

Diplôme national de master

Domaine - sciences humaines et sociales

Mention - sciences de l'information et des bibliothèques

Spécialité - sciences de l'information et des bibliothèques et
information scientifique et technique

Archives Ouvertes de la Connaissance. Valoriser et diffuser les données de recherche.

Violaine REBOUILLAT

Sous la direction d'Agnieszka Tona
Maître de conférences - Université Lyon 1

Remerciements

Je tiens à remercier :

Ma tutrice de stage, Adeline Rege, qui s'est montrée disponible, tout en me laissant libre dans mon travail, et dont les conseils et les pistes de réflexion m'ont été d'une grande aide dans la réalisation du stage comme dans la rédaction du mémoire, voire plus encore...,

Ma directrice de mémoire et maître de conférences à l'Université de Lyon 1, Agnieszka Tona, pour son encadrement pédagogique,

Madame Dominique Wolf, directrice du Service Commun de la Documentation de l'Université de Strasbourg, pour m'avoir permis de réaliser ce stage,

L'équipe projet des Archives Ouvertes de la Connaissance, Sylvain Joncour, Catherine Lourdel, Ana Schwartz et Catherine Storne notamment, qui m'ont accueillie avec sympathie et m'ont bien volontiers apporté leur aide,

Marie-Christine Jacquemot-Perbal et Françoise Cosserat, du Département de l'Offre de Services de l'Inist-CNRS, pour m'avoir très gentiment reçue et fait partager leur expérience,

Marie-Madeleine Géroudet, directrice de la Bibliothèque Numérique de l'Université Lille 1, et Coline Blanpain, chargée de mission sur le projet Archive Ouverte/Données de la recherche, pour m'avoir présenté leur travail,

Franck Borel, responsable du département eScience à la bibliothèque universitaire de Fribourg en Brisgau, pour sa patience dans nos échanges de mails,

Les enseignants-chercheurs et doctorants qui ont eu la gentillesse de m'accorder de leur temps, pour me présenter leurs recherches et répondre à mes questions,

Ainsi que toutes les personnes qui ont répondu à mes demandes d'informations par courriel ou par téléphone,

Les bibliothécaires du PEGE pour m'avoir accueillie dans leurs locaux durant quatre mois,

Yves pour ses relectures attentives.

Résumé : Projet commun de l'Université de Strasbourg, l'Université de Haute-Alsace, l'Institut National des Sciences Appliquées (INSA) et la Bibliothèque Nationale et Universitaire (BNU) de Strasbourg, les Archives Ouvertes de la Connaissance offriront aux (enseignants)-chercheurs et doctorants un service pour la valorisation de leurs données de recherche. Ce mémoire propose, dans un premier temps, de replacer le projet dans le contexte des archives institutionnelles françaises et européennes, afin d'en dégager les spécificités ; dans un second temps, sont présentés les enjeux et les modalités de mise en forme et de diffusion des données de recherche, que produisent les établissements alsaciens partenaires et qui seront liées à l'archive ouverte.

Descripteurs : recherche scientifique ; libre accès ; archive institutionnelle ; données de la recherche ; Archives Ouvertes de la Connaissance ;

Abstract : As a joint project of the Strasbourg University, Haute-Alsace University, National Institute of Applied Sciences and National and University Library of Strasbourg, the “Archives Ouvertes de la Connaissance” provides to the (teachers)-researchers and Ph. D. students a service to enhance their research data assets. Firstly, this training report seeks to place the project in the context of the French and European institutional repositories, in order to determine their particularities; secondly, the issues and methods of research data assets formatting and dissemination are presented. The research in question is produced by the Alsatian partner institutions, in relation with the open archive.

Keywords : scientific research ; open access ; institutional repository ; research data ; Archives Ouvertes de la Connaissance ;

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

SIGLES ET ABREVIATIONS	7
INTRODUCTION	9
LES ARCHIVES OUVERTES DE LA CONNAISSANCE DANS LE PAYSAGE INSTITUTIONNEL FRANÇAIS ET EUROPEEN DE LA RECHERCHE	11
1. Les archives ouvertes institutionnelles aujourd'hui.....	11
1.1. Comment sont-elles définies ?.....	11
1.2. Comment sont-elles encadrées ?.....	14
1.3. Quels sont les enjeux ?.....	18
1.4. Quelques exemples d'archives institutionnelles françaises.....	25
2. Les Archives Ouvertes de la Connaissance : environnement du projet	25
2.1. Description du projet	25
2.2. Racines du projet	28
2.3. Les objectifs spécifiques du projet.....	31
PRATIQUES DE GESTION DES DONNEES DE LA RECHERCHE SUR LE SITE ALSACIEN : ETUDE EXPLORATOIRE.....	37
1. L'ouverture des données de recherche.....	37
1.1. Questionnement autour de la définition des données de recherche	37
1.2. Evolution épistémologique de la recherche scientifique : de l'Open Access à l'Open Science.....	39
1.3. Des mouvements incitatifs	40
2. Enquêtes auprès des chercheurs du site alsacien	42
2.1. Méthodologie	42
2.2. Conclusions des enquêtes.....	45
3. Préconisations	49
3.1. Mettre en place un « écosystème » favorable à l'ouverture des données de la recherche	49
3.2. Etude de cas : le naturaliste Jean Hermann et la base de données prosopographique de ses réseaux.....	52
CONCLUSION.....	57
SOURCES	59
BIBLIOGRAPHIE	61
ANNEXES	65
TABLE DES ILLUSTRATIONS.....	81
TABLE DES MATIERES	83

Sigles et abréviations

ABES - Agence Bibliographique de l'Enseignement Supérieur
BOAI - Budapest Open Access Initiative
BIATSS - Bibliothèque, Ingénieurs, Administratifs, Techniciens, Social, Santé
BNUS - Bibliothèque Nationale et Universitaire de Strasbourg
BSN - Bibliothèque Scientifique Numérique
CC – Creative Commons
CCSD - Centre pour la Communication Scientifique Directe
CER - Conseil européen de la recherche
CINES - Centre Informatique National de l'Enseignement Supérieur
CNRS - Centre national de la recherche scientifique
DCC - Data Curation Centre
DOI - Digital Object Identifier
EA - Equipes d'Accueil
EER - Espace européen de la recherche
FACILE - Format d'archivage du CINES par analyse et expertise
GTAO - Groupe de Travail sur l'Accès Ouvert
H2020 - Horizon 2020
HAL - Hyper articles en ligne
HCERES - Haut Conseil de l'Évaluation de la Recherche et de l'Enseignement Supérieur
IFREMER - Institut Français de Recherche pour l'Exploitation de la Mer
INRA - Institut National de la Recherche Agronomique
INRIA - Institut National de Recherche en Informatique et en Automatique
INSAS - Institut National des Sciences Appliquées de Strasbourg
INSERM - Institut National de la Santé et de la Recherche Médicale
IST - Information scientifique et technique
LERU - League of European Research Universities
MESR - Ministère de l'Enseignement Supérieur et de la Recherche
MISHA - Maison Interuniversitaire des Sciences de l'Homme
OAI-PMH - Open Archives Initiative Protocol for Metadata Harvesting
OCDE – Organisation de Coopération et de Développement Économiques
ORI-OAI - Outil de Référencement et d'Indexation – Open Access Initiative
PURL - Persistent Uniform Resource Locators
ROAR - Registry of Open Access Repositories
SAM - Science Arts et Métiers
SCD - Service Commun de la Documentation

SHERPA/RoMEO - Publisher Copyright Policies and Self Archiving / Rights Metadata for Open archiving

SIGB - Système intégré de gestion de bibliothèque

TEL - Thèses-en-ligne

UHA - Université de Haute-Alsace

UNISTRA- Université de Strasbourg

URFIST - Unité Régionale de Formation à l'Information Scientifique Technique

URN - Uniform Resource Name

XML - Extensible Markup Language

INTRODUCTION

La *Stratégie numérique du Gouvernement*¹, présentée par le Premier Ministre, Manuel Valls, le 18 juin dernier, a confirmé l'engagement des pouvoirs publics français en faveur d'une « science ouverte », s'appuyant sur la « libre diffusion des publications et des données de la recherche ». Cette annonce tient lieu de préambule au projet de loi sur le Numérique, qui devrait être examiné par l'Assemblée nationale à l'automne 2015. Elle intervient dans un contexte international et européen, où l'ouverture des résultats de recherche est devenue un enjeu majeur. Si les questions de libre diffusion ne sont pas nouvelles au sein de la communauté scientifique et des institutions encadrant la recherche, elles trouvent aujourd'hui un appui dans l'intérêt public qui leur est porté – appui dont se saisissent les acteurs de la recherche pour concrétiser plus avant leurs objectifs.

C'est du moins cette ambition que poursuivent les universités de Strasbourg et de Haute-Alsace, l'INSA et la BNU de Strasbourg : proposer une archive institutionnelle qui s'intègre au paysage des archives ouvertes françaises et européennes, tout en développant une proposition de services adaptés aux attentes des chercheurs.

La plus-value du projet – baptisé *Archives Ouvertes de la Connaissance* – réside en effet dans une offre de proximité, comprenant un accompagnement juridique et technique au dépôt, des outils statistiques et bibliométriques, mais aussi des services liés à l'« édition » et la diffusion des documents déposés. Sur ce dernier volet, les porteurs du projet ont fait appel à une stagiaire, dans le but d'étudier les possibilités de valorisation du contenu des Archives Ouvertes de la Connaissance. La mission du stage s'articulait autour de deux principaux axes : l'un consacré aux publications scientifiques, l'autre aux données de recherche. Des scénarios ont été élaborés pour chacun.

Le premier axe consistait à proposer des outils et services d'édition électronique pouvant être connectés à l'archive ouverte. Les revues scientifiques en Sciences humaines et sociales ont fait l'objet d'une attention particulière : elles ouvriraient, par le biais du libre accès, une perspective nouvelle pour parer au manque de visibilité, dont souffrent les publications dans ce domaine.

Le second axe concernait la mise en forme des données de recherche, à partir d'outils d'« éditorialisation », permettant de lier des jeux de données aux publications et de créer des bases de données, des visualisations cartographiques,... qui seraient référencées dans l'archive ouverte.

Nos recherches nous ont d'abord conduits à nous interroger sur la manière dont les nouveaux services proposés par les archives institutionnelles s'adaptent aux enjeux de la recherche. En tant que lieu de production scientifique, les archives institutionnelles se caractérisent en effet par un panel de services, tendant à couvrir l'ensemble des résultats de recherche. Ces services concernent aussi bien l'édition de publications que le dépôt et la mise en forme des données de recherche, ainsi que le lien entre publications et données.

¹ *Stratégie numérique du gouvernement*. Gaité lyrique, 18 juin 2015. <<http://www.gouvernement.fr/sites/default/files/liseuse/4492/master/index.htm>> [en ligne]

Ayant choisi de restreindre le propos de ce mémoire à l'axe consacré aux données de recherche², nous en sommes venus à formuler une seconde problématique, plus spécifique : comment valoriser les données de recherche produites dans les établissements partenaires des Archives Ouvertes de la Connaissance, en tenant compte des enjeux techniques, juridiques et éthiques de leur diffusion ?

Le propos aura pour objectifs de :

- proposer une synthèse qui pose le contexte des Archives Ouvertes de la Connaissance, en esquissant les traits caractéristiques des archives institutionnelles aujourd'hui en France et en éclairant les spécificités du projet alsacien (première partie), Cette synthèse est le résultat d'un état de l'art réalisé au début du stage et de la collecte d'informations sur le projet alsacien.
- présenter les enjeux d'un service de gestion des données, en regard de la spécificité des établissements concernés par le projet d'archive ouverte, et proposer des scénarios pour la diffusion de ces données ainsi qu'une grille décisionnelle permettant d'associer chaque scénario à un modèle de données (seconde partie).

² Ce choix sera expliqué dans la transition du premier au second chapitre (p.35).

LES ARCHIVES OUVERTES DE LA CONNAISSANCE DANS LE PAYSAGE INSTITUTIONNEL FRANÇAIS ET EUROPEEN DE LA RECHERCHE

1. LES ARCHIVES OUVERTES INSTITUTIONNELLES AUJOURD'HUI

1.1. *Comment sont-elles définies ?*

1.1.1. Qu'est-ce qu'une archive ouverte institutionnelle ?

Une archive institutionnelle (en anglais *institutional archive/repository*) relève d'une institution (université, grande école, organisme de recherche ou association professionnelle) et a pour objectif de contenir, valoriser et conserver l'ensemble de la production scientifique de celle-ci³.

Par production scientifique, on entendra :

- Un article de revue, à l'état de pré-publication (*preprint*), post-publication (*postprint*) ou dans sa version finale (« version éditeur »)⁴,
- Ou tout autre document issu de recherches scientifiques, tel qu'un ouvrage ou chapitre d'ouvrage, une communication de congrès, une thèse, un rapport de recherche, un document de travail, un jeu de données⁵,...

La diffusion de cette production en archive ouverte repose sur un modèle alternatif au modèle d'édition traditionnel : celui de l'« auto-archivage », défini comme l'acte par lequel les chercheurs déposent eux-mêmes leurs productions dans une archive ouverte⁶.

1.1.2. A quels principes répondent-elles ?

Les archives ouvertes institutionnelles (tout comme les archives ouvertes nationales ou disciplinaires) sont nées du mouvement du libre accès – dit mouvement *Open Access* – dont nous rappellerons les principes fondateurs.

³ D'après le glossaire de l'Institut de l'information scientifique et technique (Inist), rassemblant une sélection de termes utilisés dans le cadre du libre accès (<http://openaccess.inist.fr/spip.php?page=glossaire>).

⁴ Le *preprint* correspond à une première version de l'article (telle que l'auteur la soumettra à l'éditeur) ; le *postprint* est la version de l'article acceptée pour publication dans une revue (c'est-à-dire relue, corrigée et certifiée par les pairs) ; la « version éditeur » ajoute au *postprint* la mise en forme de l'éditeur.

⁵ D'après la définition donnée par CODATA-ICSTI, un jeu de données est la « collection d'informations et de données connexes, généralement numériques, texte, son et/ou images, organisés pour permettre leur recherche, récupération ou traitement et réorganisation ». In : CODATA-ICSTI Task Group on Data Citation Standards and Practices. « Out of Cite, Out of Mind: The Current State of Practice, Policy, and Technology for the Citation of Data ». *Data Science Journal* 12, p. CIDCR1-CIDCR75., 2013. DOI : doi.org/10.2481/dsj.OSOM13-043 [en ligne]

⁶ D'après le glossaire de l'Inist (<http://openaccess.inist.fr/spip.php?page=glossaire>).

L'Initiative de Budapest⁷ (2002) a été la première à donner une définition publique du libre accès et à en généraliser l'appel à l'ensemble des disciplines et des pays : « Par « libre accès » à la littérature [des revues savantes] nous entendons sa mise à disposition gratuite sur l'Internet public, permettant à tout un chacun de lire, télécharger, copier, transmettre, imprimer, chercher ou faire un lien vers le texte intégral de ces articles, les disséquer pour les indexer, s'en servir de données pour un logiciel, ou s'en servir à toute autre fin légale, sans barrière financière, légale ou technique autre que celles indissociables de l'accès et l'utilisation d'Internet ». De cette définition dérive le terme « archive ouverte », dans lequel l'ouverture désigne un accès sans barrière.

À l'origine, le libre accès visait essentiellement la diffusion des articles scientifiques. Ce type de publication, traditionnellement publié dans des revues spécialisées, est devenu captif de quelques grands éditeurs privés. Sur le marché de l'édition scientifique, chaque éditeur détient le monopole de la revue qu'il publie et dont il détermine le prix. Or depuis la transition vers des offres d'édition numériques, l'augmentation de l'offre avec la création de revues de plus en plus spécialisées et la hausse constante des prix d'abonnement, conjuguée aux restrictions budgétaires des institutions, ont entraîné l'émergence d'une crise éditoriale sur le marché des revues scientifiques. En déposant leurs *preprint* en archive ouverte, les chercheurs réintroduisaient donc une communication scientifique directe et donnaient à leurs collègues la possibilité d'accéder aux articles, indépendamment des moyens financiers de leur établissement.

Deux stratégies du libre accès ont été élaborées pour les articles scientifiques :

- Le libre accès par le moyen d'archives ouvertes (*Green Open Access*) ;
- Le libre accès par le biais de revues (*Gold Open Access*).

Le *Green Open Access* repose sur le principe d'auto-archivage ; les chercheurs peuvent déposer leur article avant, pendant ou après sa publication dans une revue, selon la politique définie par l'éditeur. Ce mode de diffusion ne se substitue pas au modèle de publication traditionnel, c'est-à-dire par le biais d'un éditeur ; il lui est complémentaire car il permet d'augmenter la visibilité des productions d'un chercheur ou d'une institution.

Par ailleurs, il se distingue du *Gold Open Access*, qui suit le modèle de la chaîne éditoriale traditionnelle, dans laquelle la publication est relayée par un éditeur. Le modèle du *Gold Open Access* s'applique à la publication d'articles dans des revues libres et gratuites d'accès pour le lecteur. Celui-ci ne supportant pas le coût de l'édition, un nouveau modèle économique lui est substitué : le plus répandu est celui de l'auteur-payeur (l'auteur ou son institution d'appartenance ou de financement verse une contribution appelée *Article Processing Charges* à l'éditeur). Un autre modèle consiste en la mise en place de structures éditoriales par des institutions publiques, telles qu'OpenEdition⁸.

⁷ Infra infra : 1.2.1. « Au niveau mondial »

⁸ OpenEdition (<http://www.openedition.org/>) est une infrastructure d'édition électronique en sciences humaines et sociales, incluant une plateforme dédiée aux revues (Revue.org) ; elle est développée par le Centre pour l'édition électronique ouverte (Cléo), initiative publique à but non-lucratif promouvant le libre accès.

Un troisième modèle a émergé, appelé « modèle hybride » et qui constitue une variante du *Gold Open Access*. Il offre à l'auteur la possibilité, en échange d'une contribution qu'il verse à l'éditeur, de rendre librement et gratuitement accessible son article dans une revue sous abonnement et payante pour le lecteur. Ce modèle de publication n'est cependant satisfaisant ni pour l'auteur, ni pour le lecteur, car il donne lieu à un double paiement, le prix de l'abonnement et le montant des *Article Processing Charges*.

En 2003, la déclaration de Berlin sur le libre accès à la connaissance⁹ donne une définition précise de l'auto-archivage, conférant aux archives ouvertes un rôle essentiel dans la diffusion des publications scientifiques :

« 1. Les auteurs [d'une contribution au libre accès] et les propriétaires des droits afférents concèdent à tous les utilisateurs un droit gratuit, irrévocable et mondial d'accéder à l'œuvre en question, ainsi qu'une licence les autorisant à la copier, l'utiliser, la distribuer, la transmettre et la montrer en public, et de réaliser et de diffuser des œuvres dérivées, sur quelque support numérique que ce soit et dans quelque but responsable que ce soit, sous réserve de mentionner comme il se doit son auteur (les règles usuelles de la collectivité continueront à disposer des modalités d'attribution légitime à l'auteur et d'utilisation responsable de l'œuvre publiée, comme à présent), tout comme le droit d'en faire des copies imprimées en petit nombre pour un usage personnel.

2. Une version complète de cette œuvre, ainsi que de tous ses documents annexes, y compris une copie de la permission définie dans ce qui précède, est déposée (et, de fait, publiée) sous un format électronique approprié auprès d'au moins une archive en ligne, utilisant les normes techniques appropriées (comme les définitions des Archives Ouvertes [Open Archives]), archive gérée et entretenue par une institution académique, une société savante, une administration publique, ou un organisme établi ayant pour but d'assurer le libre accès, la distribution non restrictive, l'interopérabilité et l'archivage à long terme. »

Par ailleurs, la déclaration étend le périmètre du libre accès à l'ensemble de la production scientifique¹⁰. Elle justifie ainsi que soient déposés dans les archives ouvertes aussi bien les résultats de recherche, présentés sous une forme académique ou non (articles de revues, monographies, actes de congrès, rapports de recherche, habilitations de recherche, thèses, brevets,...), que les documents intermédiaires, ayant permis ou non d'aboutir à ces résultats. Les archives de type institutionnel sont probablement celles qui se prêtent le mieux à recueillir un champ aussi vaste de ressources scientifiques. Les institutions étant le lieu où sont menés les travaux de recherche, elles constituent en effet un vivier riche de documents susceptibles d'être diffusés et valorisés.

⁹ MAX PLANCK GESELLSCHAFT. *Berlin Declaration on Open Access to Knowledge in the Sciences and Humanities*. Berlin, 22 octobre 2003. <<http://openaccess.mpg.de/Berlin-Declaration>>. Traduction française disponible en ligne : <http://openaccess.mpg.de/68042/BerlinDeclaration_wsis_fr.pdf>.

¹⁰ « Open access contributions include original scientific research results, raw data and metadata, source materials, digital representations of pictorial and graphical materials and scholarly multimedia material » in : MAX PLANCK GESELLSCHAFT. *Berlin Declaration on Open Access to Knowledge in the Sciences and Humanities*. Berlin, 22 octobre 2003. <http://openaccess.mpg.de/Berlin-Declaration>. Traduction française disponible en ligne : http://openaccess.mpg.de/68042/BerlinDeclaration_wsis_fr.pdf

1.2. Comment sont-elles encadrées ?

En vingt-cinq ans, le libre accès est devenu une des formes essentielles de la diffusion de la production scientifique. Il n'a pas enrayé le modèle traditionnel de l'édition scientifique, mais s'est néanmoins attiré le soutien de nouveaux acteurs. C'est ce mouvement de soutien, dont nous étudierons le cadre politique, qui contribue aujourd'hui au développement des archives institutionnelles.

1.2.1. Au niveau mondial

Initié dès 1991 par les chercheurs¹¹ dans un but de communication scientifique directe, le mouvement du libre accès a été porté à l'échelle mondiale à partir de la conférence de Budapest, en 2002. Cette initiative, puis celles qui ont suivi, ont permis à la communauté scientifique de faire entendre la voix du libre accès et d'impliquer les pouvoirs publics et les institutions de recherche.

Adressée aux établissements de l'enseignement supérieur et de la recherche, ainsi qu'aux organismes de financement et aux éditeurs, l'Initiative de Budapest¹² a majoritairement rassemblé des chercheurs (en économie, biologie, sciences cognitives, sciences humaines et sociales). Sur seize signataires, deux universités seulement étaient représentées : l'Université de Montréal et le University College London. Ce n'est que par la suite que leur nombre a significativement augmenté¹³.

Faisant suite à l'Initiative de Budapest, la déclaration de Berlin (2003) promeut le développement de l'auto-archivage¹⁴. Elle s'adresse aux ministères et agences de financement de la recherche, ainsi qu'aux institutions de recherche et de conservation du patrimoine scientifique. Elle réunit pour la première fois des responsables universitaires, qui s'engagent à instaurer un encadrement institutionnel du libre accès. La mise en place de cet encadrement ne sera pas immédiate. Aussi la conférence de Berlin 3, en 2005, soumet-elle aux établissements signataires une feuille de route¹⁵, dans laquelle sont définis les axes stratégiques d'une charte institutionnelle du libre accès. Les établissements sont notamment incités à adopter une politique contraignant les chercheurs à déposer leurs publications dans une archive ouverte institutionnelle.

¹¹ La première archive ouverte, ArXiv, a été imaginée par le physicien Paul Ginsparg en 1991. Elle a inspiré la « proposition subversive » de Stevan Harnad, enseignant-chercheur en sciences cognitives, incitant les auteurs à déposer leurs *preprints* dans une archive institutionnelle. Voir : HARNAD Stevan, « Universal FTP archives for esoteric science and scholarship: a subversive proposal », *Scholarly Journals at the Crossroads: A Subversive Proposal for Electronic Publishing*. Association of Research Libraries, Washington DC, 1995. <<http://www.arl.org/scomm/subversive/toc.html>> [en ligne]

¹² « Budapest Open Access Initiative ». Budapest, 14 février 2002. <<http://www.budapestopenaccessinitiative.org/read>>. Traduction française disponible en ligne : <<http://www.budapestopenaccessinitiative.org/translations/french-translation>>.

¹³ Il n'est pas aisé d'extraire le nombre exact d'institutions de type universitaire parmi les signataires listés sur le site de la Budapest Open Access Initiative (http://www.budapestopenaccessinitiative.org/list_signatures). Nous en avons répertorié une centaine.

¹⁴ Supra : 1.1.2. « A quels principes répondent-elles »

¹⁵ MAX PLANCK GESELLSCHAFT, *The Road to Open Access: A guide to the implementation of the Berlin Declaration*. Southampton, mars 2005. <http://openaccess.mpg.de/81923/roadmap_print.pdf> [en ligne]

Dix ans après, en 2012, la progression du libre accès a conduit à la formulation de nouvelles recommandations¹⁶, lors d'une nouvelle conférence de Budapest. L'initiative a, cette fois, réuni 29 participants, dont 8 représentants des universités, soit un taux de représentation de 12,5% en 2002 et 27,6% en 2012). Elle fixe un nouvel objectif pour 2022 : « dans dix ans, le libre accès sera la solution par défaut pour la dissémination de nouvelles recherches dans tous les domaines, dans tous les pays ». Dans ce but, sont préconisés :

- l'adoption d'une politique de libre accès à l'échelle des gouvernements et des institutions, mais aussi au niveau des financeurs et des éditeurs ;
- le développement d'archives institutionnelles et disciplinaires.

A noter, l'investissement des services de documentation des établissements qui ont joué un rôle important dans ce mouvement, en menant une réflexion sur les outils nécessaires au libre accès.

Au niveau mondial, l'encadrement des archives ouvertes institutionnelles ne peut être que d'ordre incitatif ; il appartient aux gouvernements de chaque pays d'instaurer un cadre législatif pour le développement du libre accès dans les établissements nationaux de la recherche.

1.2.2. Au niveau européen

Le recours aux archives ouvertes a été préconisé par l'Union européenne pour la première fois en 2007, lorsque le Conseil européen de la recherche (CER) a recommandé la mise en libre accès des résultats des recherches qu'il finance dans des archives disciplinaires ou institutionnelles¹⁷. Cette initiative a été reprise par la Commission européenne dans les programmes-cadres de recherche qu'elle finance, sous la forme d'un plan d'action progressif d'ouverture des résultats scientifiques¹⁸. En 2008 a été intégré au 7^{ème} programme-cadre (2007-2013) un *Open Access Pilot*¹⁹. Ce projet-test, qui concernait sept disciplines²⁰, a été étendu à l'ensemble de la recherche publique européenne dans le programme-cadre suivant, baptisé « Horizon 2020 » (2014-2020). La convention de subvention²¹ engage les bénéficiaires de financements à déposer leurs articles de revue évalués par les pairs dans une archive ouverte, afin de les rendre librement accessibles, c'est-à-dire disponibles en ligne et gratuitement pour l'utilisateur. Les bénéficiaires sont

¹⁶ BOAI 10. « Ten years on from the Budapest Open Access Initiative: setting the default to open ». Budapest. 12 septembre 2012. <<http://www.budapestopenaccessinitiative.org/boai-10-recommendations>>. Traduction française disponible en ligne : <<http://www.budapestopenaccessinitiative.org/boai-10-translations/french>>.

¹⁷ EUROPEAN RESEARCH COUNCIL, « European Research Council-Scientific Council Guidelines for Open Access », 17 décembre 2007, 2 p. <<http://www.openaire.eu/fr/component/attachments/download/3>> [en ligne]

¹⁸ Sur le principe d'ouverture des résultats scientifiques, voir infra : « L'ouverture des données de recherche » (p.37).

¹⁹ EUROPEAN COMMISSION, *Open Access Pilot in FP7*. 2008, 3 p. <http://ec.europa.eu/research/sciencesociety/document_library/pdf_06/open-access-pilot_en.pdf> [en ligne]

²⁰ Santé, énergie, environnement robotique, sciences socio-économiques, sciences humaines, science en société et infrastructures numériques de recherche.

²¹ EUROPEAN COMMISSION, *General multi-beneficiary model grant agreement for the Horizon 2020 programme*. 2014. <http://ec.europa.eu/research/participants/data/ref/h2020/mga/gga/h2020-mga-gga-multi_en.pdf> [en ligne] Article 29.2 « Open Access to scientific publications »

également encouragés à y déposer d'autres types de publications scientifiques, soumis ou non à l'évaluation des pairs, par exemple des ouvrages, des actes de conférence, des rapports ou des données de recherche. Le programme-cadre Horizon 2020 (H2020) introduit, par ailleurs, un nouveau projet-test : l'*Open Research Data Pilot*, dont l'objectif est d'optimiser la réutilisation des données de recherche générées par les projets ; il devrait, lui aussi, être généralisé dans les prochains programmes-cadres. (Nous reviendrons sur les termes de l'*Open Research Data Pilot* en deuxième partie²².)

Depuis la création d'un « Espace européen de la recherche » (EER)²³, mis en place par la Commission du 17 juillet 2012²⁴, le libre accès fait partie de la stratégie de l'Union européenne visant à « optimiser la diffusion et le transfert des connaissances scientifiques, ainsi que l'accès à ces connaissances ». Il est perçu non comme une fin en soi, mais comme un moyen de dynamiser l'innovation et la croissance économique par un accès plus rapide et plus large aux résultats de la recherche. Les principes de la politique développée par la Commission européenne sont exposés dans la communication « Pour un meilleur accès aux informations scientifiques »²⁵ et dans les recommandations²⁶ qu'elle adresse aux Etats membres. La Communication soutient les voies du *Green* et du *Gold Open Access*, pour les publications scientifiques comme pour les données de recherche, excluant « tout nouveau paiement lors de l'accès à des informations déjà payées par des dépenses publiques ou de leur réutilisation ». La Recommandation encourage les Etats membres :

- à définir des politiques claires en matière de libre accès aux publications scientifiques et aux données de recherche et en matière de conservation et réutilisation de ces informations.
- à développer des infrastructures numériques de diffusion, ainsi qu'un dialogue multilatéral aux niveaux national, européen et international.

Concernant les établissements publics de recherche, les gouvernements sont chargés de veiller à ce que :

- des politiques institutionnelles assurant le libre accès aux résultats de recherche soient définies ;
- les chercheurs bénéficient d'infrastructures numériques et d'orientations juridiques pour la diffusion en libre accès de leurs travaux ;

²² Infra : 1.2. « Des mouvements incitatifs » (p.40)

²³ L'Espace européen de la recherche (*European Research Area*) est un espace de recherche unifié et ouvert au monde, reposant sur le Marché intérieur où chercheurs, connaissances scientifiques et technologie circulent librement. Il se définit comme cadre commun de principes, d'objectifs et d'infrastructures.

²⁴ COMMISSION EUROPEENNE, *Un partenariat renforcé pour l'excellence et la croissance dans l'Espace européen de la recherche*. COM(2012) 392 final. Bruxelles, 17 juillet 2012. 18 p. http://ec.europa.eu/research/era/pdf/era-communication/era-communication_fr.pdf p.15-16

²⁵ EUROPEAN COMMISSION, « Towards better access to scientific information : boosting the benefits of public investments in research », COM(2012) 401 final. Bruxelles, 17 juillet 2012, 15 p. <http://ec.europa.eu/research/science-society/document_library/pdf_06/era-communicationtowards-better-access-to-scientific-information_en.pdf> [en ligne]

²⁶ EUROPEAN COMMISSION, « Recommendation on access to and preservation of scientific information », C(2012) 4890 final. Bruxelles, 17 juillet 2012, 10p. <http://ec.europa.eu/research/science-society/document_library/pdf_06/recommendationaccess-and-preservation-scientific-information_en.pdf> [en ligne]

- les systèmes de recrutement, d'évaluation de carrière et d'octroi de subventions valorisent les chercheurs qui adhèrent à une culture de partage de leurs résultats de recherche ;
- l'ensemble de ces mesures bénéficient de financements appropriés.

Les institutions de l'Union européenne apportent donc au mouvement du libre accès les bénéfices d'un cadre juridique, de financements, d'infrastructures et de groupes de réflexion.

1.2.3. Au niveau national

En France, les politiques de soutien aux archives institutionnelles ont progressivement remonté les échelons de l'administration. Ce sont les directions des institutions de recherche qui, les premières, ont développé une stratégie nationale, avec la signature en 2006 d'un Protocole d'accord²⁷, qui a abouti à faire de HAL²⁸ la plateforme nationale partagée. L'effort de coordination nationale a été prolongé, à l'initiative du Ministère de l'enseignement supérieur et de la recherche, par la création de la Bibliothèque Scientifique Numérique (BSN)²⁹ en 2012. Cette infrastructure est chargée d'organiser la coopération entre les principaux acteurs nationaux de la recherche sur les questions de libre accès. Le segment BSN4, dédié aux archives ouvertes, veille à la cohérence technique et scientifique entre la plateforme nationale HAL et les archives institutionnelles locales.

L'implication du gouvernement français s'est inscrite dans la lignée des recommandations européennes de 2012. Un des temps forts de cet engagement a été la prise de position de la Ministre de l'Enseignement supérieur et de la Recherche, Geneviève Fioraso, lors des 5^{èmes} journées Open Access organisées par le Consortium Couperin³⁰, le 24 janvier 2013 : « Au regard de l'importance des enjeux, sur les plans scientifique, économique et sociétal, le gouvernement français réaffirme son soutien au principe du libre accès à l'information scientifique »³¹. Accélérée par le mouvement d'ouverture des données publiques³², la politique

²⁷ « Protocole d'accord, en vue d'une approche coordonnée, au niveau national, pour l'archivage ouvert de la production scientifique ». Paris, 6 juillet 2006. <<http://openaccess.inist.fr/?PROTOCOLE-D-ACCORD-en-vue-d-une>> [en ligne]

²⁸ L'archive pluridisciplinaire HAL, conçue par le Centre pour la Communication Scientifique Directe (CCSD) du CNRS pour le dépôt et la diffusion d'articles scientifiques et de thèses, est devenue plateforme nationale partagée en 2006, suite au Protocole d'accord signé par les principaux organismes de recherche français, la Conférence des présidents d'université et la Conférence des grandes écoles.

²⁹ <http://www.bibliothequescientifiquenumerique.fr/>

³⁰ Le consortium Couperin négocie les offres d'abonnements avec les éditeurs des grandes revues scientifiques (<http://www.couperin.org/>).

³¹ « Discours de Geneviève Fioraso lors des 5^{èmes} journées Open Access ». Paris, 24 janvier 2013. <<http://www.enseignementsup-recherche.gouv.fr/cid66992/discours-de-genevieve-fioraso-lorsdes-5-journees-open-access.html>> [en ligne]

³² « L'ouverture et le partage des données publiques consistent à mettre à la disposition de tous les citoyens, sur internet, toutes les données publiques brutes [issues des gouvernements, administrations centrales et collectivités locales] qui seront librement accessibles et gratuitement réutilisables. » Etalab, *Vade-mecum sur l'ouverture et le partage des données publiques*. Septembre 2013, 11p. <<http://www.modernisation.gouv.fr/sites/default/files/fichiers-attaches/vademecum-ouverture.pdf>> [en ligne]

Une page Wikipédia est consacrée aux données ouvertes (*open data*) : https://fr.wikipedia.org/wiki/Open_data.

volontariste de l'Etat se concrétisera d'ici fin 2015 par la présentation d'un projet de loi sur le Numérique devant l'Assemblée nationale. Celui-ci devrait répondre au principe suivant : « tout chercheur qui le souhaite doit pouvoir [...] rendre librement accessibles ses articles scientifiques, au terme d'une durée raisonnable après publication, en tenant compte des différences entre sciences exactes et sciences humaines »³³. « Un soutien spécifique au développement de plates-formes et d'infrastructures de dépôt » étant par ailleurs prévu, les archives institutionnelles pourront donc constituer un moyen de se conformer à ce souhait.

1.3. Quels sont les enjeux ?

Quelle est la valeur ajoutée d'une archive institutionnelle pour un établissement de recherche et ses chercheurs ? Quels enjeux (politiques, juridiques, techniques) convient-il de prendre en compte pour tirer parti de cet outil ?

1.3.1. Enjeux politiques

Garantir un patrimoine

En tant qu'outil d'archivage des résultats de recherche, une archive institutionnelle a vocation à devenir la mémoire scientifique de l'établissement qu'elle représente. Cette fonction soulève néanmoins la question de l'exhaustivité des dépôts et donc de la stratégie envisageable pour atteindre un tel objectif. L'instauration d'une obligation de dépôt est-elle justifiée ? Ce modèle a le vent en poupe : préconisé par Stevan Harnad - l'un des pionniers des archives ouvertes au sein de la communauté scientifique -³⁴ il rencontre aujourd'hui un soutien politique croissant -de la part des organismes de financement notamment. En France, les politiques d'obligation de dépôt se multiplient parmi les institutions de recherche depuis environ 2010. Elles apparaissent comme le plus sûr moyen de favoriser l'accroissement rapide du volume de documents dans les archives institutionnelles ; selon un rapport de H. Prost et J. Schöpfel³⁵, les institutions ayant opté pour une obligation de dépôt ont une croissance deux fois plus rapide que la moyenne (+329% entre 2008 et 2009). Cependant, en l'absence d'un mandat national unique, les mesures d'incitation au dépôt émises par les établissements (quel qu'en soit le degré de coercition) seront toujours contestables.

Confier l'archivage à un acteur public

Le modèle de l'édition scientifique veut que les chercheurs publient leurs travaux de recherche sans compensation financière³⁶. Or les barrières limitant

³³ *Stratégie numérique du gouvernement*. Gaité lyrique, 18 juin 2015. <<http://www.gouvernement.fr/sites/default/files/liseuse/4492/master/index.htm>> [en ligne] p.15

³⁴ Voir note de bas de page 11, p.14 ; HARNAD Stevan, « Waking OA's "slumbering giant": the university's mandate to mandate open access », *New Review of Information Networking*, vol.14, 2008, pp. 51-68. DOI:10.1080/13614570903001322

³⁵ PROST Hélène et SCHÖPFEL Joachim, *Développement et Usage des Archives Ouvertes en France. Rapport. 1e partie : Développement*. Juillet 2010. <<http://www.enssib.fr/bibliotheque-numerique/document-48600>> [en ligne] pp. 18-19.

³⁶ On distingue ici la « diffusion » de l' « exploitation », qui consiste à restreindre l'utilisation des résultats de recherche à des fins commerciales.

l'accès aux articles de revues servent les intérêts des intermédiaires (ceux des grands éditeurs privés). Ce constat est d'autant plus aberrant pour les institutions de recherche qu'elles irriguent le circuit de l'édition en amont comme en aval – elles financent les travaux des chercheurs et payent de nouveau pour avoir accès aux résultats publiés de ces travaux. L'enjeu consiste donc également à préserver la littérature grise³⁷ du système éditorial lucratif, qui détient déjà le quasi-monopole du marché des revues.

De ce constat est né un mouvement soutenu :

- d'une part par les militants favorables au libre accès (chercheurs, bibliothèques, etc.) qui, tel qu'ils le stipulent dans l'Initiative de Budapest³⁸, jugent impératif de rendre la connaissance disponible à tous ceux susceptibles d'en faire usage ;
- d'autre part par les pouvoirs publics, pour qui les investissements publics doivent bénéficier à la société.

En confiant l'archivage des travaux de recherche aux établissements, la communauté scientifique et les autorités de l'Etat s'assurent que la recherche financée sur fonds publics reste aux mains d'acteurs publics.

Assurer la visibilité du travail scientifique

Le dépôt en archive ouverte permet d'améliorer l'accessibilité aux documents déposés. Ce faisant, il augmente leur visibilité. Il a en effet été démontré³⁹ que le taux de citation d'une publication augmente mécaniquement et significativement, lorsqu'elle est diffusée en libre accès, quelle que soit la discipline à laquelle elle appartient.

Le travail du chercheur se trouve ainsi valorisé, de même que l'activité de recherche de l'établissement dans son ensemble. L'archive constitue une vitrine institutionnelle.

Parallèlement, le dépôt en archive ouverte démocratise l'accès aux publications. Face à la crise éditoriale des revues scientifiques, la situation est devenue intenable pour de nombreuses institutions de recherche. À l'Université de Strasbourg par exemple, le budget des abonnements électroniques atteint presque 3 millions d'euros, soit 75% du budget annuel des acquisitions de l'établissement. Depuis dix ans, le prix d'accès aux revues scientifiques augmente d'environ 7% par année⁴⁰. Soumis aux restrictions budgétaires de l'Etat et aux excès tarifaires des éditeurs, les services de documentation des universités se voient contraints de restreindre l'éventail de leurs abonnements de revues, au détriment des chercheurs, enseignants et étudiants. Dans ce contexte, les archives institutionnelles réintroduisent une égalité d'accès aux publications, indépendamment des ressources financières de l'utilisateur final.

³⁷ La littérature grise désigne l'ensemble des publications à caractère non-commercial ; dans la plupart des archives ouvertes, elle se limite aux publications d'ordre scientifique (vs administratif).

³⁸ « Budapest Open Access Initiative ». 14 février 2002. <<http://www.budapestopenaccessinitiative.org/read>>. Traduction française disponible en ligne : <<http://www.budapestopenaccessinitiative.org/translations/french-translation>>

³⁹ A propos de l'impact du libre accès sur le taux de citation des articles, on pourra consulter la bibliographie établie par le site *The Open citation project* : <http://opcit.eprints.org/oacitation-biblio.html>

⁴⁰ BACH Jean-François, JEROME Denis. *Les nouveaux enjeux de l'édition scientifique*. Rapport de l'Académie des Sciences. Institut de France, 24 juin 2014, 56 p. <http://www.academie-sciences.fr/pdf/rapport/rads_241014.pdf> [en ligne] p.24

Accélérer l'innovation et le retour sur investissement

Par une diffusion plus rapide et plus large des résultats scientifique, le libre accès améliore l'impact de la recherche fondamentale en termes d'innovation et de bénéfices économiques.

1.3.2. Enjeux juridiques

Contrairement aux autres agents publics, qui cèdent automatiquement à leur employeur les droits d'exploitation sur les œuvres qu'ils ont produites dans l'exercice de leurs fonctions, les chercheurs et enseignants-chercheurs conservent la plénitude de leurs droits d'auteur⁴¹. Ils ne sont donc ni soumis à un mode de publication prédéfini de leurs résultats de recherche, ni même contraints de les divulguer. Cette réglementation induit pour les archives institutionnelles de prendre en compte un certain nombre de considérations juridiques.

Assurer la publication parallèle des documents

Un des principes du libre accès consiste à rendre des documents librement accessibles et ainsi à empêcher l'exclusivité de leur exploitation, notamment par les éditeurs privés. En effet, lorsqu'un chercheur publie son travail auprès d'un éditeur, celui-ci est en droit d'en exiger la diffusion et l'exploitation exclusives. L'auto-archivage n'étant pas défini par la loi, c'est le contrat d'édition qui prime et définit si le chercheur est autorisé à diffuser la publication, *en parallèle*, dans une archive ouverte.

Nous l'avons évoqué, le modèle de publication parallèle augmente l'impact du document publié. Il suscite toutefois débat du côté des éditeurs, qui y voient une menace pour le marché éditorial scientifique. Le développement des archives ouvertes les a conduits à envisager des compromis entre publication traditionnelle et auto-archivage. Dans le domaine des revues scientifiques, la plupart des éditeurs autorisent le dépôt en archive ouverte de la version *preprint* ou *postprint* de l'article⁴². Certains conservent également l'exclusivité de sa diffusion durant une période définie (généralement six à douze mois), en imposant un embargo⁴³ ; les références bibliographiques de l'article sont visibles dans l'archive ouverte, mais l'accès au document est suspendu. Le site officiel SHERPA/RoMEO⁴⁴ répertorie, grâce à un code couleurs, la politique des éditeurs internationaux vis-à-vis du dépôt en archives ouvertes des articles publiés dans leurs revues ; il permet aux chercheurs de connaître les conditions d'un éditeur en matière de publication parallèle (quelle version est autorisée au dépôt, par exemple).

⁴¹ Ce régime dérogatoire a été instauré par la loi DADVSI du 1^{er} août 2006, en vertu du statut particulier des enseignants-chercheurs, enseignants et chercheurs –lesquels « jouissent d'une pleine indépendance et d'une entière liberté d'expression dans l'exercice de leurs fonctions d'enseignement et de leurs activités de recherche » (article L.952-2 du Code de l'Éducation).

⁴² Voir note de bas de page 4.

⁴³ « L'embargo est la période pendant laquelle un article de revue, déposé dans une archive ouverte, ne peut être accessible librement. Le plus souvent, il est le résultat d'un compromis entre l'institution demandant le dépôt et l'éditeur chez lequel est publié le document ». D'après le glossaire de l'Inist : <http://openaccess.inist.fr/spip.php?page=glossaire>.

⁴⁴ SHERPA/RoMEO : <http://www.sherpa.ac.uk/romeo/>

En France, un encadrement législatif est en phase d'élaboration, dans le but de clarifier le droit d'auteur en matière d'auto-archivage. Il s'inscrit dans un mouvement de réforme initié par l'Allemagne et l'Italie.

En Allemagne, la législation du droit d'auteur dans le domaine de la communication scientifique a été aménagée en 2013 via un amendement de la loi sur la propriété intellectuelle du 9 septembre 1965⁴⁵. En vigueur depuis le 1^{er} janvier 2014, le nouvel amendement accorde à l'auteur d'une contribution scientifique, majoritairement financée par des fonds publics et publiée auprès d'un éditeur, un « droit d'exploitation secondaire » (*Zweitveröffentlichungsrecht*), inconditionnel et inaliénable, l'autorisant à diffuser de nouveau cette contribution en libre accès (quel que soit le contrat d'édition, même s'il a cédé la totalité de ses droits à l'éditeur). Si cette révision répond à la demande de la communauté scientifique allemande, elle ne la satisfait que partiellement et suscite de nombreuses interrogations quant à son interprétation⁴⁶. En vertu des « intérêts des éditeurs », le droit d'exploitation secondaire répond en effet à des règles complexes :

1. Il se limite aux articles scientifiques : résultant d'une activité d'enseignement et de recherche financée au moins pour moitié par des fonds publics ; publiés dans des revues scientifiques, paraissant au moins deux fois par an.

2. Il prend effet au-delà de douze mois après la première publication de l'article.

3. Il s'applique à la version révisée par les pairs (*peer-reviewed*) de l'article (non à la version de l'éditeur), à des fins d'exploitation non commerciales, sous réserve de citation de la source de première publication.

4. Enfin, il ne peut s'exercer rétroactivement.

En Italie, la loi du 7 octobre 2013⁴⁷ instaure, quant à elle, un mandat qui permet de généraliser la diffusion secondaire à l'ensemble des chercheurs et enseignants-chercheurs affiliés à une institution de recherche publique : c'est en effet l'institution qui est mandatée de définir une politique de libre accès et de veiller à son application au sein de l'établissement. Si les conditions d'application de la loi sont les mêmes que dans la législation allemande, elles étendent

⁴⁵ „Der Urheber eines wissenschaftlichen Beitrags, der im Rahmen einer mindestens zur Hälfte mit öffentlichen Mitteln geförderten Forschungstätigkeit entstanden und in einer periodisch mindestens zweimal jährlich erscheinenden Sammlung erschienen ist, hat auch dann, wenn er dem Verleger oder Herausgeber ein ausschließliches Nutzungsrecht eingeräumt hat, das Recht, den Beitrag nach 12 Monaten seit der Erstveröffentlichung in der akzeptierten Manuskriptversion öffentlich zugänglich zu machen, soweit dies keinem gewerblichen Zweck dient. Die Quelle der Erstveröffentlichung ist anzugeben. Eine zum Nachteil des Urhebers abweichende Vereinbarung ist unwirksam.“ Ministère de la Justice, *Entwurf eines Gesetzes zur Nutzung verwaister Werke und zu weiteren Änderungen des Urheberrechtsgesetzes und des Urheberrechtswahrnehmungsgesetzes*. Berlin, février 2013. <<http://blog.die-linke.de/digitalelinke/wp-content/uploads/Ref.Entwurf.pdf>>. L'Inist propose une traduction française des principaux passages : <<http://openaccess.inist.fr/?Traduction-francaise-du-projet-de>>. Article 38, paragraphe 4.

⁴⁶ Comme en témoigne l'Alliance des Organisations Scientifiques Allemandes (die Allianz der deutschen Wissenschaftsorganisationen), qui a consacré une page web aux questions récurrentes soulevées par l'application du droit d'exploitation secondaire (<http://www.allianzinitiative.de/de/handlungsfelder/rechtliche-rahmenbedingungen/faq-zvr.html>).

⁴⁷ Loi n. 112 (G.U. n. 236, 8.10.2013) relative à la valorisation de la culture. Citée dans : MOSCON Valentina, « Open Access in Italy », *IRIS Legal Observations of the European Audiovisual Observatory*. IRIS 2014-1:1/32. <<http://merlin.obs.coe.int/iris/2014/1/article32.en.html>> [en ligne]

néanmoins les droits d'exploitation exclusifs de l'éditeur sur une durée plus longue (18 mois pour les articles en sciences techniques et médicales, 24 mois pour ceux des sciences humaines et sociales).

Le 1^{er} juillet dernier, les autorités néerlandaises ont pris, elles aussi, des dispositions législatives en faveur de la publication parallèle⁴⁸.

Il semblerait que le projet de loi français ressemble davantage au modèle allemand, comme en témoigne la version de travail publiée en juillet et qui prévoit d'ajouter au code de la propriété intellectuelle l'article suivant : « L'auteur d'une contribution scientifique, issue d'une activité de recherche financée au moins pour moitié par des fonds publics et publiée dans le cadre d'une collection paraissant au moins une fois par an, dispose du droit, même s'il a cédé un droit d'exploitation exclusif à l'éditeur, de rendre publiquement accessible la version acceptée de son manuscrit, au terme d'un délai de six mois pour les sciences et de douze mois pour les Sciences humaines et sociales à compter de la première publication, toute fin commerciale étant exclue »⁴⁹.

Protéger le droit d'auteur dans le cadre de la diffusion de la production scientifique et de son exploitation

Tout comme le modèle éditorial, l'archive institutionnelle garantit le respect du droit d'auteur des déposants et apporte une protection juridique à leurs travaux. Elles ont recours ou bien aux licences libres, dont les plus connues sont les licences Creative Commons (CC)⁵⁰, ou bien à une diffusion des documents restreinte à l'Intranet de l'établissement, laissant libre accès à la notice bibliographique seulement.

La proximité des équipes qui maintiennent l'archive institutionnelle – généralement des personnels du SCD - est un avantage pour les déposants, qui trouvent sur le terrain un conseil technique et juridique adapté. Un grand nombre d'entre eux craignent notamment d'enfreindre les règles du droit d'auteur – le plus souvent par méconnaissance des termes exacts du contrat signé avec leur éditeur ; par exemple, lorsque le contrat initial ne prévoit pas de conditions à l'auto-archivage, les équipes de l'archive sont en mesure de suggérer la rédaction d'une clause additionnelle spécifique (ou « addendum d'auteur »)⁵¹.

⁴⁸ Aux Pays-Bas, depuis le 1er juillet 2015, un amendement de la loi sur le droit d'auteur autorise les chercheurs à rendre leurs articles librement accessibles. Voir : HAMEAU Thérèse, « Le libre accès introduit dans la loi néerlandaise ». Billet du blog *Libre accès à l'information scientifique et technique*, 23 juillet 2015. <<http://openaccess.inist.fr/?Le-libre-acces-introduit-dans-la>> [en ligne]

⁴⁹ Article 39, section 3 « Libre accès aux travaux de recherche » du *Projet de loi relatif à l'ambition numérique de la France* (juillet 2015).

⁵⁰ <http://creativecommons.org/>

⁵¹ La Direction des Affaires Juridiques du CNRS met à disposition une documentation en ligne visant à informer les chercheurs de leurs droits en matière de publication parallèle en archives ouvertes (<http://www.dgdr.cnrs.fr/daj/propriete/droits/droits.htm#auteur>, rubrique « Les archives ouvertes et les publications scientifiques »). Un modèle de clause spécifique sur les droits de publication parallèle est proposé aux auteurs qui souhaitent modifier les termes du contrat signé avec leur éditeur.

1.3.3. Enjeux techniques

Si l'implication du chercheur fait la nature du modèle de l'auto-archivage (*Green Open Access*), elle n'en reste pas moins soumise à l'existence d'une infrastructure de dépôt dédiée, dont l'initiative de Budapest de 2012 (BOAI10) rappelle les conditions de viabilité : « une archive est acceptable si elle est en libre accès, se conforme aux règles de l'interopérabilité avec d'autres archives, et prend des dispositions pour la préservation à long terme des documents ». À ces critères nous ajouterons l'enchaînement efficace des tâches (*workflow*) dans le processus de dépôt.

Une diffusion rapide des publications

Le dépôt en archive ouverte assure la diffusion accélérée du document sur Internet. D'une part, grâce à la rapidité de communication de ce média ; d'autre part, parce que le modèle de l'auto-archivage n'exige aucune contribution financière de la part de l'auteur et n'induit qu'un nombre limité d'intermédiaires. Dans le cadre des archives institutionnelles, la diffusion du document est relayée par les services de documentation, qui assurent la validation des métadonnées et la vérification des permissions de diffusion accordées par l'éditeur –ce qui nécessite d'optimiser au mieux les étapes du processus de dépôt.

Un référencement exhaustif

La qualité d'une archive ouverte se mesure également à son degré d'ouverture ou « interopérabilité ». Une archive *ouverte* donne d'autant plus de visibilité à son contenu (notices bibliographiques et documents associés) qu'elle est capable de dialoguer avec d'autres systèmes d'information :

- des fournisseurs de services, tels que les moteurs de recherche (*Google, Google Scholar*), les bases de données bibliographiques (*Web of Science*⁵²), les répertoires d'archives (*OpenAIRE*⁵³), etc.
- mais aussi des fournisseurs de données, tels que les archives disciplinaires (*ArXiv*⁵⁴, *RePEc*⁵⁵ pour les publications ; *Dryad*⁵⁶, *Nakala*⁵⁷ pour les données de recherche), les archives institutionnelles⁵⁸ ou encore l'archive nationale *HAL*.

Les normes d'interopérabilité des archives ouvertes sont définies depuis 1999 par l'*Open Archives Initiative Protocol for Metadata Harvesting*⁵⁹ (OAI-PMH). Le protocole OAI-PMH donne une trame technique et organisationnelle commune, qui garantit l'interconnexion entre archives et supprime la notion de territoire, à

⁵² <https://isiknowledge.com/>

⁵³ <https://www.openaire.eu/>

⁵⁴ <http://arxiv.org/>

⁵⁵ <http://repec.org/>

⁵⁶ <http://datadryad.org/>

⁵⁷ <https://www.nakala.fr/>

⁵⁸ Quatre exemples d'archives institutionnelles seront donnés dans la partie 1.4. « Quelques exemples d'archives institutionnelles françaises » (p.34).

⁵⁹ OPEN ARCHIVE INITIATIVE, « The Open Archives Initiative Protocol for Metadata Harvesting ». Version 2, juin 2002. <<http://www.openarchives.org/OAI/openarchivesprotocol.htm>> [en ligne]

l'échelle nationale et internationale. Les auteurs de l'Initiative de Budapest⁶⁰ résument ainsi l'intérêt d'archives conformes au protocole OAI-PMH : « [...] des moteurs de recherche et autres outils peuvent traiter des archives distinctes comme un seul et unique fonds d'archive. L'utilisateur n'a alors plus besoin de savoir quelle archive existe, ni où elle est localisée, pour accéder à son contenu et l'utiliser ». C'est un gain de temps pour le chercheur également, qui peut se contenter d'un dépôt unique, tout en ayant l'assurance que le document ou sa notice sera consultable à partir d'autres systèmes d'information.

Un archivage pérenne

Les archives ouvertes sont à la fois un moyen de diffusion et un moyen de conservation du document déposé. À ce titre, elles se doivent d'en assurer l'archivage pérenne, dont le triple objectif est de conserver le document, de le rendre accessible et d'en préserver l'intelligibilité sur le long terme⁶¹.

L'objectif d'intelligibilité relève du travail de renseignement des métadonnées⁶², qui consiste à faire en sorte que le document reste compréhensible à travers le temps : la description précise du document, assurée par l'auteur (voire par les documentalistes) au moment du dépôt, conditionne la qualité de sa conservation.

L'objectif d'accessibilité, qui exige que l'on puisse retrouver et lire le document, consiste à lui attribuer un identifiant pérenne –unique et immuable– (URN, PURL, DOI, etc.).

Quant à la conservation des données, elle relève du stockage sur un serveur informatique, pendant une période supérieure à trente ans. Nécessitant d'importants volumes de stockage, les archives institutionnelles s'appuient le plus souvent sur les services d'un centre de données local ou national. En France, le Centre Informatique National de l'Enseignement Supérieur (CINES⁶³) est le service national dédié pour l'archivage numérique pérenne. Une convention a été établie entre le CINES et le CCSD pour la plateforme HAL, de sorte que toute donnée versée dans HAL est systématiquement archivée par le CINES ; les archives institutionnelles qui ont choisi d'être connectées à HAL bénéficient donc gratuitement de ce service.

⁶⁰ « Budapest Open Access Initiative ». 14 février 2002. <<http://www.budapestopenaccessinitiative.org/read>>. Traduction française disponible en ligne : <<http://www.budapestopenaccessinitiative.org/translations/french-translation>>

⁶¹ Définition empruntée au CINES (<https://www.cines.fr/archivage/un-concept-des-problematiques/le-concept-darchivage-numerique-perenne/>).

⁶² Les métadonnées (littéralement *données sur les données*) sont un ensemble de données structurées, constituant la carte d'identité d'un document (elles permettent de le décrire et de renseigner sur son contenu). Afin d'être rendues compréhensibles par l'audience la plus large possible (y compris par les machines), elles sont généralement élaborées selon un format défini –le format international Dublin Core par exemple.

⁶³ <https://www.cines.fr/>

1.4. Quelques exemples d'archives institutionnelles françaises

À la date du 5 août 2015, étaient répertoriés dans le *Registry of Open Access Repository* (ROAR)⁶⁴ : 2694 archives institutionnelles (« Research Institutional or Departmental »), dont 50 en France (146 en Allemagne et 149 au Royaume-Uni). Nous présentons en annexe un tableau comparatif de quatre d'entre elles :

- *ArchiMer*⁶⁵, archive institutionnelle de l'Institut Français de Recherche pour l'Exploitation de la Mer (IFREMER) ;
- *Okina*⁶⁶, archive ouverte Open Knowledge, Information, Access de l'Université d'Angers ;
- *ProdInra*⁶⁷, archive ouverte de l'Institut National de la Recherche Agronomique (INRA) ;
- *SAM*⁶⁸, archive ouverte Science Arts et Métiers de l'école des Arts & Métiers ParisTech.

Les archives institutionnelles bénéficient d'un cadre technique idéal, mais sont encore relativement démunies d'un point de vue juridique. Les universités s'en remettent donc aux financeurs et éditeurs de la recherche, qui ont plus d'influence, pour inciter la communauté scientifique au libre accès. Dans ce paysage, les Archives Ouvertes de la Connaissance ont la particularité de rassembler quatre établissements différents. Cette seconde partie sera l'occasion de présenter leurs spécificités structurelles, auxquelles devra veiller le projet.

2. LES ARCHIVES OUVERTES DE LA CONNAISSANCE : ENVIRONNEMENT DU PROJET

2.1. Description du projet

L'Université de Strasbourg, l'Université de Haute-Alsace (UHA), l'Institut National des Sciences Appliquées de Strasbourg (INSAS) et la Bibliothèque Nationale Universitaire de Strasbourg (BNUS) ont défini un projet d'archive institutionnelle commune, dont une première version devrait voir le jour au début de l'année 2016. Les Archives Ouvertes de la Connaissance serviront de plateforme de dépôt, d'archivage et de diffusion, ayant pour vocation de regrouper et valoriser la production scientifique des chercheurs, enseignants-chercheurs et doctorants de ces quatre établissements ; elles rassembleront en priorité les publications et les données de recherche. Les délibérations ont abouti au choix

⁶⁴ Registry of Open Access Repository (ROAR), <http://roar.eprints.org/>

⁶⁵ <http://archimer.ifremer.fr/>

⁶⁶ <http://okina.univ-angers.fr/>

⁶⁷ <http://prodinra.inra.fr/>

⁶⁸ <http://sam.ensam.eu/>

d'une plateforme locale, basée sur la solution logicielle Open Source *Islandora*⁶⁹, interopérable avec des entrepôts nationaux tels que HAL.

2.1.1. Genèse du projet

Les Archives Ouvertes de la Connaissance s'inscrivent dans le contexte de développement du *Green Open Access* comme voie complémentaire du modèle de publication traditionnel.

Le projet est né d'une initiative des Services Communs de la Documentation. Il a été formalisé comme ambition commune des quatre établissements alsaciens lors de l'élaboration d'un Schéma Directeur de la Documentation en Alsace, qui visait la mise en place de services pour l'enseignement et la recherche (documentation numérique, archives ouvertes, etc.)⁷⁰. Lui-même a, par la suite, été intégré au contrat de site, signé entre le Ministère de l'Enseignement Supérieur et de la Recherche et les quatre établissements alsaciens que sont l'Unistra, l'UHA, l'INSAS et la BNUS, en date du 4 juin 2013. Ce contrat quinquennal 2013-2017 a initié le passage d'un contrat d'établissement signé entre le Ministère et chaque université à un contrat unique pour l'ensemble des établissements du site régional⁷¹. Les Archives Ouvertes de la Connaissance répondent donc à une logique de coopération régionale. Dans un but de simplification, on évoquera les quatre établissements concernés sous la désignation labellisée de « Site Université de Strasbourg », également nommés « site alsacien ».

En créant une archive institutionnelle, le site Université de Strasbourg concrétise son engagement en faveur du *Green Open Access* et se met en accord avec la recommandation de juillet 2012 de la Commission européenne⁷². A l'Université de Strasbourg, le projet s'accompagnera d'une politique *Open Access*, décrivant les différents services proposés par l'établissement à ses chercheurs, enseignants-chercheurs et doctorants : une plateforme de diffusion des thèses Thèses-Unistra, une offre d'édition en libre accès associant les Presses Universitaires de Strasbourg à la plateforme *Revue.org*⁷³, une archive institutionnelle pour les publications et la littérature grise, ainsi qu'un service de gestion des données de recherche.

La mise en œuvre du projet Archives Ouvertes de la Connaissance a été validée par les Commissions Recherche des établissements en octobre 2014. Elle est soutenue par un financement de 450 000€, que le Ministère de l'Enseignement Supérieur et de la Recherche met à disposition pour une période de cinq ans dans le cadre du Schéma Directeur de la Documentation en Alsace. Le budget du projet prévoit les frais de mission et de matériel informatique –sachant que la plateforme reposera sur un logiciel Open Source- ainsi que le recrutement de quatre CDD :

⁶⁹ Le logiciel Open Source *Islandora* a été conçu à l'Université de l'île du Prince Edouard, au Canada : <http://islandora.ca/>.

⁷⁰ Le projet d'archive institutionnelle fait partie de l'axe 1 du Schéma Directeur de la Documentation : « Améliorer la visibilité de la recherche produite sur le site alsacien ».

⁷¹ Le contrat de site s'est substitué au contrat d'établissement, selon l'article 62 de la loi n°2013-660 du 22 juillet 2013 relative à l'enseignement supérieur et à la recherche.

⁷² Supra : 1.2.2 « Au niveau européen » (p.15)

⁷³ *Revue.org* (<http://www.revues.org/>) est une plateforme de revues en sciences humaines et sociale, faisant partie de l'infrastructure d'édition électronique *OpenEdition* (voir note de bas de page 8).

- un développeur informatique pour la programmation du logiciel ;
- deux ingénieurs d'étude documentalistes pour le dépôt rétrospectif des publications dans l'archive ;
- un ingénieur d'étude pour l'accompagnement des premiers projets de partage de données de recherche.

Si l'UHA envisage la possibilité de recruter un agent pour la gestion des dépôts dans l'archive ouverte, le SCD de l'Unistra a opté pour un plan de formation en interne, qui permettra l'acquisition de nouvelles compétences. Une analyse des besoins sera réalisée à partir de la rentrée 2015 pour déterminer les compétences à développer (langage XML, web sémantique, formats de données, etc.), en vue de l'élaboration d'un Schéma Directeur des Emplois et des Compétences à l'horizon 2016-2017⁷⁴.

Les principales phases du projet Archives Ouvertes de la Connaissance sont présentées en annexe⁷⁵.

2.1.2 Acteurs

- Les *Archives Ouvertes de la Connaissance* regroupent quatre établissements de l'enseignement supérieur et de la recherche :

- l'Université de Strasbourg⁷⁶,
- l'Université de Haute-Alsace (UHA)⁷⁷,
- l'Institut National des Sciences Appliquées de Strasbourg (INSAS)⁷⁸,
- et la Bibliothèque Nationale et Universitaire de Strasbourg (BNUS)⁷⁹.

Si la proximité géographique de ces établissements a favorisé entre eux le développement de partenariats (l'Unistra, l'UHA et l'INSAS comptent trois unités de recherche communes et mutualisent la gestion de leurs collections documentaires au sein d'un plan de conservation partagée), ils n'en sont pas moins différents par leur taille et leurs domaines d'enseignement et de recherche :

- l'INSAS est spécialisé dans les sciences de la société, de l'ingénieur, de l'information et de la communication et intervient au sein de 4 unités de recherche ;
- l'Université de Haute-Alsace et l'Université de Strasbourg sont multidisciplinaires, rassemblant respectivement 14 et 79 unités (dont deux sont communs) ;

⁷⁴ Le Schéma Directeur des Emplois et des Compétences s'inscrit dans un contexte plus large que le seul projet des Archives Ouvertes de la Connaissance ; il tient compte également du développement des ressources numériques, de la diminution des implantations de bibliothèques et de l'ouverture en 2019 d'un *learning center* sur le campus strasbourgeois.

⁷⁵ Annexe 2

⁷⁶ <https://www.unistra.fr>

⁷⁷ <http://www.uha.fr/>

⁷⁸ <http://www.insa-strasbourg.fr/>

⁷⁹ <http://www.bnu.fr/>

- la BNUS, quant à elle, est une bibliothèque de recherche en sciences humaines et sociales, ayant pour mission la conservation et la valorisation du patrimoine, qui n'héberge donc pas d'unités de recherche.

Les Archives Ouvertes de la Connaissance couvriront donc la production de 93 unités de recherche, caractérisées par des statuts divers –certaines font intervenir des établissements associés⁸⁰ et des organismes de recherche⁸¹ (dans le cadre des « unités mixtes de recherche » et « unités propres de recherche »), tandis que d'autres sont des unités universitaires propres (dites « équipes d'accueil »).

Le contrat de site vient ainsi renforcer une coopération existante, où la mise en place d'une archive institutionnelle unique facilitera la gestion des coproductions.

Le projet est piloté par l'Université de Strasbourg et l'Université de Haute-Alsace. Il implique principalement les Services Communs de la Documentation, les Directions Informatiques et les Directions de la Recherche de ces deux établissements, ainsi que ceux de l'INSAS ; les chefs de projet, Adeline Rege et Catherine Lourdel, sont d'ailleurs issues des départements de la documentation numérique –l'une au SCD de l'Unistra, la seconde au SCD de l'UHA.

Le comité de pilotage rassemble les représentants des différents établissements partenaires pour chacun des trois services évoqués, ainsi que les représentants de la Direction des Usages du Numérique et des Presses Universitaires de Strasbourg.

A terme, la gestion de l'archive sera assurée conjointement par la Direction informatique de l'Unistra, pour les aspects techniques (maintenance, assistance aux utilisateurs), et par les Services Communs de Documentation de l'Unistra, l'UHA et l'INSAS, pour les aspects documentaires (indexation, assistance aux utilisateurs).

Le dépôt dans l'archive institutionnelle concernera l'ensemble des chercheurs, enseignants-chercheurs, doctorants et personnels BIATSS⁸². Quant à l'accès, si la consultation de certains documents sera réservée à l'Intranet des établissements –étudiants inclus-, l'ensemble des notices bibliographiques sera visible en ligne. Les Archives Ouvertes de la Connaissance seront donc « ouvertes » aux acteurs de la recherche comme au grand public.

2.2. Racines du projet

2.2.1. Un contexte politique favorable⁸³

Le projet bénéficie, entre autres, des partenariats auxquels est liée l'Université de Strasbourg.

A l'échelle nationale, l'établissement fait, d'une part, partie du consortium Couperin⁸⁴, qui négocie les offres d'abonnement avec les éditeurs des grandes

⁸⁰ L'ENGEES et l'ENSAS.

⁸¹ Le CNRS, l'INSERM, l'INRA, l'INRIA et l'ISL.

⁸² Personnels ingénieurs, administratifs, techniques, sociaux et de santé et des bibliothèques.

⁸³ Notre recensement n'est pas exhaustif. Nous citerons, dans cette partie, les principaux atouts politiques qui concernent l'Université de Strasbourg, où a été encadré notre stage.

revues scientifiques ; il est représenté, au sein du Groupe de Travail sur l'Accès Ouvert (GTAO) du consortium, par la responsable de la Documentation scientifique numérique du SCD, également copilote du projet Archives Ouvertes de la Connaissance. D'autre part, il participe aux réflexions menées par la Bibliothèque Scientifique Numérique, puisque le vice-président Recherche délégué et porteur politique du projet d'archive ouverte, Paul-Antoine Hervieux, copilote le nouveau segment BSN10 consacré aux données de recherche⁸⁵. Selon Dominique Wolf, directrice du SCD⁸⁶, c'est un atout pour le projet du site alsacien mais aussi pour l'ensemble des universités françaises, qui sont globalement peu représentées au sein des groupes de travail sur la recherche scientifique.

L'appui politique des directions d'établissements profite également au projet. L'université de Strasbourg bénéficie du cadre de la *League of European Research Universities* (LERU), dont elle fait partie et qui s'est engagée dans le mouvement de l'Open Access. Alain Beretz, président de l'Unistra, a été sensible à la stratégie qu'elle développe en matière de libre accès et de gestion des données de recherche⁸⁷. Aussi sa désignation à la présidence de l'association en 2014 a-t-elle constitué un « pas politique décisif »⁸⁸ pour la promotion du projet d'archive institutionnelle.

2.2.2. Des expériences sur lesquelles prendre appui

Si l'engagement et la concertation des instances directrices jouent un rôle important dans la promotion du projet, sa réussite dépend également de l'implication des informaticiens et professionnels de la documentation, grâce auxquels s'élabore une réflexion technique et juridique.

Jusqu'à présent, le site Université de Strasbourg ne disposait pas d'archive institutionnelle ; des tentatives ont néanmoins préexisté aux Archives Ouvertes de la Connaissance.

Ainsi, en 2009, le Schéma Directeur du Numérique de l'Unistra a esquissé un axe *Open Access*, dans lequel figurait un projet d'archive institutionnelle. Il n'a pu émerger du fait de la fusion, la même année, des trois universités⁸⁹ dont résulte aujourd'hui l'Université de Strasbourg (faute de moyens, la priorité a en effet été donnée à l'organisation d'un SIGB commun).

Par ailleurs, des initiatives similaires ont été mises en place, à plus petite échelle, dans les différents établissements : l'Université de Haute-Alsace a mis en place dès 2009 une plateforme de dépôt des publications ; l'INSAS a créé en 2012

⁸⁴ <http://www.couperin.org/>

⁸⁵ Infra : 1.2. « Evolution épistémologique de la recherche scientifique » (p.39)

⁸⁶ Entretien du 17.06.15.

⁸⁷ La politique de libre accès de la LERU se manifeste à travers deux publications principales : la *LERU Roadmap Towards Open Access* (2011) et la *LERU Roadmap for Research Data* (2014). Voir infra : 1.3. « Des mouvements incitatifs » (p.40)

⁸⁸ Propos tenus par Dominique Wolf lors d'un entretien, le 17.06.15.

⁸⁹ L'Université Louis Pasteur, pour les Sciences techniques et la Médecine ; l'Université Marc Bloch, pour les Sciences humaines et sociales ; l'Université Robert Schumann, pour le Droit, l'Economie et les Sciences politiques.

un portail HAL pour l'une de ses unités de recherche ; l'Université de Strasbourg diffuse, quant à elle, ses thèses sur Thèses-Unistra depuis 2012.

Le SCD de l'UHA propose une solution d'archivage des publications scientifiques dans le SIGB Flora de la bibliothèque. Les unités de recherche transmettent leurs publications au SCD, qui se charge du renseignement des métadonnées et de l'archivage des fichiers. Le SCD récupère également dans l'archive nationale HAL les notices bibliographiques qui concernent les chercheurs de l'université. Si le nombre restreint d'unités de recherche à l'UHA permet aux bibliothécaires de prendre en charge une large part du processus de dépôt, cette offre de service est en réalité peu sollicitée. Seul le laboratoire *Modélisation, Intelligence, Processus, Systèmes* (EA 2332 –MIPS), dont le directeur est sensible aux valeurs du libre accès, dépose de manière systématique ses publications dans le SIGB Flora.

A l'INSAS, l'unité de recherche LGéCo (*Laboratoire de Génie de la Conception*⁹⁰) possède sa propre collection dans l'archive nationale HAL⁹¹. La collection est tenue par la responsable de la bibliothèque, qui « tamponne » les publications déjà déposées dans HAL, dont l'auteur ou co-auteur est affilié à LGéCo. Pour ces raisons, on ne peut pas véritablement parler d'archive institutionnelle, puisque l'auto-archivage relève de la pratique individuelle de quelques chercheurs ; la plupart des articles sont d'ailleurs déposés dans HAL par des co-auteurs issus de laboratoires extérieurs. Néanmoins, la collection permet à la direction de l'unité d'évaluer la « productivité » de ses chercheurs et à ces derniers d'apprécier la visibilité de leurs articles, grâce aux outils bibliométriques proposés sur la plateforme HAL.

Le SCD de l'Unistra a conçu sa propre plateforme de dépôt et diffusion des thèses et mémoires électroniques, Thèses-Unistra⁹², administrée à temps partiel par quatre agents du département Documentation Scientifique Numérique. Le dépôt s'effectue à partir de l'application nationale Star⁹³ ; la diffusion à partir de la plateforme ORI-OAI⁹⁴. L'outil développe des fonctionnalités qui seront également exploitées dans la future archive institutionnelle, notamment une modalité de diffusion restreinte en Intranet, la conformité avec le protocole OAI-PMH permettant l'interconnexion avec la plateforme TEL-HAL⁹⁵ et l'application thèses.fr de l'ABES⁹⁶, ainsi que la possibilité d'ajouter des fichiers liés au document principal de la thèse.

⁹⁰ Le laboratoire LGéCo est la seule unité propre de l'INSA ; il dénombre 14 enseignants-chercheurs et 20 doctorants.

⁹¹ La collection du laboratoire est consultable en ligne à l'adresse URL : <https://hal.archives-ouvertes.fr/LGECO>. A la date du 22.04.15, la collection rassemblait 222 notices de documents, dont 31% accompagnées du texte intégral.

⁹² <http://theses.unistra.fr/ori-oai-search/index.html>

⁹³ STAR est destinée au dépôt national des thèses électroniques françaises : <http://star.theses.fr/>

⁹⁴ ORI-OAI (Outil de Référencement et d'Indexation, réseau de portails OAI) est un logiciel pour la gestion des ressources numériques de l'enseignement supérieur : <http://www.ori-oai.org/>

⁹⁵ TEL-HAL (Thèses En Ligne) est le serveur de thèses multidisciplinaire de l'archive nationale HAL : <https://tel.archives-ouvertes.fr/>

⁹⁶ <http://www.theses.fr/>

2.3. Les objectifs spécifiques du projet

S'appuyant sur les acquis des archives institutionnelles développées ces deux dernières décennies, les Archives Ouvertes de la Connaissance sont aujourd'hui en mesure de proposer un outil qui n'ait plus seulement le rôle de plateforme de stockage et de diffusion des publications, mais aussi celui de services à valeur ajoutée. En tant qu'outil de communication scientifique directe, les archives institutionnelles participent à la valorisation des résultats de la recherche. Or les enjeux dans ce domaine ont évolué depuis la création des archives ouvertes ; ce sont ces enjeux qui justifient qu'elles continuent d'exister et, par conséquent, ceux auxquels elles doivent s'adapter. Nous verrons ainsi en quoi le projet alsacien propose d'enrichir les fonctionnalités -exposées en 1.3.- des archives institutionnelles traditionnelles.

2.3.1. Intégrer les données de recherche

De plus en plus mise en avant par les institutions et les financeurs de la recherche, les données de recherche ont été intégrées aux axes prioritaires du projet Archives Ouvertes de la Connaissance. L'objectif est de les archiver et de les identifier de manière pérenne, afin qu'elles soient « citables » et permettent de valider et de reproduire les résultats de recherche. Elles seront traitées dans la perspective d'être liées aux publications. Plusieurs études⁹⁷ ont en effet démontré que la liaison entre articles scientifiques et données contribuait au « succès » des publications et à leur citation.

Le site alsacien compte parmi les deux premières initiatives universitaires à intégrer les données de recherche à un service d'archive institutionnelle. La seconde est, elle aussi, en cours de lancement, à l'Université de Lille 1.⁹⁸

2.3.2. Couvrir l'ensemble de la production scientifique du site

Les Archives Ouvertes de la Connaissance la diffusion d'un large éventail de résultats scientifiques, en ouvrant le dépôt non pas seulement à la partie « émergée de l'iceberg » que sont les articles publiés, mais à l'ensemble de la production scientifique du site, incluant communications, rapports, prépublications, habilitations de recherche, données de recherche, etc. (seuls les mémoires et thèses continueront à être diffusés sur la plateforme dédiée Thèses-Unistra et dans la bibliothèque numérique de l'UHA). Les institutions de recherche, en tant que lieu de production, sont d'autant mieux placées pour rendre visible l'étendue des activités des chercheurs (les résultats publiés comme la littérature grise⁹⁹). Une

⁹⁷ REILLY Susan, SCHALLIER Wouter, SCHRIMPF Sabine, SMIT Eefke & WILKINSON Max. *Report on integration of data and publications*. ODE, 17 octobre 2011. <http://www.stm-assoc.org/2011_12_5_ODE_Report_On_Integration_of_Data_and_Publications.pdf> [en ligne]

CODATA-ICSTI Task Group on Data Citation Standards and Practices. *Out of cite, out of mind : the current state of practice policy, and technology for the citation of data*. Data Science Journal, Volume 12, 13 septembre 2013. DOI: 10.2481/dsj.OSOM13-043 [en ligne]

⁹⁸ Nous présenterons le projet de l'Université Lille 1 dans la seconde partie de ce mémoire. Quant aux autres initiatives concernent des organismes de recherche, à savoir l'Institut National de la Recherche Agronomique (INRA) et l'Institut Français de Recherche pour l'Exploitation de la Mer (IFREMER).

⁹⁹ Voir note de bas de page 27.

étude d'Hélène Prost et Joachim Schöpfel¹⁰⁰ établissait ainsi en 2009, à partir d'un recensement des archives ouvertes françaises, que 94% des archives institutionnelles contenaient de la littérature grise (contre 37% des archives thématiques).

En se dotant d'une archive ouverte, le site alsacien créera, par ailleurs, une égalité de diffusion entre ses chercheurs et, plus particulièrement, entre les disciplines de recherche. Il contribuera notamment à la visibilité des publications du domaine des sciences humaines et sociales, dont le référencement dans les bases de dépouillement internationales (*Scopus*, *Web of Science*) demeure faible (de l'ordre de 20%).

Les Archives Ouvertes de la Connaissance constitueront donc un outil de pilotage pour les établissements, qui auront ainsi la maîtrise de la production scientifique de leurs unités de recherche. Cette décision relève moins d'une volonté de contrôle que du choix de ne plus dépendre du joug financier des acteurs privés de l'édition.

2.3.3. Instaurer une politique d'archivage

De la définition d'une politique d'établissement dépend l'objectif de centralisation de la production scientifique dans l'archive institutionnelle. Dans le cas des Archives Ouvertes de la Connaissance, chacun des établissements décidera de la politique qu'il souhaite mettre en place. A l'Université de Strasbourg, la Commission Recherche réunie en octobre 2014 a choisi de faire de l'archive ouverte la bibliographie officielle de l'établissement : une charte écrite incitera chercheurs, enseignants-chercheurs et doctorants à y déposer l'ensemble de leurs publications (articles, ouvrages, actes de colloques, etc.), texte intégral compris. Les fonctionnalités d'export de la plateforme permettront, en effet, de faciliter grandement le travail de recensement des publications par l'administration de l'université. Cette décision implique de définir des stratégies.

La première consiste à définir les modalités du dépôt par les chercheurs : obligation ou recommandation. Le Service Commun de la Documentation prône l'obligation, en vertu de l'efficacité de l'archive –qui parfois s'avère diminuée en raison de l'insuffisance des dépôts. Alma Swan a montré, en effet, que la plupart des auteurs n'étaient pas prêts à auto-archiver leurs publications, à moins d'y être contraints par leur établissement¹⁰¹.

La seconde stratégie consiste à définir les modalités d'accès aux documents, conciliant le principe de libre accès (requis de plus en plus fréquemment par les programmes de financement) avec la juste rémunération du travail de l'auteur – voire, pour ce qui est des travaux publiés chez un éditeur, avec les conditions de publication parallèle telles que définies dans le contrat d'édition. C'est pourquoi les Archives Ouvertes de la Connaissance, proposeront, tout comme la plupart des archives institutionnelles, plusieurs niveaux de granularité pour la diffusion des documents : par défaut, celle-ci sera libre, mais le déposant pourra choisir une

¹⁰⁰ Pages 22-23 : SCHÖPFEL Joachim, PROST Hélène. *Développement et Usage des Archives Ouvertes en France. 1e partie : Développement*. 2010, 49 p. <http://archivesic.ccsd.cnrs.fr/sic_00497389> [en ligne]

¹⁰¹ SWAN Alma, *Open Access Self-Archiving : An introduction*. Mai 2005, <<http://cogprints.org/4406/>> [en ligne]

diffusion plus restreinte, soit en Intranet, soit après une période d'embargo¹⁰², soit sur demande de « tiré à part »¹⁰³.

L'intérêt, pour l'Université de Strasbourg, d'utiliser l'archive institutionnelle comme bibliographie officielle est de fournir une liste « juste » des publications qui y sont produites. Cette liste pourra servir de référence pour les répertoires bibliographiques, mais aussi pour l'évaluation des chercheurs et enseignants-chercheurs.

- par les organismes de financement de la recherche, dans le cadre de demandes de subvention
- par l'université, dans le cadre de promotions ou titularisations
- par le Haut conseil de l'évaluation de la recherche et de l'enseignement supérieur (HCERES)¹⁰⁴, qui évaluera l'impact de la recherche sur le site alsacien au second semestre 2016.

Une des possibilités envisagées par l'Université de Strasbourg en vue d'un mandat de dépôt serait de lier les procédures d'évaluation au dépôt dans l'archive institutionnelle. Ce modèle, initié par l'Université de Liège en 2007¹⁰⁵ et réitéré par l'Initiative de Budapest en 2012¹⁰⁶, veut que le dépôt des publications dans l'archive soit la seule voie pour les soumettre à l'évaluation des instances universitaires et nationales. Ainsi, en fondant l'évaluation des chercheurs sur les documents conservés dans l'archive de l'établissement, le dépôt n'est pas rendu obligatoire dans les termes, mais devient incontournable.

Si l'efficacité du mandat institutionnel a été prouvée, elle ne saurait toutefois se substituer à un dispositif de sensibilisation et d'accompagnement des chercheurs.

2.3.4. Accompagner les chercheurs

¹⁰² Voir note de bas de page 43.

¹⁰³ Par la fonction « tiré à part » (*request a copy*), l'auteur choisit de soumettre l'accès de son document à la demande spécifique de l'utilisateur : un bouton au niveau de la notice bibliographique proposera l'envoi d'un e-mail à l'auteur ou l'administrateur de la plateforme, qui pourra en retour faire parvenir une « copie » de son document.

¹⁰⁴ Le Haut conseil de l'évaluation de la recherche et de l'enseignement supérieur (HCERES) se substitue à l'Agence d'évaluation de la recherche et de l'enseignement supérieur (AERES), selon les termes de la loi n° 2013-660 du 22 juillet 2013 relative à l'enseignement supérieur et à la recherche. Cette nouvelle autorité administrative indépendante est chargée de l'expertise scientifique des établissements de l'enseignement supérieur et de la recherche ; son évaluation prend notamment en considération la valorisation des résultats de la recherche et le développement de la culture scientifique et technique de l'établissement. A ce titre, doit être fournie la liste des publications de chaque unité de recherche.

¹⁰⁵ L'Université de Liège (Belgique) a instauré dès 2007 un mandat de dépôt des publications dans son archive institutionnelle ORBi (*Open Repository and Bibliography*, <http://orbi.ulg.ac.be/>). Depuis 2014, les évaluateurs sont contraints de « ne considérer que les listes de publications provenant de ORBi pour l'évaluation des demandes de nomination, promotion ou attribution de crédit ».

¹⁰⁶ « Universities with institutional repositories should require deposit in the repository for all research articles to be considered for promotion, tenure, or other forms of internal assessment and review ». In BOAI 10, « Ten years on from the Budapest Open Access Initiative: setting the default to open ». Budapest. 12 septembre 2012. <<http://www.budapestopenaccessinitiative.org/boai-10-recommendations>>. Traduction française disponible en ligne : <<http://www.budapestopenaccessinitiative.org/boai-10-translations/french>>

En simplifiant les processus

Il semble naturel que les chercheurs se rallient d'autant plus facilement à la pratique de l'auto-archivage que la procédure est simple. La centralisation des publications sur une seule plateforme joue donc en faveur des Archives Ouvertes de la Connaissance. Les chercheurs procéderont, par ailleurs, à un dépôt unique et à une seule saisie des métadonnées, tout en ayant l'assurance que le document sera versé dans HAL ainsi que dans d'autres archives thématiques et indexé dans les moteurs de recherche (voir 1.3.3., *Un référencement exhaustif*).

Suivant la tendance des archives ouvertes à se doter de services à valeur ajoutée, les Archives Ouvertes de la Connaissance proposeront des fonctionnalités annexes, permettant de réutiliser les données de manière automatique et étendue – par exemple l'export de listes de publications, l'édition de CV, la production de statistiques et de résultats bibliométriques, etc.

En proposant un dispositif perfectible et adapté aux équipes de recherche du site

L'avantage d'une archive institutionnelle est d'assurer à l'établissement la maîtrise de la base de données et du logiciel –ceux-ci étant administrés en interne. De fait, il permet une meilleure réactivité –plus rapide et plus adaptée- aux besoins des unités de recherche et aux demandes d'évolution de la plateforme. Les procédures de dépôt (*workflows*) peuvent, par exemple, être adaptées aux spécificités des unités de recherche.

Au dispositif technique s'ajoute la proximité d'une équipe dédiée, qui s'appuiera, dans le cadre du projet alsacien, sur les compétences des Directions informatiques et des Services Communs de la Documentation. Les bibliothécaires documentalistes seront chargés notamment d'accompagner les procédures de dépôt, en aidant à la saisie des métadonnées et en renseignant les déposants sur des questions relatives aux droits de diffusion ou encore à la gestion des données de recherche.

En faisant le choix d'un outil local

La solution d'un portail HAL n'a pas été retenue car elle ne répondait pas de manière exhaustive aux besoins du site alsacien. On peut en effet s'interroger sur la pertinence d'un investissement budgétaire et humain, alors que la plate-forme HAL propose la création de collections institutionnelles propre à chaque université ou institut de recherche. Cette dernière possède de nombreux avantages, entre autres la garantie d'une gestion simplifiée de la collection institutionnelle et celle d'une conservation pérenne des publications au CINES ; cependant elle présente les inconvénients de proposer une typologie documentaire restreinte¹⁰⁷ et de priver l'institution du contrôle de sa politique de dépôt. Le site alsacien a fait le choix d'un outil local, car il permettait de gérer plusieurs niveaux d'accès aux publications (diffusion sur Internet, diffusion sur l'intranet, demande de tiré à part) et rendait possible une meilleure insertion dans le système d'information de l'établissement, grâce à l'utilisation des identifiants de connexion ENT.

¹⁰⁷ A l'époque où les Archives Ouvertes de la Connaissance ont été conçues, la plateforme HAL ne proposait pas encore de fonctionnalité pour les données de recherche ; la version HAL v3 inclut depuis octobre 2014 une nouvelle catégorie permettant de déposer des sons, images, cartes, vidéos et présentations.

La gestion de la production scientifique du site demande un investissement initial en temps et en ressources. Il sera compensé par des économies en termes d'efficacité, par une meilleure gestion des risques et par la possibilité d'exploitation complète des données.

Le défi des archives ouvertes institutionnelles réside dans la valorisation du travail scientifique ; les archives se veulent à la fois une vitrine de la recherche, telle que se la représente le débat public aujourd'hui, et un dispositif d'accompagnement de la recherche. Elles recouvrent donc divers types de documents, dont la diffusion implique de veiller à leurs spécificités. Parmi ces documents, les données de recherche sont une catégorie encore neuve des archives ouvertes. L'enjeu actuel qu'elles représentent et la découverte d'un potentiel important sur le site alsacien a conduit, comme nous l'avons évoqué en introduction, à faire évoluer la commande de stage. Le volet consacré aux données de recherche ayant rencontré un intérêt plus vif auprès des chercheurs que le volet sur les revues scientifiques, il est devenu une priorité de la mission de stage. Aussi la partie qui suit sera-t-elle consacrée spécifiquement aux données de recherche.

PRATIQUES DE GESTION DES DONNEES DE LA RECHERCHE SUR LE SITE ALSACIEN : ETUDE EXPLORATOIRE

1. L'OUVERTURE DES DONNEES DE RECHERCHE

1.1. Questionnement autour de la définition des données de recherche

Sous un angle extérieur, la notion de « données de (la) recherche » (*research data*) se définit simplement : ce sont des données numériques ou numérisées, produites au cours d'un processus de recherche. Cette notion couvre en réalité un très large périmètre. Les données de recherche peuvent en effet désigner une diversité de matériaux, qui est fonction : de la discipline scientifique considérée ; de la manière dont elles ont été obtenues, mais aussi de la forme sous laquelle elles ont été enregistrées et du traitement qu'elles ont subi pour être rendues lisibles. Ainsi l'exprimait Sylvie Fayet, conservatrice des bibliothèques à l'URFIST de Toulouse, dans un billet de blog : « 'données de la recherche' me semble être cette grande valise que me confie un voyageur bien sous tous rapports mais dont je n'ai pas vérifié le contenu »¹⁰⁸.

En atteste la typologie qui a pu être établie à partir de l'enquête réalisée sur le site Université de Strasbourg¹⁰⁹ fin 2014.

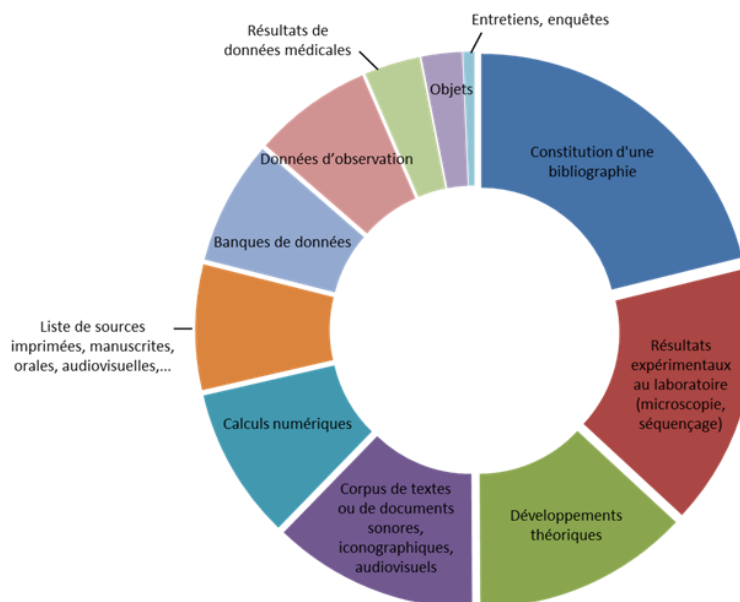


Figure 1 : Typologie des données utilisées au cours de travaux de recherche sur le site alsacien

¹⁰⁸ FAYET Sylvie, « "Données" de la recherche, les mal-nommées », 15 novembre 2013, <http://urfistinfo.hypotheses.org/2581>

¹⁰⁹ Les objectifs et la méthodologie de l'enquête seront présentés en 2.1.2. « Analyse d'une première enquête... » (p.43).

Le diagramme révèle que ces données ne sont pas toutes d'ordre expérimental, autrement dit qu'une part d'entre elles (corpus de documents, banques de données, résultats médicaux, etc.) sont très probablement soumises à des contraintes juridiques (droit d'auteur, confidentialité, etc.). Sur le plan technique, certaines de ces données ne sont pas *de facto* diffusables pour une utilisation à distance –ne répondant pas nativement au format numérique- voire perdraient leur intérêt à être exploitées sous forme numérisée.

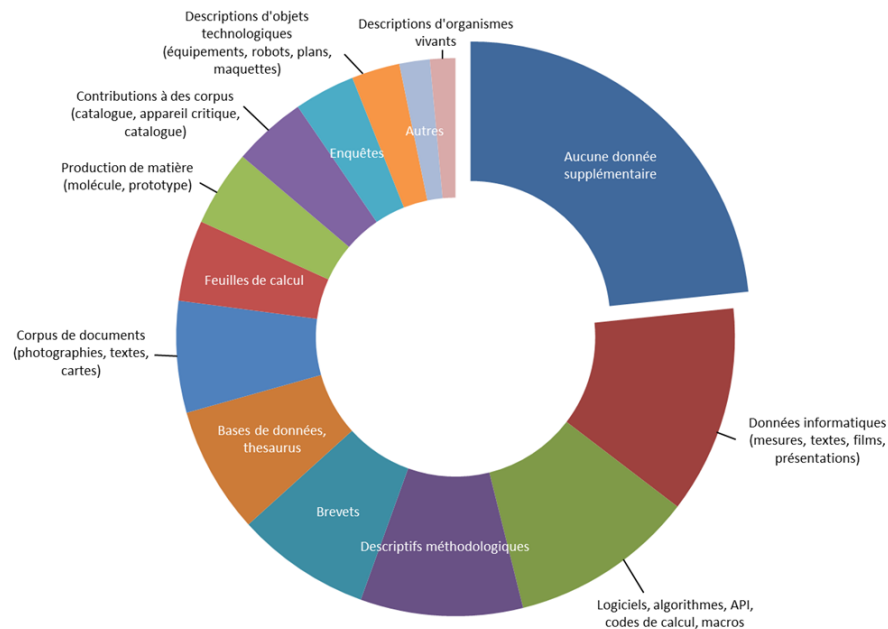


Figure 2 : Typologie des données produites sur le site alsacien en sus des publications scientifiques

Ce second diagramme laisse entrevoir que, pour 77 % des répondants, les publications (articles, ouvrages, communications, etc.) ne sont pas le seul produit de leurs recherches –lesquelles s'accompagnent d'un ensemble de données plus ou moins élaborées. La granularité du diagramme ne permet cependant pas d'estimer : leur lien avec les publications ; leur répartition par discipline ; par chercheur ou par projet de recherche. Néanmoins, considérons qu'à une échelle macroscopique, elles constituent un ensemble important, diversifié et d'une valeur certaine.

Ce mémoire ne se prêtant pas, toutefois, à une discussion sur la typologie des données de recherche, nous nous contenterons de tenir compte de leur diversité, en nous fondant sur les définitions qui ont pu être proposées dans des travaux antérieurs :

- Celle de l'OCDE¹¹⁰ notamment : « les 'données de la recherche' sont définies comme des enregistrements factuels (chiffres, textes, images et sons), qui sont utilisés comme sources principales pour la recherche scientifique et sont généralement reconnus par la communauté scientifique comme nécessaires pour valider des résultats de recherche » ;

¹¹⁰ ORGANISATION DE COOPERATION ET DE DEVELOPPEMENT ECONOMIQUES (OCDE), *Principes et lignes directrices de l'OCDE pour l'accès aux données de la recherche financée sur fonds publics*, Paris, 2007, 28 p., <http://www.oecd.org/fr/science/sci-tech/38500823.pdf>

- Celle de la Royal Society¹¹¹ : les données sont « des informations qualitatives ou quantitatives [...] qui sont factuelles. [Elles] peuvent être brutes ou primaires (directement issues d'une mesure), ou dérivées de données primaires, mais ne sont pas encore le produit d'analyse ou d'interprétation autres que de calculs ».

On pourra se reporter au mémoire de Rémi Gaillard¹¹², qui les analyse et en suggère les limites.

Le libre accès aux données de recherche a été théoriquement défini dans la déclaration de Berlin en 2003¹¹³. Dans la pratique, la notion de partage des données de recherche existe dans certaines disciplines depuis bien plus longtemps, en astronomie, génomique et cristallographie notamment¹¹⁴. Qu'est-ce qui aujourd'hui en fait un enjeu plus prégnant pour la science comme pour la société ?

1.2. Evolution épistémologique de la recherche scientifique : de l'Open Access à l'Open Science

La question des données de la recherche s'insère dans un ensemble de projets plus vaste : celui de la large diffusion des données publiques, mais aussi celui d'une science ouverte (*Open Science*).

Le terme d'*open science* a été défini par la Royal Society comme la « combinaison entre données ouvertes, publications scientifiques en libre accès et communication efficace de leurs contenus »¹¹⁵. Il repose sur des principes inhérents à la science, comme celui de la « cumulativité », qui induit que la recherche progresse grâce à la communication des travaux antérieurs. Le rythme de ces échanges a été accéléré par l'émergence du numérique. Aujourd'hui, les chercheurs ont la possibilité de partager non plus seulement leurs connaissances, mais aussi les résultats qui y ont conduit. C'est en partie pourquoi le contenu des archives ouvertes évolue vers une typologie couvrant l'ensemble du processus de recherche, depuis les données sources jusqu'à la version finale de la publication.

Le numérique offre également aux scientifiques la possibilité d'exploiter de nouveaux supports. Il permet désormais de produire, d'enregistrer et d'analyser de

¹¹¹ THE ROYAL SOCIETY. *Science as an open enterprise : summary report*. The Royal Society, juin 2012, 104 p. <http://royalsociety.org/uploadedFiles/Royal_Society_Content/policy/projects/sape/2012-06-20-SAOE.pdf> [en ligne]

« Qualitative or quantitative statements or numbers that are (or assumed to be) factual. Data may be raw or primary data (eg direct from measurement), or derivative of primary data, but are not yet the product of analysis or interpretation other than calculation » p.9

¹¹² GAILLARD Rémi. *De l'Open data à l'Open research data : quelle(s) politique(s) pour les données de recherche ?* Mémoire, Enssib, 2014, 104 p. <<http://www.enssib.fr/bibliotheque-numerique/notices/64131-de-l-open-data-a-l-open-research-data-quelles-politiques-pour-les-donnees-de-recherche>> [En ligne] p.15-21 « Data et Datasets : essai de définition »

¹¹³ Supra : 1.1.2. « A quels principes répondent-elles » (p.11)

¹¹⁴ GAILLARD Rémi. *De l'Open data à l'Open research data : quelle(s) politique(s) pour les données de recherche ?* Mémoire, Enssib, 2014, 104 p. <<http://www.enssib.fr/bibliotheque-numerique/notices/64131-de-l-open-data-a-l-open-research-data-quelles-politiques-pour-les-donnees-de-recherche>> [En ligne] p.42-43, « Logique bottom-up et culture du partage »

¹¹⁵ « Open data (available, intelligible, assessable and usable data) combined with open access to scientific publications and effective communication of their contents », p.9. In : THE ROYAL SOCIETY. *Science as an open enterprise : summary report*. The Royal Society, juin 2012, 104 p. <http://royalsociety.org/uploadedFiles/Royal_Society_Content/policy/projects/sape/2012-06-20-SAOE.pdf> [en ligne]

vastes ensembles de données ; cette forme de traitement intensif des données a été appelé *data-intensive science*. Le numérique a donc la capacité de transformer fondamentalement le développement d'une discipline scientifique, dans un contexte multidisciplinaire et international.

D'un point de vue épistémologique, la science ouverte fait du partage de l'information scientifique un gage de qualité. L'ouverture des résultats permet en effet d'encourager la collaboration entre les chercheurs et de favoriser la transparence des processus de recherche, à des fins d'évaluation et de vulgarisation. L'*Open Science* soulève, par ailleurs des enjeux d'ordre socio-économique : communiquer les résultats est un moyen de rentabiliser les coûts, parfois onéreux, des processus de recherche et d'accélérer l'innovation.

Les 6^{èmes} journées Open Access du consortium Couperin, qui se dérouleront du 12 au 14 octobre 2015, porteront sur *l'Open Access et l'Open Science*, signe de l'intérêt que suscite ce nouveau paradigme de la science.

1.3. Des mouvements incitatifs

Malgré un consensus difficile sur la définition du libre accès et des données de recherche, des efforts croissants sont entrepris au niveau national et international pour promouvoir le développement de *l'Open Research Data*. Comme l'illustre le schéma ci-dessous, les initiatives se multiplient. Nous rappellerons brièvement les principaux mouvements incitatifs qui jalonnent l'ouverture des données de recherche dans le contexte français.

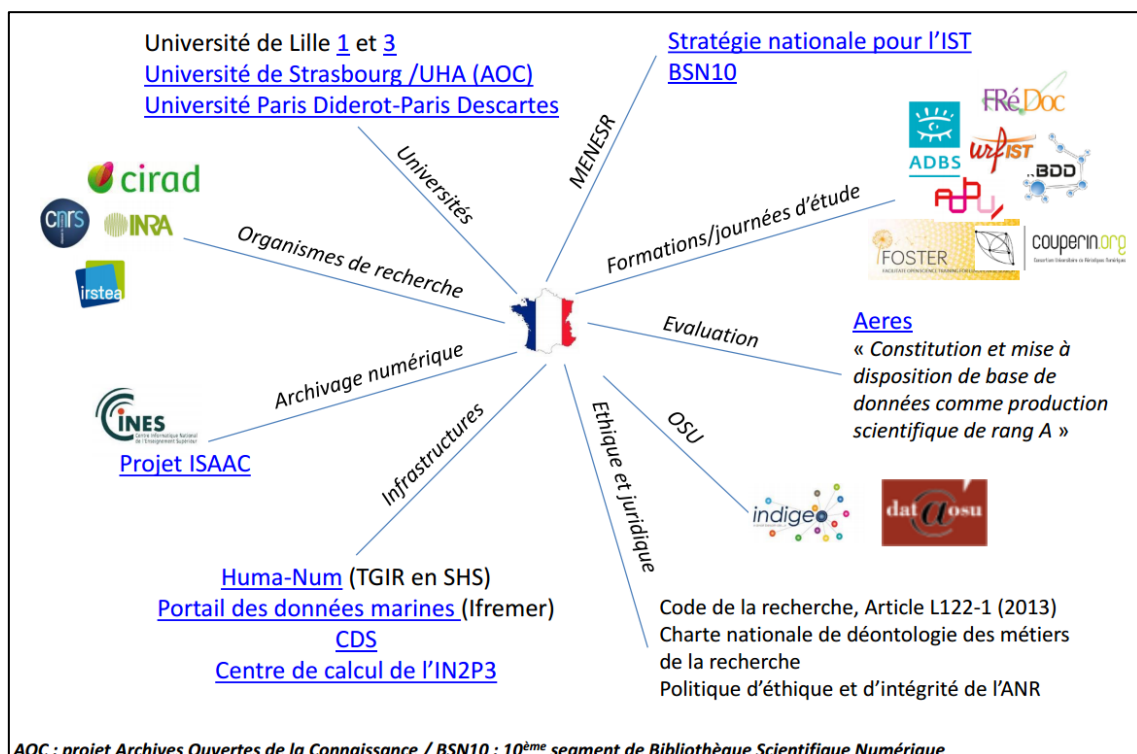


Figure 3 : Initiatives françaises autour de la gestion des données de recherche

L'impulsion vient :

- Des gouvernements

En 2004, la France signe la *Déclaration sur l'accès aux données de la recherche financée par des fonds publics*¹¹⁶ de l'Organisation de Coopération et de Développement Économiques (OCDE).

Le Ministère de l'Enseignement Supérieur et de la Recherche a récemment mis en place un cadre de réflexion sur les données de la recherche : le segment BSN10 de la Bibliothèque Scientifique Numérique.

- Des financeurs

En 2013, la Commission européenne a lancé l'*Open Research Data Pilot*¹¹⁷, un projet pilote de libre accès aux données issues de la recherche financée sur fonds publics, dans le cadre du programme H2020. Ce projet-test a pour but « d'améliorer et d'optimiser l'accès et la réutilisation de [ces] données » et repose sur les cinq principes suivants: « Les données de recherche scientifiques devraient être facilement découvrables [c'est-à-dire repérables et identifiables], accessibles, évaluables et compréhensibles, utilisables au-delà du but premier de leur collecte [c'est-à-dire stockées de manière pérenne et documentées] et interopérables selon des normes qualitatives spécifiques [c'est-à-dire échangeables et combinables] ». L'initiative porte sur sept thématiques du programme H2020, soit 20% du budget total alloué pour 2014 et 2015, ainsi que tout autre projet volontaire. « Les projets peuvent, à n'importe quel moment, se désengager du pilote ». Ils bénéficient de l'infrastructure OpenAIRE¹¹⁸, qui a mis au point un entrepôt multidisciplinaire, Zenodo¹¹⁹, chargé de recueillir leurs données de recherche ; l'infrastructure assure également le lien avec les publications. Ce projet pilote est le signe d'un engagement fort des institutions européennes en faveur de l'ouverture des données de recherche ; ce mouvement a été initié par la commission du 17 juillet 2012¹²⁰.

- Des institutions et infrastructures de recherche

Les premières institutions de recherche à s'engager dans le mouvement d'ouverture des données sont le Centre National de la Recherche Scientifique (CNRS), l'Institut National de la Santé et de la Recherche Médicale (INSERM), l'Institut National de la Recherche Agronomique (INRA) et l'Institut National de Recherche en Informatique et en Automatique (INRIA). Elles signent la Déclaration de Berlin en 2003 (CNRS et INSERM) et 2004 (INRA et INRIA). En 2013, l'INRA met en œuvre une politique de partage des données.

Le libre accès aux données de la recherche a également été promu par de nombreux rapports et feuilles de route, dont certains ont été produits dans le cadre de projets ou d'associations comme la *League of European Research Universities* (LERU).

¹¹⁶ *Déclaration sur l'accès aux données de la recherche financée par des fonds publics*. C(2004)31/REV1. Paris, 30 janvier 2004. <<http://acts.oecd.org/Instruments/ShowInstrumentView.aspx?InstrumentID=157&Lang=fr&Book=False>> [en ligne]

¹¹⁷ EUROPEAN COMMISSION, RESEARCH AND INNOVATION, *Guidelines on Data Management in Horizon 2020*, Version 1.0, 11 décembre 2013, 14 p. <http://ec.europa.eu/research/participants/data/ref/h2020/grants_manual/hi/oa_pilot/h2020-hi-oa-data-mgt_en.pdf> [en ligne] p.6

¹¹⁸ OpenAIRE, Open Access Infrastructure for Research in Europe : <https://www.openaire.eu/>

¹¹⁹ Zenodo. <<https://zenodo.org/>>

¹²⁰ Supra : 1.2.2. « Au niveau européen » (p.15)

En 2013, la LERU, qui regroupe 21 universités dont l'Université de Strasbourg, a recommandé de mettre en place une politique institutionnelle pour les données de la recherche¹²¹.

- Des éditeurs scientifiques¹²²

Les éditeurs contribuent, eux aussi, à la diffusion des données de recherche. Nombre d'entre eux se dotent de politiques pour la publication des données liées aux articles de revues ; BioMed Central¹²³ et Nature¹²⁴, par exemple. Une nouvelle offre éditoriale, dédiée à la description de jeux de données, a émergé : les *data journals*. Ils suscitent néanmoins la méfiance des institutions de la recherche publique, qui craignent que les données deviennent à leur tour un marché lucratif pour les grands éditeurs scientifiques.

2. ENQUETES AUPRES DES CHERCHEURS DU SITE ALSACIEN

2.1. Méthodologie

L'étude d'opportunité réalisée au cours du stage s'est appuyée sur une analyse de l'existant (2.1.1.), confrontée à l'analyse des besoins sur le site alsacien (2.1.2. et 2.1.3.).

2.1.1. Analyse de l'existant

En parallèle des recherches menées sur le site alsacien, j'ai entrepris d'analyser quelques-unes des initiatives existantes en matière de gestion de données. Ma démarche a consisté à explorer la documentation en ligne et à collecter des retours d'expériences ; elle avait pour objectifs de compléter les informations déjà réunies et de connaître les acteurs concernés par ces initiatives. Les informations ont été recueillies à l'occasion d'une correspondance par mails ou d'une rencontre, auprès de :

- L'Institut de l'information scientifique et technique (Inist-CNRS)¹²⁵ : l'équipe *Valorisation des données de la recherche*, dirigée par Marie Christine Jacquemot-Perbal, accompagne des équipes de recherche dans la gestion et le partage de leurs

¹²¹ LERU RESEARCH DATA WORKING GROUP. « LERU Roadmap for Research Data ». *Advice paper* No.14, décembre 2013. <http://www.leru.org/files/publications/API4_LERU_Roadmap_for_Research_data_final.pdf> [en ligne]

¹²² Sur le rôle des éditeurs scientifiques, « Incitations éditoriales » (p.35-38) & « Solutions éditoriales et data journals » (p.53-54). In : GAILLARD Rémi. *De l'Open data à l'Open research data : quelle(s) politique(s) pour les données de recherche ?* Mémoire, Enssib, 2014, 104 p. <<http://www.enssib.fr/bibliotheque-numerique/notices/64131-de-l-open-data-a-l-open-research-data-quelles-politiques-pour-les-donnees-de-recherche>> [En ligne]

¹²³ BioMed Central, « Policy on Open Data », <<http://www.biomedcentral.com/about/opendata>> [en ligne]

¹²⁴ Nature, « Availability of data, material and methods », <<http://www.nature.com/authors/policies/availability.html>> [en ligne]

¹²⁵ Dans le cadre du projet « Ingénierie des Connaissances », l'Inist, unité propre de service du CNRS rattachée à la Direction de l'information scientifique et technique (DIST), a réorganisé son offre de services aux chercheurs et à la recherche publique. A ce propos, voir : « Un nouveau départ pour l'Inist », interview du directeur de l'Inist, Raymond Bérard. *CNRS Hebdo*, 11 juillet 2014. <<http://intranet.cnrs.fr/intranet/actus/150711-inist.html>> [en ligne]

données. Son objectif premier n'est pas de leur fournir des outils logiciels, mais plutôt de les orienter vers des bonnes pratiques de gestion de leurs données. En 2013, elle a travaillé avec une unité mixte de recherche en bio-ingénierie du CNRS et de l'Université de Lorraine, ainsi qu'avec une unité du CNRS en sciences de l'environnement.

- L'Université Lille 1 : La Bibliothèque Numérique a créé le projet *Archive Ouverte/Données de la recherche*, dont une première version devrait voir le jour à l'automne 2015. Ce projet associe publications et données de recherche et se concentre, dans un premier temps, sur six unités de recherche, qui joueront le rôle de pilotes. La démarche adoptée repose sur le *Data Audit Framework*, une méthode développée par le JISC et le Data Curation Centre (DCC), permettant de localiser les données et d'identifier les politiques et pratiques existantes autour de leur gestion.
- L'Université Albert Ludwig de Fribourg en Brisgau : le département eScience de la bibliothèque universitaire met à disposition des chercheurs de l'institution des services numériques pour leur production scientifique : une archive ouverte intégrant publications et données (*FreiDok plus*¹²⁶), un environnement de travail pour les échanges entre chercheurs (*tambora.org*), des supports de base de données,... Son objectif est de créer un environnement où l'ensemble de ces services seraient connectés.

Cette analyse de l'existant m'a permis d'identifier l'offre en matière de gestion des données de recherche et d'enrichir de points de confrontation l'étude des besoins sur le site alsacien.

2.1.2. Analyse d'une première enquête sur les pratiques de publication scientifique et de production de données de recherche sur le site alsacien

Dans l'objectif d'adapter l'archive aux besoins du site Université de Strasbourg, l'équipe projet des Archives Ouvertes de la Connaissance a réalisé une enquête, diffusée en ligne entre le 1^{er} décembre 2014 et le 23 janvier 2015, auprès des chercheurs, enseignants-chercheurs, doctorants et personnels BIATSS¹²⁷ du site. Cette enquête avait pour objectif de collecter des informations quantitatives et qualitatives sur les « pratiques de publication scientifique et de production de données », afin de mieux connaître la manière dont les unités référencent, conservent et diffusent leurs résultats de recherche sur le site alsacien. Le questionnaire se structurait en deux axes distincts : le premier portait sur les pratiques de publication ; le second, auquel nous nous intéresserons dans cette partie, sur la production et l'utilisation des données de recherche. Les questions de ce dernier axe sont reproduites en annexe¹²⁸. Elles proposaient une distinction entre

¹²⁶ FreiDok plus. <http://www.freidok.uni-freiburg.de/>

¹²⁷ Personnels ingénieurs, administratifs, techniques, sociaux et de santé et des bibliothèques.

¹²⁸ Annexe 3

données d'entrée et données produites¹²⁹. Pour chaque catégorie, les réponses attendues étaient soit d'ordre quantitatif (volumétrie produite, sauvegardée et réutilisée), soit d'ordre qualitatif (mode de sauvegarde, politique de sélection, pratique de partage,...).

Un premier travail de tri sous Excel ayant été réalisé par une stagiaire, ma mission a consisté à analyser les données recueillies. L'objectif était de mettre en évidence les grandes lignes qui se dégagent des résultats de l'enquête sur le plan des pratiques de gestion des données par les chercheurs, afin de pouvoir ensuite élaborer une réflexion sur les services qu'il serait pertinent de mettre en place pour la conservation et la diffusion des données de recherche.

D'un point de vue méthodologique, les résultats ont été analysés sous Excel selon un tri à plat et présentés sous forme de statistiques illustrées par des graphiques, pour les réponses aux questions fermées, sous forme de synthèses thématiques, pour les réponses aux questions ouvertes.

Pour calculer les taux de représentation, je me suis appuyée sur le recensement des unités de recherche, publié en janvier 2015 par les établissements du site alsacien. Le taux de retours étant supérieur à 30%, les résultats sont considérés comme représentatifs du site alsacien. Les répondants sont principalement des enseignants-chercheurs, des chercheurs et des doctorants. Sur les 88 unités de recherche qu'héberge le site Université de Strasbourg, 80 sont représentées dans l'enquête. Si le nombre d'unités de recherche en Sciences humaines et sociales sur le site alsacien est moins important que celui des unités en Sciences, Technologies, Vie et Santé (35 unités contre 53), le taux de participation à l'enquête, en nombre de réponses, est sensiblement le même.

Les résultats de l'enquête ont été présentés au comité de pilotage le 27 mars 2015, ainsi qu'aux chercheurs de l'INSAS lors de la Journée de la Recherche¹³⁰. De manière globale, il ressort de l'enquête un besoin d'accompagnement et de simplification des processus.

2.1.3. Entretiens avec des enseignants-chercheurs de différentes unités de recherche

A la suite de l'enquête, j'ai entrepris de conduire des entretiens avec les chercheurs des établissements partenaires du projet. L'analyse des résultats de l'enquête a en effet servi de point de départ : elle était nécessaire pour mieux connaître le contexte et a fait naître, parallèlement, le besoin d'en affiner la compréhension, dans l'objectif d'adapter aux besoins spécifiques du site alsacien les scénarios proposés à l'issue de mon étude. Durant la phase préparatoire, j'ai repris les résultats de cette première enquête, en réalisant, cette fois-ci, une analyse bivariée : étudier les réponses de chaque unité de recherche, ayant accepté de nous rencontrer, m'a permis de me faire une idée de ses pratiques de gestion de données et ainsi de mieux cibler mes questions. Une attention particulière a été prêtée à

¹²⁹ Les données d'entrée (figure 1) sont les données existantes qu'utilisent les chercheurs, tandis que les données produites (figure 2) sont les données générées au cours du processus de recherche.

¹³⁰ Présentation proposée par Pauline Coisy, responsable de la bibliothèque de l'INSA, et centrée sur les résultats de l'institut.

chaque établissement du site alsacien hébergeant des unités de recherche : à l'UHA, nous avons contacté 6 unités ; 3 d'entre elles ont accepté de nous recevoir en entretien ; à l'INSAS, une unité nous a reçus sur les 3 contactées ; à l'Unistra, 4 unités sur 7. Souhaitant laisser aux chercheurs la possibilité de développer les points qui leur semblaient importants, j'ai choisi d'adopter une démarche exploratoire, reposant sur des entretiens semi-directifs. Un tableau, en annexe, présente la méthodologie de manière plus détaillée.

2.2. Conclusions des enquêtes

2.2.1. Découverte d'un potentiel important mais non formulé sur le site alsacien

Les enquêtes menées auprès des chercheurs ont confirmé l'intérêt d'intégrer au projet des Archives Ouvertes de la Connaissance un service pour les données de recherche. En effet, 36% des répondants au sondage en ligne se disent prêts à mettre à disposition leurs données ; 39% ne savent pas et 26% sont contre. Ces résultats sont corroborés par ceux des entretiens que nous avons menés.

Les chercheurs et équipes de chercheurs semblent majoritairement avoir la main sur la gestion des données de recherche qu'ils produisent. Celles-ci sont essentiellement conservées sur des supports personnels (disque dur, clef USB) ; elles occupent un volume moyen de 0 à 50 Go et sont souvent destinées à la réutilisation. C'est principalement à l'initiative des chercheurs qu'elles sont divulguées, selon le modèle de l'autopublication.

Des besoins en matière d'archivage se font également sentir. Ils concernent les supports de stockage mais aussi les méthodes de conservation, les chercheurs recherchant à la fois des supports adaptés –fiables, accessibles et pérennes- que des règles de tri, de nommage et de classement.

Les archives ouvertes ont donc un rôle à jouer dans la publication des données de recherche. Les éditeurs privés commençant à peine à s'intéresser à cette part du marché, l'enjeu pour le site Université de Strasbourg consiste à éviter la privatisation de leur diffusion. Par ailleurs, il semblerait que les chercheurs soient plus enclins à partager leurs données, dans la mesure où celles-ci ne sont pas soumises à validation, comme le sont les publications.

2.2.2. Les freins à l'ouverture des données de recherche

Une culture du libre accès peu ancrée dans les pratiques

Le recours aux canaux de diffusion en libre accès, le dépôt en archives ouvertes notamment, est une pratique peu répandue sur le site alsacien. L'enquête révèle que seuls 35% des répondants ont déjà déposé des publications dans une archive ouverte. Les initiatives de gestion de données sont, quant à elles, d'autant plus rares que le consensus autour de leur intérêt est encore tout récent. Il s'agit soit :

- d'initiatives individuelles : des enseignants-chercheurs diffusent leurs données avec les moyens et compétences qui sont les leurs. Le cas de figure le plus courant

que j'ai rencontré au cours des entretiens concerne des bases de données que des chercheurs ont conçues : le catalogue de bibliothèque d'un homme de lettres, en sciences humaines et sociales, le suivi de la croissance d'une solution bactérienne, en sciences de l'environnement.

- d'initiatives non valorisées : faute d'un suivi régulier des outils de diffusion, certains projets perdent en visibilité. Un enseignant-chercheur en histoire nous expliquait que la base de données qu'il avait conçue dans le cadre d'un programme de recherche cofinancé par la Maison Interuniversitaire des Sciences de l'Homme (MISHA, Unistra-CNRS), ne bénéficiait plus des services de maintenance depuis la clôture du projet. Aujourd'hui, lorsque des « coquilles » lui sont signalées par les utilisateurs, il ne peut les corriger sans faire appel à un informaticien de la MISHA.
- les entretiens que nous avons menés ont eu pour conséquence de donner aux enseignants-chercheurs une perspective de soutien pour la gestion de leurs données de recherche et, par conséquent, leur ont permis de formuler des besoins. Les enseignants-chercheurs expriment en effet souvent la demande de financements et de services dédiés pour concrétiser leur projet de valorisation de données.

Des contraintes temporelles et techniques

Les chercheurs ont souvent peu de temps à consacrer à la gestion des résultats de leurs recherches. La fonction première des chercheurs est la recherche. Son évaluation est fondée sur les publications qu'ils produisent. Aussi l'archivage et la diffusion du matériel annexe sont-ils souvent délaissés au profit d'une publication supplémentaire ou de nouvelles recherches.

2.2.3. Quelques-uns des enjeux liés à l'ouverture des données de recherche

Les enjeux énumérés ci-dessous sont la synthèse des réponses à la question ouverte de l'enquête « Etes-vous prêts à mettre à disposition vos données de recherche ? Si oui/non, pourquoi », complétée par des cas pratiques qu'ont pu rapporter les chercheurs rencontrés en entretien.

Les questions de propriété intellectuelle¹³¹

Les chercheurs ne comprennent pas toujours à qui appartiennent les jeux de données produits au cours de leur recherche et s'ils ont le droit de les diffuser librement. Dans le droit français, les données ne sont effectivement pas protégées par le droit d'auteur. Cette indétermination juridique conduit le plus souvent les

¹³¹ Sur le sujet, on pourra lire le chapitre sur « Les enjeux juridiques de l'ouverture » du mémoire de Rémi Gaillard (p.44-51). GAILLARD Rémi. *De l'Open data à l'Open research data : quelle(s) politique(s) pour les données de recherche ?* Mémoire, Enssib, 2014, 104 p. <<http://www.enssib.fr/bibliotheque-numerique/notices/64131-de-l-open-data-a-l-open-research-data-quelles-politiques-pour-les-donnees-de-recherche>> [En ligne]

chercheurs à ne pas communiquer les données qu'ils produisent. Ce sont eux, ou bien les financeurs des projets de recherche, qui décident donc principalement des modalités de diffusion des données. Lorsqu'elles sont produites dans le cadre d'un travail de recherche collectif, l'autorisation de l'ensemble de l'équipe est nécessaire pour pouvoir les rendre accessibles. Si les recherches ont été financées par un partenaire privé, celui-ci a le droit de revendiquer la propriété des résultats auxquels elles ont abouti. Un des enseignants-chercheurs que j'ai rencontrés évoquait ainsi la difficulté croissante des équipes de recherche en sciences de l'environnement à disposer librement des données qu'elles produisent, en raison du nombre croissant de contrats établis par les universités avec des partenaires privés.

La protection des données

Certains résultats de recherche relèvent de données sensibles¹³², comme le sont parfois les données médicales, les résultats d'entreprise, ou encore les études sociologiques. Comme le soulignait un des répondants à l'enquête, elles « obéissent [alors] à des règles de confidentialité qui rendent leur communication impossible sans filtre » ; elles ne peuvent être mises à disposition qu'après avoir été anonymées.

Les données sociales, judiciaires, génétiques ou biométriques peuvent également se révéler sensibles. Un enseignant-chercheur en sciences de l'environnement évoquait, par exemple, le risque (déjà constaté) de voir les données de son laboratoire utilisées de manière non objective, notamment par les médias.

Les licences *Creative Commons* (CC)¹³³, notamment celles qui définissent une condition d'attribution, sont couramment utilisées par les chercheurs ; les *waivers* ou « renoncations » sont en revanche moins connues, de même que l'encadrement juridique des bases de données, en vertu du droit *sui generis* de leurs producteurs.

La mise en forme des données

Les répondants à l'enquête ont souvent émis des doutes concernant la possible réappropriation des données par autrui : par exemple, « leur exploitation nécessite une compétence et des connaissances qui ne sont pas nécessairement acquises par les personnes qui souhaiteraient les employer ». La mauvaise utilisation des données est une crainte légitime, qui peut être limitée grâce à une description précise de ces dernières.

La mise en forme des données est, en effet, une étape préalable primordiale à leur partage. Elle suppose de les doter :

- d'un format ouvert et non-propriétaire, de manière à les rendre lisibles et réutilisables ;

¹³² D'après la définition du Correspondant Informatique et Libertés du CNRS, « les données sensibles sont celles qui font apparaître, directement ou indirectement, les origines raciales ou ethniques, les opinions politiques, philosophiques ou religieuses ou l'appartenance syndicale des personnes, ou sont relatives à la santé ou à la vie sexuelle de celles-ci » (<http://www.cil.cnrs.fr/CIL/spip.php?rubrique300>).

¹³³ <http://creativecommons.org/>

- de métadonnées descriptives voire d'une documentation détaillée, afin de les rendre intelligibles et réutilisables.

Dans des disciplines comme la sociologie ou l'anthropologie¹³⁴, dans lesquelles les chercheurs s'appuient principalement sur des enquêtes de terrain, ceux-ci expriment de fortes réserves quant aux réelles possibilités d'exploitation secondaire de ce support de recherche ; selon eux, l'interaction entre enquêteur et enquêtés est telle que les résultats sont généralement difficilement réutilisables hors contexte. Selon un enseignant-chercheur de l'INSAS, qui conçoit des enquêtes qualitatives et en exploite les résultats sous la forme de bases de données : « seuls ceux qui ont élaboré le questionnaire peuvent en exploiter les résultats ; ou bien il faudrait une journée pour expliquer comment il a été conçu (la correspondance entre les items, par exemple) »¹³⁵.

L'ouverture des données demande donc un travail d'identification, d'extraction et de documentation, permettant qu'elles soient trouvées, comprises et réutilisées.

L'investissement de temps et de compétences

Ce travail (parfois long) réclame un investissement en temps, ainsi que des compétences en documentation. Aussi est-il progressivement investi par les services de la documentation, dont les métiers évoluent vers de nouveaux profils, celui du *data librarian*¹³⁶ notamment. Si les chercheurs sont, de fait, les mieux placés pour décrire leurs données d'un point de vue scientifique, ils ne possèdent pas forcément les compétences nécessaires à leur curation et à la structuration de leurs métadonnées.

La reconnaissance du travail réalisé

Les données de recherche sont le plus souvent considérées comme personnelles ou privées par les chercheurs, en raison du travail que demande leur traitement. D'où certaines réticences à les communiquer. Les répondants qui, dans l'enquête, ont déclaré ne pas être prêts à mettre leurs données à disposition, ont par exemple apporté les justifications suivantes :

- « En dehors de collaborations scientifiques, je considère que la mise à disposition des données doit être rémunérée à l'équipe de recherche. »
- « Elles m'ont coûté trop de temps privé à les recueillir. »

Un enseignant-chercheur en lettres et langues nous a fait part, lors d'un entretien, de son indécision quant à laisser libre utilisation de ses données à un tiers. Celles-ci représentent un fonds riche d'informations, aboutissement d'un

¹³⁴ Sur la question de l'intérêt scientifique à archiver des enquêtes sociologiques : DUCHESNE Sophie & GARCIA Guillaume. « beQuali : une archive qualitative au service des sciences sociales ». M. Cornu, J. Fromageau et B. Muller. *Archives de la recherche. Problèmes et enjeux de la construction du savoir scientifique*, l'Harmattan, pp.35-56, 2014. ISBN : 978-2-343-03247-4. <halshs-00922690>

¹³⁵ Entretien mené le 19.05.15.

¹³⁶ On pourra se référer au chapitre 4 du support de la formation proposée par l'Inist-CNRS et l'Urfist de Rennes en juin dernier : JACQUEMOT-PERBAL Marie-Christine & COSSERAT Françoise. *Gestion et diffusion des données de la recherche*. Formation à l'Urfist de Rennes, juin 2015. <http://www.inist.fr/IMG/pdf/urfistrennes_20150616.pdf> [en ligne] p.129-139 « 4- Quels rôles pour les professionnels de l'IST ? »

long processus et dont il prévoit de tirer différentes publications ; c'est pourquoi il l'exploite sur plusieurs années. En diffusant ces données, il prend le risque qu'un tiers entreprenne de les exploiter et d'en publier la ou les analyse(s) avant lui. Il se dit néanmoins prêt à transmettre ses données après publication(s).

Les avis s'expriment en faveur d'une réglementation du droit d'accès, fondée, pour les uns, sur une diffusion des données restreinte au domaine de recherche, pour les autres, sur un principe de collaboration ou encore sur demande de l'intéressé.

Dans le cadre de la diffusion des données, il paraît donc essentiel de définir des permissions d'accès, soit en restreignant l'ouverture à des groupes de personnes précis, soit en imposant une période d'embargo, pendant laquelle les données ne sont pas librement accessibles.

Les enquêtes ont montré qu'il existe sur le site alsacien un besoin latent en matière de gestion des données de recherche, qui justifie la mise en place d'un service dédié au sein des Archives Ouvertes de la Connaissance.

3. PRECONISATIONS

3.1. *Mettre en place un « écosystème » favorable à l'ouverture des données de la recherche*

3.1.1. Par des moyens politiques

L'instauration d'une politique de gestion des données sur le site alsacien permettrait d'uniformiser les pratiques, grâce à des outils communs, et d'offrir aux chercheurs un cadre de confiance, fondé sur une protection juridique de leurs données.

Le projet européen RECODE¹³⁷ a réalisé cinq études de cas portant sur l'ouverture des données dans différentes disciplines de recherche. A l'issue du projet, ont été publiées des recommandations, dont le site Université de Strasbourg pourrait s'inspirer. Elles sont destinées aux institutions souhaitant établir une politique de libre accès aux données de recherche. Le rôle qu'elles leur attribuent consiste à veiller à la qualité scientifique et technique des données, grâce à la mise en place d'infrastructures, de formations, ainsi que de nouvelles modalités d'évaluation des chercheurs. Elles soulignent, par ailleurs, l'importance de consulter et d'impliquer les équipes de recherche dans la réflexion sur une politique de gestion des données. C'est à ce constat qu'a abouti ma mission de stage : l'intégration des principes du libre accès dans les processus de recherche dépend moins des outils qui sont mis à disposition que de la sensibilisation des équipes. Or les établissements représentent un lieu favorable à ce mouvement d'incitation, notamment grâce à la possible mutualisation des efforts entre chercheurs, informaticiens et documentalistes.

¹³⁷ RECODE project consortium. Policy recommendations for open access to research data. 2014, 44 p. <<http://ec.europa.eu/digital-agenda/en/news/recode-project-recommendations-open-access-research-data-are-now-available>> [en ligne]

3.1.2. Par des moyens techniques et humains

Trois schémas sont préconisés pour le partage des données prises en charge par les Archives Ouvertes de la Connaissance :

- Le dépôt de petits jeux de données dans la notice bibliographique des publications, sous forme de fichier lié. On privilégiera les formats recommandés par le CINES, en utilisant le validateur de formats FACILE¹³⁸.
- Le renvoi, grâce à un lien pérenne, vers des jeux de données déposés dans un entrepôt thématique ou bien publiés dans un *data journal* ou dans les fichiers supplémentaires associés à un article (*supplementary materials*). Le rôle des Archives Ouvertes de la Connaissance consiste, dans ce cas, à orienter vers des entrepôts fiables pour la conservation pérenne des données. Nous définirons cette notion de « fiabilité » un peu plus loin.
- La création de bases de données dédiées à un corpus spécifique, sous forme de logiciel de base de données couplé à une interface de visualisation en ligne. Les équipes informatiques pourront proposer un standard de base de données, qui puisse être facilement réadaptable à chaque nouveau projet. Cette solution a été adoptée par le département eScience de l'Université de Fribourg. La base de données *Freiburg Corpus of English Dialects* (FRED), collection d'enregistrements audio de dialectes anglais, en est un exemple ; elle est le résultat d'un travail de collaboration entre la faculté de langues et le département eScience.

Destinées à d'importants jeux de données, les bases de données créées par les Archives Ouvertes de la Connaissance nécessiteraient la mise à disposition de capacités de stockage, dont dispose par exemple les centres de calculs. La création d'un *data center* est en projet à l'Université de Strasbourg ; un partenariat avec les Archives Ouvertes de la Connaissance pourrait donc être envisagé.

Le suivi des deux premiers schémas pourra être assuré par les équipes informatiques et de la documentation chargées d'accompagner le dépôt dans l'archive ouverte ; le troisième schéma requiert la mise en place d'un comité dédié, car le modèle de base de données nécessite d'être adapté aux spécificités techniques de chaque projet ; le comité pourra être composé du porteur du projet, de l'ingénieur d'études chargé de recenser les données de recherche sur le site alsacien, d'un informaticien et d'un bibliothécaire-documentaliste. A terme, les services de la documentation pourront proposer les compétences d'un agent spécialisé dans la gestion de données¹³⁹, dans le cadre du Schéma Directeur des Emplois et des Compétences.

Dans le but d'harmoniser les pratiques, il est recommandé de s'appuyer sur des outils et des documents-ressources standardisés, qui contribueront à

¹³⁸ <http://facile.cines.fr/>

¹³⁹ Dans les pays anglo-saxons, ce profil prend le nom de *data librarian* ; ses compétences portent sur la curation des données de recherche. En France, l'offre de formation est en cours de construction, notamment du côté de l'Institut de l'information scientifique et technique (Inist) et de l'Union régionale de formation à l'information scientifique et technique (Urfist), qui travaillent en collaboration sur ce projet.

systématiser les processus à différents niveaux de la gestion des données. Parmi eux, sont préconisés :

- Un plan de gestion de données, permettant de suivre l'évolution des données et de planifier leur diffusion. D'après la définition donnée par la Commission européenne, qui le rend obligatoire dans le cadre de l'*Open Research Data Pilot*, un plan de gestion de données (*Data Management Plan*) est un « document expliquant la façon dont les données de recherche recueillies ou générées seront gérées durant le projet de recherche et après son terme, tout en décrivant quels types de données devront être recueillies/générées et suivant quelle méthodologie, quelles normes, si et comment les données seront partagées ou rendues libres, la façon dont une curation leur sera appliquée et la manière dont elles seront conservées »¹⁴⁰. C'est en quelque sorte le passeport des données. Il doit être élaboré dès la mise en place du projet de recherche et régulièrement mis à jour jusqu'au dépôt pérenne des données dans un entrepôt.

En France, plusieurs documents méthodologiques sont proposés, notamment aux équipes de recherche participant à l'*Open Research Data Pilot* du programme Horizon 2020, parmi lesquels :

Le guide de rédaction de l'Université Paris Diderot – Paris Descartes, *Réaliser un plan de gestion de données*¹⁴¹ ;

Les tutoriels multimédias de l'Inist-CNRS, « Le libre accès aux résultats de la recherche dans le cadre d'Horizon 2020 »¹⁴² ;

Un guide plus général du Cirad, *Découvrir des plans de gestion des données de la recherche, en 4 points*¹⁴³.

Dans les pays anglo-saxons, il existe également des outils d'aide à la rédaction de plan de gestion de données. Nous citerons le DMPonline¹⁴⁴ du Digital Curation Centre (DCC) au Royaume-Uni ; facilement adaptable à des projets de recherche français, il permet de générer en ligne un plan de gestion de données personnalisé et de l'exporter dans le format souhaité (csv, html, json, pdf, text, xml, docx.). Le DCC présente sur son site Internet des exemples de *Data Management Plans*¹⁴⁵.

- Un arbre de décision pour faciliter le choix du schéma de diffusion convenant le mieux à un jeu de données, parmi les trois schémas proposés ci-dessus. L'équipe

¹⁴⁰ EUROPEAN COMMISSION, RESEARCH AND INNOVATION, *Guidelines on Open Access to Scientific Publications and Research Data in Horizon 2020*. Version 1.0, 11 décembre 2013, 14 p. http://ec.europa.eu/research/participants/data/ref/h2020/grants_manual/hi/oa_pilot/h2020-hi-oa-pilot-guide_en.pdf. Traduction française disponible en ligne : http://openaccess.inist.fr/IMG/pdf/14086_lignes_directrices_la_horizon_2020_tr_fr_version-oct2014.pdf p.10

¹⁴¹ CARTIER Aurore, MOYSAN Magalie & REYMONET Nathalie. *Réaliser un plan de gestion de données : guide de rédaction*. Universités Paris Descartes – Paris Diderot, version 1, 9 janvier 2015, 30 p. [en ligne] <>

¹⁴² Inist-CNRS. Tutoriels « Le libre accès aux résultats de la recherche dans le cadre d'Horizon 2020 ». 2015. [en ligne] <<http://www.inist.fr/?-Tutorielsmultimedias-H2020->>>

¹⁴³ DEBOIN Marie-Claude. *Découvrir des plans de gestion des données de la recherche en 4 points*. Montpellier : CIRAD, 2014, 6 p. [en ligne] <<http://coopist.cirad.fr/content/download/5435/40362/version/4/file/CoopIST-plan-gestiondonnees-recherche-20140717.pdf>>

¹⁴⁴ DMPonline. <https://dmponline.dcc.ac.uk/>

¹⁴⁵ <http://www.dcc.ac.uk/resources/data-management-plans/guidanceexamples>

des Archives Ouvertes de la Connaissance souhaite inciter les chercheurs à déposer leurs jeux de données dans un entrepôt thématique, réservant l'archive institutionnelle aux fichiers de petite taille, pour lesquels il n'existe pas d'entrepôt dédié. Ces fichiers seraient joints à la notice d'une publication scientifique ; en tant que documents annexes, ils ne pourraient recevoir d'identifiant pérenne, ni être décrits par des métadonnées spécifiques. Ils ne sauraient, par conséquent, être différenciés de la publication et cités individuellement.

- Une grille de critères permettant de choisir un entrepôt de confiance pour le dépôt d'un jeu de données. Un entrepôt de confiance (*trusted repository*) est un « entrepôt répondant aux critères de qualité exigés pour obtenir une certification (format des données, qualité des métadonnées, conditions d'accès et de réutilisation, identifiant pérenne, archivage à long terme, ...) »¹⁴⁶. Il existe des outils de certification, permettant aux utilisateurs de repérer les entrepôts de confiance : le Data Seal of Approval¹⁴⁷, par exemple, qui effectue une sélection sur la base de seize critères, dont l'interopérabilité (standards de métadonnées, identifiants pérennes, formats ouverts...) ou encore les conditions de diffusion et d'utilisation des données (licences libres...).

3.2. *Etude de cas : le naturaliste Jean Hermann et la base de données prosopographique 148 de ses réseaux*

Afin d'illustrer notre propos, je me suis concentrée sur un exemple de données de recherche, dont la vocation première est d'appuyer un travail de thèse. Or il recèle des possibilités plus larges d'interprétation, que sa diffusion permettrait d'exploiter.

Plusieurs raisons ont conduit à faire porter mon choix sur un projet de sciences humaines et sociales.

En premier lieu, l'équipe des Archives Ouvertes de la Connaissance souhaite mettre l'accent sur ce domaine, dont les résultats de recherche demeurent moins visibles que ceux des sciences techniques et médicales, en raison d'habitudes d'archivage et de diffusion en libre accès moins répandues. A travers cette étude de cas, nous tâcherons donc de mettre en évidence les spécificités potentiellement réitérables dans des projets similaires de gestion de données en sciences humaines et sociales.

¹⁴⁶ JACQUEMOT-PERBAL Marie-Christine & COSSERAT Françoise. *Gestion et diffusion des données de la recherche*. Formation à l'Urfist de Rennes, juin 2015. [En ligne] <http://www.inist.fr/IMG/pdf/urfistrennes_20150616.pdf>

¹⁴⁷ <http://datasealofapproval.org/en/>

¹⁴⁸ Du grec *prosôpon* (personnage de théâtre, d'où personne en général) et *graphein* (décrire), la prosopographie signifie étymologiquement "description des personnes". En histoire, la prosopographie est une méthode d'analyse, utilisée pour l'étude des groupes sociaux qui possèdent des caractéristiques communes.

En second lieu, au sein de l'unité *Arts, civilisation et histoire de l'Europe* (ARCHE), à laquelle se rattache l'étude de cas, les chercheurs et doctorants conçoivent régulièrement des bases de données dédiées à des sources historiques¹⁴⁹. La question du partage de ces données est donc une problématique récurrente, dont les considérations relèvent tant de l'accompagnement technique, que du renouvellement des méthodes de référencement et du suivi à long terme (ces aspects ont été évoqués lors des entretiens avec les chercheurs).

3.2.1. Présentation de la base de données

La base de données prosopographique des réseaux de Jean Hermann a été élaborée par une doctorante en histoire des sciences, dans le cadre de sa thèse *-La fabrique et le commerce du savoir : le cas des pratiques et des réseaux du naturaliste Jean Hermann (1738-1800)*.

Ce travail, encore en construction, se présente sous la forme de tables de données Excel, recensant isolément :

- Les auditeurs qui ont assisté aux cours donnés par Jean Hermann (3662 personnes) ;
- Les visiteurs que celui-ci a reçus dans son cabinet de curiosités (1700) ;
- Les individus avec lesquels il entretenait une correspondance (120).

En projet figure également une table dédiée aux sociétés savantes, avec lesquelles Jean Hermann était en relation.

Les tables ont été élaborées à partir des registres manuscrits tenus par Jean Hermann et sont enrichies au fur et à mesure par le travail de recoupement et d'identification qu'effectue cette doctorante grâce au dépouillement des collections du naturaliste.

La base de données n'existe donc pas encore en tant que telle. C'est précisément à la création de l'outil support que nous réfléchissons dans cette étude. L'objectif serait de concevoir un outil de consultation en ligne de la base de données prosopographique. Une base de données prosopographique se présente sous forme de notices individuelles qui, ensemble, constituent un « dictionnaire biographique ». Appliquée aux réseaux relationnels de Jean Hermann, cette structuration permettrait d'éclairer la typologie sociale, professionnelle et géographique de ses visiteurs et auditeurs et d'étayer une réflexion sur la circulation du savoir scientifique dans sa dimension culturelle.

3.2.2. Méthodologie de l'étude de cas

Le travail de cette doctorante nous a été présenté par sa directrice de thèse. Un premier entretien a été organisé le 27 mai. Il nous a permis de prendre connaissance de l'origine du projet, des objectifs du travail de collecte, des enjeux

¹⁴⁹ En atteste le portail de la *Maison Interuniversitaire des Sciences de l'Homme – Alsace* (MISHA), qui héberge des bases de données d'histoire, d'archéologie et d'ethnologie, conçues entre 2000 et 2010 (<http://www2i.misha.fr/flora/>).

sous-jacents, ainsi que de la méthodologie de travail. Nous avons également recueilli les dispositions de la doctorante concernant la diffusion de ses bases de données : à quel type de public (grand public, personnel du site Université de Strasbourg) elle souhaiterait les rendre accessible, sous quels termes (droits d'utilisation) et à quelle étape de ses recherches de thèse.

A l'issue de l'entretien a été rédigée une fiche d'identité¹⁵⁰ des tables de données. Je me suis inspirée d'un modèle existant, utilisé par les chercheurs de l'Université de Fribourg en Brisgau¹⁵¹ pour l'élaboration d'un plan de gestion de données. Il s'agit d'une liste de vérification (*checklist*) élaborée dans le cadre du projet WissGrid, dont l'objectif était de produire des outils pour la conservation pérenne des données de recherche. Ce questionnaire permet aux chercheurs de vérifier qu'ils ont bien pris en compte l'ensemble des aspects essentiels à la gestion de leurs données. Il a été conçu pour un usage générique, adaptable à diverses disciplines ; ses auteurs spécifient qu'il doit être considéré comme une aide, en aucun cas contraignante ou exhaustive. Dans le cadre de l'étude de cas, l'Université de Strasbourg ne disposant pas actuellement de service de stockage dédié aux sciences humaines, je n'ai pu prendre en compte les questions de la liste relatives au stockage sécurisé des données et à la (ou les) structure(s) responsable(s) de la sauvegarde.

En parallèle, nous avons souhaité étudier un modèle de base de données prosopographique, afin de dresser un inventaire des fonctionnalités techniques caractéristiques de ce type d'outil. Parmi les bases existantes, toutes ne sont pas librement accessibles sur Internet : la plupart requièrent une authentification de l'utilisateur. Nous nous sommes intéressés à la base de données prosopographique *Studium Parisiense*¹⁵², consacrée aux membres de l'Université de Paris entre le XIIe et le XVIe siècle.

La faisabilité du projet a finalement été évaluée par la Direction Informatique et la responsable au SCD du système d'information documentaire. L'attribution d'identifiants a été envisagée pour les noms des visiteurs et des auditeurs, ainsi que pour les noms de villes d'origine de ces individus ; les difficultés d'identification semblent toutefois en limiter les possibilités. La Direction Informatique estime pouvoir entreprendre la réalisation de la base de données à partir du mois de mai 2016. La date et les modalités de diffusion restent néanmoins à l'appréciation de la doctorante.

3.2.3. Conclusions de l'étude de cas

L'objectif est de proposer une base de données répertoriant un ensemble de fiches biographiques, sur lesquelles seraient renseignées l'identité de la personne, son origine géographique et sociale, ainsi que le lien qui la relie à Jean Hermann. Les informations sur l'identité des individus étant à l'origine lacunaires, la base de données ne fournit pas de biographie développée ; son intérêt réside plutôt dans la dissémination des savoirs que détenait Jean Herman.

¹⁵⁰ Annexe 5

¹⁵¹ *Leitfaden zum Forschungsdaten-Management*, H. Enke & J. Ludwig (Hrsg.). Verlag Werner Hülsbusch, Glückstadt, 2013. Disponible en ligne : http://www.wissgrid.de/publikationen/Leitfaden_Data-Management-WissGrid.pdf (p.87-92).

¹⁵² <http://lamop-vs3.univ-paris1.fr/studium>

Les enjeux du projet

L'élaboration d'une base de données prosopographique des réseaux de Jean Hermann présente des enjeux scientifiques, techniques et patrimoniaux. L'enjeu premier, en relation avec les objectifs des Archives Ouvertes de la Connaissance, est de diffuser des résultats de recherche à un large public et ainsi de valoriser la production scientifique, en particulier le travail de recherche réalisé par la doctorante. D'un point de vue technique, l'interopérabilité des métadonnées semblent essentielle, si l'on souhaite lier les données biographiques à des informations extérieures qui contribueraient à les enrichir. Enfin, ce projet permettrait de donner davantage de visibilité aux collections patrimoniales universitaires. Il pourra intéresser aussi bien les chercheurs (historiens, zoologistes) que les musées en histoire des sciences ou histoire des sciences naturelles, et plus généralement tous ceux qui s'intéressent aux réseaux de savants du XVIIIe siècle.

Perspectives envisagées pour la suite du projet

Nous suggérons la création d'une base de données qui soit consultable en ligne. Celle-ci se déclinerait en :

- Une interface de saisie, accessible sur authentification des administrateurs, leur permettant d'enrichir la base. L'identification des auditeurs, visiteurs et correspondants étant un travail sur le long terme, cette seconde interface paraît essentielle ; la base de données se voudrait donc évolutive. Il peut être envisagé de concevoir un icône « faire une suggestion », qui permettrait à un utilisateur de soumettre par mail des données nouvelles aux administrateurs de la base.
- Une interface publique, composée de différents modules de recherche (recherche simple, recherche avancée, recherche par index).

La consultation de la base de données inclurait les fonctionnalités suivantes :

- l'affichage des résultats sous forme de table ;
- l'export de ces tables au format PDF, .xls, .csv ;
- un code couleur indiquant pour chaque fiche biographique son degré de complétude ;
- un outil de visualisation cartographique, permettant de localiser l'origine géographique des relations de Jean Hermann.

La base de données serait théoriquement protégée par le droit *sui generis* des bases de données. Une licence libre pourrait lui être attribuée également : l'*Open Database License (ODbL)*¹⁵³, par exemple, qui permet une diffusion libre sous conditions d'attribution et de partage sous clauses identiques.

La réalisation de cette base de données donnerait lieu à une collaboration entre la doctorante porteuse du projet, le SCD et la Direction Informatique, voire à un possible partenariat avec le Jardin des Sciences¹⁵⁴, qui élabore actuellement une plateforme collaborative et expérimentale autour des collections patrimoniales de

¹⁵³ <http://opendatacommons.org/licenses/odbl/>

¹⁵⁴ Jardin des Sciences, Université de Strasbourg. <http://jardin-sciences.unistra.fr/>

l'Université de Strasbourg. Les registres de Jean Hermann comprennent parfois la mention d'un commerce matériel (don ou promesse de don d'un ouvrage, d'un spécimen, de pierres minéralogiques, etc.). Ces informations relatives aux dons peuvent figurer en commentaire dans les fiches biographiques ; une variante plus élaborée consisterait à inclure un lien renvoyant vers l'objet ou la lettre numérisée (dans le cas des correspondants d'Hermann) lorsque ceux-ci existent. Ce point pourrait constituer une extension en articulation avec le futur portail du Jardin des Sciences.

CONCLUSION

Nous l'avons constaté en conclusion de l'étude sur les pratiques de production des données de recherche, les entretiens menés auprès des chercheurs ont fait émerger des besoins concrets. Un enseignant-chercheur en théologie nous a, par exemple, présenté une base de données en cours de construction, destinée à être mise en ligne dans le cadre d'une collaboration entre l'Université de Strasbourg et l'Université de Mayence en Allemagne ; lui manquaient cependant les financements et compétences nécessaires à la conception de l'interface publique. Cette mission a donc été présentée à l'équipe informatique du projet Archives Ouvertes de la Connaissance, qui a décidé de l'intégrer aux projets-pilotes lancés dès septembre et qui initieront le volet consacré aux données de recherche. Au nombre de trois, ces projets-pilotes concerneront respectivement une base de données d'objets archéologiques, une base de données de figurines jordaniennes et israélites, ainsi que la base de données hymnologique mise au jour lors des entretiens.

Les entretiens avec les chercheurs ont donc contribué, en partie, à accélérer la réalisation du volet sur les données de recherche. Celui-ci devait en effet être intégré aux Archives Ouvertes de la Connaissance dans une deuxième phase du projet, qui prévoyait l'ajout de fonctionnalités avancées ; il sera finalement pris en compte dès la première version de la plateforme. Représentant la partie « émergée de l'iceberg », les projets-pilotes serviront de support de communication auprès des unités de recherche du site alsacien, dans l'objectif de faire naître une dynamique autour de la gestion des données de recherche.

Marie-Christine Jacquemot-Perbal¹⁵⁵, dont l'équipe de documentalistes scientifiques de l'Inist travaille en collaboration étroite avec les chercheurs, souligne l'importance de bonnes pratiques de gestion et de conservation des données et de la responsabilisation des équipes de recherche. Cette démarche interroge, en effet, sur la possible généralisation d'un « écosystème » favorable au partage des données de recherche et sur la manière dont, au sein d'une équipe, pourrait naître une culture des données.

¹⁵⁵ Marie-Christine Jacquemot-Perbal est responsable de l'équipe *Valorisation des données de la recherche* à l'Inist. Nous l'avons rencontrée le 4 juin 2015.

SOURCES

ARCHIMER - Archive Institutionnelle de l'Institut Français de Recherche pour l'Exploitation de la Mer, [en ligne] <http://archimer.ifremer.fr/>

BIBLIOTHEQUE NATIONNALE UNIVERSITAIRE (BNU), [en ligne] <http://www.bnu.fr/>

CENTRE INFORMATIQUE NATIONAL DE L'ENSEIGNEMENT SUPERIEUR, [en ligne] <https://www.cines.fr/>

CONSORTIUM COUPERIN, [en ligne] <http://www.couperin.org/>

DIGITAL CURATION CENTRE (DCC), [en ligne] <http://www.dcc.ac.uk/resources/data-management-plans/guidanceexamples>

DMP ONLINE, [en ligne] <https://dmponline.dcc.ac.uk/>

FREIDOK PLUS - Archive ouverte de l'Université Albert Ludwig, [en ligne] <http://www.freidok.uni-freiburg.de/>

HAL - Archive pluridisciplinaire nationale, [en ligne] <https://hal.archives-ouvertes.fr>

INSAS, [en ligne] <http://www.insa-strasbourg.fr/>

INIST-CNRS [en ligne], <http://www.inist.fr>

STUDIUM PARISIENSE [en ligne] <http://lamop-vs3.univ-paris1.fr/studium>

MAISON INTERUNIVERSITAIRE DES SCIENCES DE L'HOMME (MISHA), [en ligne] <http://www2i.misha.fr/flora/>

OKINA - Archive ouverte Open Knowledge, Information, Access de l'Université d'Angers, [en ligne] <http://okina.univ-angers.fr/>

OPENAIRE, [en ligne] <https://www.openaire.eu/>

OPEN CITATION PROJECT, [en ligne] <http://opcit.eprints.org/oacitation-biblio.html>

OPEN DATA COMMONS, [en ligne] <http://opendatacommons.org>

ORBI - Open Repository and Bibliography, [en ligne] <http://orbi.ulg.ac.be/>

PRODINRA - Archive ouverte de l'Institut National de la Recherche Agronomique (INRA), [en ligne] <http://prodinra.inra.fr/>

SAM – Archive Ouverte Arts et Métiers Paris Tech, [en ligne] <http://sam.ensam.eu/>

SHERPA/RoMEO, [en ligne] <http://www.sherpa.ac.uk/romeo/>

THESES-UNISTRA, [en ligne] <http://theses.unistra.fr/ori-oai-search/index.html>

UNIVERSITE DE HAUTE ALSACE (UHA), [en ligne] <http://www.uha.fr/>

UNIVERSITE DE STRASBOURG (Unistra), [en ligne] <https://www.unistra.fr>

BIBLIOGRAPHIE

BACH Jean-François, JEROME Denis. *Les nouveaux enjeux de l'édition scientifique. Rapport de l'Académie des Sciences*. Institut de France, 24 juin 2014, 56 p. [en ligne] <http://www.academie-sciences.fr/pdf/rapport/rads_241014.pdf>

BOAI 10. « Ten years on from the Budapest Open Access Initiative: setting the default to open ». Budapest. 12 septembre 2012. [en ligne] <<http://www.budapestopenaccessinitiative.org/boai-10-recommendations>> [traduction française en ligne] <<http://www.budapestopenaccessinitiative.org/boai-10-translations/french>>

« Budapest Open Access Initiative ». Budapest, 14 février 2002. [en ligne] <<http://www.budapestopenaccessinitiative.org/read>> [traduction française en ligne] <<http://www.budapestopenaccessinitiative.org/translations/french-translation>>

CABINET DU PREMIER MINISTRE. *Stratégie numérique du gouvernement*. Gaité lyrique, 18 juin 2015. [en ligne] <<http://www.gouvernement.fr/sites/default/files/liseuse/4492/master/index.htm>>

CARTIER Aurore, MOYSAN Magalie & REYMONET Nathalie. *Réaliser un plan de gestion de données : guide de rédaction*. Universités Paris Descartes – Paris Diderot, version 1, 9 janvier 2015, 30 p. [en ligne]

COMMISSION EUROPEENNE, *Un partenariat renforcé pour l'excellence et la croissance dans l'Espace européen de la recherche*. COM(2012) 392 final. Bruxelles, 17 juillet 2012. 18 p. [en ligne] <http://ec.europa.eu/research/era/pdf/era-communication/era-communication_fr.pdf>

DEBOIN Marie-Claude. *Découvrir des plans de gestion des données de la recherche en 4 points*. Montpellier : CIRAD, 2014, 6 p. [en ligne] <<http://coopist.cirad.fr/content/download/5435/40362/version/4/file/CoopIST-plan-gestiondonnees-recherche-20140717.pdf>>

DUCHESNE Sophie & GARCIA Guillaume. « beQuali : une archive qualitative au service des sciences sociales ». M. Cornu, J. Fromageau et B. Muller. *Archives de la recherche. Problèmes et enjeux de la construction du savoir scientifique*, l'Harmattan., 2014. ISBN : 978-2-343-03247-4. <halshs-00922690>

ENKE. H & LUDWIG.J (Hrsg.).*Leitfaden zum Forschungsdaten-Management*, Verlag Werner Hülsbusch, Glückstadt, 2013. [en ligne]

<http://www.wissgrid.de/publikationen/Leitfaden_Data-Management-WissGrid.pdf>

EUROPEAN RESEARCH COUNCIL, « European Research Council-Scientific Council Guidelines for Open Access », 17 décembre 2007, 2 p. [en ligne] <<http://www.openaire.eu/fr/component/attachments/download/3>>

EUROPEAN COMMISSION, *Open Access Pilot in FP7*. 2008, 3 p. [en ligne] <http://ec.europa.eu/research/sciencesociety/document_library/pdf_06/open-access-pilot_en.pdf>

EUROPEAN COMMISSION, *General multi-beneficiary model grant agreement for the Horizon 2020 programme*. 2014. [en ligne] <http://ec.europa.eu/research/participants/data/ref/h2020/mga/gga/h2020-mga-gga-multi_en.pdf>

EUROPEAN COMMISSION, « Towards better access to scientific information : boosting the benefits of public investments in research », COM(2012) 401 final. Bruxelles, 17 juillet 2012, 15 p. [en ligne] <http://ec.europa.eu/research/science-society/document_library/pdf_06/era-communicationtowards-better-access-to-scientific-information_en.pdf>

EUROPEAN COMMISSION, « Recommendation on access to and preservation of scientific information », C(2012) 4890 final. Bruxelles, 17 juillet 2012, 10p. [en ligne] <http://ec.europa.eu/research/science-society/document_library/pdf_06/recommendationaccess-and-preservation-scientific-information_en.pdf>

EUROPEAN COMMISSION, RESEARCH AND INNOVATION, *Guidelines on Open Access to Scientific Publications and Research Data in Horizon 2020*. Version 1.0, 11 décembre 2013, 14 p. [en ligne] http://ec.europa.eu/research/participants/data/ref/h2020/grants_manual/hi/oa_pilot/h2020-hi-oa-pilot-guide_en.pdf. [traduction française en ligne] <http://openaccess.inist.fr/IMG/pdf/14086_lignes_directrices_la_horizon_2020_tr_fr_version-oct2014.pdf>

Etalab, *Vade-mecum sur l'ouverture et le partage des données publiques*. Septembre 2013, 11p. [en ligne] <<http://www.modernisation.gouv.fr/sites/default/files/fichiers-attaches/vademecum-ouverture.pdf>>

FAYET Sylvie, « "Données" de la recherche, les mal-nommées », 15 novembre 2013, [en ligne] <<http://urfistinfo.hypotheses.org/2581>>

GAILLARD Rémi. *De l'Open data à l'Open research data : quelle(s) politique(s) pour les données de recherche ?* Mémoire, Enssib, 2014, 104 p. [en ligne] <<http://www.enssib.fr/bibliotheque-numerique/notices/64131-de-l-open-data-a-l-open-research-data-quelles-politiques-pour-les-donnees-de-recherche>>

HAMEAU Thérèse, « Le libre accès introduit dans la loi néerlandaise ». Billet du blog *Libre accès à l'information scientifique et technique*, 23 juillet 2015. [en ligne] <<http://openaccess.inist.fr/?Le-libre-acces-introduit-dans-la>>

HARNAD Stevan, « Universal FTP archives for esoteric science and scholarship: a subversive proposal », *Scholarly Journals at the Crossroads: A Subversive Proposal for Electronic Publishing*. Association of Research Libraries, Washington DC, 1995. [en ligne] <<http://www.arl.org/scomm/subversive/toc.html>>

HARNAD Stevan, « Waking OA's "slumbering giant": the university's mandate to mandate open access », *New Review of Information Networking*, vol.14, 2008, pp. 51-68. DOI:10.1080/13614570903001322

Inist-CNRS. Tutoriels « Le libre accès aux résultats de la recherche dans le cadre d'Horizon 2020 ». 2015. [en ligne] <<http://www.inist.fr/?-Tutorielsmultimedias-H2020->>

JACQUEMOT-PERBAL Marie-Christine & COSSERAT Françoise. *Gestion et diffusion des données de la recherche*. Formation à l'Urfist de Rennes, juin 2015. [en ligne] <http://www.inist.fr/IMG/pdf/urfistrennes_20150616.pdf>

LERU *Roadmap Towards Open Access* (2011) et la *LERU Roadmap for Research Data* (2014).

MAX PLANCK GESELLSCHAFT. *Berlin Declaration on Open Access to Knowledge in the Sciences and Humanities*. Berlin, 22 octobre 2003. [en ligne] <<http://openaccess.mpg.de/Berlin-Declaration>>. [traduction française en ligne] <http://openaccess.mpg.de/68042/BerlinDeclaration_wsis_fr.pdf>.

MAX PLANCK GESELLSCHAFT, *The Road to Open Access : A guide to the implementation of the Berlin Declaration*. Southampton, mars 2005. <http://openaccess.mpg.de/81923/roadmap_print.pdf> [en ligne]

MINISTERE DE LA JUSTICE, *Entwurf eines Gesetzes zur Nutzung verwaister Werke und zu weiteren Änderungen des Urheberrechtsgesetzes und des Urheberrechtswahrnehmungsgesetzes*. Berlin, février 2013. [en ligne] <<http://blog.die-linke.de/digitalelinke/wp-content/uploads/Ref.Entwurf.pdf>>. [traduction française en ligne des principaux passages] <<http://openaccess.inist.fr/?Traduction-francaise-du-projet-de>>

MOSCON Valentina, « Open Access in Italy », *IRIS Legal Observations of the European Audiovisual Observatory*. IRIS 2014-1:1/32. [en ligne] <<http://merlin.obs.coe.int/iris/2014/1/article32.en.html>>

OPEN ARCHIVE INITIATIVE, « The Open Archives Initiative Protocol for Metadata Harvesting ». Version 2, juin 2002. [en ligne] <<http://www.openarchives.org/OAI/openarchivesprotocol.htm>>

ORGANISATION DE COOPERATION ET DE DEVELOPPEMENT ECONOMIQUES (OCDE), *Principes et lignes directrices de l'OCDE pour l'accès aux données de la recherche financée sur fonds publics*, Paris, 2007, 28 p. [en ligne] <http://www.oecd.org/fr/science/sci-tech/38500823.pdf>

PROST Hélène et SCHÖPFEL Joachim, *Développement et Usage des Archives Ouvertes en France. Rapport. 1e partie : Développement*. Juillet 2010. [en ligne] <<http://www.enssib.fr/bibliotheque-numerique/document-48600>>

SCHÖPFEL Joachim, PROST Hélène. *Développement et Usage des Archives Ouvertes en France. 1e partie : Développement*. 2010, 49 p. [en ligne] <http://archivesic.ccsd.cnrs.fr/sic_00497389>

SWAN Alma, *Open Access Self-Archiving : An introduction*. Mai 2005, [en ligne] <<http://cogprints.org/4406/>>

THE ROYAL SOCIETY. *Science as an open enterprise : summary report*. The Royal Society, juin 2012, 104 p. [en ligne] <http://royalsociety.org/uploadedFiles/Royal_Society_Content/policy/projects/sape/2012-06-20-SAOE.pdf>

ANNEXES

Table des annexes

ANNEXE 1 QUELQUES EXEMPLES D'ARCHIVES INSTITUTIONNELLES FRANÇAISES.....	66
ANNEXE 2 PRINCIPALES PHASES DU PROJET ARCHIVES OUVERTES DE LA CONNAISSANCE	67
ANNEXE 3 ENQUÊTE SUR LES PRATIQUES DE PUBLICATION SCIENTIFIQUE ET DE PRODUCTION DE DONNÉES DE RECHERCHE SUR LE SITE ALSACIEN	68
ANNEXE 4 METHODOLOGIE DES ENTRETIENS SEMI-DIRECTIFS	71
ANNEXE 5 FICHE D'IDENTITÉ DES BASES DE DONNÉES PROSOPOGRAPHIQUES DES RÉSEAUX DE JEAN HERMANN.....	73
ANNEXE 6 ARBRE DE DECISION POUR CHOISIR UN SCHEMA DE DIFFUSION DE SES DONNÉES	80

ANNEXE 1 QUELQUES EXEMPLES D'ARCHIVES INSTITUTIONNELLES FRANÇAISES

	Archimer	ProdINRA	SAM	Okina
Date de création	2005	2006	2013	2014
Institution	IFREMER	INRA	Arts et Métiers ParisTech	Université d'Angers
Domaines de recherche	Sciences de l'environnement et du vivant, Mathématiques et informatique appliquées	Sciences de l'environnement et du vivant, Mathématiques et informatique appliquées, Sciences sociales	Sciences de l'ingénieur, Mathématiques et informatique appliquées, Physique	Multidisciplinaire
Unités de recherche	28	235	15	29
Implantation	5 centres de recherche	17 centres de recherche	11 centres de recherche	2 campus universitaires
Nombre de personnel	1 528 (au 31.12.2012)	12 000 (au 31.12.2014)	1 100 (au 31.12.2014)	1762 + 498 doctorants (au 31.12.2014)
Obligation de dépôt¹⁵⁶	OUI	OUI	OUI	OUI
Nombre d'items¹⁵⁷	30 256	209 946	1580	12 416
Part des items en libre accès	48%	NC (accès restreint au personnel INRA)	NC	NC
Part des items avec texte intégral	100% ¹⁵⁸	11%	NC	14%

¹⁵⁶ Texte intégral compris.

¹⁵⁷ Sont regroupés sous le terme d' « items » l'ensemble des notices bibliographiques enregistrées dans l'archive, qu'elles soient associées ou non au texte intégral du document qu'elles décrivent.

¹⁵⁸ La politique de l'IFREMER impose le dépôt du texte intégral pour tout type de document. D'où la déduction – non vérifiable – que l'ensemble des références bibliographiques contient un lien vers le fichier du document.

ANNEXE 2 PRINCIPALES PHASES DU PROJET ARCHIVES OUVERTES DE LA CONNAISSANCE

	2014			2015												2016-2017	
	Octobre	Novembre	Décembre	Janvier	Février	Mars	Avril	Mai	Juin	Juillet	Août	Septembre	Octobre	Novembre	Décembre	Janvier	
Lancement du projet																	
Diffusion d'une enquête sur les pratiques de publication des chercheurs																	
Comité de pilotage - présentation des résultats de l'enquête - choix du logiciel d'archive ouverte					27/03												
Installation et paramétrage du logiciel (recrutement d'un développeur informaticien)																	
Comité de pilotage										08/07							
Formation de l'équipe projet à l'utilisation du logiciel																	
Rédaction d'une charte du publiant																	
Recrutement de deux ingénieurs d'étude en documentation pour la planification du dépôt rétrospectif des publications																	
Recrutement d'un ingénieur d'étude pour l'accompagnement des projets pilotes																	
Mise en production d'une 1 ^{er} version de l'archive ouverte																	
Evaluation du HCEERES																	
Mise en production d'une version 2, enrichie de fonctionnalités avancées																	

ANNEXE 3 ENQUÊTE SUR LES PRATIQUES DE PUBLICATION SCIENTIFIQUE ET DE PRODUCTION DE DONNÉES DE RECHERCHE SUR LE SITE ALSACIEN

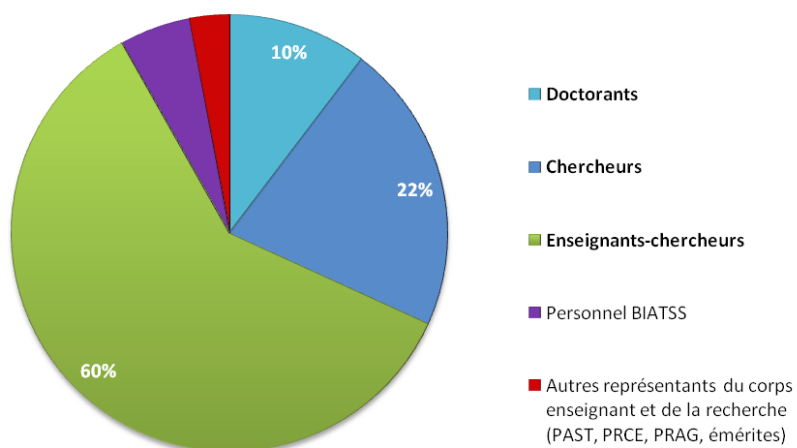
ANNEXE – QUESTIONNAIRE SUR LES PRATIQUES DE GESTION DES DONNÉES DE RECHERCHE (PARTIES 3 ET 4)

3.1 Votre recherche s'appuie sur... (Plusieurs réponses possibles)
Des corpus de textes ou de documents (photothèque, documents sonores, documents iconographiques, articles...)
Des objets
Des données d'observation (fouilles, données climatologiques,...)
Une bibliographie que vous constituez
Des résultats expérimentaux au laboratoire (acquisitions de données : microscopie, RMN, séquençage)
Des résultats de données à caractère médical (imagerie médicale, analyses biologiques)
Une liste de sources (imprimées, manuscrites, audiovisuelles, orales...)
Des banques de données (fouille de données...)
Des développements théoriques
Des calculs numériques
Autre
3.2 En plus des articles scientifiques ou des ouvrages, vous produisez des données qui sont...
Aucune donnée supplémentaire : tout est dans la publication
Des contributions à des corpus (appareil critique, catalogue...)
Un corpus de documents (photographies, textes, cartes...)
Des feuilles de calcul
Des bases de données, des thesauri
Une production matérielle (molécule, prototype, de la matière...)
Des brevets
Des descriptions d'objets technologiques (machines, robots, équipements optiques ou électroniques, plans, maquettes...)
Des descriptions d'organismes vivants (humains, animaux, végétaux)
Des données informatiques (mesures, données de simulation, textes, séquences biologiques, films, présentations...)
Des logiciels, algorithmes, API, codes de calcul, macros
Des descriptifs méthodologiques
Des enquêtes
Autre
4.1 Où sont situées les données que vous utilisez (données d'entrée) ?
Sur mon ordinateur domestique et personnel
Dans votre établissement (poste de travail à l'université, serveur de votre équipe ou de votre unité de recherche...)
Sur le terrain
Dans des centres de données (ex : centres de données disciplinaires)
Dans des bibliothèques ou des centres d'archives
Dans des musées
En ligne
Autre

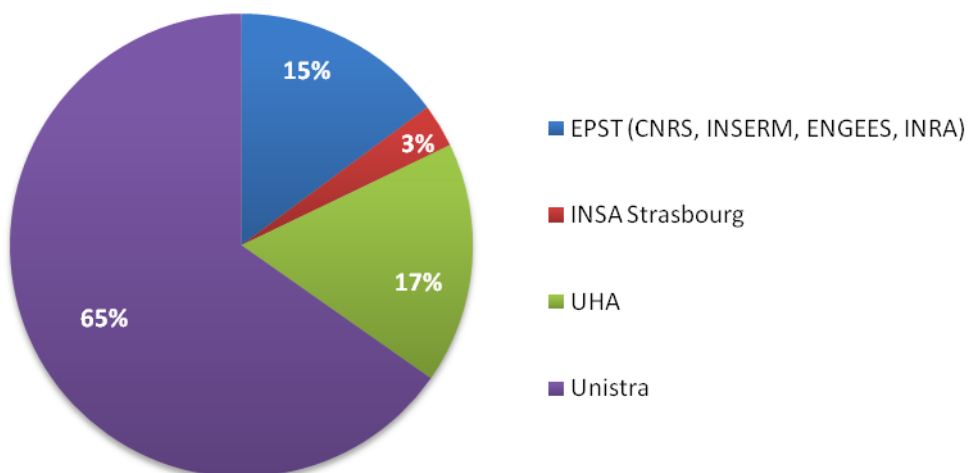
4.2 Quel volume de données conservez-vous dans votre établissement d'appartenance ?
4.3 Les données que vous utilisez sont-elles... ? (Plusieurs réponses possibles)
D'accès gratuit
D'accès payant (ex : données d'image satellite)
Accessibles grâce à la collaboration entre chercheurs
Utilisables sans restriction
Utilisables sous conditions (embargo, restrictions légales...)
À usage strictement personnel
4.4 Ces données peuvent-elles être réutilisées par d'autres personnes ?
Non
Oui, par vos proches collaborateurs
Oui, par votre domaine de recherche
Oui, par tout le monde
4.5 Les données que vous utilisez sont-elles accessibles en ligne ? (Plusieurs réponses possibles)
Oui, par vos proches collaborateurs
Oui, par votre domaine de recherche
Par tout le monde
4.6 Votre unité de recherche vous propose-t-elle un service de stockage des données ?
4.7.1 Pouvez-vous estimer le volume de données que vous produisez et qui sont nécessaires à votre production scientifique annuelle ?
4.7.2 Si oui, quel est le volume de ces données ?
4.8.1 Conservez-vous les données (annexes à la publication) que vous avez produites sous forme informatique ?
4.8.2 Quelle volumétrie (informatique) cela représente-t-il ?
4.8.3 Pendant combien de temps les conservez-vous ?
4.9.1 Où les conservez-vous ?
Un support lié à votre ordinateur personnel
Un espace dédié à votre unité de recherche
Dans le Cloud (Google, DropBox...) non académique
Dans un entrepôt disciplinaire en ligne
Un service académique dédié (ex : archivage au CINES)
Autre
4.9.2 Pour quelles raisons ?
Absence de support informatique équivalent (ex : cahier de laboratoire, j'ai produit un organisme vivant...)
Je ne sais pas où les stocker
Absence d'espace sécurisé de stockage
Autre
4.10.1 Réutilisez-vous les données produites pour des publications différentes ?
4.10.2 Pouvez-vous donner un pourcentage des données réutilisées ?
4.11 Etes-vous prêts à mettre vos données à disposition ? (Commentaire)

ANNEXE – EXTRAITS DES RÉSULTATS DE L'ENQUÊTE : PROFILS DES RÉPONDANTS

Taux de représentation par statut (sur 593 réponses connues)



Taux de représentation par établissement (sur 593 réponses connues)



ANNEXE 4 METHODOLOGIE DES ENTRETIENS SEMI-DIRECTIFS

Objectifs	Objectifs généraux - Confirmer les problématiques repérées dans les résultats de l'enquête - Mettre en lumière d'autres aspects de ces problématiques - Dégager de nouvelles problématiques
	Objectifs spécifiques - Connaître les pratiques des chercheurs concernant les données de recherche - Connaître l'état de leurs réflexions sur le sujet et leurs dispositions - Recueillir leurs besoins et leurs attentes vis-à-vis du projet des Archives Ouvertes de la Connaissance
Profil des enquêtés	-Statut : chercheurs et enseignants-chercheurs -Etablissement de rattachement : Unistra, UHA et INSAS -Discipline de recherche : physique, sciences de l'ingénieur, sciences de l'environnement, sciences sociales, lettres et langues, histoire, théologie
Mode de sélection	Ont été sélectionnées en priorité les unités de recherche ayant répondu à la première enquête. Un mail a été adressé aux directeurs des unités concernées, par l'intermédiaire des chefs de projet. Ont également été contactés les membres de ces unités avec lesquels les chefs de projet étaient régulièrement en contact. Une première série de questions était posée dans le mail de contact, destinée à introduire les thématiques de l'entretien (ces questions n'attendaient pas de réponse immédiate) : - Quel(s) service(s) vous semblerait bienvenu(s) pour simplifier ou valoriser la gestion de vos données ? - Dans quelle mesure vos données peuvent-elles être réutilisées ou partagées ? - Disposez-vous d'un ou plusieurs jeu(x) de données que vous souhaiteriez partager ?
Nombre d'entretiens menés	8 entretiens - 5 en sciences humaines et sociales - 3 en sciences dites « dures »
Type d'entretien	Entretiens semi-directifs
Durée	Une heure à une heure et demie
Période de collecte	De mai à juin 2015
Mode de collecte	Face-à-face En présence des chefs de projet (à l'exception de deux entretiens) Questions thématiques ouvertes, portant sur la typologie des données de recherche, leur traitement, leur conservation et leur diffusion.

	Prise de notes, enrichie après entretien <i>Observation : grâce à la présence des chefs de projet, les chercheurs se sont sentis plus facilement en confiance.</i>
Analyse des données	<ul style="list-style-type: none">- Découpage thématique des informations recueillies- Sélection des données répondant aux objectifs de l'étude- Croisements des données des différents entretiens- Synthèse thématique

ANNEXE 5 FICHE D'IDENTITÉ DES BASES DE DONNÉES PROSOPOGRAPHIQUES DES RÉSEAUX DE JEAN HERMANN

1. Le projet de recherche

Quelle est la nature du projet de recherche ?

Les travaux de D.R. s'inscrivent dans le cadre d'un doctorat, qui lui-même fait partie du projet de recherche intitulé « Les collections du naturaliste Jean Hermann (1738-1800). Pour une histoire sociale et culturelle des sciences naturelles au siècle des Lumières ».

Sujet de thèse : « La fabrique et le commerce du savoir : le cas des pratiques et des réseaux du naturaliste Jean Hermann (1738-1800) »

Une partie de la thèse sera consacrée aux réseaux de Jean Hermann. C'est dans cette partie que seront exploitées les registres du naturaliste, ainsi que les bases de données établies par D.R.

Objectifs

- Mettre en évidence les relations de Jean Hermann avec d'autres savants européens.

-Nature de ces relations : comment ont-elles évolué (certains visiteurs sont-ils devenus des correspondants, etc.) ?

-Interconnexions : certaines connaissances de Jean Hermann étaient-elles en relation entre elles ?

- Mettre en lumière les échanges d'objets (commerce matériel) et les transferts de savoir au sein des réseaux de Jean Hermann.

Qui sont les promoteur(s)/financeur(s)/partenaire(s) du projet ?

Le projet Hermann susmentionné a fait l'objet d'un contrat IdEX 2012-2015 ; il bénéficie donc de crédits sur 4 ans.

Quelle est la durée (planifiée) projet ?

3 à 4 ans

Date prévisionnelle de soutenance : 2nd semestre 2016

2. Les données de recherche

Le travail de recherche s'appuie-t-il sur des jeux de données existants ?

Le travail de recherche s'appuie sur les collections de Jean Hermann d'une part ; sur des documents manuscrits, d'autre part -à savoir :

- Le registre des visiteurs tenu par Jean Hermann

-3 662 personnes sont venues visiter le cabinet de curiosités entre 1764 et 1800 (du vivant de Jean Hermann).

-à sa mort, son gendre Frédéric-Louis HAMMER a continué de tenir le registre, l'allongeant de 1838 noms supplémentaires.

- Le registre des auditeurs tenu par Jean Hermann

-1700 personnes ont suivi les cours de Jean Hermann entre 1766 et 1800 (800 feuillets de notes)

-Jean Hermann a renseigné leur origine, le cours qu'ils ont suivi, le tarif appliqué (voire geste gracieux) et a parfois ajouté des commentaires.

-D.R. a pu identifier la moitié de ces auditeurs –notamment les étudiants, grâce à leur matricule. La plupart était des étudiants en médecine.

- La correspondance de Jean Hermann

-350 lettres

-120 correspondants ont été répertoriés.

Quel(s) type(s) de données sont utilisé(s)/produit(s) ?

- Données utilisées : des données de référence

Qui sont les données sources du travail de D.R. :

-le registre des visiteurs tenu par Jean Hermann

-le registre des auditeurs tenu par Jean Hermann

-la correspondance de Jean Hermann

-la bibliothèque de Jean Hermann

- Données produites : des données informatiques dérivées

-bases de données Excel

-graphiques Excel

-graphes de réseaux

-cartes géographiques dessinées sous Illustrator

Comment sont collectées et/ou produites les données : selon quel(s) critère(s), quel(s) procédé(s), avec quel(s) instrument(s) ?

Dépouillement des ouvrages et documents manuscrits des collections Hermann.

- Travail de recoupement et d'identification : qui sont les personnes mentionnées par Hermann dans ses registres ; tel ouvrage a-t-il fait l'objet d'un don à Hermann ; etc.

- Dans certains cas, travail d'enrichissement (ex : recherches biographiques sur les correspondants de Jean Hermann, afin d'établir si ces correspondants avaient également des liens entre eux).
- L'écriture et la langue de Jean Hermann (haut-allemand) sont parfois difficiles à déchiffrer.

Etant donné l'importance des registres et le temps imparti pour la réalisation de sa thèse, D.R. estime qu'une sélection sera nécessaire. Son objectif est de proposer :

- Un registre des visiteurs, renseignant :
 - Leur nom et prénom
 - Leur statut (auditeur, militaire, femme, enfant, etc.)
 - La date de leur venue
 - Leur ville et pays d'origine
 - Les remarques de Jean Hermann
 - Un éventuel commerce matériel (don ou promesse de don)
- Un registre des auditeurs, renseignant :
 - Leur nom et prénom
 - Leur statut (étudiant, professeur, médecin, etc.)
 - Leur département, ville et pays d'origine
 - Les remarques de Jean Hermann
- Un registre des visiteurs pour lesquels Jean Hermann mentionne un commerce matériel.
- Un catalogue des ouvrages annotés de la bibliothèque de Jean Hermann.

A partir des ouvrages (1000 environ) que Jean Hermann a reçus de la main de ses visiteurs, D.R. a isolé ceux qu'il avait annotés (355 actuellement).

En projet : création d'une base de données répertoriant les membres des sociétés savantes linnéennes avec lesquelles Jean Hermann était en relation –la société savante de Paris et celle de Londres.

Sous quel format informatique se présentent les données ? Subissent-elles des modifications de format au cours du processus de recherche ?

Les documents d'origine ne sont pas numérisés. Seuls 214 ouvrages de la collection de Jean Hermann ont été numérisés, dans le cadre du programme *DocNum* du SCD de l'Université de Strasbourg (<http://docnum.u-strasbg.fr/>, collections patrimoniales des bibliothèques de l'Université de Strasbourg, fonds Jean Hermann).

Les données manuscrites sont transposées sous forme de bases de données Excel.

Les résultats se présenteront sous la forme de fichiers texte, incluant des cartes et des graphiques. D.R. a déjà rédigé :

- Une analyse de la correspondance de Jean Hermann : origine géographique de ses correspondants –analyse illustrée par des graphiques de données statistiques et de réseaux (Excel), ainsi qu’une carte de l’Europe (Illustrator).
- Une analyse des auditeurs de Jean Hermann : origine géographique des personnes ayant participé à ses cours –analyse illustrée par des cartes (Illustrator) à l’échelle de l’Europe, la France et l’Alsace-Lorraine, ainsi que par des graphiques de données statistiques (Excel) s’appuyant notamment sur des regroupements par école d’origine.

Les auditeurs venaient essentiellement d’Allemagne, de Suisse et d’Angleterre (avant la Révolution française). Certains étudiants russes bénéficiaient d’un système de bourse, qui leur permettait d’effectuer une partie de leurs études à Strasbourg ; ils représentent 15% des auditeurs ayant suivi les cours de Jean Hermann.

- Une analyse semblable portant sur le catalogue des visiteurs du cabinet de curiosités est prévue. Elle mettra en évidence l’origine sociale des visiteurs et son évolution au fil du temps.

-Le cabinet de curiosités accueillait des voyageurs, notamment des militaires, et était ouvert aux femmes et aux enfants.

-A la mort de Jean Hermann, le cabinet de curiosités a été acquis par la municipalité ; il est donc passé du statut de cabinet privé à celui de musée (public). Ce changement s’est-il accompagné d’une variation sociale des visiteurs ?

[Quel volume représentent-elles approximativement ?](#)

Dossier informatique regroupant les bases de données ≈ 85 Mo

- Une cinquantaine de fichiers Excel
- D’une taille moyenne de 500 Ko

[Dans quelle mesure ces données sont-elles reproductibles : quelle charge de travail et quel coût cela implique-t-il ?](#)

Les informations rassemblées dans les bases de données sont accessibles dans les différents fonds patrimoniaux de Strasbourg. Leur collecte représente néanmoins un travail de plusieurs mois.

NB : D.R. a effectué un travail de dépouillement qui a permis de mettre au jour des documents inédits, comme la liste des auditeurs tenue par Jean Hermann.

[Quels sont les gages de leur qualité ? \(Ex : validation par les pairs, documentation des résultats et procédés de recherche, étalonnage des instruments de mesure\)](#)

3. Les données conservées

Comment sont choisies les données destinées à être conservées : par qui et selon quel(s) critère(s) ?

D.R. souhaiterait donner un exemple de ce qu'il est possible de faire à partir d'une base de données prosopographique.

- Proposer une base de données répertoriant une liste de personnes (visiteurs, auditeurs), dont serait renseignée l'origine géographique et sociale.
- Travailler un angle précis –le commerce matériel, par exemple, en montrant les échanges et promesses de dons qui se sont faits au sein du réseau.

Dans quel(s) but(s) les données sont-elles conservées :

- En tant que document de travail (on conserve une copie de sauvegarde)
- En tant que témoin des bonnes pratiques scientifiques (elles servent de fondement à la thèse développée dans la publication –en vue de la publication de la thèse notamment : les graphiques et bases de données servent de fondement au travail d'analyse et d'interprétation)
- En tant que source d'informations (elles ont un intérêt public, ici patrimonial)
- Pour être réutilisées (elles serviront de source pour des travaux de recherche ultérieurs)

A quelle étape du projet de recherche les données sont-elles prêtes à être collectées ?

Quel moyen est utilisé ? (Ex : quel logiciel ?)

4. L'archivage des données

Sur quel support souhaiteriez-vous conserver vos données ?

- L'archive ouverte de votre établissement ?
- Le serveur de votre laboratoire ?

La base de données pourrait être accessible depuis l'archive ouverte de l'Université de Strasbourg, quitte à proposer un lien sur le site de l'EA 3400 ARCHE par la suite.

Les données sont-elles liées à un outil de lecture spécifique ?

Les données sont rentrées et lisibles dans Excel.

Les listes des visiteurs, auditeurs et correspondants sont structurées en tables de données.

- A partir de ces tables sont élaborés des graphiques (analyses statistiques).
- L'outil NodeXL –plugin gratuit d'Excel- permet de générer des graphes pour l'analyse des réseaux. Exemple : réseau fondé sur les échanges matériels entre Jean Hermann et ses relations (connexions et interconnexions)

Combien de temps sont conservées les données ?

L'ensemble des bases de données produites est voué à être conservé sans limite de temps.

- Parce que ces données sont inédites (elles constituent une source rare).
- Parce qu'elles font occasionnellement l'objet d'une demande d'informations de la part de chercheurs.

Il serait intéressant de documenter :

- Le lieu de conservation des données sources (à quel fonds patrimonial elles appartiennent) ;
- La démarche de recherche (méthode, objectifs) –à la manière de la documentation consacrée à la base de données *HistCARTO* (<http://histcarto.u-strasbg.fr>).

5. L'accès aux données

Les données sont-elles soumises à confidentialité, à obligation de diffusion,... ?

Les données ne sont soumises ni à confidentialité, ni à obligation de diffusion.

Quel(le)s institutions/groupes cibles seraient potentiellement intéressés par ces bases de données ?

Le travail de D.R. présente un intérêt pour l'histoire locale et le rayonnement de Strasbourg dans l'histoire. Il a permis de recenser les collections de Jean Hermann et ainsi d'alimenter le catalogue de la BNU et de la bibliothèque Blaise Pascal.

Pourront s'y pencher aussi bien les historiens, que les minéralogistes et zoologistes (ex : description du phoque moine).

Ce travail touche également les musées d'histoire des sciences et d'histoire des sciences naturelles, ainsi que tous ceux qui s'intéressent aux réseaux de savants (notamment aux réseaux des naturalistes européens).

Les données sont-elles soumises à un embargo (leur utilisation est-elle, dans un premier temps, strictement réservée à leur auteur) ? Pendant combien de temps ?

Les bases de données pourront être mises en ligne, une fois que D.R. aura estimé leur degré d'élaboration satisfaisant. Idéalement après la soutenance de thèse.

Quel(s) type(s) d'accès aux fichiers de données est envisageable : libre accès

6. L'utilisation des données

Les données doivent-elles être soumises à une licence d'utilisation ? (ex : licence Creative Commons, licence Open Data Commons)

Oui

Quels en seraient les termes ?

- Exploitation libre (droit de création et distribution d'œuvres dérivées, à des fins commerciales ou non)

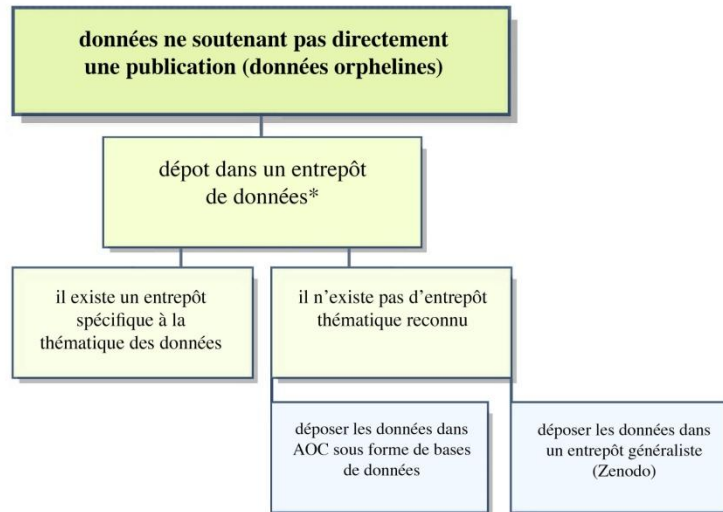
- Avec une obligation d'attribution (le nom de l'auteur devra être cité).
- L'Open Database License (ODbL)¹⁵⁹ correspondrait à ces conditions.

Sources : questionnaire élaboré à partir du guide *Leitfaden zum Forschungsdaten-Management*, H. Enke & J. Ludwig (Hrsg.). Verlag Werner Hülsbusch, Glückstadt, 2013. Disponible en ligne : http://www.wissgrid.de/publikationen/Leitfaden_Data-Management-WissGrid.pdf

¹⁵⁹ <http://opendatacommons.org/licenses/odbl/>

ANNEXE 6 ARBRE DE DECISION POUR CHOISIR UN SCHEMA DE DIFFUSION DE SES DONNÉES

Où déposer ses données?



*Recommandation: Privilégier les entrepôts publics et reconnus selon les critères proposés par *Re3data.org*

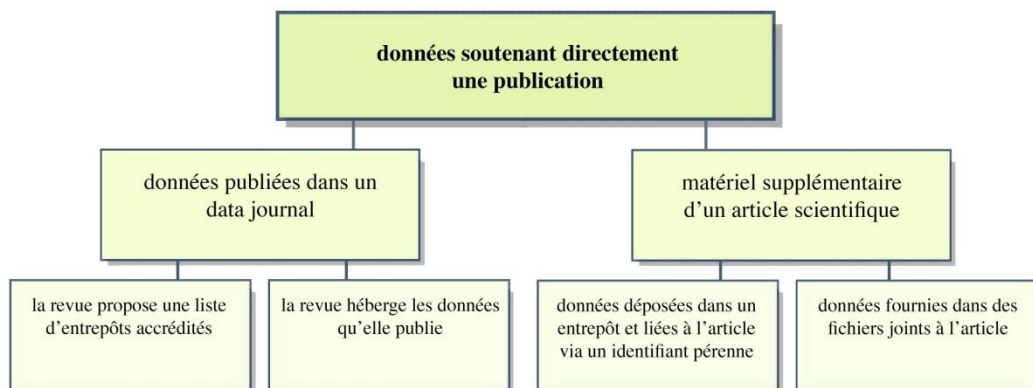


TABLE DES ILLUSTRATIONS

Figure 1 : Typologie des données utilisées au cours de travaux de recherche sur le site alsacien	37
Figure 2 : Typologie des données produites sur le site alsacien en sus des publications scientifiques	38
Figure 3 : Initiatives françaises autour de la gestion des données de recherche	40

TABLE DES MATIERES

SIGLES ET ABBREVIATIONS	7
INTRODUCTION	9
LES ARCHIVES OUVERTES DE LA CONNAISSANCE DANS LE PAYSAGE INSTITUTIONNEL FRANÇAIS ET EUROPEEN DE LA RECHERCHE	11
1. Les archives ouvertes institutionnelles aujourd’hui	11
1.1. Comment sont-elles définies ?.....	11
1.1.1. Qu’est-ce qu’une archive ouverte institutionnelle ?	11
1.1.2. A quels principes répondent-elles ?	11
1.2. Comment sont-elles encadrées ?.....	14
1.2.1. Au niveau mondial.....	14
1.2.2. Au niveau européen	15
1.2.3. Au niveau national.....	17
1.3. Quels sont les enjeux ?.....	18
1.3.1. Enjeux politiques.....	18
1.3.2. Enjeux juridiques.....	20
1.3.3. Enjeux techniques.....	23
1.4. Quelques exemples d’archives institutionnelles françaises.....	25
2. Les Archives Ouvertes de la Connaissance : environnement du projet 25	
2.1. Description du projet	25
2.1.1. Genèse du projet	26
2.1.2. Acteurs.....	27
2.2. Racines du projet	28
2.2.1. Un contexte politique favorable.....	28
2.2.2. Des expériences sur lesquelles prendre appui	29
2.3. Les objectifs spécifiques du projet.....	31
2.3.1. Intégrer les données de recherche	31
2.3.2. Couvrir l’ensemble de la production scientifique du site...	31
2.3.3. Instaurer une politique d’archivage.....	32
2.3.4. Accompagner les chercheurs	33
PRATIQUES DE GESTION DES DONNEES DE LA RECHERCHE SUR LE SITE ALSACIEN : ETUDE EXPLORATOIRE	37
1. L’ouverture des données de recherche	37
1.1. Questionnement autour de la définition des données de recherche	37

1.2. Evolution épistémologique de la recherche scientifique : de l'Open Access à l'Open Science.....	39
1.3. Des mouvements incitatifs	40
2. Enquêtes auprès des chercheurs du site alsacien	42
2.1. Méthodologie	42
2.1.1. Analyse de l'existant.....	42
2.1.2. Analyse d'une première enquête sur les pratiques de publication scientifique et de production de données de recherche sur le site alsacien	43
2.1.3. Entretiens avec des enseignants-chercheurs de différentes unités de recherche	44
2.2. Conclusions des enquêtes	45
2.2.1. Découverte d'un potentiel important mais non formulé sur le site alsacien	45
2.2.2. Les freins à l'ouverture des données de recherche	45
2.2.3. Quelques-uns des enjeux liés à l'ouverture des données de recherche	46
3. Préconisations	49
3.1. Mettre en place un « écosystème » favorable à l'ouverture des données de la recherche	49
3.1.1. Par des moyens politiques	49
3.1.2. Par des moyens techniques et humains.....	50
3.2. Etude de cas : le naturaliste Jean Hermann et la base de données prosopographique de ses réseaux.....	52
3.2.1. Présentation de la base de données	53
3.2.2. Méthodologie de l'étude de cas	53
3.2.3. Conclusions de l'étude de cas.....	54
CONCLUSION.....	57
SOURCES	59
BIBLIOGRAPHIE	61
ANNEXES	65
TABLE DES ILLUSTRATIONS.....	81
TABLE DES MATIERES.....	83