, etc.) ; d'autres s'attachent à favoriser l'adoption de plans de gestion de données (« Data Management Plan »), ou encore à améliorer la qualité des métadonnées (projet SUCCEED, ligue LERU, consortium européen PREFORMA, etc.). La préservation sur le long terme est alors comprise comme l'étape finale de la vie de la donnée, bénéficiant de la formalisation et de la standardisation du cycle de vie préalable des données. Globalement, ces divers projets s'inscrivent dans un large mouvement émergent visant à structurer les modes de production, de diffusion, de partage et d'archivage des données de la recherche, perçus par l'Europe comme un enjeu scientifique majeur auquel il est nécessaire de sensibiliser aussi bien les institutions que les producteurs. On trouvera en annexe une présentation non exhaustive de ces initiatives, qui offrent des lieux d'échanges de connaissances et de compétences, des bonnes pratiques, des recommandations et préconisations. On trouvera également en annexe une présentation des gisements d'information et des sites effectuant une veille à destination des professionnels.

### **3. Coûts et exigences techniques et réglementaires de l'archivage pérenne**

L'archivage pérenne s'inscrit dans un cadre réglementaire et technique strict. Ses pratiques sont normalisées et exigent des infrastructures spécifiques et un management dédié. Si donc comme on a coutume de le dire, les coûts du stockage ne cessent de diminuer, ceux de l'archivage pérenne ne peuvent être réduits si ce n'est, dans une certaine mesure, par une mutualisation de certains processus. Reste alors à définir sur quels acteurs faire peser ces coûts dans le cadre de l'ESR.

#### *A. Agréments et processus, standards et normes : les exigences de l'archivage pérenne*

Des opérateurs publics ont reçu l'agrément du SIAF, Service Interministériel des Archives de France<sup>69</sup>; pour les bibliothèques, il s'agit du CINES, et de la BnF.<sup>70</sup> Le CINES reçoit l'agrément du SIAF dès 2010 pour

---

<sup>68</sup> Voir THIBAULT François et JOUVE Bertrand, « Les Infrastructures de recherche en sciences humaines et sociales. Rapport du groupe "Infrastructures" de l'Alliance Athéna [Alliance nationale des sciences humaines et sociales] », 40 p., p. 14-15 [En ligne] (page consultée le 30/12/2015) <http://www.cnrs.fr/fr/pdf/allianceathena/121012-infrastructure-sh/>

<sup>69</sup> Le Service interministériel des archives de France (SIAF) est en charge de la politique nationale des archives, audite les systèmes d'archivage et donne son agrément aux tiers-archivistes publics ou privés pour la conservation d'archives publiques courantes ou intermédiaires.

<sup>70</sup> D'autres acteurs publics assurent un archivage intermédiaire ou pérenne, notamment : le Centre national d'études spatiales (CNES) avec sa plateforme STAF en production depuis 1995, l'Institut national de physique nucléaire et de physique des particules (IN2P3) au Centre national de la recherche scientifique (CNRS), la Direction générale des

l'archivage pérenne. Il est ensuite agréé par arrêté du 24 janvier 2014 (renouvellement d'agrément) pour la conservation d'archives publiques courantes et intermédiaires, sur support numérique, sur sa plate-forme d'archivage électronique, hébergée au sein des centres serveurs de Montpellier (34) et de Villeurbanne (69).

D'autres certifications existent au niveau international ; ainsi l'institut néerlandais DANS a mis en place une certification, le DSA, Sata Seal of Approval<sup>71</sup>, qui distingue les entrepôts engagés dans l'archivage numérique à long terme des données aptes à satisfaire aux exigences de la norme OAIS (Open archival information system). Le CINES, ou encore l'archive de données EASY, elle-même créée par le DANS, ont obtenu cette certification, qui garantit aux services versants la qualité de l'archivage pérenne effectué.

Cette approche qualité<sup>72</sup>, essentielle dans l'archivage pérenne, s'appuie sur un ensemble de normes et de standards émis par une grande diversité d'organismes, selon les domaines et les pays. Les normes sont émises en France par l'AFNOR, l'organisme de normalisation (qui a pour équivalent le BSI en Grande-Bretagne, l'ANSI aux États-Unis, etc.). En Europe, le CEN (Comité Européen de Normalisation) a un rôle comparable à celui de l'AFNOR, de même que l'ISO au niveau international. Les standards quant à eux peuvent être élaborés par des groupes ou consortiums, tels le W3C, le consortium Unicode, la bibliothèque du Congrès qui maintient le standard METS, ou le CCSDS (Consultative Committee for Space Data System, organisme international de normalisation des agences spatiales). Certains standards deviennent des normes, tel le format de document PDF/A, employé aujourd'hui dans le dépôt électronique des thèses par exemple, ou le format d'image PNG<sup>73</sup>.

Issu de ce travail de normalisation, le modèle de référence OAIS<sup>74</sup> (Reference Model for an Open Archival Information System, norme ISO 14721) est le résultat d'un travail du CCSDS réalisé à la demande de l'ISO<sup>75</sup>. Adopté comme modèle de processus métier de référence par les plateformes d'archivage pérenne, c'est un modèle abstrait qui se focalise sur les risques spécifiques à la préservation du numérique. Il décrit les fonctions, les responsabilités et l'organisation d'un système préservant de l'information, en particulier des données numériques mais aussi des documents papier, sur le long terme, pour en garantir l'accès à une communauté d'utilisateurs identifiés. Le modèle concerne aussi bien les aspects informatiques que l'organisation humaine et administrative ou les aspects légaux. Il suppose notamment de préserver des copies de données en des endroits distants, de vérifier régulièrement l'intégrité des fichiers archivés, l'usage effectif des formats de données, de renseigner des métadonnées décrivant les documents en

---

finances publiques (DGFIP) avec sa plateforme ATLAS lancée en 2007, l'Imprimerie nationale, et enfin l'Institut géographique national (IGN).

<sup>71</sup> Data Seal of Approval [En ligne] (page consultée le 30/12/2015) <http://datasealofapproval.org/en/>

<sup>72</sup> Sur l'approche qualité du CINES, voir ROUCHON Olivier, « La démarche qualité pour la préservation à long terme des données numériques », Dossier « Archivage numérique pérenne », *La Gazette du CINES*, février 2013, p. 26.

<sup>73</sup> Voir sur ce point, et sur les divers aspects techniques de l'archivage pérenne, BANAT-BERGER Françoise, DUPLOUY Laurent, HUC Claude, *L'archivage numérique à long terme : les débuts de la maturité ?*, op. cit., p. 78.

<sup>74</sup> Sur ce modèle, voir BANAT-BERGER Françoise, DUPLOUY Laurent, HUC Claude, *L'archivage numérique à long terme : les débuts de la maturité ?*, op. cit., p. 35-79.

<sup>75</sup> Dans la lignée de l'OAIS, le CCSDS travaille sur des normes complémentaires : PAIMAS pour les protocoles de versement entre le Producteurs et l'Archive, PAIS pour les règles de constitution du SIP.

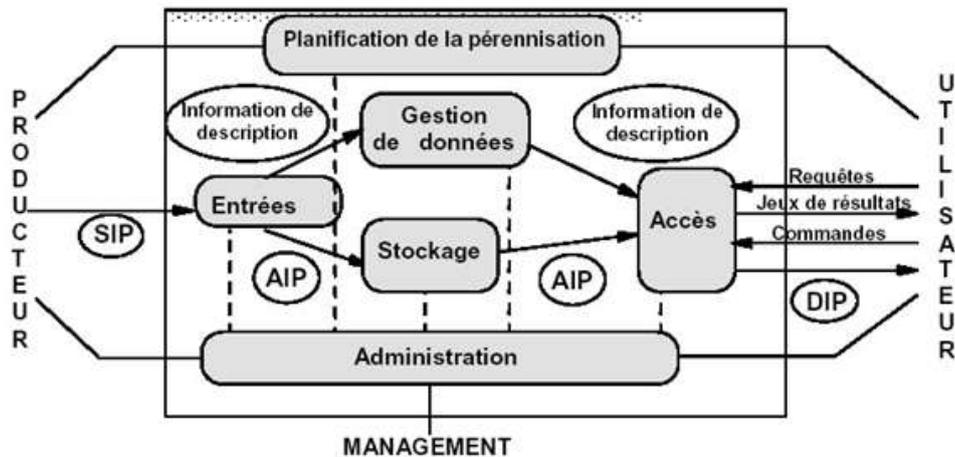
interaction avec une communauté cible d'utilisateurs. Le respect de normes rend possible l'échange de documents et de métadonnées entre logiciels distants. C'est à la fois un modèle fonctionnel (qui décrit ce qu'il faut faire) et un modèle d'information (qui décrit les métadonnées nécessaires). Le modèle OAIS repose sur l'idée que l'information constitue des paquets. Un paquet est constitué de fichiers et de métadonnées ou « informations », qui sont à la fois des « informations de représentation » (qui permettent de connaître le format et la sémantique), et des informations de « pérennisation » (décrivant la provenance, le contexte, l'intégrité, etc.). Chaque institution définit ce qu'est pour elle un paquet selon ses besoins et son contexte : par exemple dans le cas des fascicules d'un périodique, qui peuvent être traités ou non à la pièce, par année, ou encore par unité de conservation. Ces paquets ne sont pas les mêmes que l'on soit en train de produire l'information, d'essayer de la conserver, ou de la communiquer à un utilisateur. Il y a donc trois sortes de paquets, chacun accompagné de métadonnées :

- Les paquets de versement (SIP) préparés par les producteurs à destination de l'archive,
- Les paquets d'archivage (AIP) transformés par l'archive à partir du SIP dans une forme plus facile à conserver dans le temps,
- Les paquets de diffusion (DIP) transformés par l'archive à partir de l'AIP dans une forme plus facile à communiquer.

Le schéma ci-dessous représente les liens entre les différentes entités impliqués dans l'archivage pérenne selon l'OAIS. Le modèle précise les intervenants et leur rôles. Les acteurs externes comprennent les Producteurs, les Utilisateurs et le Management :

- le Producteur verse dans l'archive les objets numériques et les métadonnées.
- l'Utilisateur accède aux données, un groupe d'utilisateurs constitue la Communauté cible de l'archive.
- le Management est responsable de la politique de l'archive et du soutien financier sur le long terme.

Les acteurs internes de l'archive participent au fonctionnement des Entités et à la réalisation des Fonctions. Dans le cadre d'un projet mené en SCD, où les bibliothécaires sont à la fois producteurs et utilisateurs, c'est par exemple le versement des paquets SIP qui exigent l'intervention des personnels de bibliothèque, les deux autres types de paquet étant pris en charge par l'archive (le CINES par exemple), qui renvoie en fin de cycle les fichiers AIP (avis d'archivage) à conserver par le service versant.



**Le schéma fonctionnel du modèle OAIS.** BANAT-BERGER Françoise, DUPLOUY Laurent, HUC Claude, *L'archivage numérique à long terme : les débuts de la maturité ?*, Paris, La Documentation française, 2009, Coll. Manuels et guides pratiques, figure 10, p. 53.

Les processus impliqués par le modèle OAIS supposent des besoins en personnel au niveau de l'archive même – pour l'administration et le management, pour la gestion des entrées et des accès, pour la gestion des données et des métadonnées associées... – ainsi qu'au niveau de l'institution versante, qui doit préparer les paquets d'archives (SIP) et les jeux de métadonnées, assurer les relations avec l'archive, mais aussi transmettre les demandes de remontée d'archives des utilisateurs, anticiper la volumétrie des données...

L'archivage pérenne exige également, on l'a vu, le respect de formats de fichiers et de métadonnées normalisés. Une métadonnée est une « donnée qui définit et décrit une autre donnée » (ISO/IEC 11170-3, 2003). Les métadonnées sont de plusieurs sortes : descriptives (identification du document), techniques (caractéristiques du fichier : format, taille), administratives (gestion des droits d'accès, de modification, modalités de préservation) et sont utiles tant pour la diffusion que pour la conservation des fichiers numériques. Les SCD sont d'ores et déjà confrontées la gestion des métadonnées dans le cadre du processus STAR, l'ABES exigeant certains formats de PDF, ou des dépôts dans HAL. Les formats et métadonnées de pérennisation favorisées sont mis à jour par les instances en charge de la normalisation ; en résultent des préconisations de la part de l'archive concernée. Pour les métadonnées<sup>76</sup>, sont ainsi par exemple favorisés le DublinCore que les bibliothécaires connaissent par ailleurs en tant que jeu de métadonnées descriptives ; à ce premier niveau généraliste de métadonnées s'ajoutent des recommandations comme les dictionnaires de données PREMIS (Preservation Metadata : Implementation Strategies), définissant des éléments principaux afin d'améliorer les fonctions de conservation, TEF pour les thèses électroniques françaises ou LOM, LOMFR et SupLOMFR pour les ressources d'enseignement et d'apprentissage. Ces divers dictionnaires peuvent être combinés grâce à des standards d'empaquetage des métadonnées. Parmi les plus connus, on compte METS (Metadata Encoding and Transmission Standard - Norme de codage et de transmission de métadonnées), format d'empaquetage en XML ou le format utilisé en archivistique mais aussi de plus en plus par les bibliothécaires, EAD (Encoded Archival Description – Description archivistique encodée). Le CINES comme la

<sup>76</sup> Cf. BANAT-BERGER Françoise, DUPLOUY Laurent, HUC Claude, *L'archivage numérique à long terme : les débuts de la maturité ?*, op. cit. p. 125-146.

BnF émettent des préconisations envers les établissements versants, sous forme de listes de métadonnées retenues pour leurs systèmes d'archivage, PAC<sup>77</sup> ou SPAR.

En ce qui concerne les formats<sup>78</sup> (formats images, textes ou audiovisuels<sup>79</sup>), le CINES comme la BnF effectuent une veille technologique, et publient également une liste des formats archivables<sup>80</sup>. Les fichiers doivent être publiés dans un format ouvert<sup>81</sup> afin d'éviter de dépendre de l'existence d'un logiciel propriétaire donné<sup>82</sup>. Il est possible de tester les fichiers à déposer avant envoi dans l'outil FACILE<sup>83</sup> mis à disposition par le CINES afin de vérifier leur conformité<sup>84</sup>. Le choix d'un format aisé à traiter, ouvert, standardisé mais aussi en adéquation avec les besoins des établissements est stratégique pour un projet. Ainsi, l'adoption très large à une époque du format TIFF, qui offre une excellente qualité d'image, engendre cependant un « poids » de fichiers en TO très important. Jean-François Vincent (BIUS) note que c'est le choix du JPEG et non du TIFF pour la bibliothèque numérique Medic@ qui a permis son archivage pérenne : le coût d'archivage au CINES d'une même quantité de fichiers en TIFF aurait été près de sept fois supérieur, et donc impossible à assumer pour le SCD.

Enfin, de nombreux processus sont à l'œuvre au sein d'une archive<sup>85</sup> comme PAC afin d'assurer une lutte efficace contre les divers risques qui menacent les fichiers et données – parmi lesquels l'obsolescence des formats<sup>86</sup> – grâce entre autres à la migration et l'émulation<sup>87</sup>. Dans les conventions établies avec les BU, le

<sup>77</sup> Voir CINES, « Les métadonnées de pérennisation » [En ligne] (page consultée le 30/12/2015) <https://www.cines.fr/archivage/un-concept-des-problematiques/les-metadonnees-de-perennisation/>

<sup>78</sup> Voir PRAT Philippe, « L'expertise sur les formats de fichiers : une spécificité de l'offre d'archivage du CINES », Dossier « Archivage numérique pérenne », *La Gazette du CINES*, février 2013, p. 36.

<sup>79</sup> Parmi les formats plus couramment employés, on trouve pour les ebooks le format EPUB, qui apparaît comme le standard émergent et auquel le format PDF constitue une alternative. Pour les fichiers images, on retient généralement les TIFF, JFIF/JPEG et PNG pour les fichiers haute qualité. Les performances de JPEG 2000 en compression avec et sans perte sont supérieures à celle du JPEG. Pour l'océrisation, le format XML-ALTO, bien qu'encore assez peu utilisé, constitue un gage certain de pérennité en tant que format documenté et standard ; pour les documents bureautiques enfin, le format PDF/A (versions PDF/A-2 ou PDF/A-3) Au sein des bibliothèques numériques, « la prépondérance du JPEG et du JPEG 2000 apparaît avec évidence (...). Les formats TIFF et PDF ne sont pas en reste (...) Le PDF, favorisé comme format d'export, sert parfois aussi à la diffusion de préférence au JPEG, quand il n'est pas tout simplement le seul format mentionné (...). La résolution de ces images est fréquemment de 300 DPI (...) une résolution inférieure est plutôt réservée aux images diffusées sur le site et à celles en noir et blanc ou en niveaux de gris (des imprimés souvent), et une résolution supérieure aux manuscrits, aux documents iconographiques et de manière générale aux pièces en couleurs (...) », PÉRÉSAN-ROUDIL Alice, *Dans la forêt touffue des bibliothèques numériques patrimoniales françaises : quels choix, quelles stratégies, quelles perspectives ?*, Diplôme de conservateur des bibliothèques (DCB), sous la direction d'Aline Girard, Enssib, 2015, p. 36.

<sup>80</sup> Cf. CINES, « Liste des formats archivables » [En ligne] (page consultée le 30/12/2015) <https://www.cines.fr/archivage/des-expertises/expertise-formats/liste-des-formats-archivables/> et BnF, « Les métadonnées de préservation numérique » [En ligne] (page consultée le 30/12/2015) [http://www.bnf.fr/fr/professionnels/preservation\\_numerique\\_boite\\_outils/a.pres\\_num\\_metadonnees.html](http://www.bnf.fr/fr/professionnels/preservation_numerique_boite_outils/a.pres_num_metadonnees.html)

<sup>81</sup> Les critères d'évaluation d'un format supposent des formats largement utilisés, non propriétaires, ouverts, sans DRM ; la compression doit se faire sans perte, avec des algorithmes non propriétaires ; l'archive et le producteur du format doivent en avoir la maîtrise (Source : Stéphane Reece, BnF).

<sup>82</sup> Les normes et règles de la conservation de documents numériques sont abordées dans le manuel de Thierry Claerr et Isabelle Westeel (dir.), *Numériser et mettre en ligne*. Villeurbanne : Enssib, 2010.

<sup>83</sup> CINES, « Facile » [En ligne] (page consultée le 30/12/2015) <https://www.cines.fr/archivage/des-expertises/expertise-formats/facile/>

<sup>84</sup> D'autres outils existent, comme JHOVE et JHOVE2 ou Mediainfo, pour les ressources audiovisuelles.

<sup>85</sup> Parmi lesquels une sauvegarde sécurisée (au CINES, deux en local et une distante, à l'IN2P3) ; la copie sur plusieurs technologies de stockage (disques et bandes) ; les contrôles des métadonnées, de la qualité des données, la validation des formats ; les contrôles d'intégrité.

<sup>86</sup> Le groupe PIN fait la liste des méthodes pour éviter l'obsolescence technologique : voir la V2 (en anglais) : « Reference model for an open archival information system (OAIS) » [En ligne] (page consultée le 30/12/2015) <http://public.ccsds.org/publications/archive/650x0m2.pdf>

<sup>87</sup> Sur ces notions, voir annexes.

CINES propose ainsi un archivage pérenne incluant la garantie d'évolution des formats.

Des coûts importants résultent de l'ensemble de ces contraintes, des acteurs et des processus supposés par le modèle OAIS, et de la gestion des plateformes d'archivage elles-mêmes. Ces coûts sont à la charge des établissements en charge de l'archive – CINES, BnF- mais aussi des services versants. Ces derniers doivent en effet mobiliser des personnels, surtout pendant la phase de projet, mais également fournir au CINES une participation financière aux coûts si le choix se porte sur l'opérateur national.

### *B. Les coûts de l'archivage pérenne*

#### **Le calcul des coûts du tiers-archivage dans l'ESR**

Le calcul des coûts de l'archivage pérenne, qui nécessite la prise en compte de nombreux facteurs (volumétrie, typologie des données) est difficile à établir, mais stratégiquement nécessaire au vu de l'augmentation inévitable des volumétries. Le projet LIFE<sup>88</sup> et d'autres études<sup>89</sup> proposent des méthodes de calcul, qui soulignent toutes l'importance des économies d'échelle rendues possibles lorsque les volumes archivés sont conséquents. Le CINES note ainsi qu'une « masse critique [doit être] atteinte pour permettre des économies d'échelle par la gestion d'une volumétrie supérieure à effectifs constants. L'augmentation du nombre de « clients » (centres de recherche surtout, dont les données, météorologiques ou archéologiques par exemple, sont très lourdes, les bibliothèques étant encore marginales dans la demande d'archivage pérenne), a déjà fait baisser les coûts. La mutualisation, entre autres des infrastructures de stockage, en permet également une réduction considérable. Pierre Carbone avait proposé de « confier au groupe PIN une étude sur les modèles économiques de tiers-archivage public et privé dans le monde. Une telle étude, qui serait utile pour tous, éclairerait les travaux de la Commission, qui s'est intéressée aux coûts et aux prestations de tiers-archivage en France ». Cette étude n'a finalement pas été menée. Le groupe PIN s'est cependant penché sur les coûts que représente l'archivage pérenne pour les infrastructures et a produit plusieurs études<sup>90</sup>, notamment sur une comparaison des coûts d'archivage papier/numérique, ou sur un modèle d'évaluation des coûts de l'archivage numérique au CINES<sup>91</sup>.

#### **Les coûts au CINES**

Les infrastructures comme le CINES ou la BnF doivent en effet assurer à la fois des coûts de fonctionnement, des coûts d'investissement des infrastructures et des coûts de personnel. Ni le CINES ni la BnF ne facturent à coûts réels aux établissements, qui financent donc une participation aux coûts.

---

<sup>88</sup> Projet britannique (Life Cycle Information for E-Literature) de collaboration entre le University College London (UCL) et la British Library. LIFE [En ligne] (page consultée le 30/12/2015) <http://www.life.ac.uk/>

<sup>89</sup> Sur ces études et le calcul des coûts des plateformes d'archivage, voir BANAT-BERGER Françoise, DUPOUY Laurent, HUC Claude, *L'archivage numérique à long terme : les débuts de la maturité ?*, op. cit., p. 161 et suiv.

<sup>90</sup> Voir HUC Claude, JEAN-ANTOINE Anne, « Le coût de l'archivage numérique. Introduction au débat. Présentation Groupe PIN /CNES », Groupe PIN, 4 septembre 2003

<sup>91</sup> PIN, « Compte-rendu de la réunion du 4 janvier 2013 » [En ligne] (page consultée le 30/12/2015) [http://pin.association-aristote.fr/doku.php/public/reunion\\_pleniere/cr/cr\\_20130412](http://pin.association-aristote.fr/doku.php/public/reunion_pleniere/cr/cr_20130412)

La grille tarifaire au CINES dépend du secteur : établissements relevant du MENESR, établissements externes au ministère, ou prestataires privés. Les prestataires privés sont théoriquement autorisés à archiver au CINES, mais toute entrée nécessite une acceptation par le directeur et le CA pour les établissements hors ESR ; un seul établissement a été ainsi accepté : la Cour des Comptes. De fait, le financement a été conçu pour répondre aux besoins de l'ESR prioritairement. La différence de coût dépend du niveau de service selon les supports de stockage, selon qu'il s'agisse de stockage sur disque et bande ou sur bande uniquement (plus rapide que sur disque ; cette option sera ouverte en 2016) ; et de la volumétrie, le premier TO étant plus cher du fait des coûts de mise en œuvre. La participation financière était évaluée en 2013 en tenant compte des coûts de fonctionnement à 5000€ TTC par TO archivé et par an, avec une diminution de 20% entre 2012 et 2013 ; cette participation a de nouveau diminué en 2014-2015.

A l'instar des divers acteurs publics de l'archivage pérenne en France, le CINES essaie en effet de contenir les coûts. Ces derniers étaient jugés très élevés par les établissements, mais ont été divisés par 3 ou 4 depuis 3 ans, rendant aujourd'hui le choix de l'archivage au CINES presque concurrentiel en regard des coûts du stockage dans une DSI – sans compter le temps de mise en œuvre du projet et de développement dédié. Plusieurs SCD notent ainsi qu'au regard des services rendus, les coûts du CINES paraissent raisonnables. Ceux qui ont les moyens d'investir dans la phase initiale du projet considèrent par ailleurs que la participation annuelle requise par la convention serait supportable à long terme par l'université. La question se pose davantage pour les établissements de petite taille, comme nous le verrons (II).

### **Les financements**

Aucune source de financement ou aide spécifique n'est assurée à des programmes locaux, contrairement aux programmes nationaux (Thèses.fr, HAL, Persée, STAR). Au sein de BSN 6, s'il fut un temps question d'un financement de soutien ou d'impulsion aux établissements, il n'est plus question actuellement que de définition de stratégie, entre autres concernant la baisse des coûts des tiers-archivistes publics. Les recommandations validées par BSN, dans le cadre du document stratégique 2012-2015 (mars 2012), au titre de BSN 5, prévoyaient entre autres d'« Établir des critères de priorisation des financements pour la numérisation de documents scientifiques qui seront archivés, sur un mode pérenne, au CINES ». Au titre des appels à projet de BSN 5 (2012-2013), des établissements ont été financés pour de la numérisation, mais les financements pour la mise en œuvre de démarche d'archivage pérenne n'ont pas vu le jour. La TGIR Huma-Num finance l'archivage pérenne des données des SHS uniquement ; hors de ce périmètre précis (et hors des systèmes HAL et STAR), l'archivage pérenne des établissements reste donc à leur charge.

Demeure ainsi la question de la participation au coût de l'archivage pérenne par les établissements publics, qui a été contestée dans le principe et qui, dans les faits, limite l'accès pour certains établissements aux services du CINES. L'une des préconisations du groupe PIN proposait sa suppression. De fait, il peut sembler paradoxal que des programmes de numérisation puissent être subventionnés quand leur archivage pérenne ne l'est pas, partition qui revient à rendre aléatoire la préservation d'un patrimoine numérique acquis sur des fonds publics.

### Les budgets des établissements

Les établissements doivent donc assurer eux-mêmes le financement de leur projet, dans la logique de leur nouveau statut d'autonomie (loi LRU). Or l'intégration de l'archivage pérenne dans les budgets de fonctionnement peut poser problème dans le contexte budgétaire actuel. Les coûts de conservation, en particulier numérique, sont apparus depuis peu dans les administrations. C'est une ligne budgétaire à créer le plus souvent, dans une situation économique de plus en plus contrainte, et qui implique un coût perpétuel dont on ne peut se dédire. Les conclusions de l'étude du JISC<sup>92</sup> qui souligne que les coûts de l'archivage pérenne, faibles au regard du budget complet d'une institution, sont largement supportables par les instituts de recherche, ne peuvent être aisément appliquées à l'ensemble des universités françaises, confrontées pour certaines à de réelles difficultés budgétaires depuis la loi LRU et en particulier aux SCD aux budgets parfois très réduits.

Les coûts sont de deux sortes : les coûts d'entrée chez le tiers-archiviste et les coûts récurrents, par définition sans limite de temps définie. Il faut prendre en compte également le coût des prestations pour des tâches qui ne pourraient pas effectuées en interne. Le coût de la convention avec le CINES est à calculer sur le très long terme, et en anticipant l'accroissement de la volumétrie en TO, ce qui nécessite une gestion pluriannuelle. D'où la nécessité d'un engagement budgétaire fort sur la longue durée, ce que tous les établissements ne peuvent garantir.

On comprend donc, au vu des exigences techniques et des coûts impliqués, que l'archivage pérenne soit souvent considéré par les SCD davantage comme un objectif à moyen voire long terme qu'un projet immédiatement réalisable. Le rapport Lecoq-Micol notait ainsi en 2014 :

« La complexité scientifique et technique ainsi que le coût élevé des programmes d'archivage pérenne fait que seules quelques rares institutions y ont recours. »<sup>93</sup>

---

<sup>92</sup> Voir à ce sujet « LERU-Roadmap for Research Data – FOSTER », *op. cit.* : « However, the evidence of the JISC studies shows that the costs of archiving activities (archival storage, preservation planning and related actions) are consistently a small proportion of a research institution's overall budget ».

<sup>93</sup> *Le stockage des collections imprimées à l'heure du numérique*, *op. cit.*, p. 14.



## II. PROJETS D'ARCHIVAGE PÉRENNE EN BIBLIOTHÈQUES UNIVERSITAIRES : RETOURS D'EXPÉRIENCE

---

### 1. Méthodologie de l'enquête et bilan général

#### Méthodologie de l'enquête

Les établissements universitaires enquêtés sont aussi bien parisiens que de province ; soit BIU, soit BU, ils sont de tailles et d'importance diverses, la problématique variant fortement selon les moyens des SCD. La grille d'entretien type adressée aux interlocuteurs des SCD figurent en annexe. Deux appels à contributions ont été faits par le biais de la liste de diffusion de l'ADBU (la seconde a été relayée sur la liste de la section Aurore, de l'Association des archivistes français<sup>94</sup>), et des contacts directs établis avec les SCD (entretiens téléphoniques et par voie de mail ; le tableau-bilan des interlocuteurs se trouve en annexe).

#### Établissements et institutions enquêtées

La présentation détaillée des projets des établissements classés selon leur niveau d'avancement se trouve en annexe.

<i>Bibliothèques universitaires et inter-universitaires enquêtées</i> Bibliothèque inter-universitaire Cujas Bibliothèque inter-universitaire de santé Paris Descartes-Paris 5 (BIUS) Bibliothèque de Science Po SCD de l'université Pierre et Marie Curie (BUPMC) Bibliothèque Sainte-Geneviève (BSG) SCD de l'université de Rouen SCD de l'université de Poitiers SCD de l'université de Strasbourg BNU, Strasbourg SCD de l'université d'Angers SCD de l'université de Toulon SCD de l'université de Lille 1 SCD de l'université de Lille 2 SCD de l'université de Lille 3	SCD de l'université de Lorraine SCD de l'université de Lyon 1 SCD de l'université Paris 3 SCD de l'université de Lorraine SCD de l'université Clermont-Université Université de Toulouse  <i>Autres établissements : opérateurs de l'État, établissements publics, MENESR, autres services de l'État</i> BnF (SPAR, service numérisation, dépôt légal du Web) CINES CCSD ANRT BSN 6 SIAF CNDP
---	---

---

<sup>94</sup> Site de l'AAF [En ligne] (page consultée le 30/12/2015) <http://www.archivistes.org/>

## Bilan général

En 2015, soit six ans après le premier projet d'archivage pérenne mené avec le CINES (bibliothèque Cujas), les SCD français sont encore peu nombreux à avoir finalisé un archivage pérenne de leurs ressources. De fait, sur l'ensemble des bibliothèques universitaires françaises, seuls sept établissements ont établi ou sont en train d'établir à ce jour une convention avec le CINES, le plus souvent pour une partie restreinte de leurs fonds (à l'exception de la BIUS, qui archive la totalité de la base Medic@ et de sa banque d'images, et qui est donc le seul établissement à sauvegarder la totalité de ses collections numérisées au CINES, hors expositions virtuelles et éditions en ligne). Plusieurs SCD ont envisagé un projet d'archivage pérenne, mais l'ont abandonné pour des raisons tantôt budgétaires, tantôt liées aux ressources humaines. D'autres affirment appliquer d'ores et déjà une méthode de conservation pérenne, mais se trouvent en réalité avoir recours à un stockage sécurisé<sup>95</sup>. La problématique est émergente cependant, et commence à bénéficier d'un ensemble de retours d'expérience (groupe de travail PIN, segment BSN 6, journées utilisateurs du CINES, etc.).

## 2. Solutions adoptées et acteurs des projets

### A. Les solutions d'archivage en bibliothèque universitaire

L'archivage pérenne du numérique nécessite des compétences poussées, un accompagnement important et des outils coûteux que les établissements n'ont pas les moyens d'assumer seuls. L'une des solutions est de passer une convention avec un partenaire spécialisé proposant une offre de tiers-archivage. Quelques opérateurs privés proposent des offres de tiers-archivage ; dans leur totalité cependant, les SCD enquêtés, eux-mêmes établissements publics, n'envisagent que le recours à un opérateur public, à même de proposer une offre en adéquation avec les missions de service public dévolues aux SCD, et qui puisse assurer la pérennité et la continuité nécessaires à un archivage de nature patrimoniale.

### Les solutions intermédiaires au niveau de l'université : stockage sécurisé, Data Center, acteurs locaux

Constatons dans un premier temps que dans leur majorité, les SCD ne procèdent pas à l'archivage pérenne de leurs documents, quelles qu'en soient les raisons, et stockent leurs données grâce aux moyens informatiques du SCD ou de l'université<sup>96</sup>. Ces moyens sont divers : ils dépendent de la taille de l'université, et des investissements matériels et logiciels auxquels procède l'établissement, qui peut faire le choix d'une redondance des informations entre plusieurs disques et la duplication entre plusieurs sites géographiques, comme pour la Bibliothèque virtuelle des premiers socialismes (Poitiers), ou passer par des espaces de stockage dédiés – une solution qui ne présente pas un caractère pérenne aussi fort – sur

---

<sup>95</sup> Voir le résultat de l'enquête de PÉRÉSAN-ROUDIL Alice, *Dans la forêt touffue des bibliothèques numériques patrimoniales françaises : quels choix, quelles stratégies, quelles perspectives ?*, op. cit., p. 41, consacrée aux bibliothèques numériques : « Dans le cadre des réponses apportées au questionnaire, 19 établissements sur 33 affirment appliquer d'ores et déjà une méthode de conservation sur le long terme. 7 autres disent projeter le développement de stratégies. Toutefois, les répondants semblent mal distinguer stockage et archivage, faute de bénéficier d'une information suffisante sur ces questions ».

<sup>96</sup> Sur la stratégie d'hébergement (dans une BM), voir WESTEEL Isabelle et PLANCKAERT Nicolas, « Le cercle vertueux du numérique : la mise en ligne des collections patrimoniales de la bibliothèque de Lille », in LEBLOND Corinne (dir.), *Archivage et stockage pérenne*, op. cit., p. 141-148.

serveurs ou sur architecture NAS<sup>97</sup>. Il est courant qu'une plateforme consacrée à l'archivage de thèses et de mémoires bénéficie d'un « back up » sur des serveurs distants. Il est également généralement possible d'assurer en interne une migration de format régulière (d'un CR rom Gold à un disque dur par exemple), ou encore de procéder à des rafraichissements, le CD Rom ayant une durée de vie limitée<sup>98</sup>.

Dans l'impossibilité d'un archivage pérenne, des alternatives sont envisagées, et plusieurs établissements définissent ainsi une gestion de risque en doublant un serveur de mise en ligne répliqué avec des disques durs externes. Le SCD de Lyon 1 par exemple envisage un projet commun de stockage sécurisé sur des serveurs en interne entre la DSI de l'université, la direction des affaires juridiques, et le SCD. De ce fait, il s'agit d'un « demi-abandon » de la procédure de migration de formats propre à l'archivage pérenne. Plusieurs précautions sont prises cependant afin d'établir une assurance supplémentaire contre les risques de perte : n'inclure par exemple que des formats dont la conception permet une interprétation pérenne (PDF/A, etc.) ; par ailleurs les éventuels besoins de conversion en masse de fichiers seraient pris en charge par la DSI. Enfin, les DSI des universités sont parfois à même de proposer aux services, dont les SCD, une mise à disposition d'espaces de stockage de grande capacité, et offrent aussi de sauvegarder des données<sup>99</sup>.

Les plates-formes de dépôt, d'archivage et de diffusion de publications électroniques, libres et paramétrables (par exemple FEDORA, E-PRINT et DSPACE), qui permettent l'archivage des documents numériques d'une institution, et sont couramment utilisées par les SCD, demeurent quant à elles des logiciels à intégrer dans un fonctionnement général, et ne fournissent pas un service en tant que tel. L'archivage qu'elles proposent ne peut être qualifié de pérenne<sup>100</sup>.

Au-delà des risques liés au stockage sécurisé opéré en interne, la question se pose de la capacité de la DSI à opérer sur le long terme une migration de formats et à s'assurer de la pérennité des formats (PDF, PDF-A etc.). Dans un service informatique, la sauvegarde sécurisée peut au mieux disposer les données à trois endroits et effectuer une mise à jour synchronisée. Par ailleurs, les services n'ont généralement pas les moyens d'assurer une veille continue sur les formats, indispensable pour savoir quand et comment migrer. Enfin, si le format des documents n'était pas correct au moment du dépôt, la migration n'est ensuite plus possible. C'est donc la question de la qualité et du suivi des données qui se pose. La migration par une DSI peut donc constituer un pari risqué en l'absence de veille. Ce risque est à envisager selon l'état de danger dans lequel se trouvent la donnée ou le document – selon donc, généralement, l'existence ou pas de doubles,

---

<sup>97</sup> Sur les typologies d'architecture DAS et NAS, et sur le stockage en général, voir BANAT-BERGER Françoise, DUPLOUY Laurent, HUC Claude, *L'archivage numérique à long terme : les débuts de la maturité ?*, op. cit., p. 81-101.

<sup>98</sup> Voir les « Recommandations relatives à la gravure, à la conservation et à l'évaluation des cd-r » (2005) de la Direction des archives de France, <http://www.archivesdefrance.culture.gouv.fr/gerer/preservation/materiaux/>

<sup>99</sup> Voir par exemple: Lille 1, CRI [En ligne] (page consultée le 30/12/2015) [http://cri.univ-lille1.fr/Catalogue-des-services/Hebergements\\_informatique/](http://cri.univ-lille1.fr/Catalogue-des-services/Hebergements_informatique/)

<sup>100</sup> Conçues pour l'archivage de documents universitaires publiés, ces plateformes offrent la prise en compte des bonnes pratiques des bibliothèques et des archives concernant les métadonnées -DublinCore, MODS, METS-, la préservation, les URL persistantes (Handles, DOIs, PURLs), la validation des formats de fichiers, la vérification de l'intégrité, l'intégration de logiciels antivirus..., et l'interopérabilité (OAI-PMH, SWORD protocol, OAI-ORE). De nombreux gestionnaires de contenu fournissent des modules qui permettent l'intégration de plateformes d'archivage. Certaines solutions permettent l'utilisation conjointe d'un gestionnaire de contenu et d'une plateforme d'archivage : par exemple Islandora (construit sur le CMS Drupal, plateforme Fedora) ; le module DSpace de Drupal (qui permet l'affichage des métadonnées et du contenu d'une plateforme DSpace dans Drupal) ; le module DSpace de Joomla (J-CAR) (qui permet l'affichage des métadonnées et du contenu d'une plateforme DSpace dans Joomla). Dspace est utilisé par l'ENSAM pour SAM, par Dauphine pour BIRD et par Lille 1 pour IRIS.

ou la possibilité de renumériser – qui implique un coût. Le risque est particulièrement élevé pour les originaux numériques. L’archivage pérenne tel qu’il est effectué au CINES ou à la BnF ne double donc pas le stockage qui continue à être assuré par l’université, argument parfois avancé – puisque les données demeurent aussi disponibles en local, sur les serveurs de l’établissement. Par ailleurs ce sont les versions haute définition qui sont conservées au CINES (les scans de documents en formats TIFF, JPEG ou PNG), et non les versions basse définition (par exemple les fichiers PDF utilisés pour la diffusion). La bibliothèque universitaire quant à elle archive dans son service informatique les formats de diffusion, les services informatiques vidant les serveurs des hautes définitions pour avoir assez de place de stockage<sup>101</sup>. Certaines universités, très peu nombreuses, envisagent de procéder elles-mêmes à l’archivage de leur production. De nombreux projets de Data Center, déjà évoqués, sont ainsi en cours au sein des universités.

D’autres solutions en dehors du périmètre de l’ESR sont parfois envisagées, selon les besoins, les moyens et les contextes propres à chaque SCD, qui visent à une préservation à long terme, mais n’impliquent pas un archivage pérenne à proprement parler. Le SCD de Paris 3 par exemple a engagé un plan de numérisation de la cinémathèque centrale de l’enseignement public dont les fichiers sont transmis au fur et à mesure pour affichage et archivage au Réseau Canopé (le CNDP, dorénavant Réseau Canopé, procède à l’archivage de ses propres productions sur un système interne de disques durs et de bandes LTO ; l’archivage pérenne à la BnF y est envisagé). Le SCD de Clermont envisage quant à lui de se tourner vers le Centre régional de ressources informatiques (CRRI), lequel cherche à proposer un système d’hébergement pérenne localement. Cette solution n’est pas considérée comme une alternative équivalente à celle du CINES, mais comme le moyen d’assurer l’hébergement à moyen terme des données, « entre stockage post-numérisation et archivage définitif »<sup>102</sup> avec la nécessité à terme d’« articuler le stockage au CRRI (le service informatique du PRES Clermont Université) d’une part, et l’archivage pérenne lui-même »<sup>103</sup>. Cette solution mutualisée est favorisée par le statut interuniversitaire de l’établissement. Le CRRI, qui a une infrastructure, mais pas encore de plan d’archivage, vise à se positionner au niveau régional. Ce cas ouvre la réflexion à une mutualisation de « site », qui excéderait le périmètre universitaire.

### **Une solution mutualisée pour l’ESR : le CINES**

Les projets en production en 2015 au CINES sont la BIU Cujas, la BUPMC, la BIUS et la BSG<sup>104</sup>. Les projets en cours en 2015 sont ceux du SCD de Lille 2, de la bibliothèque universitaire de Lorraine et de la bibliothèque de Sciences Po.

Les SCD ayant fait le choix du CINES avancent tous plusieurs facteurs de décision. On relève ainsi, outre la qualité de l’archivage et de l’accompagnement, le

---

<sup>101</sup> La question ne se pose pas pour les thèses, dont la version est identique pour la diffusion et l’archivage.

<sup>102</sup> Source : Olivier Legendre, conservateur adjoint au directeur de la BCU, responsable de la bibliothèque numérique, SCD de l’université de Clermont.

<sup>103</sup> L’Alambic numérique. Le blog de la bibliothèque numérique de Clermont-Ferrand [En ligne] (page consultée le 30/12/2015) <http://alambic.hypotheses.org/tag/archivage-perenne>

<sup>104</sup> D’autres institutions que les BU archivent également au CINES. Voir CINES, « Les fonds conservés au CINES » [En ligne] (page consultée le 30/12/2015) <https://www.cines.fr/archivage/nos-references-nos-partenariats/les-fonds-conserves-au-cines/>

partage de périmètre BnF / CINES, qui oriente les SCD vers le CINES. Le CINES apparaît ainsi la plupart du temps comme l'interlocuteur naturel des établissements de l'ESR en général et des SCD en particulier, entre autres pour des raisons de ministère de tutelle. Ainsi à Cujas, outre que le choix a paru évident, il se trouvait que l'ingénieur informatique de l'équipe avait travaillé au CINES ; le choix n'a donc pas été une source de discussion. Le coût de l'hébergement y était encore très cher et aurait cependant pu poser problème à la bibliothèque lors du lancement du projet. A la BSG, le projet d'archivage pérenne, qui visait à assurer la pérennité de fichiers auparavant stockés sur des disques durs, n'a envisagé également une solution que dans le cadre du MENESR. Le CINES a été naturellement retenu, d'autant plus que la plateforme PAC était lancée de façon concomitante (2011). A Lille 1, d'autres solutions ont été envisagées mais l'opérateur public a semblé la solution la plus logique. A Lille 2, Paris 6 ou Strasbourg (projet en cours – convention non établie), il en est de même, le CINES étant déjà identifié comme un acteur incontournable du fait de l'implication des SCD dans le circuit STAR. On notera que les projets de numérisation ayant été lancés de façon concomitante à la mise en production de la plateforme PAC ont naturellement envisagé un archivage que les projets antérieurs n'ont pas pris en considération. Ainsi la bibliothèque numérique consacrée à Flaubert par le SCD de Rouen, ou Nordnum de Lille 3, sont-ils nés « trop tôt » et doivent considérer a posteriori l'archivage pérenne de leurs produits de numérisation.

### **La BnF : un rôle d'incitateur ; le dépôt légal du Web**

Certaines bibliothèques universitaires se sont également adressées à la BnF pour envisager un tiers archivage par SPAR (Lyon 1, Clermont Université). La nature des fonds (fonds anciens numérisés) pouvait rendre logique cette hypothèse. Les coûts de la BnF, à une époque assez sensiblement inférieurs à ceux du CINES, ont pu favoriser ces demandes. La ligne du service SPAR s'étant précisée entretemps (voir I 2. B), ces projets n'ont pas vu le jour.

La BnF peut jouer cependant un rôle d'incitateur à l'archivage pérenne via des conventions passées pour des opérations de numérisation impliquant en retour un archivage pérenne au CINES. Ainsi à l'université de Lorraine<sup>105</sup>, l'obligation spécifiée par la BnF de procéder à un archivage pérenne a motivé le lancement du projet : les documents patrimoniaux mis en ligne sur la plateforme Pulsar ont en effet été numérisés grâce une convention passée avec la BnF, laquelle spécifiait l'obligation de procéder à l'archivage pérenne des documents issus de la numérisation.

Le dépôt légal du web<sup>106</sup> pose question : car la collecte des sites web, considérés par la loi comme une « production intellectuelle de la société », est devenue une responsabilité de la BnF (et de l'INA) depuis la loi DADVSI de 2006<sup>107</sup>. A ce titre, il est ainsi tout à fait possible pour un SCD d'adresser une

---

<sup>105</sup> Également pour le projet Pôlib de Lille 3.

<sup>106</sup> Sur ce point, voir ILLIEN Gildas, « Le dépôt légal de l'internet en pratique », Bulletin des bibliothèques de France [en ligne], n° 6, 2008 [consulté le 30 décembre 2015] <http://bbf.enssib.fr/consulter/bbf-2008-06-0020-004> ; et BnF, « Dépôt légal des sites web » [En ligne] (page consultée le 30/12/2015) [http://www.bnf.fr/fr/professionnels/depot\\_legal/a.dl\\_sites\\_web\\_mod.html](http://www.bnf.fr/fr/professionnels/depot_legal/a.dl_sites_web_mod.html)

<sup>107</sup> L'alinéa de l'article L. 131-2 du code prévoit que : « Sont également soumis au dépôt légal les signes, signaux, écrits, images, sons ou messages de toute nature faisant l'objet d'une communication au public par voie électronique. ». Le décret publié le 19 décembre 2011 précise le périmètre de collecte : il concerne les sites enregistrés sous le nom de domaine .fr ou tout autre nom de domaine national (.re, .nc...), ainsi que les sites enregistrés par une personne domiciliée en France ou produits sur le territoire français.

demande de collecte ponctuelle de son site ou d'une partie de son site<sup>108</sup>. Cet archivage conserve aussi bien le contenu que la forme, et serait donc adapté pour des expositions virtuelles, certaines ressources pédagogiques ou des éditions en ligne par exemple. Il existe toutefois des exceptions pour certains sites difficiles à collecter : dans ce cas une instruction peut s'avérer utile. Par ailleurs, un SCD peut également s'adresser à l'un des correspondants en charge du DL du web à la BnF pour savoir si son site est d'ores et déjà collecté. La collecte sélective des sites de BU est à la charge du Département de l'Orientation et de la recherche bibliographique<sup>109</sup>.

Cette forme de coopération ne repose pas nécessairement la question des périmètres entre BnF et CINES, puisque la collecte entre dans le cadre du dépôt légal du Web, mission de la BnF uniquement. Elle offrirait une solution logique et pérenne sans coûts pour les SCD.

### **Les tiers-archivistes privés**

Françoise Banat-Berger, Laurent Duplouy et Claude Huc notent qu'« il n'existait sur le marché (en 2005) aucun produit répondant à l'ensemble des fonctionnalités requises, notamment pour la partie prise en charge des versements. En revanche quelques solutions logicielles couvrant une partie des fonctionnalités demandées commençait à émerger (...) »<sup>110</sup>. Dix ans plus tard, plusieurs solutions internationales sont apparues, certaines prenant en compte l'OAIS (voir en annexe). Aucun SCD français ne semble s'y être pour l'instant intéressé, le positionnement du CINES et la baisse des coûts régulière de l'archivage dans PAC rendant d'autres options, même équivalentes sur le plan technologique, peu fondées. Le tiers-archivage privé n'est en effet jamais envisagé ni adopté par les SCD français<sup>111</sup> et n'est pas étudié par les équipes projet, ne serait-ce que dans le cadre d'études préalables. Les garanties fournies par un prestataire externe ne peuvent en effet pas être comparées à celles du CINES, qui n'est pas en quête de rentabilité et ne limite pas le quota horaire consacré au suivi de chaque projet.

On constate que les SCD qui envisagent un archivage pérenne sont en quête de solutions assurées par des institutions publiques, perçues comme plus fiables et en adéquation avec les missions de conservation patrimoniale des établissements. La solution institutionnelle présente en effet des avantages considérables, comparables à ceux offerts par les archives ouvertes. Elle offre notamment une garantie de pérennité que des prestataires privés, qui pourraient n'être plus financés, n'offrent pas. Les SCD retiennent ainsi presque exclusivement l'offre du CINES, ou plus rarement des solutions locales. Ce que soulignait déjà la commission Bibliothèque Numérique : « Les tiers-archivistes publics offrent des garanties supérieures de pérennisation et d'accompagnement des besoins des

---

<sup>108</sup> A l'adresse générique : [depot.legal.web@bnf.fr](mailto:depot.legal.web@bnf.fr). (Ibid.).

<sup>109</sup> Source : Christine Génin, chargée de collections en littérature française contemporaine. Coordination du dépôt légal du Web et de la mise en ligne de contenus pour le département Littérature et art. Direction des Collections, Département Littérature et art, Service du Livre et de la littérature française, BnF.

<sup>110</sup> *Op. cit.*, p. 166.

<sup>111</sup> Le risque qu'elles présentent réside entre autres dans le recours à des marchés publics, qui sont conclus pour des périodes de 5 ans ; un prestataire serait ainsi potentiellement en situation de ne plus répondre au marché lors de son renouvellement, ou de ne pas proposer toutes les fonctionnalités requises. Car plusieurs solutions externalisées n'assurent pas la migration de formats ; ainsi Digital Archive d'OCLC n'assure que la vérification de formats, et non la migration. OCLC, DIGITAL ARCHIVE [En ligne] (page consultée le 31/12/2015) <https://www.oclc.org/digital-archive/features.en.html>

clients, tandis que les acteurs privés offrent surtout un archivage intermédiaire ou à valeur probante »<sup>112</sup>.

### *B. Les acteurs d'un projet d'archivage pérenne : des projets de coopération*

Un renforcement des collaborations entre informaticiens, archivistes et bibliothécaires est indispensable. L'archivage numérique pérenne se situe à la croisée de deux mondes : celui de l'informatique et celui de l'archivage, et ces deux compétences métier sont essentielles à la réussite de tout projet. Il faut donc que se crée un langage commun. (CINES)<sup>113</sup>

Les projets menés par les SCD avec le CINES impliquent trois ou quatre acteurs selon les projets : le CINES lui-même, les référents au sein du SCD, les informaticiens, et éventuellement un prestataire externe. Il s'agit donc par définition de projets de coopération, qui peuvent d'ailleurs dépasser le cadre d'un pilotage par le SCD si des mutualisations sont envisagées avec les archives de l'université par exemple.

Le rôle du bibliothécaire est de mettre en place la planification de la préservation, de la même façon qu'un plan de conservation pour les documents physiques est défini par les équipes des SCD. Il s'agira ainsi de sélectionner les documents, et de définir les tâches et responsabilités de chacun et les processus retenus en s'appuyant sur le modèle OAIS, dans une optique de gestion de risques. Les informaticiens quant à eux fournissent un service technique apte à répondre à une demande, mais ne définissent pas le besoin de préservation, tâche qui est du ressort des personnels du SCD. La dimension archivistique (et de gestion de projet le plus souvent) est ainsi assurée par le bibliothécaire, et la mise en place technique, par l'informaticien. Concrètement, l'établissement des métadonnées dans les SCD est du ressort des personnels de bibliothèque, de même que les mapping de métadonnées. Les développements informatiques sont du ressort des services informatiques, ou des développeurs de l'université. Un projet d'archivage pérenne au CINES définit un interlocuteur unique et un seul référent projet au sein de l'archive et du service versant. En SCD, le CINES fait le constat que tous les référents projet, à une exception près (Cujas), sont de profil documentaire et non informaticien. Or tous les avis des référents projet s'accordent sur la nécessité pour le SCD d'assurer une coopération constante avec les services informatiques, et de s'ouvrir suffisamment aux enjeux de l'informatique documentaire pour suivre les projets, en particulier si un dialogue avec un prestataire de service est nécessaire. Inversement, un profil « informaticien » risquera de ne pas être en contact suffisamment rapproché avec les fonds. Les projets ont donc tout à gagner d'une complémentarité entre les deux profils.

Le recours à un prestataire a parfois lieu. Cujas l'avait envisagé pour la constitution des fichiers METS, mais le coût de la prestation avait fait pencher l'équipe vers une « décision de travailler d'abord en interne »<sup>114</sup>. C'est rarement le cas pour l'établissement des métadonnées : plusieurs professionnels soulignent le

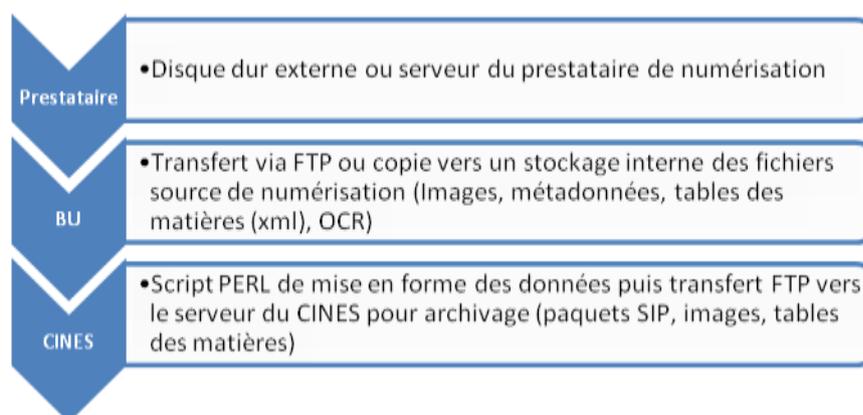
---

<sup>112</sup> *Op. cit.* p. 6.

<sup>113</sup> CINES, « Un concept, des problématiques » [En ligne] (page consultée le 30/12/2015) <https://www.cines.fr/archivage/un-concept-des-problematiques/>

<sup>114</sup> Noëlle Balley, « Numérisation et archivage pérenne à la bibliothèque Cujas : un retour d'expérience », Journées du CTLes, 20 juin 2013. CTLes [En ligne] (page consultée le 30/12/2015) <https://www.ctles.fr/sites/default/files/ctles/documents/JPro2013Balley.pdf>

risque de déléguer cette tâche spécifique à des prestataires peu au fait des formats de métadonnées des bibliothèques. Cela peut être envisagé si le rétrospectif à traiter est important (projet de Strasbourg), et dans le cadre d'une anticipation des besoins définis par l'institution. La BUPMC a ainsi fait figurer parmi les missions du prestataire de numérisation la production des métadonnées de conservation, le prestataire fournissant un dossier diffusion et des fichiers archives distincts. La « chaîne » de travail, de la production de la numérisation jusqu'au dépôt au CINES, comprend alors trois acteurs, chacun responsable d'un stade de préservation du document numérique :



A la BIUS, le prestataire est en charge des scripts<sup>115</sup> et de la procédure FTP au CINES ; le personnel de la bibliothèque n'a pas la main sur les scripts. C'est également le prestataire qui modifie à la demande la forme des métadonnées en cas de problème lors de l'envoi des fichiers au CINES – leur établissement initial étant à la charge des bibliothécaires. Dans ces divers cas, le recours au prestataire est limité à une ou plusieurs étape(s) du projet.

Le cas de la BSG est à part. La bibliothèque a fait dès le départ le choix d'un prestataire (Diadéis), pour des raisons de personnel. Le CCTP mentionne ainsi à chaque nouveau marché de numérisation les exigences liées à l'archivage pérenne, qui constituent l'un des critères de choix du prestataire. Le PPDI a été écrit par l'équipe de la BSG, et les métadonnées sont produites par la bibliothèque, qui se charge également de la conversion en Dublincore afin de limiter le risque d'erreurs sur les métadonnées. Le rôle dévolu au prestataire comprend ensuite toutes les étapes propres à l'archivage, en relation avec le CINES : la création de balises dans le Dublincore, la création de fichiers SIP, le transfert FTP au CINES, les tests, puis une fois la validation faite, l'envoi du lot sur la plateforme PAC. Une fois l'image envoyée, la bibliothèque valide l'envoi, et le prestataire y associe les paquets d'archive SIP comprenant les métadonnées. Le prestataire est en donc en relation directe avec le CINES, y compris en cas de rejets des dépôts.

Cette configuration, qui introduit véritablement un nouvel acteur dans les processus, n'est pas jugée idéale par la bibliothèque qui relève plusieurs inconvénients à ce fonctionnement : à chaque nouveau marché, il est possible de devoir reprendre intégralement les relations avec un nouveau prestataire, qui lui-

<sup>115</sup> Les scripts vérifient dans les serveurs les modifications effectuées sur les fichiers, et recensent les nouveautés à archiver ou à mettre à jour.

même doit nouer de nouvelles relations de travail avec les équipes du CINES ; la confiance dans le prestataire doit être complète ; la place de la bibliothèque n'est pas idéale, puisque située entre deux interlocuteurs qui interagissent ; la définition des responsabilités en cas d'échec des dépôts est plus complexe. Pauline Rivière (service informatique de la BSG) note ainsi qu'avec une équipe plus étoffée, le choix aurait été fait de travailler en interne. La présence d'un seul ETP sur la numérisation rendait cette option impossible, malgré la coopération des magasiniers et des conservateurs durant la phase de sélection scientifique des documents effectuée en amont. Le choix d'un prestataire vise ainsi à rendre le projet faisable, et à gagner du temps dans l'organisation interne. Les métadonnées techniques demeurent cependant la tâche des bibliothécaires, le prestataire n'étant pas qualifié pour les définir ; il serait envisageable d'avoir recours à un second prestataire, dédié à cet établissement des métadonnées, mais le fonctionnement n'en serait que plus complexe encore.

### 3. Projets d'archivage pérenne en bibliothèque universitaires : retours d'expérience

#### A. A l'initiative du projet

L'archivage pérenne figure rarement de façon explicite dans les missions des SCD. Pour autant, ce sont les SCD eux-mêmes qui sont à l'origine des projets d'archivage pérenne, en particulier pour les ressources mises en ligne depuis le lancement de la PAC du CINES et l'émergence de la problématique. Les projets sont généralement portés par les SCD eux-mêmes. A Lyon 1 par exemple, le projet d'archivage pérenne (non mené à terme à ce jour) était issu d'échanges entre la responsable du service informatique, la responsable des collections, et pour l'université la responsable des archives, sur une décision du SCD.

Si l'archivage pérenne ne figure pas dans les politiques documentaires formalisées des établissements, les projets en cours sont bien sûr mentionnés dans les rapports d'activité, comme c'est le cas à l'université de Lorraine, dont le rapport d'activité 2013 des bibliothèques universitaire de DDE (direction de la documentation et de l'édition) définit un axe « Valorisation et archivage pérenne des collections numériques ». L'archivage pérenne est ici associé à la valorisation des collections : il est perçu comme le prolongement logique de la numérisation des documents patrimoniaux et de la mise en production de Pulsar, « plateforme de consultation des livres anciens numérisés ».

La position de l'université est dans chaque cas favorable, mais le plus souvent sans appui particulier (ni en moyens ni en ressources humaines). L'impulsion de l'université même s'inscrit généralement dans un projet de mutualisation, qui englobe et dépasse alors le seul SCD. Dans le cas de l'Unistra, le vice président a été missionné sur la question de l'archivage pérenne ; le président de l'université a soutenu le projet de façon visible. Ce cas d'écoute politique assure une action efficace et coordonnée, des moyens, et une motivation des équipes bien supérieure, les enjeux du projet excédant le cadre du seul SCD. Il s'inscrit dans le cadre du « Schéma directeur de la documentation universitaire pour le site alsacien », au sein duquel le projet « AOC » découle de l'axe 1 ainsi formulé : « Définir des politiques concertées afin d'optimiser les moyens du site -

financer une étude de faisabilité pour une solution de conservation partagée à l'échelle de l'Alsace (...) »<sup>116</sup>.

Certains établissements mènent une étude de faisabilité. Ainsi à Lyon 1, une étude de faisabilité a été menée grâce au mémoire professionnel d'un stagiaire et à une analyse de fond opérée par un agent titulaire porteur de projet. Une mise en commun des besoins et des réflexions techniques a été opérée avec le service juridique. Ces études formalisées demeurent minoritaires, mais les équipes avant de s'engager auprès du CINES, après une première prise de contact, s'accordent des délais de réflexion afin d'établir la faisabilité du projet au vu des exigences techniques et budgétaires. Ce délai, parfois de plus d'un an, aboutit régulièrement selon le CINES à l'abandon du projet avant sa concrétisation, pour des raisons de coûts ou de ressources humaines insuffisantes.

### ***B. Les retours d'expérience des professionnels***

La bibliothèque Cujas, qui a été le premier établissement universitaire à mener un projet d'archivage pérenne (au CINES), est souvent consultée par les SCD qui envisagent un archivage pérenne : son expérience est considérée à la fois comme fondatrice et comme exemplaire, d'autant plus que les porteurs de projets ont formalisé leur retour d'expérience<sup>117</sup>. Globalement, les projets en production sont encore peu nombreux, au CINES ou ailleurs, et il est encore difficile de bénéficier d'un corpus significatif de retours d'expérience formalisés ; la dynamique est pour l'instant inverse : ce sont les établissements porteurs de projets en production qui ont pour ambition de partager leurs propres expériences. Pour autant, des universités se sont appuyées sur l'expérience acquises par les établissements en cours de projets, ou encore sur l'expertise de l'INIST concernant les données de la recherche. Les récents « groupes utilisateurs » du CINES offrent un lieu d'échanges des bonnes pratiques qui semble prometteur. Prévus pour les nouveaux services versants, l'objectif y est d'ajouter aux présentations du CINES des temps d'échange et de retours d'expérience (surtout sur les aspects techniques), selon les profils des utilisateurs.

Les retours d'expérience insistent sur diverses exigences méthodologiques, dont la nécessité d'opérer une sélection sur la base d'une typologie des documents , de définir un projet fort, de faire preuve de réalisme et d'anticipation, de mettre en place la gestion d'un processus complet (l'archivage pérenne doit être intégré dès le début d'une numérisation par exemple), et de témoigner une très forte exigence envers le prestataire de numérisation, avec un contrôle qualité très serré. Noëlle Balley relève divers obstacles rencontrés lors du projet de Cujas (« Un dossier beaucoup plus lourd qu'on ne le pensait au début », « Le PPDI oblige à relire en profondeur les objectifs et les méthodes de son programme », « Ambitions revues à la baisse sur les formats à archiver, à la hausse sur la rigueur des procédures », « Nécessite des compétences d'informatique documentaire (solides notions de XML, connaissance approfondie du DublinCore) ET de développeur », « Dialogue permanent indispensable entre l'informaticien et le bibliothécaire »...)<sup>118</sup>.

---

<sup>116</sup> Cf. « Schéma directeur de la documentation universitaire pour le site alsacien », mars 2014 [En ligne] (page consultée le 30/12/2015) [journal.insa-strasbourg.fr/wp-content/.../note-de-synthèse-du-SDD.pdf](http://journal.insa-strasbourg.fr/wp-content/.../note-de-synthèse-du-SDD.pdf)

<sup>117</sup> Voir ainsi la présentation de l'intervention de Noëlle Balley, Chef du département des Monographies et porteuse de projet, aux Journées du CTLes, 20 juin 2013 (*op. cit.*)

<sup>118</sup> *Op. cit.*

Globalement, c'est la technicité du projet et la rigueur exigée que les professionnels soulignent.

### *C. Volumétrie et typologie des collections retenues pour l'archivage pérenne*

Bien que les volumétries soient nécessairement appelées à augmenter régulièrement, les estimations au sein des SCD misent sur des volumes demeurant faibles. Les établissements sont conscients des coûts et de l'importance de leur soutenabilité à long terme ; beaucoup cherchent ainsi à limiter la taille des jeux de données, et favorisent les données orphelines ou de longue traîne, opérant une sélection en amont.

Il est difficile de quantifier les volumes de fichiers déposés au CINES par les bibliothèques universitaires ; ces dépôts restent cependant encore très minimes en matière de volumétrie. L'enquête que l'université de Strasbourg a menée sur les fonds à archiver de façon pérenne a par exemple conclu à une volumétrie relativement restreinte, constat partagé par la plupart des établissements. De même à l'université de Lorraine, où il est question de 11 000 documents, soit 13 000 fichiers pour un total de 100 Gigas. La plateforme de consultation des livres anciens numérisés Pulsar pour sa part contient environ 60 documents soit 75 fichiers. L'université de Lorraine a ainsi établi que les ressources qui seraient éligibles à un archivage pérenne seraient loin d'atteindre 1 TO annuel. A Lyon 1, le volume envisagé était d'environ 2 TO. La BUPM n'arrive pas encore à 1 TO, de même que Lille 2 malgré la présence d'images et de fichiers sons. Les volumétries maximum sont de 4 TO pour Medic@ (la banque d'images comptant pour 57 gigas) et 5 TO pour la bibliothèque numérique de la BSG.

Dans les faits, le CINES constate que les bibliothèques universitaires qui s'adressent au service PAC le font de façon très majoritaire :

- pour des numérisations d'ouvrages, parfois de corpus de cours, comme à Cujas ; de thèses et mémoires comme aux SCD de Lille 2 et de Lorraine ; ou d'ouvrages et de revues de presse comme à Science Po ;
- pour de la littérature grise.

C'est-à-dire donc pour des documents dont la perte engendrerait une perte financière pour l'établissement – dans le cas des numérisations, la perte du fichier numérisé implique une nouvelle numérisation ; ou qui ne peuvent être retrouvés sur d'autres supports ou à d'autres endroits ; ou pour lesquels l'établissement a une obligation de conservation (thèses d'exercice) et pour lesquels aucune solution comme STAR n'existe.

	Publications scientifiques	Mémoires, HDR, littérature grise, thèses d'exercice	Documents patrimoniaux numérisés	Données de la recherche	Documents pédagogiques	Abonnements onéreux
Cujas			x		x	
BUPMC			x			
BSG			x			
BIUS		x	x			
Lille 1	x	x	x	?	?	
Lille 2		x				
Science Po		x				
Unistra et BNUS	x		x	?		
Lorraine		x	x			
Toulouse	x					
Lyon 1		?	x			
Clermont			x			
Toulon	x					
Angers	x		x			

#### Typologie des documents retenus pour un archivage pérenne par les SCD enquêtés

Le projet doit définir en amont la nature des paquets d'archivage, qui varient selon les collections concernées. A Cujas, « L'unité intellectuelle (...) retenue était le volume physique traité (i.e. 1 volume physique = 1 paquet archivé). A celui-ci, correspond un numéro de code-barres qui (...) sert d'identifiant tout au long du processus de numérisation. Cela signifie autant de paquets d'archives que de volumes pour une œuvre en plusieurs volumes par exemple »<sup>119</sup>.

Les fichiers sont définis en accord avec le CINES ; sont généralement retenus les fichiers master (en TIFF le plus souvent pour les numérisations), le fichier table des matières au format CSV (indiquant la concordance entre les pages et les images) et les fichiers XML de métadonnées (fichier Dublin Core, fichier METS).

#### *D. Les équipes et les profils des équipes projet en bibliothèques universitaires et dans les services informatiques*

Les équipes comportent, au sein des SCD, de une à cinq personnes, pour un total rarement supérieur à quatre ETP.

A Cujas, Noëlle Balley, archiviste paléographe, est la responsable du projet numérique ; le groupe de travail a été constitué, outre le responsable du service informatique, d'un ingénieur informatique (qui n'était pas affecté à plein temps sur le projet), d'une documentaliste, et d'un PRCE, soit 5 personnes - mais moins de 5 ETP consacrés au projet. A la BIUS, le choix a été fait d'adopter des profils chartistes pour assurer le suivi du projet en production. Le lancement a nécessité deux à trois ETP. A Science Po, deux personnes ont porté le projet, un ingénieur-

<sup>119</sup> LEULMI Brouk, RENARD Catherine, « La bibliothèque Cujas : l'archivage du fonds des ouvrages numérisés en histoire du droit », Dossier « Archivage numérique pérenne », *La Gazette du CINES*, février 2013, p. 22.

développeur et la responsable du Service Numérisation, assurant ainsi la complémentarité entre les deux métiers supports.

Il arrive que le groupe projet en SCD soit composé d'une seule personne ; ainsi à la BUPMC, une seule conservatrice (à 100%) a assuré le portage du projet ; le poste était plus globalement dédié à la conservation en général et à la valorisation du patrimoine, et comprenait entre autres l'établissement d'un plan de sauvegarde. De même à l'université de Lorraine, l'équipe des personnels de bibliothèque est composée d'une seule personne, assistée des personnels de la direction du numérique (3 personnes) ; au total donc, 4 ETP mais qui sont loin d'être à 100% sur le projet (20% par exemple pour Maximilien Savoye, de la direction de la documentation et de l'édition de la bibliothèque universitaire de Metz). A Lille 2, le projet est porté par le SCD et la DSI. L'archivage au CINES, qui fait partie du projet « Pépite » (groupe de travail de six personnes), nécessite un prolongement de deux personnes : une informaticienne pour un total de 0,5 ETP, et un profil bibliothécaire pour 0,10 ETP. A Lyon 1 également, dans un premier temps, l'équipe supposée aurait compté « de 0.5 (avec externalisation) à 1 ETP (sans externalisation) pour le lancement du projet, et le suivi après mise en place aurait chargé à hauteur de 0.2 ETP environ un profil de personnel du SCD. Des ressources humaines en informatique étaient disponibles, mais il aurait sans doute fallu soit externaliser soit modifier les profils de poste pour absorber les besoins supplémentaires » selon François-Xavier Boffy, responsable Service Informatique Documentaire de la BU Sciences de l'université Claude Bernard.

Le CINES assure le soutien des projets, tant au niveau informatique qu'archivistique ; l'équipe « archivage pérenne » du CINES est rompue à la gestion de projet autant qu'à la résolution des problèmes de formats. L'outil FACILE proposé sur la plateforme du CINES permet de valider aisément les formats adéquats à l'archivage pérenne. Le rôle de conseil du CINES porte entre autres sur le choix des formats, et oriente les stratégies d'archivage des BU, permettant par exemple à Cujas de « revoir les ambitions à la baisse » et d'assurer la faisabilité du projet<sup>120</sup>. L'aide du CINES s'arrête cependant au soutien technique et archivistique : les divers développements informatiques doivent être réalisés en interne, ce qui nécessite tant des compétences spécifiques que les ETP nécessaires.

Au niveau informatique, l'investissement de départ - environ un an pour Cujas par exemple - est en effet loin d'être anodin, comme le souligne Catherine Renard, et nécessite des développements propres (XML, fichiers SIP, fichiers METS). Des développements en Java sont par exemple nécessaires, afin d'instaurer des commandes afin de constituer les paquets d'archives et lancer la procédure d'envoi par FTP. Le suivi des routines informatiques est également effectué en interne, et non par le CINES.

La plupart des SCD, quand ils ont les moyens humains de gérer le projet, font cependant souvent le constat d'un manque de ressources humaines en informatique, car ils sont régulièrement sous-dotés. De nombreuses bibliothèques dépendent de la DSI quant aux postes d'informaticiens ; or la plupart des SCD sont en demande de leurs propres postes d'informaticiens dédiés. Le plus souvent, seule une aide ponctuelle est apportée par la DSI aux projets nécessitant un développement informatique, les DSI étant impliquées par ailleurs dans d'autres projets (réinformatisation comme à l'université de Lorraine) et donc peu

---

<sup>120</sup> Noëlle Balley, *op. cit.*

disponibles. Un haut niveau technique des équipes informatiques permettent a contrario d'assurer une avancée efficace du projet, comme à Lille 2.

### *E. Le déroulement du projet*

Le CINES décrit les étapes d'un projet d'archivage pérenne dans la plateforme PAC<sup>121</sup> et définit quatre étapes génériques : phase préparatoire, phase de test, phase de développement, phase de déploiement en production. Noëlle Balley a également présenté avec précision, et vues du côté de l'établissement, les étapes du projet mené à Cujas<sup>122</sup>. Les grandes étapes du projet sont ainsi : la signature d'une convention avec le CINES ; la définition des « paquets d'archives » (SIP)<sup>123</sup> et des informations à préserver ; la création du PPD<sup>124</sup> ; la phase de tests ; les développements ; et l'automatisation des « routines ».

Les SCD qui ont mené un projet d'archivage pérenne s'accordent sur l'importance d'anticiper dès l'amont d'un processus de numérisation plusieurs besoins liés à l'archivage : réfléchir aux règles de nommage et d'adressage, documenter les métadonnées. Car dans le cas d'un projet de numérisation, l'archivage pérenne oblige aussi à accentuer le contrôle qualité, ce qui a un impact sur toute la chaîne de numérisation en amont et alourdit voire ralentit les procédures. Ce travail a par exemple été fait à Lille 2, dont la spécialiste de numérisation a défini des critères précis de nommage et d'adressage des fichiers en amont du projet afin de faciliter l'archivage pérenne au CINES pour les thèses et les mémoires. Le projet AOC (Strasbourg), en cours, a quant à lui désigné le même chef de projet technique sur son AOI, « AOC », et sur l'archivage pérenne de cette même AOI : ce choix stratégique apparaît particulièrement pertinent, puisque les décisions effectuées pour l'AOI engageront le futur archivage pérenne. Ainsi le choix des métadonnées et la constitution des collections seront par exemple établis de façon cohérente, garantissant la faisabilité de l'archivage pérenne ultérieur.

Le projet d'archivage pérenne est par nature complexe, et il est difficile de déterminer une date précise de mise en production dès l'orée du projet. La planification d'un projet en amont s'avère par ailleurs difficile, car du temps s'écoule généralement entre la prise de contact avec le CINES et le début réel du projet. A Cujas, la durée de l'avant-projet a été de 8 mois, suivie d'une phase de tests de 6 mois. S'agissant uniquement de documents PDF et de thèses numérisées, la tâche peut être assez rapide si le service dispose de suffisamment d'ETP d'informaticien. S'agissant d'images, le projet peut être nettement plus long : de 2 à 3 ans, selon la taille du service (trois années à la BUMPC par exemple, entre le lancement de la Jubilothèque et la mise en production de l'archivage au CINES). Les données de la recherche impliquent quant à elles une réflexion sur le long terme.

---

<sup>121</sup> CINES, « Comment archiver au CINES » [En ligne] (page consultée le 30/12/2015) <https://www.cines.fr/archivage/comment-archiver-au-cines/le-deroulement-dun-projet-darchives/>

<sup>122</sup> N. Balley, *op. cit.*

<sup>123</sup> L'unité d'archivage est le paquet d'archive, qui correspond à un volume physique. Un dépôt de paquet SIP comprend un bordereau d'envoi (en XML), par exemple une image en format TIFF, une table des matières (en format METS). Le bordereau XML décrit précisément l'envoi. Le CINES compare ensuite le bordereau avec le contenu pour en vérifier l'authenticité et l'intégrité.

<sup>124</sup> Document de nature archivistique qui décrit la nature des fonds service versant, les éléments techniques (circuits de production : OCR...), le contenu et les objectifs de l'archivage, les droits d'accès aux documents, la structure de l'unité d'archivage, les règles de saisie de la notice Dublincore, du Mets, etc.

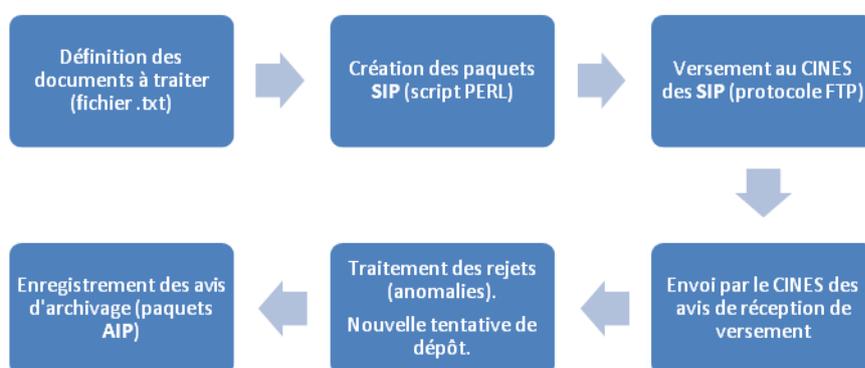
Selon Noëlle Balley, la planification à Cujas fut la suivante :

- « • Contacts informels avec le CINES dès 2008
- 2009 : évaluation des besoins dans le cadre du contrat quadriennal 2010-2013 (...)
- Réunion de lancement le 1er octobre 2009
- Signature de la convention le 10 mai 2010
- Réunions téléphoniques et échanges par mail avec le CINES
- Comptes rendus réguliers en interne par mail, par écrit et par oral
- Entrée en production le 22 février 2011 à 10 h 18
- Réunions annuelles de bilan ».

L'université de Lorraine vise une mise en production fin 2016, soit au total un an de conduite de projet (le projet a été interrompu un an), de même qu'à Lille 2. A Lyon 1, la durée envisagée était d'un an à 18 mois.

### *F. La gestion des routines : une fois en production, quelle gestion ?*

Une fois en production, la gestion suppose à la fois diverses routines informatiques, le traitement des mails d'anomalie envoyés au service versant en cas d'erreurs lors des versements (dites anomalies pré-dépôts), et une anticipation des besoins en volumétrie selon l'accroissement des collections.



A Cujas, « les informations nécessaires à la génération du SIP ainsi que les fichiers à archiver sont récupérés automatiquement dans les répertoires sources de (la) bibliothèque. Une fois les fichiers sip.xml et METS générés et après un dernier contrôle qualité, le SIP est envoyé automatiquement au CINES. Le certificat d'archivage (AIP) est récupéré automatiquement dans une base de données qui permet de gérer les différentes versions des archives, si nécessaire ».

Les SCD affectent généralement une seule personne (qui n'est pas toujours à temps complet sur le projet) à la gestion des tâches dites de « routine » - sans compter le recours au service informatique, parfois nécessaire. A Cujas, la constitution d'un paquet d'archive est évalué à environ 10 minutes, et le lancement de la procédure d'envoi par FTP à 30 minutes de temps humain et une heure d'envoi au minimum. Les routines informatiques sont encadrées par l'ingénieur

informatique, également développeur en Java, qui gère les divers « bugs ». L'université de Lorraine estime qu'un versement par mois sera suffisant pour adresser au CINES les documents académiques (environ 1000 documents produits par an). La mise en place d'un web service contrôlant la conformité des fichiers y évitera de plus le contrôle manuel dans FACILE, qui sert à anticiper le dépôt initial en évitant le rejet de fichiers, de même qu'à Science Po. A la BIUS, une seule personne, dont le profil de poste est dédié à l'archivage pérenne, suit le traitement des anomalies au fur et à mesure de leur envoi. Le service informatique intervient en soutien sur les aspects techniques, en relation avec le prestataire ou le CINES.

A plus long terme, suivre et anticiper la volumétrie, en coopération avec le tiers-archiviste, est indispensable à la fois pour permettre à ce dernier d'évaluer les besoins d'hébergement et éventuellement revoir les termes de la convention, et pour avoir une vision large sur les collections archivées et leur devenir. Le projet de Science Po établit ainsi une projection sur cinq ans (2015-2019) pour la numérisation patrimoniale, les archives électorales et les dossiers de presse.

#### **4. Les freins et les causes des abandons des projets**

Selon le CINES lui-même, le premier frein aux projets d'archivage pérenne des SCD auprès du CINES est financier ; le second est lié à la disponibilité des personnels, que ce soit au SCD et/ou au sein du service informatique ; le troisième, aux compétences informatiques disponibles dans les établissements. Il peut arriver que ces trois facteurs (personnel, budget, compétences informatique) se cumulent. Demeure alors soit la possibilité d'avoir recours à un prestataire ; soit, si les coûts ne le permettent pas, l'abandon temporaire ou définitif du projet. Le tiers-archiviste public rapporte ainsi que beaucoup de projets sont en « stand-by » : après avoir pris un premier contact, les porteurs de projets suspendent le processus, le plus souvent par manque de ressources, financières et humaines. Par exemple, l'université Paris 1 a contacté le CINES dans l'objectif de mettre en place un projet d'archivage pérenne, afin de profiter de l'archivage pérenne en cours à la BSG pour y adjoindre l'archivage pérenne du SCD et optimiser la volumétrie. Après ce premier contact en 2010, le projet a été suspendu. Dans d'autres cas, les porteurs de projets sont entre temps appelés sur d'autres projets jugés plus prioritaires et n'ont plus de temps à consacrer à l'archivage pérenne. Catherine Renard note également que plusieurs BU sont venues à Cujas s'inspirer de l'expérience de la bibliothèque, mais ont ensuite renoncé à leur projet initial au vu du coût pluriannuel à assumer. Quant aux BU enquêtées, elles placent également la question du coût devant celles de la disponibilité et de la technicité des projets.

##### ***A. Le coût***

Sans financement national, l'archivage pérenne est supporté par le budget de fonctionnement général des SCD. C'est l'obstacle le plus mentionné par les établissements eux-mêmes (Toulouse (OATAO), Lyon 1, Clermont Université). Même si les coûts ont beaucoup baissé depuis le lancement de la PAC au CINES, ils demeurent, même dans le cas de volumétries faibles, trop élevés pour des établissements aux moyens restreints ou des établissements soumis à une diminution de leur budget. Or de nombreux SCD constatent qu'entre HAL et

STAR, une partie importante de leurs ressources sont déjà au CINES. Pour les documents restants, pour lesquels il s'agit de déterminer le caractère d'urgence de la conservation, le niveau de risque d'un simple stockage en interne, et le coût d'un archivage pérenne par convention au CINES, la tentation existe de considérer une conservation a minima. Dans le cas de l'AOI OATAO (Toulouse) par exemple, les coûts du CINES sont jugés élevés ; au vu des nombreuses publications déjà présentes dans HAL et de la présence des thèses de l'INT au CINES via STAR, l'archivage pérenne des documents restants paraît onéreux par rapport à la quantité documentaire (moins d'un TO annuel). Ces documents sont donc conservés sur les serveurs de la DSI. L'archivage pérenne était pourtant prévu dans le projet de l'AOI ; mais le projet<sup>125</sup> a été abandonné pour des raisons financières. Lyon 1 évoque également le coût annuel du service rendu par le CINES, considéré comme trop élevé. Un engagement envers l'archivage pérenne au SCD aurait été sans doute pris par l'université, mais le projet, suite à une étude préalable, est abandonné. A Clermont université, la bibliothèque numérise depuis 2011 ; la question de l'archivage pérenne se pose tôt, mais le volume est faible au début, tandis que les coûts au CINES sont à l'époque de 6000 euros/TO. L'équipe considère après réflexion que le corpus, présent à la bibliothèque, peut être renumérisé en cas de perte des fichiers : le critère appliqué est donc la disponibilité de l'original. L'impossibilité d'avoir recours à un archivage pérenne fait ainsi envisager un projet a minima, et fait courir le risque économique aux établissements d'avoir à renumériser à partir de leurs originaux. A cela s'ajoutent les risques en termes de conservation des originaux eux-mêmes, que les opérations de numérisation ne laissent pas toujours indemnes, et la potentielle perte de cohérence des collections ainsi numérisées.

### *B. La disponibilité*

En terme de management, les SCD rappellent souvent qu'un projet d'archivage pérenne est une activité parmi d'autres : le suivi de ces projets s'inscrit dans des contextes où la dotation en ETP est limitée et doit donc composer, comme les retours d'expérience le soulignent, avec peu de ressource RH. Plusieurs SCD (Clermont université, BUPMC, université de Lorraine...) mentionnent ainsi la difficulté de maintenir des équipes stables et suffisantes pour la conduite de projets ; en cas de gels de postes, ou de diminution des ETP, la mission archivage peut être compromise ou différée. Suite au manque de postes, il est difficile d'entamer une réflexion de long terme sur la problématique et d'envisager porter un projet technique aux procédures lourdes. A cela s'ajoute le fait qu'un projet sans impact visible et immédiat a tendance à ne pas être prioritaire.

### *C. La technicité*

Un projet d'archivage pérenne est reconnu comme particulièrement exigeant par les établissements qui s'y consacrent ou s'y sont consacrés, sur le plan de la maîtrise technique et de la rigueur des pratiques d'informatique documentaire. Cette technicité nécessite une montée en compétence des équipes des SCD, qui doivent être aptes à

---

<sup>125</sup> Voir la présentation de Sandrine Malotiaux et Jean-Marie Le Béhec : « Une archive ouverte institutionnelle : OATAO », 9 février 2010 [En ligne] (page consultée le 30/12/2015) <http://fr.slideshare.net/sandrinemalotiaux/oatao-urfist-bordeaux-2010>, p. 38 : « Assurer l'archivage pérenne » et p.57 : « Sauvegarde externe et conservation : le CINES » (2010).

dialoguer avec le CINES et les informaticiens. Les établissements mentionnent ainsi le grand nombre de procédures écrites complexes fournies par le CINES (plusieurs centaines de pages dont il faut extraire les données pertinentes au regard du projet mené), le vocabulaire et les notions informatiques complexes, la nécessité d'adapter les cahiers des charges des éventuels prestataires de numérisation, la rigueur du nommage et de l'adressage des fichiers à adopter sous peine de rejet par le serveur du CINES, l'adaptation des métadonnées (le suivi de la norme ISO oblige à l'ajout ou à la modification de certains champs), la documentation exacte des dates de production des fichiers pour savoir quoi et quand sauvegarder... A Clermont, les tests entrepris avec CINES ont ainsi placé les équipes face à des exigences très fortes quant à la qualité des images et des métadonnées. Améliorer la qualité des versements est apparu comme une tâche très lourde, en particulier en ce qui concernait le diagnostic des erreurs lors des versements – qui exige une interprétation des logs afin de déterminer la cause de l'erreur. Selon les collections, des problèmes spécifiques peuvent se poser. Ainsi à Science Po, l'archivage pérenne des dossiers de presse a nécessité un renommage des dossiers important pour les fichiers SIP ; la structuration de paquet archivable a également posé problème, le sip.xml n'étant « pas très riche »<sup>126</sup> compte tenu du fait que les métadonnées existantes étaient faibles.

Globalement, l'archivage pérenne implique une discipline rigoureuse qui s'impose à plusieurs entités d'un SCD – services en charge de la numérisation, de la conservation, etc. – et qui peut contredire une autre priorité – par exemple, assurer une mise en ligne rapide d'un projet, quitte à accepter certaines petites imperfections. Or une numérisation peu rigoureuse peut faire obstacle à un projet d'archivage pérenne, les exigences en matière de nommage s'avérant insurmontables si le nombre de fichiers à renommer est trop important et les moyens insuffisants en interne. Il en est de même pour l'adressage des paquets SIP (il est nécessaire d'indiquer au CINES le chemin des fichiers, les métadonnées devant donc être parfaitement conformes sous peine de rejet par le serveur). A Cujas, la numérisation avait commencé dès 2007, mais à petite échelle : il a ainsi été possible de reprendre le nommage des fichiers. Noëlle Balley revient sur les difficultés engendrées par les « libertés » prises au début du projet de numérisation :

« Les libertés prises au début du projet ont obligé à un très gros travail de vérification et de réfection avant archivage : un an de travail sur les données mises en ligne depuis 2008, avec interruption du travail de numérisation interne pendant 1 an. Il a fallu refaire TOUT ce qui avait été mal fait au départ... Chaque changement majeur implique de reprendre l'archivage ».<sup>127</sup>

Ainsi, l'INHA a renoncé à son projet suite à un nommage inadéquat de ses fichiers. A Lille 1 également, l'archivage pérenne n'a pas été envisagé dès le départ, en matière de nommage et d'organisation des fichiers en particulier. Le problème s'est surtout posé pour la bibliothèque numérique patrimoniale. Depuis le projet d'archivage pérenne, le processus de numérisation a été adapté, mais le problème se pose toujours pour les fichiers à traiter de façon rétrospective. D'où l'intérêt des solutions qui intègrent dans une même chaîne de travail la diffusion et

---

<sup>126</sup> Source : Oleséa Dubois, responsable du Service Numérisation, Science Po.

<sup>127</sup> *Op. cit.*

## II. Projets d'archivage pérenne en bibliothèques universitaires : retours d'expérience

l'archivage pérenne (tels Gallica ou HAL) : l'archivage y est alors anticipé, et sa gestion, prise en compte très en amont, pour un gain de temps et une efficacité très supérieurs. Pour autant, l'archivage pérenne peut constituer en lui-même une « clef de voûte »<sup>128</sup> du système de numérisation, qui permet de tester et certifier la cohérence et la justesse de l'ensemble des processus.

Se pose aussi la question des développements informatiques à mener en interne. Chaque établissement doit en effet mettre en place un outil pour construire automatiquement les paquets d'archive (soit un fichier et ses métadonnées) à envoyer au CINES. Suite à plusieurs demandes de SCD, le CINES réfléchit à fournir un tel outil aux services versants. Lorène Béchard, archiviste au département Archivage et Diffusion du CINES, remarque cependant que chaque système d'information ayant sa propre organisation, un seul et unique outil ne pourra pas répondre à tous les besoins ; les services versants devront continuer à fournir des données afin d'être identifiés par l'outil. Ces développements, jugés assez peu complexes à programmer, dépendent cependant des ressources informatiques des SCD.

Sur le plan stratégique enfin, un projet de numérisation, qui accorde une visibilité à l'établissement et à ses ressources à court ou moyen terme, offre à première vue davantage d'opportunités qu'un projet d'archivage pérenne par définition peu visible, conduit sur le long terme et qui immobilise de façon constante voire croissante au gré des téraoctets archivés une part du budget du SCD, de ce fait non disponible pour des besoins plus immédiats. Il peut donc être difficile de défendre par exemple une demande de poste en vue de la mission, qui semblera moins opérationnelle que d'autres (une modernisation des services par exemple). Le risque serait de ce fait d'assurer la valorisation et la diffusion des ressources en reportant ultérieurement la conservation, ce qui on l'a vu expose à divers risques – non prise en compte de l'archivage pérenne dès le début du projet, coûts non calculés, risque de perte de données selon les solutions transitoires adoptées. L'enjeu est alors de définir le niveau de priorité accordé à l'archivage pérenne au sein des SCD, ses coûts demeurant malgré tout supportables au sein des budgets universitaires, si on les rapporte au coût d'abonnement à des ressources électroniques ou au coût consacré au stockage de collections papier. C'est ainsi, au-delà du coût lui-même, la légitimité des projets qui est à renforcer.

L'enquête menée a permis de souligner que les projets mis en œuvre concernent très majoritairement les collections patrimoniales numérisées et les documents académiques qui ne sont pas pris en charge dans les dispositifs mutualisés (STAR ou HAL). Les difficultés, en particulier budgétaires, que rencontrent les BU dans l'archivage pérenne de ces documents poussent à délimiter leurs missions et responsabilités et à s'interroger sur une mutualisation qui pourrait assurer le succès d'une politique de l'archivage pérenne en France pour l'ensemble des établissements.

---

<sup>128</sup> L'expression est de Jean-François Vincent (responsable du service d'histoire de la santé, BIUS).

### III. SITUATION DES BIBLIOTHÈQUES UNIVERSITAIRES : VERS UNE NÉCESSAIRE MUTUALISATION

---

#### 1. Situation des SCD dans l'université et dans l'ESR

Un projet d'archivage pérenne au sein d'une université naît de certaines prises de conscience de la part des acteurs en charge des collections. Parmi ces acteurs, le SCD n'est pas seul en charge des productions de l'établissement : le service juridique, les archives lorsqu'elles sont présentes dans l'université, les quelques missions Patrimoine scientifique existantes peuvent également être amenées à concevoir un archivage pérenne, en lien avec les services informatiques. Par ailleurs, si le SCD doit assumer l'archivage de certaines publications, les périmètres doivent être définis entre ce qui relève de sa responsabilité légale, et ce qui relève d'une décision du SCD lui-même. De façon implicite, il apparaît cependant qu'au sein de l'université, c'est au SCD qu'échoit le plus souvent l'archivage pérenne des nombreux contenus, pour diverses raisons : il est déjà en charge de plusieurs circuits de gestion et d'archivage de documents en lien avec des opérateurs publics (HAL, et STAR qui est mis en place dans tous les SCD français ou voué à l'être) ; il est souvent en relation avec les producteurs de données –chercheurs, doctorants... – dans le cadre de ses activités (formation ou valorisation) et possède déjà une expertise dans le domaine des métadonnées et des publications de contenus ; il est identifié comme une instance de nature patrimoniale, surtout quand il est déjà en charge d'un fonds ancien. De ce fait, les SCD enquêtés n'ont, pour aucun d'entre eux, remis en cause la responsabilité qui s'est trouvée la leur et qui leur semblait naturelle ; ils n'ont que rarement évoqué l'implication effective ou potentielle d'autres acteurs de l'université, ou d'autres projets équivalents dans l'université. Pourtant, nous verrons que la mutualisation de l'archivage, au sein de l'université mais aussi au niveau d'une COMUE, peut faciliter ce type de projet, entre autres sur le plan budgétaire. Cet état de fait peut expliquer, parmi d'autres raisons, le faible nombre de SCD engagés dans ce type de projet, puisque le cadre est peu directif et peu contraignant : très peu d'obligations leur en sont faites ; le soutien financier est limité ; le SCD se retrouve généralement porteur de projet, ce qui implique de disposer du personnel et de la disponibilité nécessaires.

#### *A. Responsabilités et périmètres au sein de l'université*

Dans un premier temps, notons que le périmètre de responsabilités légales des SCD face aux documents, données et productions de l'établissement universitaire auxquels ils appartiennent est réduit : outre la responsabilité patrimoniale des fonds anciens, rares ou précieux, les SCD ont une responsabilité légale pour le dépôt des thèses de doctorat<sup>129</sup>. Au-delà, il s'agira d'une

---

<sup>129</sup> Le cas particulier des thèses est régi par 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 présentes en soutenance en vue d'un doctorat. » « Article 10 : L'établissement de soutenance procède au dépôt de la version validée de la thèse dans ses formats de diffusion et d'archivage, ainsi que du bordereau électronique, dans l'application nationale Star, gérée par l'Agence bibliographique de l'enseignement supérieur (ABES), qui assure les fonctions suivantes : - enregistrement du dépôt de la version de diffusion et de la version d'archivage de la thèse ainsi que de ses métadonnées ; - signalement dans le catalogue et le portail Sudoc ; - attribution d'un identifiant permanent ; - envoi de la version d'archivage, y compris

responsabilité « décidée par rapport à des besoins et des usages et qui relèvent donc d'une politique documentaire »<sup>130</sup>. Parmi leurs activités traditionnelles, les SCD assument déjà la gestion et l'archivage de publications scientifiques ou de numérisations patrimoniales ; l'archivage pérenne de ces ressources peut alors sembler le prolongement naturel de la mission initiale d'acquisition ou de collecte, de diffusion et de valorisation des contenus, puisque dans ces cas précis, les professionnels sont directement confrontés à la problématique de l'archivage et doivent définir les risques encourus par leurs collections. Dans les faits d'ailleurs, on a déjà relevé que la plupart des projets d'archivage pérenne menés en université sont issus de la volonté propre des SCD de procéder à la préservation des ressources dont ils ont la charge. Dans le cadre de ce périmètre défini, la décision est alors motivée par la responsabilité des documents et de leur conservation. Car si l'archivage pérenne ne fait pas explicitement partie des missions des SCD, ils ont cependant la mission, de portée générale, de « conserver les documents »<sup>131</sup>. Lorsque les SCD participent à des programmes nationaux qui sont en charge de certains contenus (STAR, HAL, DUMAS), les responsabilités et missions des acteurs sont définies en amont et ne nécessitent plus de prise de décision à proprement parler. Mais certains contenus, auxquels sont confrontés certains SCD, telles les données de la recherche produites par les laboratoires par exemple, ne sont pas de son ressort direct : c'est la stratégie globale de l'université en matière d'archivage qui est en jeu, et l'on verra qu'il est difficile pour un acteur isolé de faire face aux difficultés engendrées par la préservation des données de la recherche. D'autres productions sont plus largement du ressort de l'ESR en général ; on pensera par exemple aux archives des abonnements onéreux de périodiques, acquises au sein du programme ISTEEX. D'autres instances enfin peuvent être en charge de l'archivage pérenne de certaines ressources de la recherche (comme le CNRS), ce qui pose également la question de l'articulation entre les organismes de recherche et l'université, que BSN fait justement dialoguer. On pourra alors demander s'il est nécessaire pour le SCD de conserver par lui-même « des contenus numériques qui sont déjà préservés ou qui, réglementairement, doivent être préservés par d'autres »<sup>132</sup>.

Au-delà des responsabilités patrimoniale et légale qui motivent de nombreux projets d'archivage pérenne, les établissements ont donc à établir par eux-mêmes leur périmètre d'intervention, en concertation avec les autres acteurs (archives, etc.). Une liberté d'appréciation est laissée au SCD, qui peut dans une certaine mesure décider de ce qu'il soumettra ou non à un archivage pérenne. Plusieurs SCD enquêtés font d'ailleurs état de décisions prises en dehors de tout cadre légal

---

dans le cas d'une thèse non diffusable, au Centre informatique national de l'enseignement supérieur ; - le cas échéant, à la demande de l'établissement, envoi des métadonnées ou de la version de diffusion de la thèse vers les sites désignés par celui-ci. » (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 du doctorat, JORF n°195 du 24 août 2006 page 12471, texte n° 24). Un prochain arrêté sur le doctorat, qui oriente vers le dépôt numérique des thèses et à la suite duquel plus aucun dépôt papier ne serait possible, seules les bibliothèques universitaires conservant un exemplaire papier, est en préparation.

<sup>130</sup> LADANT Julie, « La conservation pérenne des collections de l'ESR : enjeux, actions et perspectives au niveau national » [En ligne] (page consultée le 30/12/2015) <https://www.ctles.fr/sites/default/files/ctles/documents/JPro2013Ladant.pdf>

<sup>131</sup> Voir à ce titre le Code de l'éducation, qui précise : « Le service public de l'enseignement supérieur (...) participe à l'étude et à la mise en valeur des éléments du patrimoine national et régional. Il assure la conservation et l'enrichissement des collections confiées aux établissements » (article L 123-6 modifié par la loi du 22 juillet 2013).

<sup>132</sup> CLAERR Thierry, MOUFFLET Jean-François, « Gestion de la conservation des collections numériques », in *Intégrer des ressources numériques dans les collections*. Villeurbanne, France: Presses de l'Enssib, 2014, p. 155.

ou réglementaire, selon des critères établis en interne et selon les besoins et les capacités des services. Ce sont les documents les moins encadrés sur le plan réglementaire qui sont bien sûr soumis le plus à cette latitude de jugement, telles les productions pédagogiques ou les expositions virtuelles. Pour ces contenus, les SCD doivent en appeler, comme vu précédemment, à un critère de patrimonialisation, lui-même difficile à établir. De fait, la Charte de la conservation dans les bibliothèques précise qu'« est dit patrimonial un document, un objet ou un fonds auquel est attachée une décision de conservation sans limitation de durée » : la réglementation en matière de patrimoine nécessite donc une interprétation de la part des acteurs qui en sont chargés. Dans certains cas, où le statut du document ou de la donnée est incertain, la décision pourra être collégiale et excéder le cadre de la seule BU, comme le note Hélène Richard : « Le statut patrimonial des fonds doit donc être, à l'exception des ensembles incontestables tels que nous les avons décrits plus hauts, le résultat d'une discussion entre les responsables du fonds, les autres institutions patrimoniales de référence et le monde de la recherche »<sup>133</sup>. Ainsi la décision de « patrimonialiser » telle ou telle collection numérique ou ensemble de données pourra relever d'une concertation entre le SCD, les UFR et les laboratoires et tiendra d'une politique générale de conservation de l'établissement, laquelle peut mener à une stratégie de mutualisation de l'archivage. Faire le choix de l'archivage pérenne fait en tout cas peser de nouvelles responsabilités sur les SCD, raison pour laquelle le choix du CINES s'impose le plus souvent, le tiers-archivage garantissant la qualité et la pérennité de l'archivage aux institutions versantes et leur offrant ainsi une garantie. De plus, s'impliquer dans l'archivage pérenne de certains projets, par exemple de projets de recherche, peut faire naître de nouvelles responsabilités, d'ordre contractuel entre autres, puisque de l'archivage, on attendra alors qu'il puisse assurer la valeur probante des données qui justifient une publication par exemple. C'est alors à chaque SCD de définir sa politique, selon ses ambitions et ses moyens.

### **Le rôle des services d'archives et des autres acteurs : des périmètres encore peu définis**

Les SCD ne sont pas les seuls à envisager un archivage pérenne au sein des universités : les services d'archives peuvent également être amenés à en envisager le projet<sup>134</sup>. La mission d'archivage au sein des universités est cependant généralement peu soutenue, peu d'universités ayant un service d'archives dédié. Comme le souligne Charlotte Maday :

« La présence d'archivistes en universités est relativement récente et (...) tout le travail de collecte et de conservation reste à faire dans la majorité des universités qui n'ont pas encore de service d'archives. Le réseau d'archivistes dans les établissements d'enseignement supérieur et recherche, appelé AURORE, est désormais bien structuré et s'est équipé d'outils fonctionnels élaborés à partir des expériences de collecte sur le terrain. Ces outils permettent aux professionnels arrivant en poste de débiter rapidement leur activité sans avoir à tout créer. Comptant actuellement plus de 110

---

<sup>133</sup> RICHARD Hélène, *La formation aux questions patrimoniales dans les bibliothèques. Rapport à monsieur le ministre de la Culture et de la Communication*, IGB, 2010, p. 14.

<sup>134</sup> Sur les archives en université, voir *La gestion des archives au sein d'un établissement d'enseignement supérieur et de recherche*, Les dossiers de l'agence, AMUE, 2010.

### III. Situation des bibliothèques universitaires : vers une nécessaire mutualisation

membres, on ne peut qu'espérer son développement de manière à améliorer l'efficacité de l'action de ces archivistes »<sup>135</sup>.

L'archivage du papier est ainsi souvent problématique, et l'archivage électronique l'est plus encore. Depuis la LRU, les universités sont par ailleurs confrontées à des difficultés en termes de personnel et de moyens financiers. Au sein des universités, il est pour l'instant difficile d'avoir le recul nécessaire sur la conservation pérenne numérique, aucun SAE (Système d'Archivage Électronique) n'ayant plus de 5 ans.

Les archives définies comme administratives sont traditionnellement du ressort du service des archives de l'université, quand il y en a un, ou des services administratifs ; leur archivage peut cependant se poser aux SCD dans le cas où aucun service ne se trouverait en charge de certaines archives définies comme administratives, et dont le SCD est également producteur. La spécificité des archives administratives, qui impliquent un travail sur les DUA et sur la durée de l'archivage (intermédiaire ou définitif) suppose une réflexion en parallèle de celle à mener sur les publications scientifiques. Les services d'archives sont aussi en charge des archives pédagogiques (supports de cours, données et documents issues de l'activité d'enseignement) et des archives scientifiques ou de la recherche (documents produits par les chercheurs dans le cadre de leur activité de recherche : cahiers de laboratoires, résultats d'expériences, calculs)<sup>136</sup>. La question de la répartition des missions se pose pour ces ressources ; certains SCD assurent par exemple la diffusion de contenus pédagogiques sur des portails (par exemple à Lille 2) et se posent la question de leur archivage. Les données de la recherche sont également problématiques ; c'est pour l'instant le SCD qui en est souvent en charge, ou qui mène une réflexion à ce sujet. Il peut également s'agir de la DSI, selon les établissements. Certaines ambiguïtés demeurent également concernant le périmètre des missions entre les archives publiques et le SCD. Par exemple dans le cas des archives des chercheurs : s'agit-il d'archives publiques ou d'œuvres de l'esprit, qui pourraient alors être traitées par les bibliothèques ?<sup>137</sup>

Concrètement, les pratiques varient selon les universités. De même pour les documents pédagogiques produits par l'université : légalement, il n'y a pas de textes qui définissent précisément si un support pédagogique doit être archivé par tel ou tel service ; dans ce cas aussi, les pratiques fluctuent. Certains périmètres sont donc encore à clarifier entre le SCD et les services d'archives au sein des établissements ; dans le cadre de BSN également, la discussion est en cours à ce sujet, dans un contexte spécifique puisque les établissements sont autonomes depuis la LRU. Une mutualisation au sein de l'université pourrait constituer une voie de clarification. L'université de Strasbourg envisage ainsi de mutualiser l'archivage pérenne de ses archives publiques et des productions scientifiques de l'établissement ; les deux types de productions feraient cependant selon Frédéric Blin, directeur de la conservation et du patrimoine de la BNUS, l'objet de deux circuits parallèles, les problématiques étant spécifiques en termes de durée de

---

<sup>135</sup> « L'archivage des données administratives : 5 questions à Charlotte Maday », Dossier « Archivage numérique pérenne », *La Gazette du CINES*, février 2013, p. 38. Voir aussi MADAY Charlotte, « Les archives en université, un si long chemin... », *Arabesques*, janv.-fév.-mars 2013, p. 18.

<sup>136</sup> Voir sur cette question MEYSTRE-SCHAEREN Nicole, « Quel patrimoine pour l'enseignement et la recherche ? », in Université de Lausanne, *Les archives dans l'université : actes du colloque international, Université de Lausanne, 30 et 31 octobre 2008*, Olivier Robert (dir.), p. 121-135.

<sup>137</sup> Même question pour les archives privées de chercheurs qui pose la question du statut de ces archives : s'agit-il d'archives privées ou publiques ? Selon la réponse qu'on apporte à cette question, le traitement des archives sera différent.

conservation des documents ou de droits. Ainsi, une fois collectées, les archives des établissements sont transmises par les services d'archives aux Archives Départementales ou Archives Nationales pour une conservation pérenne : le circuit et les solutions d'archivage ne sont pas les mêmes que ceux retenus par les SCD (le CINES le plus souvent), ce qui suppose également des traitements différenciés.

D'autres acteurs peuvent être également impliqués dans la conservation du patrimoine dans les universités. Les universités qui possèdent un patrimoine important ont ainsi instauré des Missions Culture (ou Patrimoine) Scientifique (université de Bourgogne, de Marseille, de Lille 1, ou de l'UPMC) qui demeurent principalement en charge de collections physiques (objets scientifiques)<sup>138</sup>. Le service juridique peut être confronté à la problématique de l'archivage, mais selon des modalités sensiblement différentes des SCD. A Science Po, les archives des centres de recherche de l'université feront également l'objet d'un archivage pérenne, projet piloté parallèlement à celui de la bibliothèque par les centres eux-mêmes. Si des coopérations sont possibles dans l'adoption d'une solution commune, les circuits d'archivage sont souvent spécifiques, les contenus variant dans leurs exigences légales de conservation (archives privées ou publiques, données personnelles, DUA).

### ***B. Les circuits d'archivage des documents en SCD***

Déterminer les acteurs en charge des documents selon les périmètres – ESR, circuits gérés par des acteurs publics, université et autres acteurs universitaires en coopération – permet de définir ce qui demeure à la charge du SCD en dernier recours, et qui ne dépend du périmètre d'aucun autre acteur. Nous envisagerons les ressources qui relèvent du périmètre de l'ESR et des acteurs nationaux (BnF, CNRS), auxquels ne prennent pas part les SCD ; puis les circuits gérés par l'ABES (STAR) et le CCSD (HAL), au sein desquels les SCD tiennent un rôle défini, et n'assument pas seuls la responsabilité de l'archivage ; puis le cas des contenus qui peuvent relever d'une décision de conservation du SCD, et donc de sa responsabilité directe : thèses numérisées, thèses d'exercice, mémoires, documents patrimoniaux numérisés, ressources pédagogiques numériques produites par l'université, expositions virtuelles et éditions mises en ligne sur le site du SCD.

#### **Les ressources onéreuses (périodiques et ebooks)**

Les SCD consacrent un budget croissant et considérable à l'achat de ressources numériques : ebooks (livres numériques natifs) et abonnements de périodiques. Les éditeurs des revues en ligne assurent la préservation des publications. Un SCD qui souhaiterait garantir par lui-même leur conservation doit obtenir les fichiers de l'éditeur, lesquels doivent être dépourvus de DRM. Or « les agrégateurs ne fournissent pas systématiquement les PDF des ouvrages aux bibliothèques, mais un système complexe supposé permettre aux clients de conserver l'accès aux e-books achetés même en cas de fermeture de la plateforme. Ils fournissent parfois les données des ouvrages mais omettent volontairement de fournir les métadonnées »<sup>139</sup>. Le rapport « Schéma numérique des bibliothèques » fait donc la recommandation suivante : « Dans le cas d'un achat définitif de

---

<sup>138</sup> A notre connaissance, aucun ne semble envisager de projet d'archivage pérenne, pour des numérisations patrimoniales par exemple.

<sup>139</sup> LE FOLL Anne, *Une offre d'e-books à la Bibliothèque Sainte-Geneviève. Dossier d'aide à la décision*, sous la direction de Benoît Epron et Franck Smith, mémoire d'étude Enssib, 2009, p. 29.

collections ou de titres individuels ou de souscription à des collections, l'accès perpétuel aux ressources doit être assuré, c'est-à-dire que l'éditeur doit prévoir des solutions d'archivage et d'accès pérennes, et rendre possibles la récupération des données par les acheteurs ou les abonnés, le dépôt à un tiers de confiance (ABES, CINES, BnF, PORTICO ...); ou la participation à des projets d'archivage pérenne partagés et répartis comme LOCKSS ou CLOKSS ».<sup>140</sup>

Au sein de l'ESR, l'INIST, en charge de l'acquisition de ressources électroniques, fait de la pérennité des droits d'accès une priorité et consacre une part de son budget à l'achat d'archives pérennes; certaines des ressources sont disponibles en licence nationale. Quant au projet ISTEEX, il a « pour principal objectif d'offrir, à l'ensemble de la communauté de l'enseignement supérieur et de la recherche, un accès en ligne aux collections rétrospectives de la littérature scientifique dans toutes les disciplines (...) Le CINES assure la conservation à long terme et l'INIST garantira l'accès sur la plateforme qu'il développe. Dès lors, la politique nationale d'achat pérenne de documentation électronique peut déboucher sur des décisions locales relatives à la gestion des collections imprimées (désherbage, déménagement, etc.) »<sup>141</sup>. Par ailleurs, un projet associant le portail de diffusion de publications scientifiques Persée<sup>142</sup> et le CINES<sup>143</sup> participe également à la préservation des revues en ligne. Conçu, développé et maintenu par l'UMS 3602, Persée est un service de l'université de Lyon, du CNRS et de l'ENS de Lyon qui bénéficie du soutien du MENESR. Une initiative datant de 2008 vise à archiver les revues de SHS du portail qui sont issues de la numérisation de revues papier, et ainsi à « valoriser et à préserver des collections rétrospectives originales au format papier dont certaines ont près de cent ans ». La chaîne de numérisation mise en place par l'université de Lyon 2 a ainsi été intégrée à la plateforme PAC; les documents numériques créés sont déposés au CINES depuis le printemps 2008 pour une préservation à long terme. Fin 2013, plus de 18 To sont archivés<sup>144</sup>.

Enfin, concernant les ebooks<sup>145</sup>, hors du périmètre de l'ESR, la BnF se fixe dans son contrat de performance l'objectif de « parvenir, à l'horizon 2016, à collecter, mais aussi à signaler, conserver et communiquer l'édition nationale de livres numériques. En parallèle, elle s'attachera à rendre plus facilement accessibles les collections numériques qu'elle a déjà constituées (archives du web,

---

<sup>140</sup> Voir également le rapport Salençon : « La pérennité d'accès aux revues électroniques sur la période d'abonnement doit impérativement être assurée, même en cas de rupture d'abonnement ». [En ligne] (page consultée le 30/12/2015) <http://www.Enssib.fr/bibliotheque-numerique/documents/1825-rapport-du-comite-information-scientifique-et-technique.pdf>.

<sup>141</sup> Punktokmo, Blog technique de l'ABES [En ligne] (page consultée le 30/12/2015) <http://punktokomo.abes.fr/2014/02/25/politique-de-signalement-des-licences-nationales-istex/>; ISTEEX [En ligne] (page consultée le 30/12/2015) <http://www.istex.fr/>

<sup>142</sup> Persée [En ligne] (page consultée le 30/12/2015) [www.persee.fr](http://www.persee.fr). Persée assure la diffusion de collections patrimoniales et « propose des outils d'exploitation similaires à ceux présents sur les portails d'édition courante (indexation au niveau de l'article, plan des documents, exports des références bibliographiques, format détachable, référencement croisé, etc.). Ouvert en 2005, persee.fr diffuse actuellement plus de 170 collections, soit plus de 530 000 documents en texte intégral et en libre accès. Le portail accueille en moyenne 25 millions de visites par an » cf. <http://www.persee.fr/web/guest/a-propos>.

<sup>143</sup> Le CINES est chargé du volet « archivage pérenne » de [licencesnationales.fr](http://licencesnationales.fr).

<sup>144</sup> CINES, « Revues numérisées Persée » [En ligne] (page consultée le 30/12/2015) <https://www.cines.fr/archivage/nos-references-nos-partenariats/les-fonds-conserves-au-cines/star-larchivage-du-programme-persee/>

<sup>145</sup> Les souhaits de la cellule ebook du consortium Couperin (la CeB) dans ses « 10 commandements à l'éditeur d'ebooks : pour une offre idéale », stipule que dans le cas d'un achat définitif de collections ou de titres individuels ou de souscription à des collections, l'accès perpétuel aux ressources doit être assuré, c'est-à-dire que l'éditeur doit prévoir des solutions d'archivage et d'accès pérennes, parmi lesquelles la récupération des données par les acheteurs ou les abonnés et le dépôt à un tiers de confiance (ABES, CINES, BnF, PORTICO, etc.).

ebooks acquis de façon pérenne, archives numériques reçus par dons...), en mettant au point des modes de signalement qui favorisent l'usage par les publics de ce patrimoine documentaire, pour partie inexploré »<sup>146</sup>. A cette collecte des ebooks s'ajoute aussi la collecte de contenus web en accès protégé (sites web de presse quotidienne nationale ou régionale)<sup>147</sup> et l'activité des bibliothèques de dépôt légal imprimeur, qui « doublent leur activité traditionnelle de collecte et de conservation de collections imprimées d'une implication de plus en plus marquée dans des collaborations innovantes avec la BnF : dépôt légal du web, archivage des sites électoraux et Grande collecte – Europeana 1914-1918 »<sup>148</sup>.

Dans le cas de ressources dont les archives sont conservées de façon pérenne par d'autres institutions publiques françaises ou étrangères et accessibles librement et sans coûts<sup>149</sup>, il semble donc que le SCD n'ait pas mission d'une conservation coûteuse et redondante. Dans le cas où la publication ne serait archivée par aucune des initiatives ou institutions mentionnées – Persée par exemple étant limité au domaine des SHS, et la BnF concentrant son action sur les ebooks français-, il est possible d'avoir recours à une solution mutualisée (comme LOCKSS, CLOKSS ou PORTICO), ou de faire appel à l'archivage ciblé du web.

### **Le circuit d'archivage des thèses électroniques**

Depuis 2006, les établissements de l'ESR habilités à délivrer le doctorat peuvent faire le choix du dépôt électronique de leurs thèses comme mode de dépôt officiel<sup>150</sup>. Le dépôt se fait alors par l'application web STAR (Signalement des Thèses électroniques et leur Archivage) de l'ABES. Le circuit des thèses électronique<sup>151</sup> est propre à l'organisation et aux outils de chaque établissement. Il implique automatiquement en fin de cycle l'archivage pérenne au CINES, qui est assuré par l'ABES, premier partenaire du CINES pour l'archivage<sup>152</sup>. Le SCD, qui est chargé du signalement et de la diffusion des thèses soutenues à l'université, assume le plus souvent l'un des « rôles » du processus Star, dit « BIBL ». La mission du SCD est prioritairement axée sur la gestion du document et l'attribution des métadonnées descriptives. L'archivage pérenne des thèses électroniques

---

<sup>146</sup> « Missions et projets de la BnF », « Contrat de performance » [En ligne] (page consultée le 30/12/2015) [http://www.bnf.fr/fr/la\\_bnf/missions\\_bnf/s.contrat\\_performance.html?first\\_Art=non](http://www.bnf.fr/fr/la_bnf/missions_bnf/s.contrat_performance.html?first_Art=non)

<sup>147</sup> BnF, « Collecte de contenus web en accès protégé » [En ligne] (page consultée le 30/12/2015) [http://www.bnf.fr/fr/professionnels/archivage\\_web\\_bnf/a.dlweb\\_collecte\\_acces\\_protege.html](http://www.bnf.fr/fr/professionnels/archivage_web_bnf/a.dlweb_collecte_acces_protege.html)

<sup>148</sup> Ibid. Voir à ce sujet les références bibliographiques de la BnF : BnF, « Bibliographie sélective sur le dépôt légal de l'internet » [En ligne] (page consultée le 30/12/2015) [http://www.bnf.fr/documents/bibliographie\\_dl\\_web.pdf](http://www.bnf.fr/documents/bibliographie_dl_web.pdf)

<sup>149</sup> Voir III sur la mutualisation entre institutions.

<sup>150</sup> Selon l'arrêté du 7 août 2006.

<sup>151</sup> Il s'agit des étapes administratives et scientifiques par lesquelles une thèse doit passer, depuis son dépôt jusqu'à sa diffusion. Ce circuit officiel est régi par 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 du doctorat; il annule le précédent décret de 1986 et inclut les thèses électroniques. Les modalités précises de chacune des étapes (services concernés, dates, format des thèses, etc.) dépendent de chaque université. Une thèse nativement déposée sous format électronique est systématiquement reversée dans le système STAR (Signalement des Thèses électroniques, Archivage et Recherche) développé par l'ABES, qui permet son signalement automatique dans le catalogue SUDOC (Système Universitaire de Documentation). Les thèses papier font quant à elles l'objet d'un signalement classique dans le catalogue de l'université ainsi que dans le catalogue collectif des thèses françaises, Docthèses (lui-même intégré au SUDOC).

<sup>152</sup> Les premiers documents archivés dans la plateforme PAC furent les thèses électroniques délivrées en France et collectées par l'ABES auprès des BU par l'intermédiaire de l'outil STAR. Fin 2013, plus de 22 500 thèses sont archivées dans la plateforme PAC. Voir GILLOUX Marianne, « L'agence bibliographique de l'enseignement supérieur (ABES) : l'archivage des thèses dans PAC », Dossier « Archivage numérique pérenne », *La Gazette du CINES*, février 2013, p. 18, et le site du CINES [En ligne] (page consultée le 30/12/2015) <https://www.cines.fr/archivage/nos-references-nos-partenariats/les-fonds-conserves-au-cines/archivage-des-theses/>

n'implique donc pas de réflexion stratégique à proprement parler pour les SCD, puisqu'il est légalement nécessaire et qu'il relève pour partie de ses missions, comme défini dans le circuit établi par l'ABES.

#### **Les publications publiées dans les archives ouvertes nationales HAL ou dans des archives ouvertes institutionnelles**

Il existe trois types d'archives ouvertes<sup>153</sup> :

- Les archives disciplinaires : ArXiv en physique, RePEc en économie, PubMedCentral en sciences du vivant, HAL- SHS ;
- Les archives institutionnelles : OATAO (Toulouse), Okina (Angers), etc.
- Les archives centrales, en l'occurrence la plateforme HAL (voir I). Cette solution est spécifique au contexte français.

La décision par les établissements de créer une archive ouverte institutionnelle qui redouble ou remplace l'archive HAL peut être motivée par plusieurs intérêts : accroître la visibilité des productions de l'établissement, avoir une vision exhaustive de la production scientifique, permettre une évaluation précise de la production des chercheurs via des listes de publication, faciliter le dépôt, ou avoir des indicateurs pour mieux orienter la politique scientifique. Le choix d'ajouter au dépôt dans HAL la création d'une AOI peut être dicté par l'impossibilité de déposer dans l'archive nationale HAL un fichier en accès restreint permanent ou sous embargo de plus de 2 ans. Par conséquent certains laboratoires, confrontés aux réticences des chercheurs à publier en *open access*, ont tendance à ne verser dans HAL que les métadonnées, et non les textes eux-mêmes. Par ailleurs, la connexion à HAL se fait souvent pour une partie des publications (un connecteur est employé pour « pousser » dans HAL les documents) ; mais les recensions ou les données par exemple, qui ne sont pas éligibles, n'y figureront pas. Cette situation explique le recours à des AOI qui permettent de rendre accessibles davantage de documents.

Le choix stratégique du type d'archive dans laquelle seront conservées les documents conditionnent fortement les solutions d'archivage pérenne ultérieures. Il en est ainsi dans le cas des archives ouvertes institutionnelles, qui remplacent ou doublonnent le dépôt dans HAL dans certaines universités : si l'archive est directement créée dans HAL (portail thématique), ou reliée à HAL, l'archivage pérenne au CINES est assuré automatiquement ; si ce n'est pas le cas, l'établissement doit assurer par lui-même l'archivage de ses données, au CINES ou ailleurs. Mais même dans l'hypothèse où l'AOI est connectée à HAL, les documents pour lesquels les chercheurs refusent le versement dans HAL ne bénéficient plus alors d'un archivage qu'au sein de l'AOI, et se trouvent faire l'objet uniquement d'un stockage sécurisé sur les serveurs de la DSI (voir le cas d'OATAO par exemple pour cette configuration). Par conséquent, les établissements sont nombreux à cumuler un dépôt dans HAL et un projet d'archivage pérenne pour les publications hors HAL – qu'elles soient publiées au sein d'une AOI ou non ; c'est ainsi le cas par exemple de l'université de

---

<sup>153</sup> Voir à ce sujet SCHÖPFEL Joachim, PROST Hélène, *Développement et Usage des Archives Ouvertes en France*, Université Charles-de-Gaulle Lille 3, Laboratoire GERIICO, 2010 [En ligne] (page consultée le 30/12/2015) [http://archivesic.ccsd.cnrs.fr/sic\\_00497389/document](http://archivesic.ccsd.cnrs.fr/sic_00497389/document) et SCHÖPFEL Joachim, STOCK Christiane, « Les archives ouvertes en France », *Distances et savoirs* 3/2009 (Vol. 7), p. 443-456.

Strasbourg et de l'archive AOC (voir II), pour laquelle un archivage pérenne global est prévu.

De façon générale, la publication en AO favorise souvent la préservation numérique : le stockage en local (sur microfiches ou CD Rom) de documents non diffusés (HDR, etc.) est un facteur de risque quant à la préservation à long terme. Les AO favorisent ainsi à la fois la diffusion et la conservation. La promotion de l'Open Access, courante dans le monde des bibliothèques, peut donc entraîner indirectement celle de l'archivage pérenne.

### **Les données de la recherche : une problématique émergente qui appelle une mutualisation**

La question des données de la recherche est celle qui interroge le plus la place et le rôle des SCD en regard des productions scientifiques de l'université. Les données de la recherche constituent des gisements de connaissance de formes variées (il peut s'agir d'observations, de simulations, de mesures, etc.<sup>154</sup>) dont la masse tend à croître de façon exponentielle. Considérées comme le « matériau brut » de la recherche, et donc souvent non destinées à un archivage en bonne et due forme, elles sont régulièrement perdues car mal conservées<sup>155</sup>. D'où l'émergence de stratégies de gestion consistant à proposer des plans de gestion de ces données (« data management ») et permettre leur accès et leur réutilisation les plus larges possible<sup>156</sup>. Selon le bilan que dresse Rémi Gaillard dans son mémoire consacré aux données de la recherche<sup>157</sup>, « de nombreuses bibliothèques étrangères [en particulier américaines et anglo-saxonnes] se saisissent de la problématique mais la France accuse un retard considérable, exception faite d'initiatives pionnières auxquelles participent des professionnels de l'IST, comme à l'INRA ou au Centre de données astronomiques de Strasbourg. L'expérience prouve pourtant que dans la majorité des cas étudiés, au Royaume-Uni par exemple, les bibliothèques sont en première ligne ». Or les politiques institutionnelles imposent généralement aux projets de recherche des périodes minimales de conservation de leurs données ; grâce aux plans de gestion des données, qui tendent à se développer en autres sous l'impulsion de projets européens, il est possible de déterminer les données ayant de la valeur à long-terme et devant donc bénéficier d'une conservation pérenne dans un entrepôt. Afin de garantir cette pérennité, diverses universités, majoritairement britanniques et américaines, mettent en place leur

---

<sup>154</sup> Voir sur ce point JACOBSON Michel, LARROUSSE Nicolas, MASSOL Marion, « La question de l'archivage des données de la recherche en SHS (Sciences Humaines et Sociales) », *op. cit.* : « Les SHS sont très hétérogènes en France avec des diversités de pratiques importantes. De même, si on les compare aux sciences dites « dures », les objets manipulés peuvent être de nature très variée », p. 4.

<sup>155</sup> Cf. « The STM Report An overview of scientific and scholarly journal publishing, Fourth Edition, March 2015 », mis en ligne en avril 2015 [En ligne] (page consultée le 30/12/2015) [www.stm-assoc.org/2015\\_02\\_20\\_STM\\_Report\\_2015.pdf](http://www.stm-assoc.org/2015_02_20_STM_Report_2015.pdf) ; et GUINDON Alex, « La gestion des données de recherche en bibliothèque universitaire », *Documentation et bibliothèques*, vol. 59, n° 4, 2013, p. 190 : « 28 % des répondants [de l'enquête] (archivistes, bibliothécaires, responsables des technologies de l'information (TI), mais aussi producteurs de données) indiquent avoir perdu des données (29 % des répondants n'étaient pas en mesure de répondre à la question) alors que 48 % craignent la disparition de données ».

<sup>156</sup> Sans compter les enjeux juridiques : « [Il existe] une forte demande sociale quant à l'administration de la preuve scientifique et à la réutilisation et l'archivage des données Huma-Num (...) Est-on capable de mettre en place une chaîne de responsabilités garantissant l'archivage et la transmission des données numériques sur plusieurs dizaines d'années ? », POUYLLAU Stéphane (TGIR Huma-Num), *Les Digital Humanities*, Open Access Week, Paris (UPMC) [En ligne] (page consultée le 30/12/2015) [https://scd-collections.u-strasbg.fr/wiki/aoc/documentation\\_generale](https://scd-collections.u-strasbg.fr/wiki/aoc/documentation_generale)

<sup>157</sup> GAILLARD Rémi, *De l'Open data à l'Open research data : quelle(s) politique(s) pour les données de recherche ?*, Diplôme de conservateur des bibliothèques (DCB), sous la direction de Monique Joly, Enssib, 2014.

entrepôt de données<sup>158</sup>, dont certains ont obtenu le DSA, Data seal of approval, du DANS.

Quelle peut être la place des SCD dans la gestion et la conservation des données de la recherche produites par les établissements - qu'ils assument déjà mais très rarement encore avec la formation à la gestion des données, la création d'entrepôts institutionnels ou encore l'identification pérenne des données ? Certes ces missions sont en continuité naturelle avec les compétences des bibliothécaires, en ce qui concerne en particulier les métadonnées, et en droite ligne de leur implication historique dans le mouvement de l'open access. Comme le note par ailleurs Charlotte Maday, responsable du Bureau des archives de l'université Paris Diderot et présidente de la section AURORE, « les SCD sont bien plus armés que les archivistes sur les questions de conservation de l'électronique », notamment parce qu'ils « disposent de bien plus de moyens »<sup>159</sup>. Les SCD auraient donc théoriquement toute légitimité à s'emparer de la question des données de la recherche, qui en tant que production issue d'une activité de recherche peuvent être considérées comme de la littérature grise ou des archives ; ainsi selon Charlotte Maday, « il semble important d'inclure, dans les plans de gestion des données, la dimension archivistique, ou du moins celle du records management, des données de la recherche »<sup>160</sup>. Par ailleurs, le Code de la recherche, modifié en juillet 2013, précise que « l'organisation de l'accès libre aux données scientifiques » fait partie des objectifs de la recherche publique française et donc des missions des établissements de recherche français<sup>161</sup>. Mais le postulat, surtout anglo-saxon, qui tend à voir dans les BU les responsables de la conservation numérique (« digital curation ») dans ses diverses étapes (sélection, préservation, archivage)<sup>162</sup> et à leur déléguer une activité qui est également celle des producteurs des données et des laboratoires, UMR et UMS, est-il fondé et surtout pertinent, étant donné que la question des données mobilise des acteurs à divers niveaux : infrastructures et plateformes des laboratoires, établissements, réseaux disciplinaires, etc. ?

Les SCD sont en tout cas assez nombreux à envisager l'archivage pérenne des données de la recherche. Ainsi à Lyon 1, le SCD a participé au pilotage d'un groupe projet Enssib mandaté par le SCD de Lyon 2 sur les données de la recherche, et veille sur le sujet depuis un à deux ans. Les SCD se heurtent cependant à plusieurs difficultés, qui expliquent que peu de solutions soient

<sup>158</sup> Voir sur ce point R. Gaillard, *op. cit.*, p. 57 et suiv.

<sup>159</sup> Citée par R. Gaillard, *op. cit.* p. 70.

<sup>160</sup> Records management is the new black, « Les données de la recherche sont des archives comme les autres », 24 septembre 2013 [En ligne] (page consultée le 30/12/2015) <https://lotteauxfraises.wordpress.com/2013/09/24/les-donnees-de-la-recherche-sont-des-archives-comme-les-autres-2/>

<sup>161</sup> Code de la recherche, « chap. II Objectifs et moyens institutionnels de la recherche publique », article L112-1 modifié par la loi n° 2013-660 du 22 juillet 2013, article 16. [En ligne] (page consultée le 30/12/2015) [http://www.legifrance.gouv.fr/affichCode.do?jsessionid=1B91C4435EFC974490B5CC8CE2295188.tpdjo14v\\_1?idSectionTA=LEGISCTA000006151273&cidTexte=LEGITEXT000006071190&dateTexte=20131113](http://www.legifrance.gouv.fr/affichCode.do?jsessionid=1B91C4435EFC974490B5CC8CE2295188.tpdjo14v_1?idSectionTA=LEGISCTA000006151273&cidTexte=LEGITEXT000006071190&dateTexte=20131113)

<sup>162</sup> Voir sur ce point LORD Philip, MACDONALD Alison, « Data curation for e-Science in the UK: an audit to establish requirements for future curation and provision », The JISC Committee for the Support of Research (JCSR), 2003 : « There is an increasing trend for universities to bring IT and libraries under one umbrella. While this would perhaps tend universities to see digital curation as the responsibility of the merged, information services entity, responses did not indicate that the situation is or would be clear cut. We heard conflicting opinions regarding the role of libraries (and the university libraries in particular) in curating primary research data, both views stated with emphasis. One view is that they are the natural heirs for data, and that their role should adapt to acquire necessary technical skills to enable them to become custodians of data; they were already skilled practitioners at organising information and disseminating it. The opposing view was that they are stuck in a mindset which does not make them able to adopt this role in the near future and they currently lack some of the requisite skills, such as IT, and that they may also lack the necessary imagination to change » [En ligne] (page consultée le 30/12/2015) [http://www.nesc.ac.uk/news/articles\\_and\\_calls/290503.html](http://www.nesc.ac.uk/news/articles_and_calls/290503.html)

opérationnelles à ce jour. Tout d'abord, si de plus en plus d'établissements initient leur réflexion sur la question, tel Lille 1, il n'existe pas encore de politique générale d'archivage des données de la recherche par les universités françaises, encore peu au fait des enjeux pour la plupart, ce qui laisse les SCD démunis sur le plan stratégique. Par ailleurs, d'autres interlocuteurs dans les universités sont aussi en charge des données de la recherche ; les auteurs des recherches peuvent ainsi bénéficier au sein de leur laboratoire des services de leurs propres informaticiens. De ce fait la liaison avec le SCD n'est pas toujours perçue comme une évidence. A ce titre, il est possible d'effectuer un travail d'information<sup>163</sup>, effectué par certains SCD auprès des laboratoires ; le CNRS (via l'INIST) informe pour sa part les producteurs de données sur les nécessités d'un archivage et sur les solutions possibles<sup>164</sup>. Mais il semble difficile d'aboutir à une politique globale et systématique qui puisse légitimer l'action des SCD sans en passer par une décision prise au niveau de l'université – voire de l'ESR<sup>165</sup>. De plus, la masse des données peut être très importante selon les domaines ; les données produites sont donc conservées sur des plateformes sécurisées que les chercheurs considèrent comme suffisant à leurs besoins ; la procédure d'archivage pérenne, lourde, leur apparaît complexe voire inadaptée. Ajoutons qu'il existe peu de services d'archives en université consacrés aux données de la recherche (quand un service d'archives existe en université, il est généralement administratif). Le cas de Paris 7, où un plan de gestion des données mutualisé entre Paris 5 et 7 a été mis en place, reste une exception. Enfin, si certains jeux de données sont déjà structurés et publiables, d'autres nécessitent une structuration, aussi bien pour la diffusion que pour l'archivage. Certains fichiers posent ainsi la question des formats de données ; à Lille 2 par exemple, Anne Flahaut-Steiner remarque que les fichiers multimédia ne seront pas archivés au CINES pour des questions de format, trop compliqués à traiter en interne<sup>166</sup>. Or les données produites par les chercheurs sont très variées, surtout en SHS : il apparaît complexe pour un SCD d'en acquérir une maîtrise exhaustive<sup>167</sup>, et d'être à même d'assumer leur gestion.

Outre ces difficultés liées à la gestion des données, quelles solutions envisager au niveau de l'université ? Le stockage intermédiaire de données est une réponse de premier niveau mais n'est pas un archivage pérenne au sens strict. La construction de Data Centers, actuellement en projet dans de nombreuses universités, pose le problème du coût, initial et de fonctionnement ; le MENESR essaie d'en limiter le développement<sup>168</sup> et favorise la mutualisation, l'IN2P3

---

<sup>163</sup> Voir à ce sujet GÉROUDET Marie-Madeleine, « Sur le terrain des données de la recherche à Lille 1 », *Arabesques* n° 73, janvier-février-mars 2014.

<sup>164</sup> Voir par exemple le tutoriel en ligne « Archivage des données » :CNRS, « DMP-Archivage de données » [En ligne] (page consultée le 30/12/2015) [http://www.inist.fr/formations/DMP-Archivage-des-donnees/story\\_html5.html](http://www.inist.fr/formations/DMP-Archivage-des-donnees/story_html5.html)

<sup>165</sup> Sur ce point, voir la partie sur la mutualisation et la redéfinition des périmètres institutionnels : III A.1

<sup>166</sup> Les annexes « multimédia » des mémoires d'orthophonie (objets en 3D (jeux, cartes...), ou vidéo) ne sont pas traités avec la version électronique du mémoire au CADIST en orthophonie de Nancy, mais sont signalés dans la notice et conservés en document physique (éventuellement conditionnés dans des boîtes en magasin). C'est également ce qui se fait à Lille 2 (source : Solenn Bihan, chargée de l'information scientifique : bibliométrie et open access, SCD & Service de la Recherche (SeRVIS)). Ce constat est largement partagé, voir par exemple *e-Science Curation Report, op. cit.*, p. 25 : « The UK Data Archive at The University of Essex expressed the view that no-one has any real experience yet with preserving multi-media formats. Difficult formats are not addressed, nor is dynamic information ».

<sup>167</sup> Voir sur ce point LEONELLI Sabina, SMIRNOFF Nicholas, MOORE Jonathan, COOK Charis et BASTOW Ruth, « Making open data work for plant scientists », *Journal of Experimental Botany*, septembre 2013, p. 4114 [En ligne] (page consultée le 30/12/2015) <http://jxb.oxfordjournals.org/content/early/2013/09/15/jxb.ert273.full>

<sup>168</sup> Face au nombre important de projets de Data Centers présentés par les universités dans le cadre des contrats de plan État-Régions (un rapport publié par l'IGF et l'IGAENR à l'été 2014 recense une trentaine de projets, parfois concurrents), un comité de pilotage a été mis en place à la DGRI pour proposer à « l'ensemble des décideurs des établissements publics un schéma rationnel d'infrastructure garantissant une fiabilité et rapidité des données », signale

### III. Situation des bibliothèques universitaires : vers une nécessaire mutualisation

assurant déjà la double mission de calcul intensif et d'archivage pérenne. La solution la plus viable est probablement nationale. Les cadres du CINES et de la TGIR, qui proposent des solutions mutualisées, semblent plus adaptés, d'autant plus que divers services sont proposés, pour un archivage intermédiaire ou définitif selon les données. On notera que les projets de recherche soutenus par la Commission Européenne, dans le cadre du projet H2020 par exemple, ne désignent pas de tiers-archivageur précis lorsqu'ils formulent l'obligation d'un archivage pérenne. Charge donc aux producteurs des données de définir leurs besoins et de choisir un tiers de confiance.

Le CINES, pour qui la problématique est devenue centrale ces deux dernières années, a mis en place dans le cadre du projet ISAAC<sup>169</sup> (« Information scientifique archivée au CINES ») un service d'archivage intermédiaire des données scientifiques à l'attention de la communauté des chercheurs. ISAAC comporte un important volet « conservation », avec la préservation à moyen terme (3 à 5 ans) des données versées et la possibilité de les constituer en archive définitive dans PAC. Sont par ailleurs archivées de façon pérenne certaines données définies comme patrimoniales : celles de la photothèque de l'Inserm par exemple. Le projet Eudat propose quant à lui une solution de préservation pérenne aux plateformes de dépôt de données dont les capacités seraient insuffisantes, ou dont le stockage ne serait pas garanti, ou encore non pérenne. Toutes les institutions peuvent y déposer leurs données, à condition d'avoir l'infrastructure et les métadonnées adaptées<sup>170</sup>. Les problèmes posés par les données de la recherche sont bien sûr les volumes, mais aussi les formats, l'accès, et la gestion des documents associés aux données. Les équipes du CINES travaillent à prendre en compte les nouveaux formats, et s'interrogent actuellement sur l'outil à favoriser : faut-il conserver les données sur la plateforme PAC, qui héberge l'archivage pérenne des autres types d'archives, ou faut-il envisager un autre outil plus adapté ? Car les données de la recherche ne nécessitent pas toutes un archivage sur le long terme, la priorité étant davantage d'ouvrir les données aux chercheurs. Le service Nakala<sup>171</sup>, de la TGIR Huma-Num, qui permet l'exposition des données, peut s'avérer dans de nombreux cas plus adapté. Quant à la TGIR Huma-Num, elle dépose d'ores et déjà au CINES en archivage pérenne les données 3 D des laboratoires en archéologie et a mené une étude sur les données exprimé dans le format TEI, afin que les corpus demeurent intelligibles. La réflexion des institutions spécialistes est ainsi encore en cours, tant sur les formats des données de la recherche que sur les modalités de leur archivage.

---

Alain Abécassis au 45ème congrès de l'ADBU, le 1er octobre 2015 (ADBU, « Retour sur la matinée politique du congrès 2015 de l'ADBU: Vers un domaine public des données ? » [En ligne] (page consultée le 30/12/2015) <http://adbu.fr/retour-sur-la-matinee-politique-du-congres-2015-de-ladbu-vers-un-domaine-public-des-donnees/>

<sup>169</sup> Voir THERON Jean-Pierre, « L'archivage des données scientifiques : ISAAC, une solution du CINES », Dossier « Archivage numérique pérenne », *La Gazette du CINES*, février 2013, p. 46.

<sup>170</sup> « Any community and departmental data repositories that have a proper repository infrastructure supporting PIDs and metadata describing the properties and context of the data being replicated can participate in the B2SAFE service. The replication service currently relies on iRODS but also supports other federation technologies. Participating repositories can choose to use iRODS and link it with their local repository to enable a tight integration to the EUDAT infrastructure, or use other easy-to-integrate federation approaches and client libraries », EUDAT [En ligne] (page consultée le 30/12/2015) <http://eudat.eu/>

<sup>171</sup> Huma-Num, « Exposer-Nakala » [En ligne] (page consultée le 30/12/2015) <http://www.huma-num.fr/service/nakala>

Dans ce contexte encore mouvant et technique, les SCD peuvent s'inscrire dans la « chaîne » de production de la donnée, mais peu probablement la soutenir dans sa totalité.

Le cycle de vie des données, qui se rapproche davantage du travail des archivistes que de celui des bibliothécaires, commence en effet dès le début du processus de recherche. Plusieurs étapes peuvent être définies dans ce cycle : découverte et planification, collecte initiale de données, préparation des données et analyse, publication et partage, gestion à long terme comme archive de données. Au vu des divers acteurs – chercheurs et autres producteurs, instituts de recherche et laboratoires, opérateurs de l'État, tiers-archivistes... - impliqués dans ces étapes du « cycle », il semble peu probable qu'un SCD puisse assumer seul une tâche aussi lourde. Comme le note Alex Guindon<sup>172</sup>, « il s'agit plutôt d'une entreprise qui doit être menée par un ensemble d'acteurs incluant chercheurs, assistants de recherche, bibliothécaires, personnel des dépôts de données et personnel du soutien technique »<sup>173</sup>. La coopération renforcée avec les producteurs des données, aptes à en déterminer la valeur et le besoin de conservation qui en découle, s'avère nécessaire, et exige la création d'une culture de collaboration entre bibliothécaires et équipes de recherche, entre autres au sein des UMS, qui ont leurs bibliothèques spécialisées. Plus encore que pour d'autres types de productions, l'archivage – qui peut ou pas être pérenne- des données de la recherche implique de connaître les besoins des chercheurs, et de favoriser la solution la plus adaptée à leur pratique. D'où le choix de certains SCD comme celui de Strasbourg qui envisagent de recenser les entrepôts à recommander selon la nécessité. Plus largement, le défi que cet archivage représente semble excéder les capacités de la plupart des institutions individuelles. Coordonner les acteurs nationaux (ABES pour le signalement, INIST pour l'accès, et CINES pour l'archivage<sup>174</sup>) comme c'est par exemple le cas au sein d'ISTEX<sup>175</sup>, qui vise à la réalisation d'un socle documentaire numérique pérenne commun à l'ensemble des Initiatives d'excellence, permet de proposer des solutions à toutes les SCD, y compris ceux, de petite taille, qui n'ont pas les moyens d'envisager de solutions locales.

### **Le cas des thèses numérisées, des thèses d'exercice, des HDR et des mémoires**

Plusieurs universités ont pris la décision de ne pas procéder à la numérisation rétrospective des thèses produites avant le lancement du service STAR, au vu des coûts financiers et humains (Strasbourg par exemple). Ainsi à Toulouse, quelques-unes de ces thèses papier peuvent parfois être scannées à la demande, et déposées

---

<sup>172</sup> *Op. cit.*, p. 191.

<sup>173</sup> Voir également la conclusion de Philip Lord et Alison Macdonald, *op. cit.* : « In practice the organisational focus of curation may be a local implementation depending on local circumstances and skills. We would argue that the role of researchers and discipline specialists remains at the heart of curation, but many different professional skills including those of librarians and IT providers will also need to be brought to bear on the curation process », p. 12.

<sup>174</sup> Voir BÉRARD Raymond, « ISTEEX. Vers des services innovants d'accès à la connaissance », *Arabesques*, revue de l'ABES, n° 66 (avril-mai-juin 2012), p. 6-8 : « La répartition au sein d'ISTEX des rôles des partenaires du projet, travaillant ensemble de longue date, est assise sur leurs domaines respectifs de compétences : le consortium Couperin (...) pour le recueil des besoins des communautés de recherche, (...) l'ABES pour la négociation et l'acquisition des ressources, leur signalement, la gestion des accès et des droits (en partenariat avec le CNRS-INIST et en relation avec les opérateurs locaux) ; le CNRS, en s'appuyant sur l'INIST, pour la conception et l'hébergement de la plate-forme d'agrégation du texte intégral (...); l'université de Lorraine pour la recherche et les services. L'archivage pérenne des données sera assuré par le Centre informatique national de l'enseignement supérieur (CINES) ».

<sup>175</sup> ISTEEX [En ligne] (page consultée le 30/12/2015) <http://www.istex.fr/les-partenaires/>

### III. Situation des bibliothèques universitaires : vers une nécessaire mutualisation

en PDF sur l'AOI, OATAO, ce qui reste très ponctuel : pour des questions de moyens, le rétrospectif n'est pas traité. Ces thèses échappent donc aux champs d'action respectifs de l'ABES et du CINES<sup>176</sup>.

Les thèses d'exercice, pour lesquelles il existe une obligation de conservation, et les mémoires de Master, pour lesquels il n'existe pas de dépôt légal, ne sont pas signalées par STAR. De même que dans le cas des thèses produites antérieurement au dépôt numérique, les SCD doivent donc décider à la fois du mode de signalement et de conservation de ces productions universitaires. Certains établissements signalent ces thèses d'exercice et mémoires sur leurs plateformes (voir par exemple les universités de Lorraine, Lille 1, ou encore Lille 2). L'archivage pérenne des ressources est alors conditionné à l'archivage pérenne de la plateforme dans son ensemble. La préservation de ces publications de l'université peut d'ailleurs être à l'origine d'un projet d'archivage pérenne. C'est par exemple le cas de Lille 2, qui procède à la mise en ligne des thèses d'exercice et des mémoires avec une gestion fine des droits d'accès sur Pépité, et qui procède à leur archivage pérenne au CINES : en passant à la version électronique comme version de référence (en 2014 pour les thèses d'exercice en santé) la version papier n'est plus conservée ; or seules les thèses de doctorat sont archivées par le CINES. Notons que certains services versants qui archivent des thèses d'exercices et des mémoires au CINES dans une faible volumétrie peuvent bénéficier de la gratuité car les documents concernés constituent un complément à la mission nationale d'archivage des thèses de doctorat du CINES. Cette décision relève davantage de négociations entre établissements que d'une règle généralisable. L'une des solutions qui s'offrent aux SCD pour la diffusion en open access et l'archivage pérenne de ces documents est la plateforme DUMAS, du CCSD, archivée au CINES, dédiée aux mémoires de Master, aux thèses d'exercice, de pharmacie et d'orthophonie entre autres.

#### **Les bibliothèques numériques : produits de numérisation et documents nativement numériques (hors abonnements onéreux)**

La question de la préservation à long terme des données numériques reste encore souvent mal connue des porteurs de projets de numérisation. Or l'archivage des données nécessite d'être pris en compte dès le début d'un projet de numérisation. Certains logiciels proposent un système de gestion d'archives pour les bibliothèques numériques<sup>177</sup> - Invenio du CERN, la suite ORI-OAI du consortium ORI-OAI, Content d'OCLC, Eprints de l'université de Southampton ; toutes les exigences de l'archivage pérenne ne s'y rencontrent pas cependant. Les plates-formes nationales de dépôt proposent quant à elle un archivage pérenne associé au dépôt des numérisations : HAL (via le CINES) ; les plateformes du CNRS (CCSD/CINES) ; Gallica Marque Blanche (BnF).

Internet Archive<sup>178</sup>, E-corpus<sup>179</sup> et Hathitrust<sup>180</sup> offrent la possibilité à toute bibliothèque, y compris universitaire, de diffuser du contenu en ligne. Ces

---

<sup>176</sup> De même pour les HDR qui sont « généralement stockées en bibliothèque, microfichées. Leur diffusion reste très faible », in *Pérenniser le document numérique : séminaire INRIA*, 2-6 octobre 2006, ADBS, 2006, p. 138.

<sup>177</sup> Pour un recensement des bibliothèques numériques en bibliothèques universitaires (sous réserve de mise à jour du site) : page « Bibliothèques numériques » de Bibliopédia [En ligne] (page consultée le 30/12/2015) [http://www.bibliopedia.fr/wiki/Biblioth%C3%A8ques\\_num%C3%A9riques#Universit.C3.A9s](http://www.bibliopedia.fr/wiki/Biblioth%C3%A8ques_num%C3%A9riques#Universit.C3.A9s)

<sup>178</sup> Internet Archive [En ligne] (page consultée le 30/12/2015) <https://archive.org/index.php>

<sup>179</sup> E-corpus [En ligne] (page consultée le 30/12/2015) <http://www.e-corpus.org/>

plateformes de dépôts mutualisées<sup>181</sup> offrent, entre autres avantages, des opportunités certaines de rendre les contenus visibles ; pour autant, aucune ne fournit un véritable service d'archivage pérenne. E-corpus et Internet Archive mettent à disposition uniquement un stockage sécurisé ; ainsi la BSG, qui assure la diffusion de sa bibliothèque numérique à la fois sur E-corpus et sur Internet Archive, assure par ailleurs l'archivage pérenne de ses documents au CINES. Hathitrust, qui annonce plus de 6 millions de livres numérisés (dont 1 million en libre accès), se conforme aux standards PREMIS et OAIS, et se prévaut de la certification du CRL (Center for Research Libraries)<sup>182</sup>. Pour autant, l'offre de service ne semble pas pouvoir être comparée à celles du CINES ou de la BnF (en terme de moyens humains, de suivi des projets, de veille technologique, de migration de formats, etc.).

### **Ressources numériques produites par l'université ou en coopération avec le SCD**

**Ressources pédagogiques numériques produites par l'université.** Exercices et corrigés, ressources interactives, cours en ligne... Très peu d'université consacrent un archivage pérenne à ces ressources, qui sont parfois signalées par les SCD, mais rarement hébergées par eux. La question des ressources pédagogiques se pose au sein de BSN 6, et interroge la répartition des missions entre le MENESR et le MCC, puisqu'il pourrait être question d'un archivage ciblé du web assuré par les services de la BnF.

Dans les cas où ce type de ressource est moissonné sans distinction par le SCD parmi les productions pédagogiques de l'université pour figurer dans une AOI ou sur une plateforme de diffusion, l'absence de critères de qualité définis est un obstacle à un archivage pérenne, qui présuppose une sélection. Le problème des documents non libres de droits et la complexité de certains formats constituent également des obstacles. Il semble que pour la plupart des SCD concernés, l'archivage pérenne des ressources pédagogiques ne nécessite pas de mettre en place un workflow, la valeur des documents n'étant pas perçue comme suffisante pour justifier un archivage pérenne.

**Les expositions virtuelles et éditions mises en ligne sur le site d'une bibliothèque numérique.** Plusieurs SCD produisent des expositions virtuelles, voire, en partenariat avec des enseignants-chercheurs, des éditions en ligne, par exemple la BIUS. Les expositions virtuelles ne sont pas archivées de façon pérenne par les SCD : les images des collections étant sauvegardées par ailleurs, archiver l'exposition virtuelle n'est pas ressenti comme une urgence.

La BIUS l'a pourtant envisagé : les expositions virtuelles ont un coût, en temps et en expertise, de la part des équipes de la bibliothèque ; par ailleurs, le

---

<sup>180</sup> Hathitrust [En ligne] (page consultée le 30/12/2015) <https://www.hathitrust.org/community>

<sup>181</sup> Pour une présentation, voir ANDRO Mathieu, ASSELIN Emmanuelle, MAISONNEUVE Marc, *Bibliothèques numériques : logiciels et plateformes*, op. cit., p. 293 et suiv.

<sup>182</sup> « Hathitrust has been certified as a trustworthy digital repository by the Center for Research Libraries (CRL) through their rigorous Trustworthy Repositories Audit and Certification (TRAC) assessment program », cf. Hathitrust, « Hathitrust Certified as Trustworthy Repository » [En ligne] (page consultée le 30/12/2015) <https://www.hathitrust.org/hathitrust-certified-as-trustworthy-repository>. Les membres du CRL sont exclusivement américains, canadiens, anglais ou québécois.

SCD produit des revues en ligne, ainsi que des éditions critiques, auxquelles les chercheurs ont parfois consacré des années de travail<sup>183</sup>. L'archivage pérenne serait ici pleinement justifié, pour des raisons scientifiques évidentes. L'archivage du web par la BnF a été envisagé, mais pas encore demandé ; il constituerait une solution mutualisée rapide, efficace et sans coût pour sauvegarder ce type de contenus.

**La production documentaire numérique.** En tant qu'administration enfin, une BU produit des documents dans le cadre de ses activités. Une partie de ces documents peut avoir une utilité administrative ou juridique. « La préservation numérique d'une partie d'entre eux peut se poser : par exemple, la conservation du fichier de récolement conçu lors d'une opération de numérisation est précieuse pour garantir la bonne gestion et la bonne conservation des ressources issues de la numérisation »<sup>184</sup>. La responsabilité de ces archives, qu'elles soient ou non considérées comme archives publiques, peut relever ou pas des missions de la BU seule, ou du service des archives de l'université. La mutualisation de l'archivage au sein de l'université constitue une autre solution à la gestion de ces documents.

On constate ainsi qu'en dehors des circuits STAR et HAL, des solutions au sein desquelles l'archivage pérenne est intégré (DUMAS, ISAAC) et des documents pour lesquels il est possible de faire reposer l'archivage pérenne sur d'autres acteurs, la mutualisation – au sein de l'université, ou plus largement – apparaît comme l'horizon naturel des projets d'archivage pérenne.

## 2. De la mutualisation à la réflexion sur les périmètres institutionnels

### A. La mutualisation, l'horizon des projets universitaires d'archivage pérenne ?

#### Stratégies et recommandations

Dans le cadre budgétaire contraint qui est celui des universités, et au vu de la situation actuelle des finances publiques, fort peu de moyens supplémentaires sont à attendre des universités mêmes ou de l'État, comme en témoigne la Communication à la commission des finances du Sénat de juin 2015 sur l'autonomie financière des universités<sup>185</sup>. La mutualisation apparaît ainsi inévitable et indispensable, comme l'observait Pierre Carbone, alors président de la Commission Bibliothèques Numériques :

---

<sup>183</sup> Voir par exemple l'édition en ligne « La Fabrique de Vésale et autres textes », BIU Santé [En ligne] (page consultée le 30/12/2015) <http://www3.biusante.parisdescartes.fr/vesale/debut.htm>, et l'« Edition critique de la correspondance de Guy Patin », BIU Santé [En ligne] (page consultée le 30/12/2015) <http://www2.biusante.parisdescartes.fr/wordpress/index.php/correspondance-patin-par-loic-capron/>

<sup>184</sup> CLAERR Thierry, MOUFFLET Jean-François, « Gestion de la conservation des collections numériques », *op. cit.*, p. 61.

<sup>185</sup> « L'autonomie financière des universités : une réforme à poursuivre », Communication à la commission des finances du Sénat, Juin 2015 : « Compte tenu du poids de la masse salariale, de la faiblesse des ressources propres, et de l'absence de maîtrise de déterminants majeurs de leur activité, les marges de manœuvre financières des universités apparaissent en réalité limitées » ; « Dans la situation actuelle des finances publiques, l'augmentation des moyens financiers des universités ne peut être attendue des dotations de l'État. La loi du 29 décembre 2014 de programmation des finances publiques pour les années 2014 à 2019 ne prévoit pas d'augmentation pour la mission MIREs, qui comprend le programme 150 de financement des universités, alors que, d'après les analyses démographiques du ministère, une croissance de la population étudiante de 3,9 % est attendue entre 2014 et 2017 ».

« Étant donné les moyens humains et matériels mobilisés par les opérateurs les plus importants (des équipes d'une dizaine d'ETP, des mémoires mesurées en téraoctets ou pétaoctets) ainsi que les coûts d'investissement et de fonctionnement, la mutualisation de l'archivage, comme sa gestion sur la longue durée, est indispensable. Pour la Commission, il serait justifié de créer un centre de compétences mutualisé pour la préservation de l'information numérique, qui serait compétent pour la veille sur les formats d'archivage, les supports, les formats de métadonnées et pour la formation. La mutualisation constitue dans ce cadre une opportunité majeure, offrant la possibilité de mener des projets autrement limités par les moyens des services ». <sup>186</sup>

C'est d'ailleurs vers la mutualisation des moyens que s'oriente la conservation des documents au sein de l'ESR. Elle est au cœur de l'activité des CollEx (anciens CADIST), qui assure la réflexion sur la conservation des ressources papier, et du groupe de travail piloté par le CTles et l'ABES sur les états de collection et la conservation partagée au niveau national. Cette réflexion sur un archivage pérenne des collections papier doit être logiquement prolongée par une semblable réflexion sur l'archivage pérenne des ressources numériques. A BSN 6, on estime ainsi que « le coût à venir de l'archivage numérique pérenne est tel qu'il nécessite une politique nationale portée au niveau interministériel et facilitant des coopérations et une mutualisation des efforts. BSN doit constituer un des lieux visibles de cette synergie. Dans ce cadre, plusieurs points sont étudiés parmi lesquels la typologie de données à conserver sur le très long terme, les moyens de rationaliser les pratiques des opérateurs existants et ainsi de réduire les coûts, la prise en compte de l'archivage pérenne dans les contrats des établissements et son financement dans la durée, les conditions de l'obtention de l'agrément des Archives de France, et l'organisation et l'assise juridique de l'ensemble » <sup>187</sup>.

Au sein des instances universitaires, l'ADBU et le groupe de travail EPRIST se sont saisi de la problématique de l'archivage pérenne, et ont proposé parmi plusieurs préconisations concernant les politiques de sites et l'IST de :

« Mutualiser pour les données de la recherche comme pour d'autres dossiers (conservation numérique pérenne par exemple), quand cela est pertinent, les infrastructures à un niveau national, et confier au niveau local d'une part la mise en œuvre de projets à dimension régionale ou locale, d'autre part, l'accompagnement scientifique, technique et juridique des communautés de recherche, via l'allocation de ressources humaines mutualisées au niveau des sites. A minima, il conviendrait de former les professionnels de l'IST et de les intégrer officiellement dans la chaîne de traitement des données. » <sup>188</sup>

Ces diverses recommandations font émerger les orientations majeures qui devront structurer la stratégie de l'archivage pérenne universitaire : mutualisations à divers niveaux (dont celle des plateformes d'archivage nationales) qui permettent une réduction des coûts, liens à nouer avec la recherche et formation des

---

<sup>186</sup> CARBONE Pierre, « Numérique et archivage pérenne : regard du président de la Commission Bibliothèques Numériques », *op. cit.*

<sup>187</sup> BSN, « BSN6. Archivage pérenne » [En ligne] (page consultée le 30/12/2015) <http://www.bibliothequescientifiqnumerique.fr/?BSN-6-Archivage-perenne>

<sup>188</sup> ADBU, « Politiques de sites et IST : les préconisations ADBU – EPRIST » [En ligne] (page consultée le 30/12/2015) <http://adbu.fr/politiques-de-sites-et-ist-les-preconisations-adbu-eprist/>

professionnels (voir infra sur ces deux points), mais aussi réflexion sur les impulsions à donner au niveau de l'État.

#### Les niveaux de la mutualisation

La mutualisation peut s'opérer à divers niveaux : au sein même d'une université, au sein d'une fusion d'établissements ou d'une COMUE, ou en ayant recours à des solutions mutualisées ; elle concerne aussi les plateformes d'archivage, et peut ainsi poser la question des périmètres institutionnels définis jusqu'à présent. Actuellement, à ces divers niveaux, la mutualisation est encore à l'état de projet.

Au niveau des établissements, plusieurs SCD envisagent ou ont envisagé de mener des projets en collaboration avec d'autres services de l'université, eux aussi confrontés à la problématique de l'archivage. Le SCD d'Angers témoigne par exemple d'une réflexion en ce sens : lorsque la question de l'archivage pérenne s'était posé aux équipes, un archivage groupé au CINES avait été considéré, qui aurait englobé aussi bien les documents de l'archive ouverte Okina que d'autres documents « à l'échelle de l'Université »<sup>189</sup>. Un nouvel archiviste, rattaché depuis peu au SCD, pourrait travailler sur les questions d'archivage numérique au niveau de l'ensemble de l'université ; il n'est donc pas impossible que l'orientation initiale soit reprise. A Lyon I, un projet mutualisé entre la DSI de l'université, la direction des affaires juridiques et le SCD est à l'étude mais ne concerne que le stockage sécurisé.

Dans le cadre politique et institutionnel particulièrement favorable de la fusion d'établissements, Strasbourg envisage une stratégie commune pour l'archivage pérenne qui est en cours d'étude, et qui bénéficierait à tous les établissements partenaires (INSA, université de Mulhouse, etc.). Il s'agirait à la fois d'une mutualisation inter-services (à l'Unistra, seraient ainsi concernés le SCD, le service des archives pour les archives administratives et l'équipe du projet GED, une réflexion commune sur le cycle de vie du document et les DUA étant prévue) et d'une mutualisation inter-établissements.

Au niveau des COMUE, la mutualisation est souvent mentionnée davantage comme un horizon que comme projet réalisable à court terme. Par exemple par l'université de Toulouse, où la question de l'archivage pérenne des données de la recherche a été abordée avec certains laboratoires ; mais du fait du caractère mixte des laboratoires et des volumétries concernées, c'est au sein de la COMUE que le SCD envisagerait de traiter la question, afin de faire dialoguer SCD et instances en charge de la recherche. Deux COMUE parisiennes cependant ont initié des projets d'archivage pérenne mutualisé : SPC et PSL.

Au sein de la COMUE SPC<sup>190</sup>, le constat est fait que chaque BU de la COMUE suit en local un même ensemble de processus (du catalogage dans le Sudoc, ou dans Calames, jusqu'au dépôt dans PAC) ; l'hypothèse est alors d'un outil de back office commun pour la numérisation, la diffusion et l'archivage, qui permettrait de récupérer les notices bibliographiques dans les SIGB, de composer des lots pour la numérisation, puis une fois les images validées, de procéder automatiquement à la création des paquets SIP pour un archivage pérenne au CINES. Ce schéma commun, qui répondrait au besoin de prendre en charge toute

---

<sup>189</sup> Source : Stéphanie Bouvier, chargée de mission Open access (LabUA - Direction du développement du numérique).

<sup>190</sup> USPC [En ligne] (page consultée le 30/12/2015) <http://www.sorbonne-paris-cite.fr/>

la chaîne de production des données et métadonnées depuis la préparation avant la numérisation jusqu'à la réception des contenus numérisés et la préparation à la diffusion et à l'archivage, assurerait de plus l'adoption par tous les établissements concernés de règles comparables concernant les métadonnées par exemple. Le choix des métadonnées et la qualification précise des données demeurerait de la responsabilité des établissements, car chaque fonds ou collection est spécifique. Ces spécificités pourraient être conservées par la création de collections différentes dans PAC. La mutualisation serait alors celle des outils et des moyens, financiers et informatiques principalement (développements logiciels, etc.), quand le travail archivistique serait à conserver en interne. Partant de ce constat d'un besoin « d'automatiser les échanges avec le CINES, pour rendre les processus de numérisation (incluant la partie archivage pérenne) plus fluides afin de pouvoir travailler de façon moins artisanale »<sup>191</sup>, le projet de back-office « Plateforme de gestion et de diffusion des contenus numérisés »<sup>192</sup>, mené conjointement dans le cadre de la COMUE par la BSG, la BULAC et Sciences Po, a pour objectif de construire une « passerelle » vers la plateforme d'archivage numérique pérenne du CINES en optimisant les ressources et en mutualisant les coûts d'investissement des établissements partenaires. Le CINES recevra d'un seul service versant (la COMUE) les paquets de versement<sup>193</sup>. Tous les établissements de la COMUE ne sont pas encore impliqués dans le projet ; il n'y a pas d'établissement pilote, mais la bibliothèque Sainte Geneviève, déjà déposante au CINES, constitue une référence pour le groupe projet. Chaque établissement disposera d'un bibliothécaire qui assumera le rôle de correspondant<sup>194</sup>. La fonctionnalité « archivage pérenne » comprendra « la création des paquets à archiver, le dépôt, la gestion des rejets et des certificats d'archivage, mais aussi la gestion des préférences (filtres, affichages, tri par défaut ...), l'import, la conversion des formats des divers SIGB, la vérification des doublons »<sup>195</sup>. Le projet est ancien

---

<sup>191</sup> Cette citation et les suivantes renvoient aux propos d'Alexandra Gottely (chef de programme numérisation, Cujas), Olessea Dubois (responsable du Service Numérisation, Science Po) et Geneviève Cron (responsable du projet « PGCN » (plate-forme de gestion de contenu numérisé), SPC) qui ont apporté leurs éclairages sur ce projet.

<sup>192</sup> Voir la présentation de la plateforme par PÉRÉSAN-ROUDIL Alice, *Dans la forêt touffue des bibliothèques numériques patrimoniales françaises : quels choix, quelles stratégies, quelles perspectives ?*, op. cit., p. 34 : « Depuis quelques années, Science-Po et la BSG mettent à profit une subvention attribuée par la ville de Paris et la région Île-de-France pour la création d'une plateforme de diffusion de contenus numériques commune aux établissements de la COMUE SPC intéressés (la BULAC, Paris III et V se sont attachés au projet). Les réunions de travail ont visé jusqu'à présent à définir les modalités de mise en ligne (des plateformes comme Gallica Marque blanche ou Internet Archive, par exemple, ont été considérées) et les caractéristiques relevant du back office d'une solution logicielle commune, tandis que certains des partenaires ont mis à profit les quelques années déjà écoulées pour créer ou enrichir leurs propres bibliothèques numériques, à l'instar de la BSG et de la BIUS. Un tel projet constitue un réel défi pour plusieurs raisons. Bien sûr, les collections présentées seront hétéroclites, ce qui nécessite aussi la définition d'une image commune lisible par le public (cette plateforme serait destinée à une audience aussi diversifiée que possible). Cette caractéristique crée des difficultés de diffusion car chaque type de documents – notamment audiovisuels – requiert une interface particulière. Cela pourrait obliger les bibliothèques partenaires à ne pas y diffuser tous leurs contenus. Enfin, l'existence de plusieurs partenaires rend indispensable l'accord sur quelques formats de fichiers et de métadonnées communs pour permettre l'interopérabilité des contenus et leur mise en commun. Cette plateforme pourrait être mise en ligne à l'horizon 2017, sans empêcher les bibliothèques partenaires de garder leurs propres sites ».

<sup>193</sup> « Ces paquets se conformeront à la grammaire XML qui fera l'objet d'un accord entre le CINES et la COMUE. En cas d'invalidation, chaque établissement s'occupera de ses corrections. En cas de validation, le paquet sera archivé avec le numéro unique (AIP). » (G. Cron)

<sup>194</sup> « Chaque établissement sera son propre service versant. Car les services qui versent déjà ne souhaitent pas changer de service versant, surtout dans le cas où ils changent de rattachement de COMUE, par exemple » (G. Cron)

<sup>195</sup> « Un gestionnaire de données est prévu pour identifier les utilisateurs habilités à créer, modifier, valider et supprimer des documents pour une bibliothèque donnée. Chaque bibliothèque aura au moins un compte administrateur qui lui permettra de créer des comptes pour les agents de sa bibliothèque. Pour chaque document les droits d'usages pourront être paramétrés. Les licences *creative commons* pourront être dynamiquement incorporées sous forme d'API sur le modèle de ce qui est proposé par Internet Archive. Ce statut pourra être modifié par notice, par lots de notices (identifiants séparés par des virgules), par collection, par bibliothèque. » (Olessea Dubois)

### III. Situation des bibliothèques universitaires : vers une nécessaire mutualisation

(2010), mais du fait de nombreux changements n'est pas encore lancé<sup>196</sup>. La numérisation patrimoniale et les archives sont concernées : il semble que le rapprochement entre archives administratives et productions scientifiques et patrimoniales tende à s'imposer, l'ouverture d'une convention avec le CINES offrant la possibilité à d'autres services que les SCD de s'inscrire dans une démarche commune. Le traitement des collections peut exiger ensuite une approche spécifique selon les fonds, en particulier concernant les aspects juridiques (droits, DUA...) et les métadonnées. Les données de la recherche (dans les cas du CDSP, du Cevipof, etc.) seront également prises en compte.

Au sein de la COMUE Paris Sciences et Lettres (PSL)<sup>197</sup>, le Liberlabo faisait en 2013 le constat du lien étroit entre diffusion et préservation dans le domaine des données numériques, et se fixait comme mission « l'archivage pérenne pour les contenus et supports numériques, grâce à un laboratoire de conservation moderne et une réflexion sur les pratiques mêmes de conservation »<sup>198</sup>. PSL poursuit et élargit cette réflexion<sup>199</sup>, en initiant une réflexion sur l'archivage pérenne au sein d'une GED (adoption de formats pérennes) et d'un SAE communs à tous les établissements de la COMUE, afin de favoriser « une approche mutualisée, [ou] de bénéficier d'économie d'échelle par la mutualisation ». Le rapport 2014 du Schéma directeur numérique de PSL<sup>200</sup> qui « initie une politique de mutualisations efficaces et pertinentes, qui vise à rationaliser les actions, les établissements bénéficiant ainsi des effets de taille et de coordination offerts par la COMUE »<sup>201</sup> note que la réflexion concernera les archives administratives (papier et numériques) et éventuellement l'archivage des données scientifiques et souligne qu'une solution d'archivage pérenne interne à la COMUE n'est pas envisageable, et « devrait être au minimum mutualisée à l'échelle Île-de-France (voir avec UNR-IDF) ou au niveau national (services du CINES ? Projet de TGIR HUMA-NUM) ». Le programme « Infrastructures » définit dans la « population cible du projet » tous les établissements de PSL ayant des besoins en stockage ou en archivage, ce qui engloberait les BU. Un archivage pérenne serait également requis au vu du programme « Capitalisation et diffusion des productions de la Recherche », qui propose plusieurs scénarios de collecte des AO des établissements<sup>202</sup>. Parmi les

<sup>196</sup> « Pour le moment les métadonnées des fonds à archiver dans leur ensemble ne sont pas documentées, il n'y a donc pas encore de PPDI. » (G. Cron)

<sup>197</sup> PSL [En ligne] (page consultée le 30/12/2015) <https://www.univ-psl.fr/fr>

<sup>198</sup> « La « boucle » patrimoniale ou les cycles de traitement des données courantes destinées à devenir patrimoniales (et non d'emblée patrimoniales comme le sont certaines des acquisitions auprès des libraires anciens par exemple) sont métamorphosés par le concept de « diffusion continue » où les missions de conservation et de diffusion ne sont plus séparées. Selon le truisme « la science en train de se faire » sera demain le précieux témoignage de la science accomplie ou de « la science telle qu'elle se faisait », la valorisation du patrimoine scientifique au sens large est consubstantielle au concept du Liberlabo, d'où les missions suivantes : mission de préservation de pointe, y compris l'archivage pérenne pour les contenus et supports numériques, grâce à un laboratoire de conservation moderne et une réflexion sur les pratiques mêmes de conservation (PSL offre en son sein des compétences de haut niveau aussi variées que chimie, biologie, informatique, histoire du livre, bibliothéconomie, etc.) (...) », Laure Léveillé, « Le Liberlabo de PSL », *Revue Sciences/Lettres* [En ligne], n. 2, 2014, mis en ligne le 07 octobre 2013, consulté le 04 décembre 2015. URL : <http://rsl.revues.org/486> ; DOI : 10.4000/rsl.486

<sup>199</sup> Dès 2012 ; voir le rapport de Jean-Luc Gautier-Gentès et Benoît Lecoq, « Le volet documentaire de Paris Sciences Lettres. L'occasion de l'exemplarité », Rapport n° 2012-004, IGB, juin 2012.

<sup>200</sup> Le schéma directeur numérique de PSL (SDN) est l'un des jalons de son contrat quinquennal 2014-2018. « Initié en juillet 2013 dans le cadre d'une convention conclue avec la Caisse des dépôts et consignations, il a été approuvé par les conseils d'administration de la FCS et de la COMUE le 06 octobre 2014 », rapport 2014 du Schéma directeur numérique de PSL, en ligne sur le site de la COMUE : <https://www.univ-psl.fr/fr/search/node?search=sdn&op=Search>

<sup>201</sup> MENESR, « Contrats pluri-annuels de site » [En ligne] (page consultée le 30/12/2015) <http://www.enseignementsup-recherche.gouv.fr/cid81351/contrat-pluriannuel-du-site-paris-sciences-et-lettres-p.s.l.-2014-2018.html>

<sup>202</sup> « Scénario 1 : la mise en œuvre d'une « collection PSL » dans HAL, sur la base des dépôts déjà effectués par les établissements dans HAL. Scénario 2 : la mise en œuvre d'une solution (de type ORI-OAI du consortium ESUP) de

acteurs pilotes du projet, on compterait les archivistes, le groupe projet SI, le pôle Ressources et savoirs de PSL, et l'UNR d'Île-de-France<sup>203</sup>. Très tôt, PSL a pris contact avec Cujas dans le cadre d'un projet fédérateur de numérisation, afin d'anticiper les futurs processus de l'archivage pérenne. Contact a été également pris avec le CINES. Comme le projet de SPC, l'archivage pérenne mutualisé à PSL est à envisager à moyen terme<sup>204</sup>.

Le Schéma de développement universitaire (SDU), présenté par le Grand Lyon et l'Université de Lyon en 2010, parmi plusieurs projets d'équipements de pointe, envisageait un « Learning Center », et un « grand centre de documentation » afin de développer la capacité globale de stockage des documents<sup>205</sup>. L'échec à l'appel à projets des IDEX en 2010 a été suivi d'un nouveau projet d'« Infrastructure d'accès et de traitement des savoirs ». Le « Très grand instrument documentaire » (TGID) y supposait entre autres un silo pour l'archivage numérique pérenne des documents numériques. L'échec du PRES « Université de Lyon » à l'IDEX 236 a « mis un point d'arrêt aux attentes » selon le rapport Lecoq-Micol qui note que « sous les mutations, l'ambition documentaire du site de Lyon semble bien avoir reflué »<sup>206</sup>. L'ajournement du projet lyonnais met en lumière la complexité des projets mutualisés, dont le pilotage demeure l'un des points clefs.

## ***B. Projets de coopération et solutions mutualisées***

### **En France**

Il n'existe pas de plate-forme de diffusion numérique pour les bibliothèques de l'enseignement supérieur assurant de façon intégrée un archivage pérenne : la « chaîne » logique qui est offerte aux partenaires de Gallica et qui implique automatiquement un archivage pérenne sans surcoût financier ou technique n'a pas d'équivalent dans l'ESR. Les projets de coopération et les systèmes mutualisés permettent cependant d'assurer une mutualisation entre les principaux acteurs.

Ainsi notons qu'un projet de coopération comme l'ancienne base Liber Floridus devenue catalogue Initiale, déjà évoquée (voir I 2. C), qui met l'expertise du CNRS (IRHT) et du CINES au service des collections des BU et se trouve coordonné par la Sous-direction des bibliothèques et de la documentation du MENESR, opère une forme de mutualisation des compétences, des moyens et du pilotage qui ouvre une piste de réflexion pertinente au moment où la problématique

---

diffusion et de collecte automatisée des archives ouvertes existantes dans les établissements (dans HAL ou dans les entrepôts d'établissement). Scénario 3 : la mise en œuvre d'un entrepôt de gestion et de diffusion propre à PSL, avec saisie directe dans l'outil ou collecte automatisée des archives ouvertes dans les établissements », *ibid.*

<sup>203</sup> La construction des universités numériques en région repose sur un contrat d'objectifs signé par l'ensemble des partenaires de chaque projet : État, établissements d'enseignement supérieur, CROUS, Région (et parfois diverses collectivités territoriales). L'AMUE y propose le développement de services d'infrastructure technique de base et communs à l'ensemble des métiers exercés au sein d'un établissement, entre autres concernant les systèmes de stockage, et la sécurité.

<sup>204</sup> Comme nous le confirme Hélène Chaudoreille, responsable Ressources et savoirs à PSL.

<sup>205</sup> Université du Grand Lyon [En ligne] (page consultée le 30/12/2015)

[http://www.economie.grandlyon.com/fileadmin/user\\_upload/fichiers/site\\_eco/20101020\\_gl\\_universite\\_sdu\\_lyon\\_2010\\_2020.pdf](http://www.economie.grandlyon.com/fileadmin/user_upload/fichiers/site_eco/20101020_gl_universite_sdu_lyon_2010_2020.pdf)

<sup>206</sup> LECOQ Benoît et MICOL Charles, « Le stockage des collections imprimées à l'heure du numérique », *op. cit.*, p. 37.

des données de la recherche exigent un rapprochement de la recherche, du monde universitaire et des infrastructures nationales.

Dans ce domaine des données de la recherche justement, le nouveau segment de BSN (BSN 10 « Données de la recherche »<sup>207</sup>) vise à faire émerger un cadre au sein duquel il sera possible de réunir les professionnels de l'IST, les chercheurs (organismes et universités) et les opérateurs nationaux « sur la base d'exemples existants à l'étranger (ex. les plateformes américaines et anglaises), (...) en coordonnant les opérateurs nationaux : ANR (financement), ABES (signalement), INIST (accès), CINES (archivage des données) »<sup>208</sup>.

De même, les systèmes mutualisés, qui font reposer les processus sur plusieurs acteurs, coordonnent les actions et permettent des économies d'échelle semblent les mieux à même de garantir une efficacité et une continuité des projets. Les systèmes STAR et HAL sont ainsi reconnus par les professionnels pour leur pertinence et leur fonctionnement exemplaire. Un tel système pour l'archivage pérenne assurerait une politique cohérente et globale de préservation des documents qui ne sont pas déjà pris en charge dans les circuits évoqués supra.

#### A l'international

Plusieurs solutions mutualisées internationales existent d'ores et déjà et proposent des modèles divers de mutualisation de l'archivage. Aux États-Unis, partant du constat que l'archivage pérenne implique un coût qui peut se révéler insurmontable pour de nombreux établissements, la Mellon Foundation<sup>209</sup> a lancé plusieurs projets de coopération sur la production et la gestion d'archives digitales de périodiques. Sont ainsi nées les solutions LOCKSS, CLOCKSS, et PORTICO<sup>210</sup>, présentées en annexe, qui permettent aux bibliothèques d'archiver de façon pérenne les publications en ligne et les e-books. Ces solutions sont basées sur la collaboration entre les éditeurs et les bibliothèques et ont vocation à archiver tous les périodiques électroniques scientifiques payants et Open Access. PORTICO et LOCKSS ont étendu récemment leur couverture à la préservation des e-books, et permettent aussi, selon des modalités un peu différentes, la préservation d'autres contenus numériques détenus par les bibliothèques. Elles sont différentes dans leur technologie : LOCKSS et CLOCKSS reposent sur la préservation distribuée alors que PORTICO est une archive centralisée. Par ailleurs, PORTICO et LOCKSS proposent un accès post-abonnement alors que CLOCKSS se limite à l'archivage pérenne<sup>211</sup>. En 2008, le rapport du JISC concluait qu'aucun des projets ne peut être considéré comme assurant complètement contre le risque de potentielle disparition des documents scientifiques produits sous forme digitale (« digital scholarly

---

<sup>207</sup> BSN, « BSN 10. Données de la recherche », [en ligne] (page consultée le 31/12/2015) <http://www.bibliothequescientifiquenumerique.fr/bsn-10-donnees-de-la-recherche/>

<sup>208</sup> COLAS Alain, « Quelle politique nationale pour les données de la recherche ? », MESR -Mission de l'information scientifique et technique et du réseau documentaire (MISTRD), 2013 [en ligne] (page consultée le 31/12/2015) [renatis.cnrs.fr/IMG/pdf/Colas\\_FREDOC\\_2013\\_Presentation.pdf](http://renatis.cnrs.fr/IMG/pdf/Colas_FREDOC_2013_Presentation.pdf)

<sup>209</sup> The Andrew W. Mellon Foundation [En ligne] (page consultée le 30/12/2015) <https://mellon.org/>

<sup>210</sup> Pour une comparaison de ces trois solutions, cf. FITZPATRICK Kathleen, *Planned Obsolescence: Publishing, Technology, and the Future of the Academy*, NYU Press, 2011, p. 150.

<sup>211</sup> Pour une présentation exhaustive du panorama de l'archivage pérenne des périodiques électroniques, et sur les coûts des solutions évoquées, voir ALEXANDRE Aude, *Archivage papier et/ou électronique et accès pérenne aux ressources documentaires électroniques en texte intégral en Fédération Wallonie-Bruxelles*, Bibliothèque interuniversitaire de la Communauté française de Belgique, 2014, p. 25-89.

record ») : « None of the current initiatives is likely to yet fulfil all the access and archival needs of a modern library »<sup>212</sup>.

Mentionnons enfin E-Depot<sup>213</sup>, le système d'archivage pérenne mis en place par la Bibliothèque nationale des Pays-Bas (KB), à l'origine en lien avec le dépôt légal, qui est un système centralisé, mais qui propose un archivage pérenne des périodiques comparable à celui de CLOCKSS. Plusieurs archives institutionnelles universitaires déposent dans E-Depot.

Outre ces solutions internationales mutualisées, à ce jour non utilisées par des SCD français (mais LOCKSS l'est par l'INIST), d'autres pistes seraient à envisager, tels des accords avec des institutions étrangères, par exemple des bibliothèques nationales. Ainsi, les bibliothèques nationales des pays d'origine des éditeurs conservent parfois les fichiers des revues ; par exemple la Bibliothèque royale des Pays-Bas conserve toutes les revues en accès libre recensées par le Directory of Open Access Journal, soit plus de 10 000 titres publiés dans plus de 120 pays. L'Eblida et l'IFLA incitent à des systèmes d'accords entre bibliothèques nationales pour avoir accès aux archives scientifiques des autres pays.

Sur les offres logicielles et les autres solutions externalisées internationales, voir en annexe.

### ***C. Les grandes orientations stratégiques nationales. Vers de nouveaux périmètres ?***

L'archivage pérenne excède largement les moyens d'un seul établissement ; seule la mutualisation, par une plate-forme d'archivage, qui peut être nationale, peut permettre d'assumer les coûts de la plate-forme et du management de l'archive<sup>214</sup>. Le coût de chaque plateforme étant conséquent, la mutualisation peut aussi être celle des différentes plateformes d'archivage publiques.

Le groupe PIN, de l'association Aristote, a fait porter sa réflexion sur cette orientation stratégique qui conditionne les fonctionnalités et les coûts de l'archivage proposé aux SCD, envisageant un archivage partagé entre le CINES et la BnF. Ont par exemple été proposés, parmi plusieurs partenariats possibles, une réplique croisée des données. Cette option, qui nécessitait de la part de la BnF une sélection des données répliquées au CINES, s'est heurtée à des obstacles techniques majeurs, du fait de la conception des collections dans SPAR. La réplique croisée des données du CINES se fait actuellement auprès de l'IN2P3.

Pour autant, la volonté de mutualisation demeure valable, impliquant potentiellement la redéfinition des périmètres institutionnels définis entre l'ESR et le MCC. Le 45<sup>ème</sup> congrès de l'ADBU (Besançon, 2015) évoque ainsi un possible resserrement des liens entre la BnF et l'ESR : la BnF pourrait « jouer le rôle de tiers de confiance sur la question de la conservation des données », et le rôle de

---

<sup>212</sup> MORROW Terry, BEAGRIE Neil, JONES Maggie, CHRUSZCZ Julia, « A Comparative Study of e-Journal Archiving Solutions », A JISC Funded Investigation, 2008.

<sup>213</sup> KB-National Library of the Netherlands, « Long-term usability of digital resources » [En ligne] (page consultée le 30/12/2015) <https://www.kb.nl/en/organisation/research-expertise/long-term-usability-of-digital-resources>

<sup>214</sup> A titre d'exemple, selon BANAT-BERGER Françoise, DUPLOUY Laurent, HUC Claude, *L'archivage numérique à long terme : les débuts de la maturité ?*, op. cit., p. 168, le coût du TO archivé par un « service producteur assurant lui-même son archivage » en regard du coût du TO archivé par une « plate-forme nationale pour l'ensemble des administrations centrales de l'État » serait près de 20 fois supérieur pour les coûts externes, multipliés par près de 66 pour les coûts liés aux agents de catégorie A, et par plus de 32 pour les agents de catégorie B.

SPAR dans le paysage, évoluer en conséquence<sup>215</sup>. Cette question est également abordée dans le cadre de BSN5 et BSN6. La réflexion, avancée sur la question de la conservation physique et de la numérisation, est en cours au sujet de l'opérateur de conservation numérique pour l'ESR, qui demeure le CINES, et se pose avec acuité pour les données de la recherche.

Le futur programme VITAM (Valeurs Immatérielles Transmises aux Archives pour Mémoire) constitue également un nouvel horizon dans le paysage institutionnel. Ce programme interministériel constituera la plateforme mutualisée d'archivage de documents électroniques produits par l'administration. VITAM a vocation à collecter et conserver la production documentaire numérique de l'ensemble des ministères, des administrations, des collectivités et des opérateurs publics et concerne potentiellement tous les documents numériques que l'administration souhaite conserver à des fins d'archivage courant, intermédiaire ou définitif. Le programme développe une brique logicielle open source d'archivage électronique qui sera générique et pourra ensuite être implémentée par les acteurs qui le souhaiteraient selon leurs besoins. Elle sera d'abord implémentée aux Archives nationales (projet Adamant), au ministère des Affaires étrangères (projet Saphir) et au ministère de la Défense (projet Garde-Archipel) pour l'archivage définitif ainsi qu'au ministère de la justice et au ministère de l'écologie pour l'archivage intermédiaire. A terme, d'autres ministères, collectivités, établissements publics, etc. pourront implémenter Vitam. « L'enseignement supérieur pourrait [également] être concerné. Des contacts ont été établis entre Vitam et le CINES pour voir quels pourraient être les rapprochements »<sup>216</sup>. Une version bêta de VITAM est attendue pour l'été 2016.

L'intérêt d'un outil en open source avait été perçu par la BnF, laquelle avait préconisé l'ouverture de SPAR en open source et proposé une étude pour évaluer l'effort à fournir<sup>217</sup>. L'open source permet en effet à une plateforme de ne plus supporter seule le système d'archivage. La coopération avec le CINES n'a pas abouti, mais le programme VITAM sera en production au moment où les systèmes d'archivage pérenne pourraient nécessiter un profond renouvellement technologique. VITAM est pour l'instant davantage conçu pour les archives que pour les bibliothèques, mais pourrait connaître des évolutions. Projet très ambitieux, qui cumulerait de façon novatrice en une seule application des fonctionnalités de diffusion et de préservation, il pourrait constituer un bouleversement des pratiques et des périmètres.

### 3. De nouvelles missions et de nouvelles compétences ?

Bien que les contextes soient variés et probablement destinés à évoluer, la nécessité demeure pour les professionnels d'être en lien avec les producteurs de

---

<sup>215</sup> Selon Sylviane Tarsot-Gillery, directrice générale de la BnF, lors du 45ème congrès de l'ADBU, organisé à Besançon du 29 septembre au 1er octobre 2015, « l'ouverture tant juridique que technique aura toute sa portée avec un décloisonnement institutionnel » ; « la constitution d'un domaine public des données implique que les acteurs travaillent entre eux » ; ainsi, « la BNF entend collaborer davantage avec le monde de l'ESR ». ADBU, « Retour sur la matinée politique du congrès 2015 de l'ADBU: Vers un domaine public des données ? » [En ligne] (page consultée le 30/12/2015) <http://adbu.fr/retour-sur-la-matinée-politique-du-congrès-2015-de-ladbu-vers-un-domaine-public-des-données/>

<sup>216</sup> Source : Stéphanie Roussel, chef du bureau de l'archivage numérique, des normes et des référentiels à la Direction générale des Patrimoines du SIAF.

<sup>217</sup> Source : Laurent Duploux, chef du service numérisation, BnF.

données et d'acquérir les compétences nécessaires à l'archivage pérenne, de son anticipation à sa mise en œuvre.

### *A. Le soutien aux chercheurs : anticiper l'archivage pérenne*

La conservation numérique est appelée à être de plus en plus être intégrée clairement dans le cycle de vie du document : la prise de conscience des enjeux de la perte de données, les diverses incitations à documenter les données et les métadonnées et la connaissance par les professionnels de l'information de la problématique de l'obsolescence vont à court terme inciter à la constitution de versions archivables répondant aux normes et formats utilisés par les plateformes d'archivage. Quel peut être le rôle des SCD auprès des chercheurs ? Quelle est leur vocation en regard de la production scientifique, et comment peuvent-ils s'inscrire dans le « cycle de vie » des données ?

Dans le cas des publications scientifiques, les SCD tentent par une information à destination des chercheurs de favoriser les dépôts dans les AOI, puisque le dépôt entraîne alors l'archivage pérenne dans les cas, majoritaires, où l'AOI est reliée à HAL. Certains arguments peuvent être porteurs : la visibilité des travaux, mais aussi la conservation des données dans un entrepôt fiable peuvent convaincre de l'utilité du dépôt<sup>218</sup>. Car si l'archivage pérenne n'est pas la priorité des chercheurs, qui visent en premier lieu la diffusion et le signalement de leurs travaux, l'argument patrimonial peut cependant porter. Les SCD, en plus des formations aux doctorants, procèdent à des démonstrations au sein des laboratoires, ou à des opérations ponctuelles comme les Open Access Weeks. A Toulon par exemple, une journée de formation, organisée avec l'Urfist, est soutenue par le vice président de la recherche et les représentants des laboratoires. La situation demeure souvent contrastée selon les pratiques des laboratoires et la motivation des chercheurs ou leur compréhension des enjeux. Plusieurs professionnels témoignent ainsi de la difficulté de réunir l'ensemble des chercheurs lors de leurs interventions dans les laboratoires, et recommandent la présence d'un bibliothécaire référent qui puisse se charger du dépôt à la place des chercheurs qui n'en auraient pas la disponibilité, comme au SCD de l'INPT de Toulouse. L'accent est mis sur la responsabilité des chercheurs eux-mêmes, qui en tant que producteurs des documents doivent être les porteurs des actions ; laisser seul le SCD en charge de la collecte et des dépôts comporte le risque de marginaliser sa démarche en le coupant de la recherche, à qui doit demeurer l'initiative. Revient le constat que le SCD, s'il peut être acteur et « opérateur », ne peut assurer seul le pilotage de tels projets, dont le pilotage politique revient à la direction de l'université. Dans le cadre très peu contraignant qui est celui de l'université – les mandats de dépôt, à l'œuvre depuis peu à Angers<sup>219</sup>, demeurent

---

<sup>218</sup> « Rien n'est plus important pour un chercheur que de savoir ses travaux visibles, et ainsi les voir éventuellement cités, preuve de leur qualité et de la reconnaissance de ses pairs. Toutefois, beaucoup de chercheurs sont encore réticents face à l'OA par peur du plagiat et des procédures d'autoarchivage jugées trop lourdes. C'est ici que les bibliothèques universitaires peuvent intervenir en servant de relais actifs d'une politique de l'OA, en les sensibilisant et en les formant aux procédures », Jean-Michel SALAÛN , « L'Open Access, quels enjeux pour les bibliothèques universitaires ? », site Économie du document [En ligne] (page consultée le 30/12/2015) <http://archinfo24.hypotheses.org/2392>, consulté le 21/10/2014

<sup>219</sup> « L'Université d'Angers s'est lancée dans la voie de l'open access, le « libre accès ». Objectif : permettre à tous de prendre gratuitement connaissance des derniers résultats de recherche de ses laboratoires. Fin mars 2013, l'UA a officiellement décidé de créer une archive ouverte institutionnelle, sous mandat de dépôt : les chercheurs devront y « signaler toutes leurs publications depuis 2008, et y adjoindre leurs articles parus depuis 2012, en version intégrale », rappelle Stéphanie Bouvier, en charge de ce projet », Université d'Angers, « Okina, la plateforme de l'UA » [En ligne] (page consultée le 30/12/2015) <http://www.univ-angers.fr/fr/recherche/actualites/open-access.html>

très rares<sup>220</sup> -, certains professionnels recommandent d'initier les actions avec des laboratoires pilotes.

Dans le cas des données scientifiques, promouvoir les bonnes pratiques (entrepôts de données, standards de métadonnées, outils et terminologie) en termes de gestion et de conservation des données avant toute diffusion est essentiel pour un potentiel archivage pérenne et permet une montée en compétence des producteurs de données et une amélioration de la qualité des données produites. Il existe en effet de façon générale un « manque d'expertise chez les chercheurs qui, bien qu'ayant une connaissance poussée et spécialisée des méthodes de collecte de données et de l'utilisation de celles-ci au sein de leur discipline scientifique, ne sont généralement pas au fait des questions entourant l'emploi des métadonnées, l'organisation intellectuelle de l'information et la préservation des documents numériques »<sup>221</sup>. Par ailleurs, il est rare que les chercheurs eux-mêmes aient la disponibilité pour anticiper l'archivage pérenne de leurs productions. Le constat que peu de chercheurs décident de partager ou d'archiver leurs données est régulièrement établi. La TGIR Huma-Num assume une mission de formation et d'information à ce sujet. Dans les établissements, les SCD, en charge des échanges avec la communauté de recherche dans le cadre des AOI, sont également aptes à dialoguer pour encourager à la structuration des données grâce à leur connaissance des procédures et des outils. Plusieurs SCD sont engagés dans des dialogues avec les laboratoires ; à Lille 1 par exemple, le soutien aux chercheurs est en projet<sup>222</sup>, le SCD cherchant à créer une dynamique autour de cet enjeu. Le SCD propose aux laboratoires de les aider à cartographier leurs données, « une mission de portée globale que les directions de laboratoire ne proposent généralement pas ». Il s'agit d'aider à définir les critères motivant l'archivage, démarche de sélection qui ne peut pas être faite sans l'appui et le point de vue des directions de laboratoire. Ni le chercheur individuel, ni le seul SCD ne peuvent décider de la valeur des données et de l'intérêt de leur archivage pérenne ; le rôle des directions de laboratoires est ici essentiel<sup>223</sup>. Le SCD assume alors une mission d'information et d'incitation, et peut signaler les outils adaptés, en termes de stockage à moyen terme<sup>224</sup> ou de conservation pérenne.

Enfin, certaines institutions ont désormais recours à un plan de gestion des données ou « data management plan » (DMP), qui pense la totalité de la chaîne de production, de valorisation et de conservation des données au sein d'un document formalisé qui décrit la manière dont seront obtenues, documentées, analysées, diffusées et utilisées les données produites au cours et à l'issue d'un processus ou d'un projet de recherche. Peu d'universités françaises, au contraire des universités anglo-saxonnes et américaines, en ont institué à ce jour (parmi elles, les universités Paris Diderot et Paris Descartes<sup>225</sup>). Le DMP précise les normes de métadonnées,

<sup>220</sup> « En France, les mandats, défendus par le CNRS et dont l'efficacité a été reconnue et approuvée à un niveau officiel, sont encore plutôt le fait d'organismes nationaux que d'universités », Laurent Jonchère, « Synthèse sur les politiques institutionnelles de libre accès à la recherche », 2013, pp.1-35. <sic 00794497v2>

<sup>221</sup> GUINDON Alex, « La gestion des données de recherche en bibliothèque universitaire », *op. cit.*, p. 189-200.

<sup>222</sup> Voir GÉROUDET Marie-Madeleine, « Sur le terrain des données de la recherche à Lille 1 », *op. cit.*, p. 14-15, sur les besoins des chercheurs en termes de structurations et de conservation de leurs données et les démarches de pédagogie et d'information qui peuvent être envisagées par les BU.

<sup>223</sup> Source : Marie-Madeleine Géroudet, responsable du service Bibliothèque numérique, Lille 1.

<sup>224</sup> Comme le Purdue University Research Repository (PURR, <https://purr.purdue.edu/>) qui guide le producteur, du Data management plan au stockage sécurisé.

<sup>225</sup> Voir CARTIER Aurore, MOYSAN Magalie, REYMONET Nathalie, « Réaliser un plan de gestion de données : guide de rédaction » (V1), Université paris-Descartes, 09/01/2015. [En ligne] (page consultée le 30/12/2015) [www.univ-paris-diderot.fr/DocumentsFCK/.../Realiser\\_un\\_DMP\\_V1.pdf](http://www.univ-paris-diderot.fr/DocumentsFCK/.../Realiser_un_DMP_V1.pdf)

les formats des bases de données, les méthodes et la sécurité d'accès, les durées d'archivage, ainsi que les coûts associés à la gestion des données, le sort des données à l'issue du projet, les critères de sélection des données, le volume final des données et la plateforme d'archivage envisagée, afin que l'archivage soit inclus dès le départ dans le cycle de vie de la donnée. Car si l'archivage pérenne n'est pas anticipé, il est ensuite très complexe voir impossible à mettre en œuvre. Reste à déterminer à quel stade de ce cycle les SCD peuvent –et doivent– intervenir. La rédaction d'un DMP « fait en effet appel à diverses compétences bien souvent présentes au sein de nos universités mais généralement disséminées entre plusieurs acteurs, voire plusieurs services qui n'étaient pas, jusqu'ici, nécessairement habitués à travailler ensemble »<sup>226</sup> : les chercheurs, les ingénieurs projets, les services informatiques, les archivistes, et les référents IST. Dans le cadre de ce type de workflow, les périmètres et les responsabilités sont cependant définis, ce qui permet de délimiter clairement l'action du SCD, et ses missions – information et formation, promotion des bonnes pratiques, ou encore rôle de référent.

La difficulté pour les SCD tient une nouvelle fois à définir ce qui revient à l'activité de recherche ou aux professionnels de l'IST. La « dichotomie entre enseignement supérieur et recherche, héritage de l'histoire » selon Raymond Bérard est à l'origine d'un paysage complexe, au sein duquel l'émergence de nouvelles problématiques exige probablement un pilotage qui excède la seule BU<sup>227</sup>.

### ***B. Une nécessaire montée en compétences. Les impacts sur le métier***

La prise de conscience de la spécificité de l'archivage pérenne est encore récente, les acteurs publics du domaine ne proposant des solutions opérationnelles que depuis peu, et la notion n'étant pas bien connue de tous les professionnels. C'est ainsi que stockage sécurisé et archivage pérenne sont parfois mal distingués.

Outre cette prise de conscience, l'archivage pérenne implique non seulement la maîtrise d'outils, de processus, de compétences informatiques et archivistiques, mais aussi comme on l'a vu une démarche pro-active envers les producteurs de contenus dont on souhaite *in fine* la préservation. L'ensemble de ces compétences a inévitablement des répercussions sur la formation des professionnels, mais aussi plus largement, à l'instar d'autres évolutions récentes de l'IST, sur la conception même du métier. La question de la formation des bibliothécaires est ainsi un enjeu décisif, qui contribuera également à déterminer la place des SCD dans le futur paysage de l'IST. Parmi les multiples nouvelles missions de plus en plus souvent attribuées aux SCD, on compte par exemple la veille sur les projets de recherche, le rôle de conseil entre autres sur les questions de droits, la bibliométrie,

---

<sup>226</sup> Ibid., p. 11.

<sup>227</sup> « En France, le principal problème est celui de la dichotomie entre enseignement supérieur et recherche, héritage de l'histoire. Les dispositifs documentaires se sont constitués et développés en parallèle sans coordination. Les modèles d'organisation documentaire des universités et des organismes de recherche divergent, reflétant les différences de missions (recherche d'une part, enseignement et recherche d'autre part), de publics (chercheurs, étudiants et enseignants chercheurs) ; de filières professionnelles et de formations ; de représentation du métier. Ajoutons le paysage morcelé des 85 universités françaises autonomes face à des organismes de recherche peu nombreux et puissants. Jusqu'en 2008, ces différences se reflétaient dans la tutelle des établissements avec une direction pour l'enseignement supérieur et une autre pour la recherche qui disposaient chacune de structures spécifiques pour le pilotage de l'IST ». BÉRARD Raymond, *op. cit.*, p. 3.

l' « advocacy » pour l'Open Access ou les AOI, ou l'analyse de métadonnées<sup>228</sup>. L'archivage pérenne est l'une de ces potentielles nouvelles missions, qui nécessite l'acquisition de nouvelles compétences.

Des notions d'archivistique sont ainsi essentielles, notamment pour la rédaction du PPDI, ou pour la gestion du cycle de vie du document. En l'absence d'un service d'archives, le conservateur doit s'en imprégner. Les postes sont ainsi parfois profilés : à Lille 1, une personne avec des compétences en archivistique aurait pu être privilégiée lors du recrutement sur un profil « données de la recherche » ; à la BIUS, le profil de poste « Archivage pérenne » est conçu pour les chartistes ; et à la tête du projet de Cujas se trouvait une ancienne élève de l'école des Chartes, assistée par une documentaliste. Au niveau stratégique, le conservateur doit mettre en place la « planification de la préservation », définir les acteurs, les tâches de chacun et les processus, et établir une gestion de risques en suivant le modèle OAIS, ce qui suppose une maîtrise de la gestion de projet. Afin de négocier les clauses des conventions ou des contrats avec les tiers-archivistes, ou pour maîtriser la réglementation en vigueur dans le domaine de la propriété intellectuelle, des notions juridiques seront également requises. Enfin, la maîtrise de l'informatique documentaire est fondamentale. Une part du travail de « routine » peut en effet être automatisée une fois un projet en production, mais cette automatisation est limitée : il incombe donc au management des bibliothèques de définir les compétences nécessaires et éventuellement des besoins en formation pour acquérir l'expertise métier. Les bibliothécaires ont déjà la maîtrise des GED (comme Nuxeo), ou des logiciels type ORI-OAI, mais l'archivage pérenne, à part, exige aussi une maîtrise minimum du XML et des métadonnées de conservation ; le catalogage ne se limite plus au descriptif et nécessite des métadonnées techniques et administratives (concernant les droits, l'historique des données, etc.). Le chef de projet doit ainsi savoir relire des fichiers METS ou DublinCore, être familier des tables de mapping, de la conversion de données, etc.

Les projets d'archivage pérenne se mènent en outre avec un accompagnement perçu comme plus réduit que dans le cadre des projets communs avec l'ABES, bien que le CINES dédie un référent par projet. Or au sein des établissements enquêtés, aucun référent n'avait reçu de formation spécifique, malgré des besoins réels exprimés par certains professionnels. Les demandes semblent rares : l'apprentissage se fait au fil du projet, grâce également aux retours d'expériences des autres établissements. La commission Bibliothèques Numériques remarquait sur le sujet de la formation : « On constate sur cette question parmi les bibliothécaires un déficit de culture professionnelle tant sur les procédures et les coûts que sur les niveaux de service en matière de tiers-archivage »<sup>229</sup>.

La situation semble cependant évoluer : les dispositifs de formation continue prennent davantage en compte la problématique ; des stages ou des journées de

---

<sup>228</sup> Voir la grille établie par COX Andrew, VERBAAN Eddy, SEN Barbara, « Upskilling Liaison Librarians for Research Data Management », *Ariadne*, 2012 : « Lead on institutional data policy, Bring data into undergraduate research-based learning, promoting data information literacy, Teach data literacy to postgraduate students, Develop researcher data awareness, Provide an advice service to researchers (and research administrators) Eg on writing Data Management plans or advice on RDM within a project, Advice on licensing data, Advice on data citation, Develop and manage access to data collections, Develop local data curation capacity (...) ». [En ligne] (page consultée le 30/12/2015) <http://www.ariadne.ac.uk/issue70/cox-et-al#sthash.paQShNpI.dpuf>

<sup>229</sup> Rapport 2012-2013, *op. cit.*

sensibilisation, par le groupe PIN, Huma-Num, l'IN2P3<sup>230</sup> ou encore l'École des Chartes<sup>231</sup> informent les professionnels. Globalement, le défi demeure important pour les bibliothèques universitaires, concernant l'archivage pérenne mais aussi plus généralement la maîtrise des pratiques innovantes des sciences de l'information qui dessinent de nouveaux métiers comme « data scientist », « data librarian », « bibliothécaire numérique », etc.

Comme dans le cas d'autres pratiques récentes dans le domaine de l'IST, l'archivage pérenne participe à la redéfinition des missions des bibliothécaires, et ce pour tous les corps du métier. Marie-Madeleine Géroutet, responsable du service Bibliothèque numérique au SCD de Lille 1, note ainsi que « la question se pose d'impliquer les magasiniers, l'équipement numérique des documents ne différant pas dans le principe de la tâche d'équipement de documents en général », pour certaines tâches - nommage ou renommage de fichiers par exemple. Il s'agirait alors d'une nouvelle mission du corps des magasiniers, appelé par ailleurs à voir évoluer ses missions<sup>232</sup>. Plus largement, la communauté de pratique entre archivistes, documentalistes et bibliothécaires est particulièrement perceptible dans les projets d'archivage pérenne, suggérant une convergence des métiers et une polyvalence des acteurs qui repose sur une formation en adéquation avec les évolutions récentes de l'IST<sup>233</sup>.

---

<sup>230</sup> Journées de sensibilisation à la pérennisation des données par l'IN2P3. CNRS-INIST, LaLIST [En ligne] (page consultée le 30/12/2015) <http://lalist.inist.fr/?p=13464>

<sup>231</sup> Par exemple, l'École Nationale Des Chartes a proposé en novembre 2015 un stage de formation continue « Dématérialisation et fiabilité de l'information numérique » en coopération avec le groupe PIN. OpenEdition, Calenda [En ligne] (page consultée le 30/12/2015) <http://calenda.org/331471>.

<sup>232</sup> Cf. CARBONE Pierre, AROT Dominique, CLAUD Joëlle, LECOQ Benoît et MICOL Charles, « Quels emplois dans les bibliothèques ? État des lieux et perspectives », Rapport de l'IGB, mars 2013.

<sup>233</sup> Ce que souligne par exemple la cartographie des métiers proposée par l'ADBS. ADBS, « Carte métiers 2013 » [En ligne] (page consultée le 30/12/2015) [http://www.adbs.fr/carte-metiers-adbs-2013-132079.htm?RH=MET\\_REFMETIER&RF=1374074004500](http://www.adbs.fr/carte-metiers-adbs-2013-132079.htm?RH=MET_REFMETIER&RF=1374074004500)





## CONCLUSION

---

La situation des BU face à la question de l'archivage pérenne est contrastée. Elle dépend de la taille et des moyens de l'université auxquelles elles appartiennent, de l'inscription de cette dernière dans une COMUE ou au sein d'une politique de site, de la politique générale de l'établissement universitaire concernant l'Open Access et la valorisation des données de la recherche, de la stratégie d'archivage globale incluant ou pas les archives et supposant ou non la future présence d'un Data Center. Les entretiens menés soulignent cependant que ce sont bien souvent les BU elles-mêmes qui sont à l'origine des projets d'archivage pérenne, afin de répondre à des besoins de conservation qui sont perçus comme de son ressort, l'institution étant en charge d'un patrimoine, qu'elle a produit (numérisations patrimoniales) ou pour lequel elle joue le rôle de tiers de confiance ou d'acteur du circuit d'archivage (HAL, STAR, ressources pédagogiques...). La BU tend donc à s'attribuer tout ou partie d'une mission qui n'est pour l'instant que rarement prise en charge par d'autres acteurs de l'université. Pour autant, le développement exponentiel des données de la recherche pourrait inciter à modifier cette situation, puisque comme nous l'avons vu, l'archivage de ces données nécessite une chaîne d'acteurs et une redéfinition des périmètres qui excèdent la seule BU.

C'est alors la question du pilotage des projets et la mutualisation des solutions qui se pose, ce qui implique probablement une politique nationale. Ainsi pour Alain Abécassis, chef du service de la coordination des stratégies de l'ESR (MENESR – DGESIP, DGRI), le MENESR a « une responsabilité publique à assurer la publicité, l'archive, l'enregistrement et la pérennité des publications de la recherche nationale » et « le fera en complémentarité avec l'ensemble des acteurs, comme accompagnateur vigilant et facilitateur des passerelles »<sup>234</sup>. Dans cet esprit, le ministère a décidé le renforcement du rôle de HAL. Face à des besoins de coordination des actions et de réduction des coûts, c'est une stratégie nationale incitative qui serait à même d'orienter les projets des établissements, par exemple en favorisant la prise en compte par BSN de l'archivage pérenne dans les contrats d'établissements et son financement dans la durée, ou dans les contrats de site. Un décroisement entre Culture et Enseignement Supérieur ouvrirait également la voie à de nouvelles solutions pour les établissements, qui se heurtent à des problèmes de coûts et de technicité des projets d'archivage pérenne.

Une solution mutualisée qui intégrerait à la fois le signalement, la diffusion et l'archivage au CINES permettrait à toutes les BU, y compris celles dont les moyens sont limités, de procéder à l'archivage pérenne des ressources dont elles ont la charge. STAR constitue à cet égard un modèle organisationnel et de coopération pertinent, qui repose sur les acteurs nationaux (ABES, CINES, etc.) et définit une place et un rôle clairs pour les BU. Un modèle de même nature rendrait cohérente une action concertée au sein de l'ESR. En tant que prolongement logique de leur rôle dans la diffusion de la production de l'établissement, dans l'archivage des thèses et dans les archives ouvertes, l'archivage pérenne pourrait alors s'inscrire explicitement de façon légitime et supportable dans les missions des BU.

---

<sup>234</sup> Cf. ADBU, « Retour sur la matinée politique du congrès 2015 de l'ADBU: Vers un domaine public des données ? » [En ligne], *op. cit.*



## SOURCES

### LISTE DES ÉTABLISSEMENTS AYANT RÉPONDU À L'ENQUÊTE

Bibliothèque universitaire Cujas	inter-	Catherine Renard, coordinatrice Numérisation et traitement documentaire
Bibliothèque universitaire de santé Descartes (BIUS)	inter-Paris	Jean-François Vincent, responsable du Service d'histoire de la Santé Claire Ménard, adjointe technique à la bibliothèque numérique, au Service d'histoire de la Santé
Bibliothèque de Science Po		Olesea Dubois, responsable du Service numérisation
SCD de l'université Pierre et Marie Curie (BUPMC)		Rémi Gaillard, responsable du pôle Politique documentaire et valorisation des collections
SCD de l'université Sorbonne Nouvelle - Paris 3		Romain Le Nezet, directeur-adjoint de la Direction des Bibliothèques Universitaires
Bibliothèque Sainte-Geneviève (BSG)		Pauline Rivière, chef de projet numérisation
SCD de l'université de Poitiers		Anne-Sophie Traineau-Durozoy, conservateur du Fonds ancien, responsable du Pôle Moyen Âge
SCD de l'université de Strasbourg		Adeline Rege, chef de projet Archives ouvertes de la connaissance
BNU, Strasbourg		Frédéric Blin, directeur de la conservation et du patrimoine
SCD de l'université d'Angers		Stéphanie Bouvier, chargée de mission Open Access, LabUA, direction du développement du numérique
SCD de l'université de Toulon		Daniel Eymard, directeur du SCD
SCD de l'université de Lille 1		Marie-Madeleine Géroutet, responsable du service Bibliothèque numérique
SCD de l'université de Lille 2		Anne Flahaut-Steiner, responsable de la Bibliothèque numérique, BU Santé Recherche Solenn Bihan, chargée de l'information scientifique : bibliométrie et open access, SCD & Service de la Recherche (SeRVIS)
SCD de l'université de Lille 3		Isabelle Westeel, directrice du SDC

	Cécile Malleret, responsable des services numériques
SCD de l'université de Lyon 1	François-Xavier Boffy, responsable Service Informatique Documentaire, BU Sciences
SCD de l'université de Lorraine	Maximilien Savoye, responsable du projet d'archivage pérenne, Direction de la documentation et de l'édition, Bibliothèque universitaire de Metz
SCD de l'université Clermont-Université	Olivier Legendre, conservateur adjoint au directeur de la BCU, responsable de la bibliothèque numérique
SCD de l'INPT (Institut National Polytechnique de Toulouse), Université de Toulouse	Jean-Marie Le Béhec, responsable Ingénierie documentaire & Direction du Système d'Information. Référent Études
Centre Flaubert, Laboratoire CÉRÉDI, Université de Rouen	Yvan Leclerc, membre permanent du CÉRÉDI
BnF	Laurent Duploux, chef du Service Numérisation  Stéphane Reecht, expert de préservation numérique, Département de la Conservation, Service Numérisation  Christine Génin, chargée de collections en littérature française contemporaine. Coordination du dépôt légal du Web et de la mise en ligne de contenus pour le département Littérature et art. Direction des Collections, Département Littérature et art
CINES	Lorène Béchard, archiviste au département Archivage et Diffusion
CCSD	Raphaël Tournoy, en charge de l'archivage à long terme, HAL et moteur de recherche Solr
ANRT	Joachim Schöpfel, directeur de l'ANRT
BSN 6	Julie Ladant, adjointe au chef du Département de l'information scientifique et technique et réseau documentaire (DISTRD). Sous-direction du pilotage stratégique et des territoires.
SI AF	Stéphanie Roussel, chef du bureau de l'archivage numérique, des normes et des référentiels, Direction générale des Patrimoines, Sous-direction de la politique archivistique
CNDP	Dominique Armand, documentaliste

	audiovisuelle, DIDFP
COMUE SPC	Geneviève Cron, responsable du projet « PGCN » (plate-forme de gestion de contenu numérisé)
COMUE PSL	Hélène Chaudoreille, responsable Ressources et savoirs



## BIBLIOGRAPHIE

---

### Généralités

CARBONE Pierre, CAVALIER François, *Les collections électroniques, une nouvelle politique documentaire*, Paris : Éditions du Cercle de la librairie, 2009

CAVALIER François, POULAIN Martine (éd.), *Bibliothèques universitaires : nouveaux horizons*, Paris, France : Éditions du Cercle de la librairie, 2015

CHARTRON Ghislaine, EPRON Benoît, MAHÉ Annaïg, *Pratiques documentaires numériques à l'université*, Villeurbanne : Presses de l'Enssib, 2012

CLAERR Thierry, MOUFFLET Jean-François, « Gestion de la conservation des collections numériques », in BARRON Géraldine, LE GOFF-JANTON Pauline (éd.), *Intégrer des ressources numériques dans les collections*, Villeurbanne, France : Presses de l'Enssib, 2014

COQ Dominique, *Apprendre à gérer des collections patrimoniales en bibliothèque*, Villeurbanne : Presses de l'Enssib, 2012

SCHÖPFEL Joachim, *La publication scientifique : analyses et perspectives*, Paris : Hermès Science Publications : Lavoisier, 2008

### Généralités sur l'archivage pérenne

ALEXANDRE Aude, *Archivage papier et/ou électronique et accès pérenne aux ressources documentaires électroniques en texte intégral en Fédération Wallonie-Bruxelles*, Bibliothèque interuniversitaire de la Communauté française de Belgique, 2014

BALLEY Noëlle, « Numérisation et archivage pérenne à la bibliothèque Cujas : un retour d'expérience », Journées du CTLes, 20 juin 2013 [En ligne] <https://www.ctles.fr/sites/default/files/ctles/documents/JPro2013Balley.pdf>

BANAT-BERGER Françoise, DUPLOUY Laurent, HUC Claude, *L'archivage numérique à long terme : les débuts de la maturité ?*, Paris, La Documentation française, 2009, Coll. Manuels et guides pratiques

BECHARD Lorène, « La conservation pérenne des données de recherche », CINES, 30 juin 2015

BÉCHARD Lorène, PRAT Philippe, « Une grille pour l'archivage intermédiaire et à long terme des données numériques en SHS », Journées OAS pour l'archivage de la recherche en SHS, 23-24 octobre 2012, CINES/TGE Adonis <http://slideplayer.fr/slide/1142284/>

BLIN Frédéric « La numérisation des bibliothèques », *Pour la science*, novembre 2013, no 433, p. 48-51

BONE Daniela, BURNS Peter, « An Overview of Content Archiving Services in Scholarly Publishing », Allen Press, Inc, 2011

BORGHOFF Uwe, RÖDIG Peter, SCHEFFCZYK Jan, SCHMITZ Lothar, *Long-Term Preservation of Digital Documents : Principles and Practices*, Springer, 2010

CARBONE Pierre, « Numérique et archivage pérenne : regard du président de la Commission Bibliothèques Numériques », in *Bulletin des bibliothèques de France* [en ligne], n°5, 2013

[://bbf.Enssib.fr/consulter/bbf-2013-05-0043-010](http://bbf.Enssib.fr/consulter/bbf-2013-05-0043-010)

CINES, « La plateforme PAC et ses évolutions fonctionnelles et techniques », 3ème journée des utilisateurs de l'archivage, 9 juin 2015

<https://www.cines.fr/wp-content/uploads/2015/06/PAC-Journ%C3%A9eUtilos-2015-06-09aprem.pdf>

DAUMAS Francis, MASSOL Marion « Au cœur d'un centre d'archivage », *Pour la science*, novembre 2013, no 433, p. 42-47

DERROT Sophie, FAUDUET Louise, OURY Clément, PEYRARD Sébastien, « Preservation is knowledge : a community-driven preservation approach », *Proceedings of the 9th International Conference on Preservation of Digital Objects (iPRES)*, Toronto, Canada, octobre 2012, p. 1-8

<https://ipres.ischool.utoronto.ca/sites/ipres.ischool.utoronto.ca/files/iPres%202012%20Conference%20Proceedings%20Final.pdf>

*D-Lib Magazine*, volume 19, numéro 1/2, janv.-fév. 2013, National Library of Australia Doi :10.1045/january2013-webb

Dossier « Archivage numérique pérenne », *La Gazette du CINES*, février 2013

FAUDUET Louise, « La conservation numérique. Enjeux, risques, solutions », Service de la conservation, Bibliothèque nationale de France

[www.bnf.fr/documents/conservation\\_numerique.pdf](http://www.bnf.fr/documents/conservation_numerique.pdf)

FITZPATRICK Kathleen, *Planned Obsolescence : Publishing, Technology, and the Future of the Academy*, NYU Press, 2009

LEBLOND Corinne (dir.), *Archivage et stockage pérenne*, Lavoisier, 2009

MASSOL Marion, « L'archivage pérenne », in « Littérature "grise" : de l'ombre à la lumière », *I2D*, Volume 52, n° 1, mars 2015, p. 62-63

MORLEY Chantal, BIA FIGUEIREDO Marie, GILLETTE Yves, *Processus métiers et systèmes d'information : gouvernance, management, modélisation*, Paris, France : Dunod, 2011

*Pérenniser le document numérique : séminaire INRIA*, 2-6 octobre 2006, ADBS, 2006

WEBB Colin, PEARSON David, KOERBIN Paul, « Oh, you wanted us to preserve that?! Statements of Preservation Intent for the National Library of Australia's Digital Collections », 2013

<http://www.dlib.org/dlib/january13/webb/01webb.html>

## Archives

BÉCHARD Lorène, FUENTES HASHIMOTO Lourdes, VASSEUR Édouard, *Les archives électroniques*, Paris : Association des archivistes français, 2014

*La gestion des archives au sein d'un établissement d'enseignement supérieur et de recherche*, Les dossiers de l'agence, AMUE, 2010

UNIVERSITÉ DE LAUSANNE, *Les archives dans l'université : actes du colloque international, Université de Lausanne, 30 et 31 octobre 2008*, Olivier Robert (dir.), Bern, Suisse, Allemagne, Belgique 2009

### **Archives ouvertes**

SCHÖPFEL Joachim, PROST Hélène, *Développement et Usage des Archives Ouvertes en France*, Université Charles-de-Gaulle Lille 3, Laboratoire GERIICO, 2010

[http://archivesic.ccsd.cnrs.fr/sic\\_00497389/document](http://archivesic.ccsd.cnrs.fr/sic_00497389/document)

SCHÖPFEL Joachim, STOCK Christiane, « Les archives ouvertes en France », *Distances et savoirs* 3/2009 (Vol. 7), p. 443-456

[www.cairn.info/revue-distances-et-savoirs-2009-3-page-443.htm](http://www.cairn.info/revue-distances-et-savoirs-2009-3-page-443.htm)

### **Bibliothèques numériques**

ANDRO Mathieu, ASSELIN Emmanuelle, MAISONNEUVE Marc, *Bibliothèques numériques : logiciels et plateformes*, Tosca consultants, Paris : ADBS Éditions, 2012

CLAERR, Thierry, WESTEEL, Isabelle (dir.), 2010. *Numériser et mettre en ligne*, Villeurbanne : ENSSIB

LE FOLL Anne, *Une offre d'e-books à la Bibliothèque Sainte-Geneviève. Dossier d'aide à la décision*, sous la direction de Benoît Epron et Franck Smith, mémoire d'étude Enssib, 2009

### **Données de la recherche**

GAILLARD Rémi, *De l'Open data à l'Open research data : quelle(s) politique(s) pour les données de recherche ?*, Diplôme de conservateur de bibliothèque (DCB), sous la direction de Monique Joly, 2014, Enssib

GÉROUDET Marie-Madeleine, « Sur le terrain des données de la recherche à Lille 1 », *Arabesques* n° 73 janvier-février-mars 2014, p. 14-15

GUINDON Alex, « La gestion des données de recherche en bibliothèque universitaire », *Documentation et bibliothèques*, vol. 59, n° 4, 2013, p. 189-200

JACOBSON Michel, LARROUSSE Nicolas, MASSOL Marion, « La question de l'archivage des données de la recherche en SHS (Sciences Humaines et Sociales) », *Archives et données de la recherche (ICA/SUV 2014)*, juillet 2014

LARROUSSE Nicolas, « Le projet d'archivage des données exprimées en TEI & en 3D », *Journée des Utilisateurs du CINES*, 2013

LERU, « LERU Roadmap for Research Data –FOSTER », *Advice Paper*, n. 14, novembre 2013

<https://www.yumpu.com/en/document/view/36353062/1kdbrob/27>

LORD Philip, MACDONALD Alison, « Data curation for e-Science in the UK : an audit to establish requirements for future curation and provision », *The JISC Committee for the Support of Research (JCSR)*, 2003

[http://www.nesc.ac.uk/news/articles\\_and\\_calls/290503.html](http://www.nesc.ac.uk/news/articles_and_calls/290503.html)

POUYLLAU Stéphane (TGIR Huma-Num), *Les Digital Humanities*, Open Access Week, UPMC, Paris, 2013

[https://scd-collections.u-strasbg.fr/wiki/aoc/documentation\\_generale](https://scd-collections.u-strasbg.fr/wiki/aoc/documentation_generale)

Records management is the new black, « Les données de la recherche sont des archives comme les autres », 24 septembre 2013

<https://lotteauxfraises.wordpress.com/2013/09/24/les-donnees-de-la-recherche-sont-des-archives-comme-les-autres-2/>

Site d'information sur les Données de la Recherche : <http://www.donneesdelarecherche.fr/spip.php?rubrique15>

### **Infrastructures et institutions**

LADANT Julie, « La conservation pérenne des collections de l'ESR : enjeux, actions et perspectives au niveau national », Journées professionnelles du CTLes, vendredi 21 juin 2013

<https://www.ctles.fr/sites/default/files/ctles/.../JPro2013Ladant.pdf>

SCHÖPFEL Joachim, « L'Atelier National de Reproduction des Thèses en cinq questions », *Arabesques*, 2013, p.22-23

THIBAUT François et JOUVE Bertrand, « Les Infrastructures de recherche en sciences humaines et sociales. Rapport du groupe "Infrastructures" de l'Alliance Athéna [Alliance nationale des sciences humaines et sociales] », p. 14-15

<http://www.cnrs.fr/fr/pdf/allianceathena/121012-infrastructure-sh/>

### **Commissions et rapports**

CARBONE Pierre, AROT Dominique, CLAUD Joëlle, LECOQ Benoît, MICOL Charles, « Quels emplois dans les bibliothèques ? État des lieux et perspectives », rapport de l'Inspection générale des bibliothèques, mars 2013

Commission Bibliothèques numériques, Rapport 2012-2013

<http://www.enseignementsup-recherche.gouv.fr/cid75370/commission-bibliotheques-numeriques.html>

« L'autonomie financière des universités : une réforme à poursuivre », Communication à la commission des finances du Sénat, Juin 2015

<https://www.ccomptes.fr/.../20150930-autonomie-financiere-universites.pdf>

LECOQ Benoît, MICOL Charles, « Le stockage des collections imprimées à l'heure du numérique », rapport n° 2014-029 de l'Inspection générale des bibliothèques, octobre 2014

« Politiques de sites et IST : les préconisations ADBU - EPRIST », ADBU – Association des directeurs et des personnels de direction des bibliothèques universitaires

<http://adbu.fr/politiques-de-sites-et-ist-les-preconisations-adbu-eprist/>

RICHARD Hélène, « La formation aux questions patrimoniales dans les bibliothèques. Rapport à monsieur le ministre de la Culture et de la Communication », rapport de l'Inspection générale des bibliothèques, 2010

ROUCHON Olivier, « Le modèle d'évaluation des coûts de l'archivage numérique », CINES, Réunion plénière, 14 Mars 2013

[http://pin.association-aristote.fr/doku.php/public/reunion\\_pleniere/cr/cr\\_20130412](http://pin.association-aristote.fr/doku.php/public/reunion_pleniere/cr/cr_20130412)

RACINE Bruno, « Schéma numérique des bibliothèques », rapport remis le 22 mars 2010 à M. Frédéric Mitterrand, Ministre de la culture et de la communication, élaboré dans le cadre du Conseil du Livre

<http://www.ladocumentationfrancaise.fr/rapports-publics/104000143/index.shtml>

« The STM Report : An overview of scientific and scholarly journal publishing », Fourth Edition, avril 2015

[www.stm-assoc.org/2015\\_02\\_20\\_STM\\_Report\\_2015.pdf](http://www.stm-assoc.org/2015_02_20_STM_Report_2015.pdf)

### **Sites web (dernière date de consultation : 30/12/2015)**

Archivage électronique au Service Interministériel des Archives de France : <http://www.archivesdefrance.culture.gouv.fr/>

Blog du CCSD : <http://blog.ccsd.cnrs.fr/2013/01/vos-fichiers-lisibles-dans-30-ans/>

BnF : <http://www.bnf.fr/fr/acc/x.accueil.html>

CINES : [www.cines.fr](http://www.cines.fr)

Conservation électronique à la Library of Congress : <http://www.digitalpreservation.gov/>

Digital Curation Centre : <http://www.dcc.ac.uk/>

Digital Preservation Europe : <http://www.digitalpreservationeurope.eu/>

Groupe PIN (Pérennisation de l'Information numérique) : <http://pin.association-aristote.fr/>

TGIR Adonis (CNRS – Universités) : <http://www.tge-adonis.fr/service/archivage-perenne>

Huma-Num : <http://www.huma-num.fr/>



# ANNEXES

---

## *Table des annexes*

<b>ANNEXE 1 : PRÉSENTATION DES PROJETS DES ÉTABLISSEMENTS ENQUÊTÉS (EN PRODUCTION, EN COURS, À L'ÉTUDE, EN STAND-BY).....</b>	<b>104</b>
<b>ANNEXE 2 : GRILLE GÉNÉRALE D'ENTRETIEN.....</b>	<b>110</b>
<b>ANNEXE 3 : INITIATIVES EUROPÉENNES.....</b>	<b>111</b>
<b>ANNEXE 4 : GISEMENTS D'INFORMATION ET VEILLE PROFESSIONNELLE .....</b>	<b>113</b>
<b>ANNEXE 5 : SOLUTIONS D'ARCHIVAGE PÉRENNE PROPRIÉTAIRES, EN OPEN SOURCE, NATIONALES ; PLATEFORMES D'ARCHIVES OPEN ACCESS .....</b>	<b>114</b>

## ANNEXE 1 : PRÉSENTATION DES PROJETS DES ÉTABLISSEMENTS ENQUÊTÉS (EN PRODUCTION, EN COURS, À L'ÉTUDE, EN STAND-BY)

### Projets en production

#### Bibliothèque inter-universitaire Cujas

Le projet de bibliothèque numérique<sup>235</sup> de la bibliothèque Cujas est lancé en 2009, et entré en production en février 2011<sup>236</sup>. La bibliothèque Cujas est la première BU française à avoir mis en place, en partenariat avec le CINES, un système d'archivage pérenne de ses documents numérisés. Le projet comprend l'archivage pérenne d'ouvrages anciens numérisés des 19<sup>ème</sup> et 20<sup>ème</sup> siècles, proposés en mode texte. Les documents plus anciens sont disponibles en mode image. La bibliothèque numérise également des documents pédagogiques (cours photocopiés de Jean Carbonnier) avec l'accord des ayants droits<sup>237</sup>. La volumétrie est de 436 titres, soit 1 500 volumes, ou 893 000 pages. Le stockage nécessaire est estimé à environ un TO par an. Début 2013, tous les documents déjà en ligne sont archivés pour un total de 3 TO archivés fin 2013. Démarche pionnière, menée en interne par les équipes de la bibliothèque, l'expérience de Cujas demeure une expérience de référence pour les BU qui souhaitent initier un projet d'archivage pérenne, d'autant plus que Noëlle Balley, Chef du département des Monographies et en charge du projet d'archivage pérenne, a diffusé en ligne le retour d'expérience de l'établissement<sup>238</sup>.

#### SCD de l'université Pierre et Marie Curie (BUPMC)

La bibliothèque numérique de la BUPMC, la Jubilotheque, a ouvert début 2007. Le projet de numérisation, initié dans le cadre du contrat quadriennal 2005-2008 de l'Université Pierre et Marie Curie par la Bibliothèque universitaire Pierre et Marie Curie, en coopération avec certaines bibliothèques de laboratoire de l'université, comprend principalement « des monographies imprimées en langue française, quelques tirés à part et des manuscrits ou ouvrages ronéotypés, s'échelonnant entre la fin du XVIII<sup>ème</sup> siècle et le début du XX<sup>ème</sup> siècle »<sup>239</sup>. Fin 2009, un projet a posteriori d'archivage pérenne est initié ; une réunion de lancement du projet a lieu en janvier 2010. En septembre 2010, une convention avec le CINES est signée, et le premier document est archivé en février 2011. La bibliothèque numérique comprenait en 2013 plus de 380 000 pages numérisées. En 2015, la volumétrie est encore inférieure à un TO.

#### Bibliothèque Sainte-Geneviève (BSG)

Le programme de numérisation patrimoniale lancé par la Bibliothèque Sainte-Geneviève a été mis en place en 2010 et répondait à un double objectif de diffusion et de préservation du patrimoine. Dès l'origine, une conservation numérique pérenne est prévue. La Bibliothèque Sainte-Geneviève offre sur sa

---

<sup>235</sup> Bibliothèque Cujas, « Bibliothèque numérique » [En ligne] (page consultée le 30/12/2015) <http://cujasweb.univ-paris1.fr/>

<sup>236</sup> Sur ce projet, Noëlle Balley, *op. cit.*

<sup>237</sup> Sur les corpus numérisés, voir le site de la bibliothèque Cujas, « Numérisation du fonds ancien » [En ligne] (page consultée le 30/12/2015) <http://bcujas-cms.univ-paris1.fr/fr/node/206>

<sup>238</sup> BALLEY Noëlle, *op. cit.*

<sup>239</sup> Site de la BUPMC [En ligne] (page consultée le 30/12/2015) <http://jubilotheque.upmc.fr/pages/about.html>

bibliothèque numérique<sup>240</sup> 300 incunables, 250 récits de voyages nordiques, 400 livres rares du 19<sup>e</sup> siècle, 900 factums (dans le cadre de l'appel à initiatives pour la numérisation en sciences juridiques BnF / Cujas), 5 périodiques en histoire de l'art (dans le cadre de l'appel à initiatives pour la numérisation en art BnF / INHA), les manuscrits et cartes du fonds Dezos de la Roquette, un échantillon de son fonds d'estampes, et une collection de reliures estampées à froid (12<sup>ème</sup>-18<sup>ème</sup> siècle). Au total, 1120 documents patrimoniaux sont numérisés par la société Diadéis, pour un total de 5 TO, et tous sont archivés au CINES (la collection de reliures estampées à froid constitue un rétrospectif à traiter). La base Liber Floridus, qui présente les enluminures issues de manuscrits à décor, est gérée par l'IRHT dans le cadre d'un partenariat (voir I), et également archivée au CINES par l'IRHT. L'ensemble de ces ressources est accessible, outre le site de la bibliothèque, sur Internet Archive, E-corpus et Liber Floridus, aujourd'hui catalogue Initiale. La Bibliothèque Sainte-Geneviève a fait le choix d'une sous-traitance pour opérer l'archivage pérenne au CINES.

### **Bibliothèque inter-universitaire de santé Paris Descartes (BIUS)**

Lancée à l'automne 2000, Medic@<sup>241</sup> est la bibliothèque numérique réalisée par le Service d'histoire de la BIU Santé, qui vise à proposer un accès à des ressources numériques intéressant l'histoire de la médecine, de l'odontologie et de la pharmacie. Sont numérisés des monographies, revues, articles de revues, thèses, dictionnaires, rapports, manuscrits. Quelques documents fragilisés, en mauvais état physique, sont numérisés afin de bénéficier d'un support de substitution adéquat. Depuis 2011, la totalité des documents en ligne est archivée de façon pérenne au CINES.

### **Bibliothèque de Science Po**

La Direction des Ressources et de l'Information Scientifique (DRIS) de Sciences Po est à l'origine de ce projet pour la Fondation (FNSP). Le service Numérisation est en charge du projet. Selon Olessea Dubois, responsable du service, l'archivage pérenne a été conçu très tôt, comme « une suite dans la chaîne de numérisation », qui permette de lancer la validation interne des documents pour contrôler leur cohérence et la compatibilité des formats. Le CINES a été retenu, même si le choix initial était la BnF. Des fonds numérisés de monographies, des archives électorales, et des dossiers de presse seront archivés au CINES ; les archives des centres de recherches de l'université feront également l'objet d'un archivage pérenne. La mise en production est prévue pour décembre 2015.

### **Projets en cours**

#### **SCD de l'université de Lille 2**

Inaugurée en mai 2014, Pépite est une plateforme de diffusion en ligne destinée à accueillir toutes les thèses de doctorat (la plateforme moissonne celles qui sont déposées dans STAR) et d'exercice (médecine, pharmacie, chirurgie dentaire) de Lille 2, des ressources pédagogiques (cours, conférences, tutoriels), et

<sup>240</sup> Site de la bibliothèque Sainte-Geneviève, « Mieux connaître la bibliothèque numérique » [En ligne] (page consultée le 30/12/2015) <http://www-bsg.univ-paris1.fr/ressources/bibliotheque-numerique-de-la-bsg>

<sup>241</sup> Site de la BIUS [En ligne] (page consultée le 30/12/2015) <http://www.biusante.parisdescartes.fr/histoire/medica/presentation.php>. Voir l'article de Jean-François Vincent : « Les 12 ans de Medic@ », *Arabesques*, janv.-fév.-mars 2013, p. 9.

des mémoires d'orthophonie<sup>242</sup>. Pépite répond entre autres aux nécessités de diffuser les thèses en accès restreint non diffusées dans HAL, et de conserver les thèses d'exercice, qui ne sont plus déposées à Lille 2 sous format papier depuis 2014 et qui ne rentrent pas dans le circuit de STAR. L'archivage pérenne au CINES a été conçu dès le départ comme un prolongement naturel de Pépite. Ne sont archivées cependant au CINES que les thèses et mémoires de Pépite, et non les ressources pédagogiques. Lille 2 est la première université à ainsi archiver des thèses au CINES. Les tests sont actuellement en cours ; la mise en production est prévue pour 2015, et une automatisation des processus pour 2016.

### **SCD de l'université de Lorraine**

Une convention signée avec la BnF, stipulant l'obligation d'un archivage pérenne en échange d'une contribution pour assurer la numérisation des fonds patrimoniaux<sup>243</sup> de l'université, initie un projet d'archivage au SCD. Parallèlement, le PRES de l'Université de Lorraine lance en 2011 une plateforme de diffusion des thèses et mémoires numériques, PETALE<sup>244</sup> (Publications Et Travaux Académiques de Lorraine), qui propose plus de 1500 thèses de doctorat et environ 200 thèses d'exercice, plus de 1 000 documents venant alimenter PÉTALE chaque année (thèses de doctorat, HDR, thèses d'exercice et mémoires), auxquels s'ajoutent plusieurs centaines d'anciennes thèses qui font l'objet de programmes de numérisation. En 2013 est lancé par la direction du SCD le projet d'un archivage pérenne au CINES concernant à la fois les fonds patrimoniaux et les documents numériques produits par la direction de la documentation et de l'édition (thèses, mémoires et livres anciens numérisés). Nommé « Archivage des fonds numériques de l'université de Lorraine », il a été conçu comme un projet large, en capacité d'englober dans le futur l'archivage de ressources d'autres composantes de l'université, et excédant le périmètre du seul SCD (archives de l'université, données de la recherche, etc.). La mise en production de l'archivage au CINES est prévue pour 2016.

### **Projets à l'étude**

#### **SCD de l'université de Lille 1**

Le projet d'archivage pérenne du SCD de Lille 1, à l'étude, englobe la bibliothèque numérique patrimoniale IRIS, proposant 2500 documents (le projet a été initié en 2007, et ne bénéficie pas encore d'archivage pérenne) ; et les thèses numérisées depuis les origines (de 1856 à 2007, les thèses postérieures à cette date étant prises en charge par l'ABES) dans le cadre d'une numérisation de masse. IRIS, bibliothèque numérique en histoire des sciences, contient « des monographies et des périodiques à partir du début du XIX<sup>e</sup> siècle, des documents libres de droits, à l'exception des collections récentes de la Société Géologique du Nord pour lesquelles les autorisations des ayants-droits ont été recueillies, et des documents conservés au Service Commun de Documentation de

<sup>242</sup> Site de l'université de Lille 2, « Une nouvelle pépite à Lille 2 : toutes les thèses accessibles en ligne » [En ligne] (page consultée le 30/12/2015) <http://www.univ-lille2.fr/actualites/detail-article/art/une-nouvelle-pepite-a-lille-2-toutes-les-theses-accessibles-en-ligne.html>

<sup>243</sup> La collection numérisée par l'université de Lorraine est constituée de 70 ouvrages anciens faisant partie d'un corpus de droit lorrain et numérisés grâce à des crédits de la BnF. De nouvelles collections patrimoniales sont appelées à être numérisées dans les prochaines années.

<sup>244</sup> Université de Lorraine [En ligne] (page consultée le 30/12/2015) <https://petale.univ-lorraine.fr/index.html>

l'Université Lille 1, issus de collections personnelles de chercheurs ou possédés par d'autres institutions et prêtés dans le cadre de conventions »<sup>245</sup>.

### **SCD de l'université de Strasbourg**

La BNUS travaille avec l'Unistra dans une coopération documentaire qui s'inscrit au sein du « Schéma directeur de la documentation universitaire pour le site alsacien » (contrat quinquennal 2013-2017). Dans ce cadre, plusieurs projets sont à l'étude, dont celui d'une bibliothèque numérique patrimoniale pour l'ensemble du site universitaire alsacien, et d'une AOI, « AOC »<sup>246</sup>, destinée à assurer l'archivage et la diffusion des publications scientifiques de l'université. Une mission d'analyse du paysage de l'archivage numérique a été envisagée, afin de déterminer la meilleure stratégie commune pour ces ressources. A l'initiative de cette mission, la direction de la BNUS, avec le soutien du comité de pilotage du schéma directeur, auquel participent tous les établissements partenaires (INSA, université de Mulhouse, etc.). De strictement patrimonial, le projet initial de valorisation et d'archivage pérenne s'est donc élargi, englobant sous la notion de patrimoine le patrimoine scientifique numérique ou numérisé produit par l'université. La mission d'étude, qui a pour vocation d'établir la typologie et la volumétrie des contenus à archiver et de dresser un panorama des solutions et des acteurs du domaine, doit être menée en 2016. Si elle voit le jour, elle devra envisager également le rôle du futur Data Center de l'établissement, en projet pour 2018, et dont les missions (initialement le calcul intensif, mais peut-être aussi l'archivage) sont encore à clarifier. Enfin, un projet d'archivage pérenne de la documentation administrative est également envisagé dans l'étude sus-citée, mais ferait l'objet d'une gestion différenciée.

### **Projets en stand-by**

#### **SCD de l'université de Lyon 1**

Le SCD a envisagé en 2012-2013 un archivage au CINES pour deux typologies différentes de documents, selon le coût et les moyens qui auraient été nécessaires : les fonds anciens numérisés uniquement, ou les fonds anciens numérisés et les productions universitaires (mémoires, thèses d'exercice, thèses intranet), pour un total d'environ 2 TO. Une étude de faisabilité préalable a été conduite. Pour des raisons de coûts, cet archivage pérenne est actuellement différé. Un projet mutualisé entre la DSI de l'université, la direction des affaires juridiques, et le SCD est à l'étude. La DSI prendrait ainsi en charge « les éventuels besoins de conversion en masse des fichiers pour assurer la pérennité de leur conservation » selon François-Xavier Boffy, responsable du Service Informatique Documentaire du SCD (Université Claude Bernard Lyon 1 - BU Sciences), et ne retiendrait que les formats pérennes.

#### **SCD de l'université Clermont-Université**

<sup>245</sup> Université de Lille 1, « IRIS » [En ligne] (page consultée le 30/12/2015) <http://iris.univ-lille1.fr/>

<sup>246</sup> Le projet, en cours, d'archives ouvertes institutionnelles AOC, « Archives ouvertes de la connaissance », est porté politiquement par l'UHA et l'Unistra (universités de Strasbourg et de Mulhouse (UHA), l'INSA et la BNU (Bibliothèque nationale et universitaire) ; il vise à rassembler sur un même portail l'ensemble de la production scientifique des établissements. Les objectifs sont ainsi formulés : « Regrouper, diffuser, valoriser, *archiver de manière pérenne*, et protéger la production scientifique du site », avec l'avantage pour les chercheurs de « garantir la pérennité des fichiers et des accès ». Une connexion est prévue à HAL pour une partie des publications. L'archivage pérenne de l'ensemble des contenus de l'AO a été prévu dès le départ du projet, et est donc en cours de définition.

Le SCD de Clermont envisage très tôt l'archivage pérenne d'un corpus numérisé constitué tout d'abord de carnets de géologues, premier apport de la numérisation entreprise à l'université, et qui bénéficie d'un site dédié. Mis en ligne, le corpus est moissonné par Gallica. S'y ajoute le site EVE, constitué de dessins d'enfants réalisés dans des contextes de guerre<sup>247</sup>. La volumétrie, très faible au départ, s'accroît : s'ajoutent des volumes de droit, des planches d'herbier, et de la littérature de jeunesse de la fin du 19<sup>ème</sup> siècle. Pour des raisons de coûts, cet archivage pérenne est actuellement différé.

### Université de Toulouse

OATAO (« Open Archive Toulouse Archive Ouverte »<sup>248</sup>) s'inscrit dans le cadre du PRES « Université de Toulouse ». C'est une AOI créée et gérée par des professionnels de la documentation de plusieurs établissements de l'université de Toulouse, sur une impulsion initiale du SCD de l'INPT. L'INPT (Institut National Polytechnique de Toulouse), l'ENVT (École Nationale Vétérinaire de Toulouse), l'ISAE (Institut Supérieur de l'Aéronautique et de l'Espace), l'ENFA (École Nationale de Formation Agronomique) et certains laboratoires de l'université Paul Sabatier participent au projet. L'objectif est de faire connaître, valoriser et conserver la production scientifique des établissements et de permettre un suivi des publications et la mise en place d'indicateurs de la recherche. Tous les types de documents de nature universitaire (articles, conférences, livres ou chapitres de livres, thèses, HDR, rapports...) sont potentiellement concernés (à l'exception des pré-prints). OATAO est connectée à HAL ; la majorité des documents de l'AOI y est présente. Un archivage pérenne a été envisagé au CINES pour les documents d'OATAO non versés dans HAL, mais le coût et l'investissement importants pour des ressources restreintes ont découragé le projet.

### SCD de l'université d'Angers

Okina (Open Knowledge, Information, Access)<sup>249</sup>, l'AOI de l'université d'Angers, ne bénéficie pas d'archivage pérenne, malgré les interrogations des équipes du SCD qui avaient envisagé lors du projet un archivage au CINES<sup>250</sup>. Le projet d'AOI est actuellement porté par la direction du numérique. « La question de l'archivage des fichiers déposés dans Okina devrait se reposer prochainement » selon Stéphanie Bouvier, chargée de mission Open access.

### SCD de l'université de Toulon

Un projet d'AOI pour diffuser les productions de la recherche de l'université est lancé fin 2005 au SCD de Toulon, afin de compléter le portail, alors limité aux documents pédagogiques. Dans le cadre d'un établissement de petite taille, le choix est fait de se tourner vers les outils nationaux ; le portail HAL Utoulon<sup>251</sup> est ainsi créé, qui couvre l'essentiel des besoins d'archivage pérenne du SCD. Un autre mode d'archivage pérenne n'est pas envisagé.

<sup>247</sup> Bibliothèque Clermont Université, « Collections numérisées » [En ligne] (page consultée le 30/12/2015) <http://bibliotheque.clermont-universite.fr/ressources-numeriques?q=node/240>

<sup>248</sup> Sur OATAO, voir Sandrine Malotaux, « Pour des bibliothèques engagées dans la diffusion des savoirs de l'université », *Bulletin des bibliothèques de France* [en ligne], n° 1, 2011 [consulté le 30 novembre 2015] <http://bbf.Enssib.fr/consulter/bbf-2011-01-0054-011>

<sup>249</sup> Université d'Angers [En ligne] (page consultée le 30/12/2015) <http://okina.univ-angers.fr/>

<sup>250</sup> Université d'Angers, blog « Projet AO » [En ligne] (page consultée le 30/12/2015) [http://blog.univ-angers.fr/projetao/#.Vk4G\\_tLRZdg](http://blog.univ-angers.fr/projetao/#.Vk4G_tLRZdg)

<sup>251</sup> HAL, Archive ouverte HAL Toulon [En ligne] (page consultée le 30/12/2015) <https://hal-univ-tln.archives-ouvertes.fr/>

### SCD de l'université de Lille 3 et de Rouen

NordNum<sup>252</sup>, bibliothèque numérique de l'Université Lille 3 consacrée à l'histoire de la région Nord-Pas-de-Calais, donne accès à plus de 500 documents numérisés (livres et revues) du 19<sup>ème</sup> et du début du 20<sup>ème</sup> siècle. Projet du SCD, NordNum date de 2001, « époque où les questions d'archivage pérenne étaient encore très peu présentes et où l'on parlait davantage de stockage et de conservation. La question de l'archivage pérenne pour NordNum est à construire » selon Isabelle Westeel, directrice de la BU de Lille 3.

Notons qu'on retrouve la même problématique pour la BU de Rouen (bibliothèque numérique consacrée aux écrits de Gustave Flaubert, gérée entre autres par le SCD de l'université de Rouen, et qui a ouvert en 2001<sup>253</sup>).

Dans le cas de PôLib, bibliothèque numérique qui contient des ouvrages du 16<sup>ème</sup> au début du 19<sup>ème</sup> siècle, dont la plupart sont conservés dans la Réserve patrimoniale commune des universités Lille 1, Lille 2 et Lille 3<sup>254</sup>, Isabelle Westeel note qu'« il a été envisagé au moment du dossier de numérisation des ouvrages en sciences juridiques en lien avec la BnF de travailler avec le CINES sur un archivage pérenne. Il a été décidé de mettre fin au projet PôLib avant la réalisation de ce dossier. »

---

<sup>252</sup> Université de Lille 3, Nordnum [En ligne] (page consultée le 30/12/2015) <http://nordnum.univ-lille3.fr/>

<sup>253</sup> Centre Flaubert CEREdI [En ligne] (page consultée le 30/12/2015) <http://flaubert.univ-rouen.fr/bibliotheque/>

<sup>254</sup> POLIB [En ligne] (page consultée le 30/12/2015) <http://polib.univ-lille3.fr/?q=fr/Presentation>

## ANNEXE 2 : GRILLE GÉNÉRALE D'ENTRETIEN

Les entretiens menés avec les professionnels ont abordés les points suivants, selon les spécificités des établissements :

1) Qui est à l'origine du projet ? Quels étaient initialement les objectifs ? Le produit final correspond-il aux attentes initiales, ou avez-vous dû renoncer à une part du projet ? Est-ce un projet qui a reçu le soutien/une impulsion de la direction de l'université ? Des financements particuliers ?

2) Quelles solutions avez-vous envisagées pour l'archivage pérenne de vos collections ? (hébergement sur les serveurs de l'université, portail HAL, CINES, prestataires externes, BnF, autre ?) Quelle solution a été retenue, et pourquoi ? D'autres ressources de l'université font-elles également l'objet d'un archivage pérenne ?

3) Avez-vous bénéficié de retours d'expérience ? Vous êtes-vous appuyé sur un réseau ? Le projet s'est-il intégré dans l'environnement numérique local ou régional ? Avez-vous connaissance de projets mutualisés d'archivage pérenne dans votre COMUE / PRES... ? Avez-vous collaboré avec d'autres services de l'université (archives...) ?

4) Quelle est l'équipe qui a porté le projet ? (ETP, postes d'informaticiens, compétences requises...). Quels sont les rapports entre la BU et la DSI en termes de répartition des missions et des tâches ? A la DSI, quels sont les personnels dédiés au projet, et s'il est opérationnel, à la maintenance de la plateforme et aux relations avec le tiers archiveur pour la gestion de l'archivage pérenne ?

5) Quelles sont la typologie et la volumétrie des documents/données retenu(e)s pour l'archivage pérenne ?

6) Quel est le coût du projet ? Quels sont les coûts de l'hébergement auprès de votre tiers archiveur ? Vous ont-ils parus raisonnables, et supportables sur le (très) long terme ? La création de cette ligne budgétaire a-t-elle posé problème ?

7) Combien de temps le projet a-t-il nécessité jusqu'à sa mise en production ? L'archivage pérenne a-t-il été pris en compte dès le début du projet de numérisation ? Avez-vous par exemple établi des critères précis de nommage des fichiers en amont du projet afin de faciliter l'archivage pérenne ? Utilisez-vous un logiciel de numérisation qui prenne en compte cette problématique ?

8) La qualification de vos équipes en matière informatique a-t-elle posé problème ? Avez-vous eu recours à des formations ?

9) Le projet d'archivage pérenne vous a-t-il semblé très technique, chronophage... ? A quels obstacles les équipes se sont-elles heurtées ? Certaines étapes sont-elles plus particulièrement difficiles ?

### ANNEXE 3 : INITIATIVES EUROPÉENNES

Le programme de financement Européen H2020<sup>255</sup> demande dorénavant à chaque projet de recherche financé de décrire le devenir des données produites durant son déroulement. Il comporte également l'obligation d'assurer le libre accès aux publications issues des recherches qu'il aura contribuées à financer.

Le projet SUCCEED<sup>256</sup>, qui est financé par la Commission européenne sous le thème FP7-ICT, préconise quant à lui des formats de données et de métadonnées (« Recommendations for digitisation projects »), en accord par exemple avec le JISC, et envisage entre autres dans sa réflexion la préservation sur le long terme (« long-term preservation »), en proposant des recommandations pour l'archivage pérenne (« Recommendations for long-term preservation »).

L'infrastructure DARIAH-EU<sup>257</sup>, « Digital Research Infrastructure for the Arts and Humanities » est un ERIC, « European Research Infrastructure Consortium ». Le cadre juridique applicable aux ERIC (European Research Infrastructure Consortium) date de 2009. Il a été conçu pour faciliter la mise en place d'infrastructures européennes de recherche. « L'objectif de DARIAH est de développer et de soutenir la recherche dans toutes les disciplines des SHS qui travaillent sur des objets tels que les textes, les sons, les images, la vidéo. Elle offre aux communautés de recherche un ensemble de services et souhaite jouer un rôle phare dans le développement des collaborations internationales. La création de l'ERIC DARIAH représente une étape majeure, qui va participer à la visibilité du travail en SHS à long terme, l'ERIC étant établi pour vingt ans »<sup>258</sup>. La France fait partie des membres fondateurs, et accueille le siège social du Consortium.

La LERU<sup>259</sup> (Ligue universitaire de recherche européenne, dont sont membres en France l'UMPC, Paris Sud et l'Unistra) joue également un rôle d'incitation ; très active, entre autres par son activité d'advocacy auprès de la Commission européenne, elle encourage les politiques d'open access ou à la prise en compte des données de la recherche, et sensibilise les établissements à ces questions.

Le Projet PÉRICLÈS<sup>260</sup> (Promoting and Enhancing Reuse of Information throughout the Content Lifecycle taking account of Evolving Semantics) est un projet lancé sur des fonds européens (2013-2017) qui vise à préserver les données scientifiques afin de les rendre accessibles et réutilisables par la recherche en s'appuyant sur les fruits de la recherche sémantique. Des ateliers (« workshops ») sont organisés ; des projets s'attachent dans ce cadre à intégrer la préservation sur le long terme au management numérique des archives digitales, des collections en ligne et des dépôts de données numériques.

<sup>255</sup> MENESR, « H2020 » [En ligne] (page consultée le 31/12/2015) <http://www.horizon2020.gouv.fr/>

<sup>256</sup> SUCCEED [En ligne] (page consultée le 31/12/2015) <http://www.succeed-project.eu/?q=node/1> ; <http://www.digitisation.eu/training/recommendations-for-digitisation-projects/>

<sup>257</sup> « Thereby it aims to enhance data quality, preservation, and deep interoperability, as well as furthering a culture of data sharing in the Arts and Humanities ». DARIAH [En ligne] (page consultée le 31/12/2015) <http://dariah.eu/>

<sup>258</sup> CNRS, « Création de l'ERIC DARIAH pour les humanités numériques » [En ligne] (page consultée le 31/12/2015) <http://www.cnrs.fr/inshs/recherche/creation-eric-dariah.htm>

<sup>259</sup> LERU [En ligne] (page consultée le 31/12/2015) <http://www.leru.org/index.php/public/home/>

<sup>260</sup> PERICLES, « Publications » [En ligne] (page consultée le 31/12/2015) <http://pericles-project.eu/publications>

Le consortium européen PREFORMA <sup>261</sup>, PREservation FORMAts for culture information/e-archives, est un projet co-fondé par la Commission européenne sous le Programme FP7-ICT; lancé en 2014, il a pour but de favoriser la standardisation de bonne qualité des formats de préservation de données sur le long terme, en aidant les institutions à avoir la pleine maîtrise des tests de conformité des documents ayant pour vocation à intégrer leurs archives.

L'Open Preservation Foundation <sup>262</sup>, fondée en 2010 en tant que Open Planets Foundation ayant pour vocation de soutenir les résultats de la R&D menée sur des fonds européens, promeut les technologies et les compétences nécessaires à la gestion sur le long terme de l'« héritage culturel numérique » (« long-term management of digital cultural heritage ») ; la fondation, dont sont membres entre autres la BnF, la British Library et le JISC propose un lieu d'échanges de connaissances et de compétences, des bonnes pratiques....

Le projet européen Eudat <sup>263</sup> (European Data Infrastructure) a pour objectif de mettre en place une infrastructure européenne d'échange et de conservation de données, en travaillant avec les différentes communautés scientifiques structurées au niveau européen. Il est cofinancé par le 7e Plan de la Communauté Européenne. L'objectif de ce projet est de répondre aux besoins futurs des chercheurs en matière d'accès et de préservation de données scientifiques. Les partenaires du projet vont recenser les besoins et exigences des communautés de recherche afin de « développer, déployer et sécuriser une infrastructure commune et persistante de données, la « Collaborative Data Infrastructure » (CDI) » <sup>264</sup>. UDAT comprend actuellement 25 partenaires. Il travaille particulièrement avec 5 communautés scientifiques (CLARIN, EPOS, ENES, Life Watch, VPH) et 12 grands centres de données, dont le CINES, qui est devenu « l'unique nœud français de l'infrastructure de données européenne mise en place dans le cadre du projet EUDAT » <sup>265</sup>.

L'organisation European Alliance for Permanent Access <sup>266</sup>, créée en 2008, rassemble en Europe des laboratoires de recherche, des bibliothèques et des éditeurs ; elle a pour vocation de soutenir la préservation numérique et l'accès aux ressources digitales.

---

<sup>261</sup> PREFORMA [En ligne] (page consultée le 31/12/2015) <http://www.preforma-project.eu/>

<sup>262</sup> OPENPRESERVATION.ORG [En ligne] (page consultée le 31/12/2015) <http://openpreservation.org/>

<sup>263</sup> EUDAT [En ligne] (page consultée le 31/12/2015) <http://www.eudat.eu>

<sup>264</sup> CINES, « Le projet européen » [En ligne] (page consultée le 31/12/2015) <https://www.cines.fr/archivage/le-projet-europeen-eudat/>

<sup>265</sup> *ibid.*

<sup>266</sup> PRELIDA [En ligne] (page consultée le 31/12/2015) <http://www.prelida.eu/partners/apa-alliance-permanent-access>

## ANNEXE 4 : GISEMENTS D'INFORMATION ET VEILLE PROFESSIONNELLE

De nombreux gisements d'information sont disponibles pour les professionnels sur l'archivage pérenne, et proviennent de divers organismes qui rendent accessibles en ligne leurs bonnes pratiques ou recommandations. On recensera, parmi d'autres :

- les Conférences annuelles sur la préservation numérique de l'iPRES<sup>267</sup> ;
- les blogs sur la préservation numérique de la bibliothèque de la bibliothèque du Congrès<sup>268</sup>, qui est pilote du National Digital Information Infrastructure and Preservation Program<sup>269</sup> ;
- les ressources sur le site du CINES<sup>270</sup> ;
- les pages pour les professionnels sur le site de la BnF<sup>271</sup> ;
- certains blogs tenus par des professionnels de l'information tels ArchivEngines<sup>272</sup> ;
- les nombreux groupes de travail européens évoqués ci-dessus.

Certains sites proposent des veilles sur des types de ressources particulières :

- sur les archives de l'Internet par exemple, le site de l'IIPC (International Internet Preservation Consortium)<sup>273</sup> ;
- pour les archives audiovisuelles, l'organisation Prestocentre, qui s'adresse entre autres aux universités et aux bibliothèques nationales<sup>274</sup>.

<sup>267</sup> IPRES [En ligne] (page consultée le 31/12/2015) <http://ipres2015.web.unc.edu/>

<sup>268</sup> Library of Congress [En ligne] (page consultée le 31/12/2015) <http://blogs.loc.gov/digitalpreservation> et <http://www.digitalpreservation.gov/formats/> .

<sup>269</sup> Id., <http://www.digitalpreservation.gov/>

<sup>270</sup> CINES [En ligne] (page consultée le 31/12/2015) <http://www.cines.fr/spip.php?rubrique219>

<sup>271</sup> BnF, « La préservation à l'heure du numérique » [En ligne] (page consultée le 31/12/2015) [http://www.bnf.fr/fr/professionnels/innov\\_num\\_preservation\\_numerique.html](http://www.bnf.fr/fr/professionnels/innov_num_preservation_numerique.html)

<sup>272</sup> Blog rédigé par Pierre Couchet, ingénieur d'étude au CNRS. ArchivEngines [En ligne] (page consultée le 31/12/2015) <https://archivengines.wordpress.com/page/2/>

<sup>273</sup> IIPC [En ligne] (page consultée le 31/12/2015) <http://netpreserve.org/>

<sup>274</sup> PRESTOCENTRE [En ligne] (page consultée le 31/12/2015) <http://prestocentre.eu/>

## ANNEXE 5 : SOLUTIONS PROPRIÉTAIRES, EN OPEN SOURCE, NATIONALES ; PLATEFORMES D'ARCHIVES OPEN ACCESS

Les logiciels de préservation numérique propriétaires existent depuis plusieurs années. DIAS (IBM)<sup>275</sup> a été utilisé dès 2003 par e-Depot et KOPAL. La solution de Tessella, la Safety Deposit Box, est utilisée par les Archives nationales du Royaume-Uni. Rosetta<sup>276</sup> (ExLibris) a été développée en collaboration avec la Bibliothèque nationale de Nouvelle-Zélande et mis en production en 2009, et est utilisée par des bibliothèques, archives et d'autres grandes organisations en Australie, Europe (ETH-Zürich, Bibliothèque Nationale de Bavière, KULeuven) et Amérique du Nord.

« Les logiciels de ce type proposent des fonctionnalités larges et largement paramétrables, destinées à préserver de grandes collections d'objets numériques de différents formats et structures, et sont dès lors très adaptés pour des organisations telles que les bibliothèques nationales, ou des collections numérisées. Cette robustesse demande en contrepartie un investissement financier non négligeable. Par ailleurs, il s'agit d'outsourcing à un logiciel propriétaire, ce qui, dans le cas de la préservation pérenne, constitue un risque ».<sup>277</sup>

Les logiciels de préservation numérique Open Source se sont développés récemment, tels Archivematica, Preservica et DAITSS.

Le logiciel Archivematica<sup>278</sup> est développé par Artefactual System, en collaboration avec le sous-comité Technologie de l'Unesco, et différentes grandes organisations canadiennes et américaines du monde des bibliothèques, archives et musées (dont le MoMa et plusieurs bibliothèques universitaires). Archivematica s'adresse surtout à des archivistes et bibliothécaires ayant des moyens limités, et fournit aux institutions un outil Open Source qui respecte l'OAIS, et utilise les standards METS, PREMIS, Dublin Core.

Preservica<sup>279</sup>, surtout utilisé pour la gestion d'archives (la version 1 a été créée pour les Archives nationales de Grande-Bretagne en 2003), et également fondé sur l'OAIS, offre une version Cloud.

DAITSS<sup>280</sup> est développé par le Florida Center for Library Automation (FCLA) et utilisée par la Florida Digital Archive (FDA). Dépôt numérique utilisé par les 11 universités publiques de Floride (Florida virtual campus), il est lancé en

<sup>275</sup> IBM, « DIAS » [En ligne] (page consultée le 31/12/2015) <http://www-304.ibm.com/partnerworld/gsd/solutiondetails.do?&solution=10393&lc=en>

<sup>276</sup> EXLIBRIS, « Rosetta » [En ligne] (page consultée le 31/12/2015) <http://www.exlibrisgroup.com/files/Products/Preservation/RosettalowresA4.pdf>

<sup>277</sup> ALEXANDRE Aude, *Archivage papier et/ou électronique et accès pérenne aux ressources documentaires électroniques en texte intégral en Fédération Wallonie-Bruxelles*, op. cit., p.29.

<sup>278</sup> ARCHIVEMATICA [En ligne] (page consultée le 31/12/2015) <https://www.archivematica.org/en/>

<sup>279</sup> PRESERVICA [En ligne] (page consultée le 31/12/2015) <http://preservica.com/preservica-works/>

<sup>280</sup> DIATSS [En ligne] (page consultée le 31/12/2015) <https://daitss.fcla.edu>

2005 et réécrit en 2009. Qualifié de « OAIS based preservation repository system »<sup>281</sup>, il respecte les standards OAIS et PREMIS.

« Ces logiciels ont l'avantage d'être Open Source et gratuits, mais, seuls, souffrent de la même limitation que les logiciels propriétaires : il s'agit de solutions avant tout techniques (...) »<sup>282</sup>.

**Les plateformes d'archives Open Access**, telles PubMed Central<sup>283</sup>, ArXiv<sup>284</sup> ou Persée<sup>285</sup> proposent certaines garanties d'archivage à long terme : serveurs et localisations multiples, missions établies dans la durée (PubMed Central), financement national (PubMed Central), archivage pérenne garanti (Persée, dont l'archivage pérenne est garanti par le CINES). Cependant elles ont certaines limites, en termes de couverture et surtout de contrôle : « Leurs missions ne sont pas définies par rapport aux bibliothèques, et celles-ci n'ont pas de poids dans la gouvernance ou les décisions de l'archive. Ces solutions sont donc plutôt des solutions d'accès alternatives (...) »<sup>286</sup>.

Enfin, l'archivage pérenne peut être fourni par des entités nationales, comme DANS<sup>287</sup> et le 3TUDatacenter<sup>288</sup> aux Pays-Bas. La plateforme DANS est à l'origine du DSA (Data Seal of Approval) ; les services qu'ils fournissent sont proches de ceux du CINES mais ne semblent pas ouverts à une communauté plus large que les Pays-Bas.

## LOCKSS, CLOKSS, PORTICO : des solutions avec mandat dédié

### LOCKSS

La coordination de LOCKSS (Lots Of Copies Keeps Stuff Safe) est assurée par les bibliothèques de l'université de Stanford. Le projet se définit comme une « initiative de la communauté internationale » (« international community initiative ») rassemblant des centaines de bibliothèques universitaires dans le monde. Chaque bibliothèque installe le système LOCKSS, gratuit et en disponible en Open Source, sur un poste informatique appelé « LOCKSS Box ». La LOCKSS box maintient ensuite la communication avec l'ensemble du réseau, comparant les contenus collectés avec ceux des archives des autres bibliothèques, et « réparant » toute différence ou lacune trouvée dans les fichiers. Comme le nom du projet l'indique, la redondance de ses fichiers distribués crée un filet de sécurité pour les contenus. Par ailleurs, Lockss préserve les contenus dans leur format originel et migre de façon dynamique les contenus vers des formats plus récents, quand un lecteur en demande la consultation. Cette « migration à l'accès » permet aux fichiers d'être préservés dans leur format originel de présentation – et de ce fait permet une conformité avec les exigences de l'archivage, tout en fournissant les

<sup>281</sup> Voir CAPLAN Priscilla, « DAITSS, an OAIS based preservation repository », Florida Center for Library Automation [En ligne] (page consultée le 31/12/2015) [http://daitss.fcla.edu/sites/daitss.fcla.edu/files/DAITSS%20in%20ACM%20rev\\_0.pdf](http://daitss.fcla.edu/sites/daitss.fcla.edu/files/DAITSS%20in%20ACM%20rev_0.pdf)

<sup>282</sup> ALEXANDRE Aude, *op. cit.*, p. 30.

<sup>283</sup> PMC [En ligne] (page consultée le 31/12/2015) <http://www.ncbi.nlm.nih.gov/pmc/>

<sup>284</sup> ARXIV.ORG [En ligne] (page consultée le 31/12/2015) <http://arxiv.org/>

<sup>285</sup> PERSEE [En ligne] (page consultée le 31/12/2015) <http://www.persee.fr/>

<sup>286</sup> ALEXANDRE Aude, *op. cit.*, p. 30.

<sup>287</sup> DANS [En ligne] (page consultée le 31/12/2015) <http://www.dans.knaw.nl/nl/actueel/nieuws/3tudatacentrum-en-dans-bundelen-krachten>

<sup>288</sup> 3TU.DATACENTRUM [En ligne] (page consultée le 31/12/2015) <http://datacentrum.3tu.nl/home/>

moyens de prévenir l'obsolescence technologique des formats, ce qui dégage les bibliothèques de la responsabilité d'avoir à assumer la migration. LOCKSS est un projet emblématique des *community driven preservation programs* - programmes de préservation assurés par la communauté ; le coût du hardware est faible, le software est gratuit, et le réseau assure l'auto-correction des fichiers. Le système est maintenu par la souscription des membres du réseau LOCKSS – lesquels contrairement aux utilisateurs qui ne paient pas, ont la possibilité de collecter et de préserver des contenus « premium » non accessibles au reste de la communauté LOCKSS.

### CLOKSS

L'Alliance LOCKSS a lancé un second projet coopératif nommé CLOCKSS (pour Controlled Lockss), au sein duquel un nombre réduit de bibliothèques membres archivent non seulement les périodiques auxquelles elles sont abonnées, mais aussi tous les périodiques auxquels les éditeurs autorisent l'accès, à la fois les fichiers de présentation et les fichiers sources ; l'objectif est de proposer une « dark archive » distribuée géographiquement qui puisse assurer la préservation à long terme des publications universitaires en ligne pour le bénéfice de toute la communauté des chercheurs. En cas d'« événement déclencheur » (« trigger event ») – un éditeur cessant son activité, une interruption de publication, etc. – CLOKSS rendra les archives accessibles non seulement à ses membres mais à l'ensemble de la communauté savante.

### PORTICO

PORTICO est une solution commerciale tournée vers le monde académique, qui synthétise les services du CINES et de l'ABES pour les USA. Y sont archivées beaucoup de publications scientifiques dont Jstor. Le service est payant. Projet d'Ithaka<sup>289</sup>, organisation parente de J-Stor, initialement fondé par la Mellon Foundation, PORTICO<sup>290</sup> est actuellement géré par un comité de bibliothécaires et d'éditeurs, et soutenu par les contributions des éditeurs et la cotisation annuelle des bibliothèques. PORTICO est un système centralisé qui produit des « dark archives » de périodiques électroniques ; les bibliothèques délèguent à PORTICO le soin d'assurer l'archivage des fichiers sources et de les normaliser en un format d'archivage standard permettant l'archivage pérenne. Cette migration initiale peut être suivie par d'autres migrations afin d'éviter l'obsolescence technologique des formats.

<sup>289</sup> ITHAKA [En ligne] (page consultée le 31/12/2015) <http://ithaka.org/>

<sup>290</sup> PORTICO [En ligne] (page consultée le 31/12/2015) <http://www.portico.org/digital-preservation/>

## GLOSSAIRE

---

Archives : 1) ensemble de documents produits ou reçus par une personne physique ou morale ou par un service ou un organisme dans l'exercice de son activité ; 2) organisme chargé du contrôle de la conservation, du tri, du classement, de l'inventaire et de la communication des archives courantes, intermédiaires ou définitives.

Archives définitives : archives conservées sans limitation de durée.

Émulation : imitation du comportement physique d'un matériel par un logiciel pour pouvoir exécuter un programme ou lire à nouveau un fichier.

Format ouvert : format de fichier pour lequel il existe une spécification librement accessible et sans restriction d'accès ni de mise en œuvre.

Intégrité : caractéristique d'une information qui n'a subi aucune destruction, altération ou modification intentionnelle ou accidentelle.

Migration: la migration est un transfert d'informations numérique dans un objectif de pérennisation de cette information. Elle consiste en pratique en la conversion d'un format de fichier jugé non pérenne en un autre format jugé pérenne (migration des formats), ou à transférer des données d'un support d'enregistrement vers un autre (migration des supports = rafraichissement), de façon à assurer la lisibilité des informations contenues dans le fichier. La dégradation des supports ou l'obsolescence technologique sont des facteurs incitant à procéder à une migration.

Système d'Archivage Électronique (SAE) : aussi appelé Plateforme d'archivage électronique ; système consistant à recevoir, conserver, communiquer et restituer des archives.

Versement : opération matérielle et administrative par laquelle des archives sont transférées du service qui les a rassemblées au service d'archives compétent pour les recevoir.<sup>291</sup>

---

<sup>291</sup> Définitions inspirées du lexique de *La gestion des archives au sein d'un établissement d'enseignement supérieur et de recherche*, op. cit., et des définitions de BANAT-BERGER Françoise, DUPLOUY Laurent, HUC Claude, *L'archivage numérique à long terme : les débuts de la maturité ?*, op. cit. p. 118 et suiv.











# TABLE DES MATIÈRES

---

<b>SIGLES ET ABRÉVIATIONS .....</b>	<b>9</b>
<b>INTRODUCTION.....</b>	<b>11</b>
<b>I. STRATÉGIES ET ENJEUX DE L'ARCHIVAGE PÉRENNE DANS L'ESR</b>	<b>14</b>
.....	
1. <i>Un nouvel enjeu patrimonial ? .....</i>	14
A. Un enjeu émergent, de nouveaux risques .....	14
B. Un prolongement de la vision patrimoniale ? .....	17
C. Qu'archiver de façon pérenne ? Critères de sélection et typologie des documents à pérenniser .....	18
2. <i>Le paysage institutionnel et les stratégies nationales et internationales .....</i>	24
A. Les impulsions et orientations stratégiques de l'État .....	24
B. Les opérateurs de l'État .....	27
C. Les infrastructures de recherche du CNRS : IRHT, CCSD et TGIR- Huma-Num.....	30
D. Les initiatives européennes .....	32
3. <i>Coûts et exigences techniques et réglementaires de l'archivage pérenne .....</i>	32
A. Agréments et processus, standards et normes : les exigences de l'archivage pérenne.....	32
B. Les coûts de l'archivage pérenne .....	37
<b>II. PROJETS D'ARCHIVAGE PÉRENNE EN BIBLIOTHÈQUES UNIVERSITAIRES : RETOURS D'EXPÉRIENCE .....</b>	<b>41</b>
1. <i>Méthodologie de l'enquête et bilan général .....</i>	41
2. <i>Solutions adoptées et acteurs des projets .....</i>	42
A. Les solutions d'archivage en bibliothèque universitaire .....	42
B. Les acteurs d'un projet d'archivage pérenne : des projets de coopération.....	47
3. <i>Projets d'archivage pérenne en bibliothèque universitaires : retours d'expérience.....</i>	49
A. A l'initiative du projet .....	49
B. Les retours d'expérience des professionnels .....	50
C. Volumétrie et typologie des collections retenues pour l'archivage pérenne.....	51
D. Les équipes et les profils des équipes projet en bibliothèques universitaires et dans les services informatiques .....	52
E. Le déroulement du projet .....	54
F. La gestion des routines : une fois en production, quelle gestion ?	55

4. <i>Les freins et les causes des abandons des projets</i> .....	56
A. Le coût.....	56
B. La disponibilité.....	57
C. La technicité.....	57
<b>III. SITUATION DES BIBLIOTHÈQUES UNIVERSITAIRES : VERS UNE NÉCESSAIRE MUTUALISATION</b> .....	<b>60</b>
1. <i>Situation des SCD dans l'université et dans l'ESR</i> .....	60
A. Responsabilités et périmètres au sein de l'université .....	60
B. Les circuits d'archivage des documents en SCD .....	64
2. <i>De la mutualisation à la réflexion sur les périmètres institutionnels</i> 75	
A. La mutualisation, l'horizon des projets universitaires d'archivage pérenne ?.....	75
B. Projets de coopération et solutions mutualisées .....	80
C. Les grandes orientations stratégiques nationales. Vers de nouveaux périmètres ?.....	82
3. <i>De nouvelles missions et de nouvelles compétences ?</i> .....	83
A. Le soutien aux chercheurs : anticiper l'archivage pérenne.....	84
B. Une nécessaire montée en compétences. Les impacts sur le métier .....	86
<b>CONCLUSION</b> .....	<b>91</b>
<b>SOURCES</b> .....	<b>93</b>
<b>BIBLIOGRAPHIE</b> .....	<b>97</b>
<b>TABLE DES ANNEXES</b> .....	<b>103</b>
<b>GLOSSAIRE</b> .....	<b>117</b>
<b>TABLE DES MATIÈRES</b> .....	<b>123</b>