

Diplôme de conservateur de bibliothèque (DCB)

Mémoire d'étude / janvier 2014

**De l'Open data à l'Open research data :
quelle(s) politique(s) pour les données
de recherche ?**

Rémi Gaillard

Sous la direction de Monique Joly
Directrice du SCD de l'INSA de Lyon, responsable du département Études &
Prospective du consortium Couperin

Remerciements

Je remercie très vivement, bien entendu, Monique Joly, pour son accompagnement attentif, sa bienveillance et son enthousiasme, et pour toute l'aide apportée au cours de la rédaction.

Merci, bien sûr, à Grégory Colcanap et à l'ensemble du consortium Couperin d'avoir proposé ce sujet passionnant.

Merci à Odile Hologne, directrice de l'IST à l'INRA, d'avoir pris le temps de me donner son point de vue sur l'orientation de ce mémoire.

Je souhaite également remercier, pour leur disponibilité, Olivier Legendre, responsable de la bibliothèque numérique de Clermont Université, son collègue Léonard Bourlet, Thierry Chanier, directeur du Laboratoire de recherche sur le langage de l'université Blaise Pascal de Clermont-Ferrand. J'adresse un remerciement tout particulier à Marie-Madeleine Géroutet, dans les pas de laquelle je me suis inscrit, qui a bien voulu me communiquer le rapport sur les données de recherche qu'elle a rédigé avant d'intégrer le SCD de Lille 1. Merci également à André Dazy, Christine Okret-Manville, Charlotte Maday, Magalie Moysan, Lourdes Fuentes-Hashimoto, Thierry Claerr, Sébastien Respingue-Perrin, Serge Fdida, Sébastien Peyrard et Emeline Dalsorg, pour leurs conseils et le temps consacré à répondre à mes questions.

Merci à mes amis et collègues, notamment Mathilde, Solen, Antoine, Marc, Sophie, Élise, Lucie, Julien, Aurore et Claire, qui savent si bien accompagner les moments de pause – avec ou sans houblon...

... et à ma famille pour sa présence et son soutien. Et en premier lieu, Marine, qui a de la patience quand j'en manque, qui n'ignore aucun de mes questionnements philosophiques sur le véritable sens des politiques de données anglo-saxonnes, et que je remercie aussi pour tout le reste.

Résumé :

Le mouvement du libre-accès aux publications scientifiques s'élargit de plus en plus aux données de la recherche. Des initiatives pour garantir l'accessibilité et la complète réutilisation de ces données sont prises par une grande diversité d'acteurs – États, agences de financement de la recherche, éditeurs, communautés scientifiques. L'ouverture des données de la recherche est rendue possible par la définition de politiques incitatives ou contraignantes, l'adoption de solutions juridiques et techniques, mais repose avant tout sur de bonnes pratiques de gestion des données. Tandis que la France s'insère progressivement dans la dynamique de l'*Open research data*, les universités sont appelées à définir leur politique de données. Les bibliothécaires ont un rôle majeur à jouer dans l'élaboration de ces politiques, peuvent contribuer à identifier les besoins des chercheurs et les assister sur le volet « métadonnées ». Aussi, la question de l'ouverture des données de recherche offre une opportunité unique à ces professionnels de la documentation : celle de remodeler, à l'échelle des établissements de recherche, leur(s) lien(s) avec la communauté des chercheurs.

Descripteurs :

Données de la recherche*

Politique scientifique

Sciences -- Vulgarisation

Édition en libre accès

Bibliothèques universitaires

Abstract :

Open access movement is increasingly expanding from scientific publications to research data. Initiatives to make research data broadly accessible and fully available for reuse are emerging from a variety of stakeholders at international and European scale – states, funding agencies, publishers and scientific communities themselves. Research data openness is achieved through different policies, technical and legal mechanisms, but also lie on good data management practices. While France is beginning to get into the Open research data movement, it is important to consider that data policies are also needed at university level. Librarians have a serious part to play in leading on institutional data policy, understanding researchers needs and helping in metadata creation. Open access to research data is thus a unique opportunity for librarians to redesign their connection with scientific communities inside research institutions.

Keywords :

Research data*

Open access publishing

Science and state

Science news

Academic libraries

Droits d'auteurs



Cette création est mise à disposition selon le Contrat : « **Paternité-Pas d'Utilisation Commerciale-Pas de Modification 2.0 France** » disponible en ligne <http://creativecommons.org/licenses/by-nc-nd/3.0/deed.fr> ou par courrier postal à Creative Commons, 171 Second Street, Suite 300, San Francisco, California 94105, USA.

Sommaire

SIGLES ET ABREVIATIONS	9
INTRODUCTION.....	11
DEFINITIONS & JALONS POLITIQUES	15
Data et datasets : essai de définition	15
<i>Qu'est-ce qu'une donnée ?</i>	<i>15</i>
<i>De la donnée brute à la donnée dérivée</i>	<i>17</i>
<i>La donnée et son cycle de vie.....</i>	<i>19</i>
L'ouverture des données, un engagement européen	21
<i>L'apparition d'une préoccupation.....</i>	<i>21</i>
<i>L'engagement progressif de l'Union européenne en faveur de l'ouverture des données</i>	<i>22</i>
Les politiques des agences de financement : un phénomène international	29
<i>Les cas américain, britannique et canadien.....</i>	<i>29</i>
<i>Du chercheur à l'institution de recherche, vers une responsabilité partagée</i>	<i>32</i>
OUVRIR LES DONNEES : JEU D'ACTEURS, SOLUTIONS JURIDIQUES ET TECHNIQUES	35
Les chercheurs face à leurs données : impulsions éditoriales, institutionnelles et disciplinaires	35
<i>Incitations éditoriales.....</i>	<i>35</i>
<i>Les mandats institutionnels : l'importance de la gestion des données et des DMP.....</i>	<i>38</i>
<i>Logique bottom-up et culture du partage.....</i>	<i>42</i>
Les enjeux juridiques de l'ouverture	44
<i>Données de recherche et propriété intellectuelle.....</i>	<i>44</i>
<i>... « un cauchemar juridique »</i>	<i>46</i>
<i>Permettre la réutilisation : licences & waivers.....</i>	<i>48</i>
Des infrastructures complémentaires ? Études de cas.....	51
<i>La diversité des solutions disciplinaires.....</i>	<i>52</i>
<i>Solutions éditoriales et data journals.....</i>	<i>53</i>
<i>Fédérer des entrepôts interopérables : l'ambition d'OpenAIRE.....</i>	<i>54</i>
DONNEES, ETABLISSEMENTS DE RECHERCHE ET BIBLIOTHEQUES.....	57
La prise en main des données par les établissements de recherche : méthodologie.....	57
<i>Des responsabilités des institutions en matière de gestion et de conservation</i>	<i>57</i>

<i>Faut-il un entrepôt institutionnel ?</i>	60
Bibliothèques et données : ce qu’elles font et comment elles doivent s’y préparer	64
<i>L’implication des bibliothèques</i>	65
<i>Les bibliothécaires sont-ils les bons interlocuteurs ? Compétences et formations.</i>	68
Bilan : qui faut-il mobiliser, en France, pour participer à l’élaboration de politiques de données ?	73
<i>À l’échelle des établissements, le levier de la gouvernance et de la coordination</i>	73
<i>Capitaliser sur l’expérience acquise : le CINES</i>	74
<i>Une impulsion nationale ?</i>	76
CONCLUSION	79
BIBLIOGRAPHIE	81
Généralités	81
Jalons politiques	82
Synthèses et rapports	84
Les politiques de données : initiatives éditoriales et disciplinaires	85
Les données, aspects juridiques et techniques	87
Les universités et la gestion des données de recherche	90
Bibliothèques et données de la recherche	91
TABLE DES ANNEXES	93
TABLE DES ILLUSTRATIONS	101
TABLE DES MATIERES	103

Sigles et abréviations

AAF	Association des archivistes français
ADBU	Association des directeurs et personnels de direction des bibliothèques universitaires et de la documentation
BIUS	Bibliothèque interuniversitaire de santé
BSN	Bibliothèque scientifique numérique
BUPMC	Bibliothèque de l'université Pierre et Marie Curie
CC0	<i>Creative Commons Zero</i>
CDS	Centre de données astronomiques de Strasbourg
CER	Conseil européen de la recherche
CINES	Centre informatique national de l'enseignement supérieur
CNRS	Centre national de la recherche scientifique
CPU	Conférence des présidents d'université
DCC	<i>Digital Curation Center</i>
DMP	<i>Data management plan</i>
DSI	Direction des systèmes d'information
EER	Espace européen de la recherche
EML	<i>Ecological Metadata Language</i>
EPST	Établissement public à caractère scientifique et technologique
INRA	Institut national de la recherche agronomique
IST	Information scientifique et technique
JISC	<i>Joint Information Systems Committee</i>
MESR	Ministère de l'enseignement supérieur et de la recherche
MISTRD	Mission de l'information scientifique et technique et du réseau documentaire
NIH	<i>National Institutes of Health</i>
NSF	<i>National Science Foundation</i>
OAIS	<i>Open Archival Information System</i>
OCDE	Organisation de coopération et de développement économiques
OSTP	<i>Office of Science and Technology Policy</i>
PCRD	Programme-cadre de recherche et de développement
RDA	<i>Research Data Alliance</i>
RDM	<i>Research data management</i>
SCD	Service commun de documentation
TGE	« Très grand équipement »
TGIR	« Très grande infrastructure de recherche »

INTRODUCTION

L'année 2014 sera sans doute, en France, l'année des données de la recherche. Depuis le printemps 2012 et l'organisation par le CNRS d'une journée d'étude intitulée *Données de la recherche : enjeux, perspectives, politique(s)*, une vraie réflexion a été engagée sur la question, suscitant même la création d'un site d'information national.¹ L'année 2013 a aussi été particulièrement riche, avec l'annonce du lancement d'un nouveau chantier consacré aux données de recherche dans le cadre du projet Bibliothèque Scientifique Numérique (BSN),² l'organisation d'une journée d'étude lors du congrès annuel de l'ADBU en septembre,³ celle des journées FRÉDOC du 7 au 10 octobre 2013 à Aussois, sur le thème *Gestion et valorisation des données de la recherche*.⁴ En novembre, le Groupement Français de l'Industrie de l'Information (GFII) devait également organiser une journée d'étude sur *Les données de la recherche dans l'écosystème des publications scientifiques*, finalement reportée au 12 février 2014.⁵

À l'échelle internationale et plus particulièrement européenne, l'actualité « brûlante »⁶ du sujet s'est incarnée dans le lancement par la Commission Européenne, le 16 décembre 2013, d'un projet pilote pour le libreaccès aux données de la recherche. Cette impulsion, qui constitue un jalon politique essentiel, témoigne de la vitalité de la question de *l'ouverture des données de la recherche* qui fait l'objet de ce mémoire, et que la Commission Européenne résume en ces termes :

« Les chercheurs œuvrant aux projets qui participent à cette initiative pilote sont invités à mettre à la disposition des autres chercheurs, des industries innovantes et des citoyens les données scientifiques qu'ils auront obtenues, et notamment les données nécessaires pour valider les résultats présentés dans des publications scientifiques. Cette initiative sera un facteur de progrès scientifique, d'efficacité et de transparence, ce qui profitera aux citoyens comme à la société. Elle contribuera aussi à la croissance économique, en rendant l'innovation plus accessible »⁷

Cet idéal d'ouverture n'est pas nouveau et a été décrit en détail en juin 2012 dans un rapport de la Royal Society de Londres intitulé *Science as an open enterprise*.⁸ Il répond à des objectifs variés et complémentaires qui, pour partie,

¹ <http://www.donneesdelarecherche.fr/> (consulté le 28 décembre 2013)

² Lors des cinquièmes Journées Open Access Couperin, organisées à Paris les 24 et 25 janvier 2013.

³ *Données de la recherche : quel rôle pour la documentation ?*, journée d'étude du congrès annuel de l'ADBU, Le Havre, 19 septembre 2013.

⁴ Cf. leur compte-rendu : <http://www.donneesdelarecherche.fr/spip.php?article402> (consulté le 28 décembre 2013)

⁵ Le programme de cette journée est disponible : <http://www.gfii.fr/fr/evenement/journee-d-etude-les-donnees-de-la-recherche-dans-l-ecosysteme-des-publications-scientifiques-12-fevrier-2014> (consulté le 28 décembre 2013)

⁶ Pour reprendre l'expression de Christophe Pérales, président de l'ADBU, lors de la journée d'étude du 19 septembre 2013 au Havre.

⁷ COMMISSION EUROPEENNE, « La Commission lance un projet pilote de libre accès aux données issues de la recherche financée sur fonds publics », 16 décembre 2013, http://europa.eu/rapid/press-release_IP-13-1257_fr.htm (consulté le 28 décembre 2013)

⁸ THE ROYAL SOCIETY, *Science as an open enterprise : summary report*, The Royal Society, juin 2012, 104 p.,

font écho aux justifications données au mouvement de l'*Open data*, comme la nécessité de faire profiter l'ensemble des acteurs économiques du potentiel commercial et industriel des données collectées sur fonds publics.⁹ C'est d'ailleurs dans le contexte de la recherche que l'expression même d'*Open data* a été forgée, en 1995, dans un rapport américain du *National Research Council* sur « l'échange complet et ouvert des données scientifiques ». ¹⁰ Outre cet argument d'« efficacité » des investissements publics, qui ne va pas sans susciter pour les données de recherche les mêmes questionnements éthiques que pour l'ensemble des données publiques,¹¹ l'importance de permettre la vérification et la reproductibilité des résultats de recherche, la nécessité d'éviter les fraudes et de favoriser la réutilisation – dans d'autres disciplines ou pour d'autres recherches – de données disponibles, sont souvent mis en avant. L'intérêt de certaines données à long terme, au-delà des projets au cours desquels elles ont été produites, est bien connu dans certaines disciplines. Comme l'a par exemple récemment rappelé Françoise Genova, directrice du Centre de données astronomiques de Strasbourg (CDS), « l'utilisation des données de l'archive du satellite IUE (1978-1996) [a] produit cinq fois plus de publications scientifiques que l'exploitation initiale des données par les équipes » qui les ont obtenues.¹² De même, le nombre d'articles basés sur la réutilisation des données archivées du télescope Hubble, en activité depuis 1990, dépasse celui des articles basés sur les observations initiales.¹³ Les cas de fraudes scientifiques constituent également des arguments de poids pour les partisans de l'*Open research data*.¹⁴ La controverse suscitée en avril 2013 par un article publié en 2010 par les économistes Kenneth Rogoff et Carmen Reinhart, « l'immense supercherie », révélée en 2009, d'un chercheur en cristallographie qui publiait des structures moléculaires factices,¹⁵ font partie des exemples souvent mobilisés pour convaincre de la nécessité que la totalité des données associées à des articles soient systématiquement rendues publiques. Un éditorial paru en 2012 dans *Nature*

http://royalsociety.org/uploadedFiles/Royal_Society_Content/policy/projects/sape/2012-06-20-SAOE.pdf

(consulté le 27 septembre 2013)

⁹ CARTIER Aurore, *Bibliothèques et Open data. Et si on ouvrait les bibliothèques sur l'avenir ?*, mémoire DCB sous la dir. de Frédérique Schlosser, 2013, 107 p., p. 21-24, <http://www.enssib.fr/bibliotheque-numerique/documents/60401-bibliotheque-et-open-data-et-si-on-ouvrait-les-bibliotheques-sur-l-avenir.pdf> (consulté le 15 octobre 2013)

¹⁰ LACOMBE Romain, BERTIN Pierre-Henri, VAUGLIN François & VIEILLEFOSSE Alice, *Pour une politique ambitieuse des données publiques. Rapport remis au ministre de l'Industrie le 13 juillet 2011*, 2011, 111 p., p. 25, <http://www.ladocumentationfrancaise.fr/var/storage/rapports-publics/114000407/0000.pdf> (consulté le 30 décembre 2013)

¹¹ Un chercheur peut-il accepter de « libérer » ses données au profit d'une entreprise qui en fera une réutilisation commerciale ? Sur cette question, cf. HAGEDORN Gregor & alii, « Creative Commons licenses and the non-commercial condition : Implications for the re-use of biodiversity information », *ZooKeys*, novembre 2011, n° 150, p. 127-149, <http://www.ncbi.nlm.nih.gov/pmc/articles/PMC3234435/> (consulté le 30 octobre 2013)

¹² GENOVA Françoise, « Le CDS : des données au service de la communauté scientifique », *Documentaliste*, vol. 50, n° 3, octobre 2013, p. 47-49.

¹³ LAGERSTROM Jill, « Measuring the Impact of the Hubble Space Telescope : open data as a catalyst for science », *76th IFLA General Conference and Assembly, 10-15 August 2010, Gothenburg, Sweden*, 2010, 11 p., <http://conference.ifla.org/past/2010/155-lagerstrom-en.pdf> (consulté le 30 décembre 2013)

¹⁴ DOORN Peter, DILLO Ingrid, VAN HORIK René, « Lies, Damned Lies and Research Data : Can Data Sharing Prevent Data Fraud ? », *The International Journal of Digital Curation*, vol. 8, n° 1, 2013, p. 229-243, <http://www.ijdc.net/index.php/ijdc/article/view/8.1.229/308> (consulté le 31 décembre 2013)

¹⁵ HODSON Simon, « Seven rules of successful research data management in universities », *The Guardian Higher Education Network blog*, juillet 2013, <http://www.jisc.ac.uk/blog/seven-rules-of-successful-research-data-management-in-universities-16-jul-2013> (consulté le 30 décembre 2013); FOUCART Stéphane, « L'immense supercherie d'un chercheur en cristallographie », *Le Monde*, 25 décembre 2009, http://lemonde.fr/planete/article/2009/12/25/l-immense-supercherie-d-un-chercheur-en-cristallographie_1284874_3244.html (consulté le 30 décembre 2013); THE ROYAL SOCIETY, *op. cit.*, 2012, p. 16.

alerterait encore sur le nombre « choquant » d'articles, en particulier dans le domaine de la recherche médicale, dont les conclusions ne pouvaient être vérifiées et reproduites faute de données suffisantes ou suffisamment documentées.¹⁶ Il relevait également l'importance de publier la totalité des résultats d'expériences, et pas seulement les données « positives » corroborant une hypothèse de travail ou appuyant une démonstration.¹⁷ Ainsi que l'ont récemment exposé dans *Le Monde* les fondateurs de *Deuxième Labo*, *think tank* français sur les mutations du monde de la recherche,¹⁸ et Yvan Stroppa, ingénieur au CNRS :

« Le 13 février [2013], Daniele Fanelli, chercheur à l'université d'Edimbourg spécialiste de l'intégrité scientifique, proposait dans la revue *Nature*¹⁹ d'élargir la définition de la fraude scientifique à toute omission ou déformation de l'information nécessaire et suffisante pour évaluer la validité et l'importance d'une recherche. [...] Ainsi, la lutte contre la fraude scientifique se jouerait plus sur le terrain de la communication des résultats que sur celui du comportement des chercheurs. **La culture de la reproductibilité est une alliée de l'intégrité scientifique** »²⁰

« Accompagner les publications scientifiques des jeux de données et codes sources qui permettront [à d'autres] de reproduire les résultats » fait donc partie des solutions avancées par les tenants d'une « science ouverte » (*open science*), irriguée par la libre circulation des informations et des connaissances. Cette nouvelle manière d'envisager la science, parfois définie comme la « science 2.0 » ou l'« e-Science », met les données de la recherche au premier plan : entrée dans l'ère des « grandes masses de données » (*big data*), la science ne serait plus seulement le fruit d'observations expérimentales, d'hypothèses ou de théories, mais naîtrait désormais aussi de la manipulation et de l'agrégation des données, par exemple par les technologies du *data mining*.²¹ La possibilité d'entreprendre de nouvelles recherches à partir de grands ensembles de données fait donc partie des promesses de la *Data-driven science*,²² mais requiert la mise en place de mécanismes de gestion, de conservation et de partage des données. Pour toutes ces raisons, les données produites par les chercheurs, dans de nombreuses disciplines, font désormais l'objet de toutes les attentions ; la nécessité de leur « ouverture » - les rendre accessibles, intelligibles et réutilisables - est progressivement intégrée par les acteurs politiques, les organismes de financement et les établissements de recherche à l'échelle internationale.

¹⁶ NATURE, « Must try harder », mars 2012, vol. 483, p. 509, <http://www.nature.com/nature/journal/v483/n7391/pdf/483509a.pdf> (consulté le 26 décembre 2013)

¹⁷ Sur ce point, voir également : COUZIN-FRANKEL Jennifer, « The Power of Negative Thinking », *Science*, vol. 342, n° 6154, p. 68-69, octobre 2013, <http://www.sciencemag.org/content/342/6154/68.full> (consulté le 30 décembre 2013)

¹⁸ <http://www.deuxieme-labo.fr/> (consulté le 30 décembre 2013)

¹⁹ FANELLI Daniele, « Redefine misconduct as distorted reporting », *Nature*, 13 février 2013, <http://www.nature.com/news/redefine-misconduct-as-distorted-reporting-1.12411> (consulté le 30 décembre 2013)

²⁰ BLANCHARD Antoine, SABUNCU Elifsu & STROPPA Yvan, « Pour une recherche reproductible, publiez vos codes et données », *Le Monde*, 15 juillet 2013, http://lemonde.fr/sciences/article/2013/07/15/pour-une-recherche-reproductible-publiez-vos-codes-et-donnees_3447825_1650684.html (consulté le 30 décembre 2013)

²¹ MARX Vivien, « Biology : The big challenges of big data », *Nature*, n° 498, juin 2013, p. 255-260, <http://www.nature.com/nature/journal/v498/n7453/full/498255a.html> (consulté le 31 décembre 2013)

²² INSTITUT NATIONAL DE LA RECHERCHE AGRONOMIQUE [INRA], *Rapport du groupe de travail sur la gestion et le partage des données*, Paris, juin 2012, 62 p., http://www.pfl-cepia.inra.fr/uploads/gdp_docs/Rapport-GestionDonnees-web.pdf (consulté le 7 octobre 2013)

Malgré l'effervescence de ces derniers mois, la France accuse un retard considérable sur ces questions. Des organismes de recherche, toutefois, défrichent le terrain, comme l'Institut national de la recherche agronomique et, plus récemment, le Centre national de la recherche scientifique. Ce dernier a en effet annoncé, en décembre 2013, avoir défini « une nouvelle stratégie en matière d'information scientifique et technique », adoptée par le collège de direction du CNRS en novembre.²³ Un volet important de cette stratégie concerne la valorisation et le partage des données de recherche.²⁴ Quoi qu'il en soit, la question de l'ouverture ne peut pas s'envisager qu'à l'échelle des établissements, les données se situant au croisement des préoccupations de nombreux acteurs : celles des organismes de recherche, certes, mais aussi celles des agences de financement, des éditeurs et, en premier lieu, des communautés scientifiques qui les produisent, les manipulent, les exploitent et, surtout, décident ou non de les « ouvrir ».

Est-il pertinent, dès lors, que des politiques de données, entendues comme les outils de gouvernance par lesquels les chercheurs peuvent être incités à (ou contraints de) gérer, documenter, archiver et partager leurs données, soient définies à l'échelle, institutionnelle, d'un organisme de recherche ou d'une université ? Si oui, dans quel écosystème s'intègrent-elles, quelles questions juridiques et techniques doivent-elles prendre en compte ? Plus concrètement, enfin, à l'échelle d'un établissement, comment peuvent-elles être élaborées, sur quelles infrastructures doivent-elles reposer, les compétences des professionnels de l'IST doivent-elles être mobilisées ?

De nombreux exemples internationaux peuvent être développés pour répondre à ces questions. Nous commencerons, dans une première partie, par définir ce que sont les « données de recherche » et par évoquer les différentes étapes de l'histoire politique de l'*Open research data*, dans le contexte européen et sous l'influence, à l'échelle internationale, des organismes de financement de la recherche. Ensuite, nous analyserons les « conditions » de l'ouverture, sous l'angle concret des impulsions éditoriales, institutionnelles et disciplinaires qui conduisent les chercheurs à développer une « culture du partage ». Nous nous arrêterons sur le statut juridique problématique de la « donnée » et sur les outils disponibles pour en faciliter la réutilisation, ainsi que sur les solutions techniques (identifiants pérennes, entrepôts de données, infrastructures de signalement) qui garantissent la visibilité des données de recherche, notamment en les reliant ou en les intégrant aux publications. Enfin, nous élargirons la focale sur la manière dont les politiques de données, dans ce contexte où l'ouverture se joue aussi en dehors des établissements de recherche, se structurent à l'échelle de ces établissements. Ce sera l'occasion de voir la place qu'occupent les bibliothèques dans ces dispositifs, la manière dont l'expertise des professionnels de l'IST est sollicitée et, sous un angle plus prospectif, qui sont les acteurs mobilisables, en France, pour que les universités ne laissent pas passer le train de l'*Open research data*.

²³ « CNRS : vers une science publique ouverte », 13 décembre 2013, <http://www.gfii.fr/fr/document/cnrs-vers-une-science-publique-ouverte> (consulté le 31 décembre 2013)

²⁴ CENTRE NATIONAL DE LA RECHERCHE SCIENTIFIQUE, *Schéma d'orientation stratégique de l'information scientifique et technique (IST)*. « Mieux partager les connaissances », novembre 2013, 49 p., p. 34-37, <http://www.cnrs.fr/dist/docs/131119-orientation.pdf> (consulté le 31 décembre 2013)

DEFINITIONS & JALONS POLITIQUES

DATA ET DATASETS : ESSAI DE DEFINITION

Qu'est-ce qu'une donnée ?

La principale difficulté que représente la définition, dans le champ de la recherche scientifique, de la notion de « donnée de recherche »²⁵, est qu'il faut s'efforcer d'identifier ce qui peut rassembler des éléments aussi divers qu'un cliché de coléoptère prélevé à Madagascar, un spectrohélogramme produit à Meudon, des informations sur les gènes d'une moisissure, les relevés météorologiques d'un vaisseau ayant traversé l'Atlantique au XVIII^e siècle ou l'enregistrement d'un dialecte rare²⁶. Tous ces éléments ont en commun d'être (désormais) des données numériques, produites au cours d'un processus de recherche et pouvant servir de support à une démonstration scientifique : elles diffèrent néanmoins par la manière dont elles ont été obtenues, la forme sous laquelle elles ont été enregistrées et le traitement qu'elles ont subi pour être formatées, lisibles. La diversité des objets que rassemble l'étiquette de « données de la recherche » pourrait l'assimiler, comme l'écrit Sylvie Fayet, à « cette grande valise que me confie un voyageur bien sous tous rapports mais dont je n'ai pas vérifié le contenu »...²⁷

Des essais de définition ont été proposés dans de nombreuses publications. Au Royaume-Uni, l'université de Bristol, qui s'efforce depuis 2011 à travers le projet *Data.Bris* d'établir une politique institutionnelle pour les données de ses chercheurs, a produit *Une introduction à la gestion des données de recherche*²⁸ et un glossaire dans lesquels les données de recherche sont définies comme :

« Les données, ou unités d'information, qui sont créées au cours d'une recherche, subventionnée ou non, et qui sont organisées ou formatées de telle sorte qu'elles soient communicables, interprétables et adaptées à un traitement souvent informatisé. »

²⁵ Dans les pages qui suivent, nous utiliserons alternativement les notions de « donnée de recherche » et « données de la recherche » comme recoupant celle, anglo-saxonne, de *research data*.

²⁶ Je fais respectivement référence à des jeux de données que l'on peut trouver dans le *Biodiversity Data Journal* (<http://biodiversitydatajournal.com/articles.php?id=1016>), dans les bases de données BASS 2000 (<http://bass2000.obsppm.fr/home.php?lang=fr>), AspGD (<http://www.aspgd.org>), PANGAEA (<http://www.pangaea.de>) ou dans le *Speech and Language Data Repository* (<http://crdo.up.univ-aix.fr>), consultés le 4 décembre 2013.

²⁷ FAYET Sylvie, « « Données » de la recherche, les mal-nommées », 15 novembre 2013, <http://urfistinfo.hypotheses.org/2581> (consulté le 4 décembre 2013)

²⁸ « *Data, or units of information which are created in the course of funded or unfunded research, and often arranged or formatted in a such a way as to make them suitable for communication, interpretation, and processing, perhaps by a computer.* » *An Introduction to Managing Research Data*, août 2013, 5 p., http://data.bris.ac.uk/research/introduction/files/2013/08/Introduction-to-research-data-management-for-researchers-v3_0.pdf (consulté le 5 décembre 2013)

Les universités américaines, comme celle de l’Oregon,²⁹ ont assimilé la définition délivrée par le Bureau de la gestion et du budget du gouvernement fédéral américain dans une circulaire de 1993 amendée en 1999 :

« La donnée de recherche est définie comme l’enregistrement factuel couramment considéré dans la communauté scientifique comme nécessaire à la validation des résultats de la recherche »³⁰

Cette définition a été reprise en 2007 par l’OCDE dans ses *Principes et lignes directrices pour l’accès aux données de la recherche financée sur fonds publics*.³¹ Aussi, avec Marie-Madeleine Géroutet,³² on peut effectivement retenir comme définition la plus large des *données* de la recherche qu’elles consistent en des enregistrements factuels nécessaires à la validation des *résultats* de la recherche, ces derniers étant habituellement délivrés sous la forme de publications. Cependant, « données » et « résultats » entretiennent une relation ambiguë, l’une des caractéristiques du mouvement d’ouverture étant justement de ne pas tenir pour intéressantes les seules données servant de justification à un résultat publié ; par ailleurs, il repose sur une définition plus large de la notion de « résultat » qui peut englober, outre les publications, les données produites au cours du processus de recherche. Si la donnée n’est donc pas une publication, elles peuvent être publiées (dans le corps d’un article, ou sous la forme de *supplementary files*), et font bien partie, avec les publications, des « produits » de la recherche (*research outputs*).

Les choses se compliquent lorsque l’on prend en compte l’ensemble des éléments produits par les chercheurs dans le cours de leur activité scientifique, comme les carnets de laboratoire, la correspondance, les carnets de terrain (*field notebooks*), les analyses préliminaires, les projets et les rapports de recherche.³³ Tous ces éléments, qui s’apparentent plus volontiers à des archives, sont pourtant parfois considérés comme des « données de recherche » à part entière, notamment dans la définition qu’en donnent certaines universités britanniques ou australiennes.³⁴ Pourtant, ces objets sont explicitement exclus de la définition fournie par le gouvernement américain dans la circulaire déjà citée.³⁵ Une approche archivistique, qui fait de ces documents, au même titre que les publications et les

²⁹ « Defining Research Data », <http://library.uoregon.edu/datamanagement/datadefined.html#one> (consulté le 5 décembre 2013)

³⁰ « Research data is defined as the recorded factual material commonly accepted in the scientific community as necessary to validate research findings ». OFFICE OF MANAGEMENT AND BUDGET, *Circulaire A-110 amendée le 30/09/99*, http://www.whitehouse.gov/omb/circulars_a110#36 (consulté le 5 décembre 2013)

³¹ 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> (consulté le 29 mai 2013)

³² GEROUDET Marie-Madeleine, *Étude prospective sur les données de la recherche*, rapport sous la dir. d’Isabelle Le Bescond, mai 2013, SCD Univ. Lille 1, 62 p. [document non public, communiqué par l’auteur expurgé des éléments qui ne pouvaient être diffusés qu’en interne]. Je remercie à nouveau l’auteur de m’avoir donné accès à ce document.

³³ *Référentiel de gestion* pour le « traitement et [la] conservation des archives des laboratoires de recherche, des chercheurs et des enseignants-chercheurs dans les universités, les organismes de recherche et les agences sanitaires », septembre 2012, 16 p., p. 12

<http://f.hypotheses.org/wp-content/blogs.dir/1176/files/2013/02/tableaugestionarchivesaurores.pdf> (consulté le 3 décembre 2013)

³⁴ Cf. la définition donnée par l’université de Melbourne, citée par l’AUSTRALIAN NATIONAL DATA SERVICE, « What is research data ? », <http://ands.org.au/guides/what-is-research-data.html> (consulté le 5 décembre 2013)

³⁵ GEROUDET Marie-Madeleine, *op. cit.*, 2013, p. 11.

enregistrements factuels effectués, les différents « jalons d'une même chaîne de recherche », ³⁶ n'est toutefois pas satisfaisante si l'on se positionne dans la perspective qui nous intéresse ici et qui est celle de l'*Open Research Data*. Plutôt que d'en donner une définition typologique, il est donc préférable d'aborder la notion sous l'angle plus utilitaire de la *validation* (ce qui est nécessaire à la validation des résultats) et de la *réutilisation* (ce qui constitue un « matériau primaire » de la recherche et qui est donc potentiellement utile à d'autres).

Dès lors, en suivant la *Royal Society* de Londres, on peut considérer les données comme :

« Des informations qualitatives ou quantitatives [...] qui sont factuelles. Ces données 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** » ³⁷

Il s'agit donc de l'enregistrement de « faits donnés », sous une forme numérique, descriptive ou visuelle, et « sur [lequel] un argument, une théorie, une hypothèse ou tout autre produit de la recherche est basé. Ces données peuvent être brutes, nettoyées ou traitées, et peuvent être enregistrées sous tout format et tout support ». ³⁸

De la donnée brute à la donnée dérivée

À partir de quel moment une donnée n'est-elle plus un simple enregistrement, objectif, du réel ? La définition est délicate à trancher et la notion de donnée « brute » (*raw data*) doit être maniée avec précaution. ³⁹ Certains chercheurs sont en effet d'avis que la notion de « donnée » doit être exclusivement réservée à la caractérisation de données brutes, tout en reconnaissant que la ligne de fracture qui les sépare des données « traitées » (*processed*) ou « dérivées » (*derived*) est difficile à repérer :

« Pour quelques-uns le terme "data" doit être limité aux données brutes, pour d'autres la notion inclut n'importe quel type d'information ou d'opération qui aboutit à une idée. Nous préférons limiter l'usage du terme aux **données brutes, neutres, objectives, qui ne dépendent pas de leur contexte de création, d'une analyse ou de leur producteur**. Dès lors qu'elles sont délimitées, filtrées et sélectionnées, elles acquièrent ou se voient donner un sens particulier dans le contexte auquel elles s'appliquent. C'est là une partie du

³⁶ FAYET Sylvie, *op. cit.*, 2013.

³⁷ « 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 ». THE ROYAL SOCIETY, *Science as an open enterprise : summary report*, The Royal Society, juin 2012, 104 p., p. 9.

³⁸ Sans doute l'une des définitions les plus claires, empruntée à la *Queensland University of Technology*. Citée par par l'AUSTRALIAN NATIONAL DATA SERVICE, « What is research data ? », <http://ands.org.au/guides/what-is-research-data.html> (consulté le 5 décembre 2013)

³⁹ Certains allant jusqu'à considérer qu'une donnée n'est jamais brute. Cf. GITELMAN Lisa éd., *Raw Data is an Oxymoron*, 2013, MIT Press.

processus qui transforme les données en information. Il n'y a pas de point clair de transition. »⁴⁰

Pourtant, même brute, une donnée est rarement dissociable du contexte dans lequel elle a été produite, des conditions techniques qui ont abouti à son prélèvement, etc. En repartant de notre définition initiale des données comme « enregistrements factuels », différentes catégories de données peuvent être identifiées, comme le précise l'*Australian National Data Service*, l'initiative australienne dédiée à la construction d'une infrastructure de données :

*« Les données [de recherche] peuvent être des données brutes, des données non traitées d'observations de phénomènes particuliers. D'autres sont des données traitées, données produites après formatage ou correction de données brutes. D'autres des données dérivées, qui présentent un résumé ou une présentation spécifique des données brutes. »*⁴¹

Une classification canonique des différents types de données de recherche peut être proposée, les différenciant suivant la manière dont elles sont produites et leur valeur supposée⁴² :

- **données d'observation**, collectées à l'instant T, nécessitent un appareil descriptif conséquent (conditions, méthodologie, équipement, etc.). Indissociables d'un contexte donné, et donc uniques et *impossibles à reproduire*. Ont vocation à être conservées de façon pérenne.

Exemples : neuroimagerie, relevés de concentration en phytoplanctons, cliché astronomique.

- **données expérimentales**, obtenues à partir d'équipements en laboratoire, suivant une méthodologie bien définie. *Potentiellement reproductibles*, mais à des coûts parfois prohibitifs. Leur conservation doit donc dépendre des investissements engagés dans leur production et de leur possible reproductibilité.

Exemples : chromatogrammes, puces à ADN, cinétique chimique.

- **données computationnelles ou de simulation**, issues de simulations à partir de modèles informatiques. *Potentiellement reproductibles* si le modèle informatique est correctement documenté.

Exemples : modèles de simulation sismique, modèles météorologiques.

⁴⁰ « For some it is limited to raw data, for others the term widens to include any kind of information or process that leads to insights. We prefer to limit the term to neutral, objective, raw data that are largely independent of context, analysis or observer. As data become constrained, filtered and selected, they acquire or are assigned a meaning in the context of what they apply to. This is part of the process that transforms data into information. There is no clear point of transition ». THESSEN Anne E., PATTERSON David J., « Data Issues in the Life Sciences », *ZooKeys*, novembre 2011, n° 150, p. 15-51, p. 17, <http://www.ncbi.nlm.nih.gov/pmc/articles/PMC3234430/> (consulté le 28 octobre 2013)

⁴¹ ANDS, « What do we mean by research data collections ? », <http://ands.org.au/guides/research-data-australia.html> (consulté le 5 décembre 2013)

⁴² On retrouve cette typologie dans la définition des *research data* que donnent de nombreuses universités britanniques ou américaines. Elle est notamment empruntée à un rapport de référence du *National Science Board* de la NATIONAL SCIENCE FOUNDATION, *Long-Lived Digital Data Collections : Enabling Research and Education in the 21st Century*, septembre 2005, 87 p., p. 19, <http://www.nsf.gov/pubs/2005/nsb0540/nsb0540.pdf> (consulté le 5 décembre 2013)

Quant aux données « dérivées », elles sont issues du traitement, de la combinaison ou de la réorganisation de données brutes, pour les rendre plus lisibles ou les présenter sous une forme canonique.⁴³

Dans ce cadre, la notion de « jeu de données » (*dataset*) peut être définie comme l'agrégation, sous une forme lisible, de données brutes ou dérivées présentant une certaine « unité », rassemblées pour former un ensemble cohérent. Toutefois, l'échelle à laquelle ces données assemblées acquièrent leur unité pour former un « jeu » varie selon les disciplines, les types de données (cf. supra), les projets, les raisons pour lesquelles ces données sont agrégées.⁴⁴ Sous l'angle spécifique de l'« ouverture » des données de recherche, on peut définir le jeu de données comme un enregistrement de données sous la forme d'un ou plusieurs fichiers électroniques, téléchargeables, citables (notamment par l'intermédiaire d'un DOI)⁴⁵ et intelligibles – ce jeu étant accompagné des métadonnées descriptives suffisantes.⁴⁶

La donnée et son cycle de vie

Si l'accent est souvent mis sur les données brutes, c'est l'ensemble des données de recherche, brutes ou dérivées, qui peuvent être concernées par l'ouverture. C'est donc le chercheur qui, en premier lieu, est appelé à *définir* ce qui constitue une donnée unique et/ou réutilisable devant être conservée, ainsi que le stade de son cycle de vie (sous une forme « brute » ou traitée) qui doit être préservé et potentiellement diffusé. Dans un article récent, des chercheurs américains en écologie ont souligné la nécessité de partager les données sous leur forme la plus primaire, correspondant au tout début de leur cycle de vie, tout en soulignant l'intérêt que pouvait présenter la diffusion simultanée des données brutes et des données dérivées :

« Il peut se révéler très compliqué d'agréger des données de plusieurs sources qui ont, chacune, été traitées d'une manière différente. Aussi, pour rendre vos données aussi utiles que possible, le mieux est de les partager sous leur forme la plus brute, [...] la plus proche des observations et des mesures de terrain à partir desquelles vous avez bâti votre analyse. Cela ne signifie pas forcément que vos données se prêtent le mieux à l'analyse sous leur forme brute, mais les diffuser sous cette forme donne à l'utilisateur plus de flexibilité. [...] Diffuser simultanément les données sous leur forme brute et dérivée, en expliquant les différences dans les métadonnées, est un moyen simple de faire profiter des avantages respectifs des données sous ces deux formes. »⁴⁷

⁴³ NSF, *op. cit.*, 2005, p. 17.

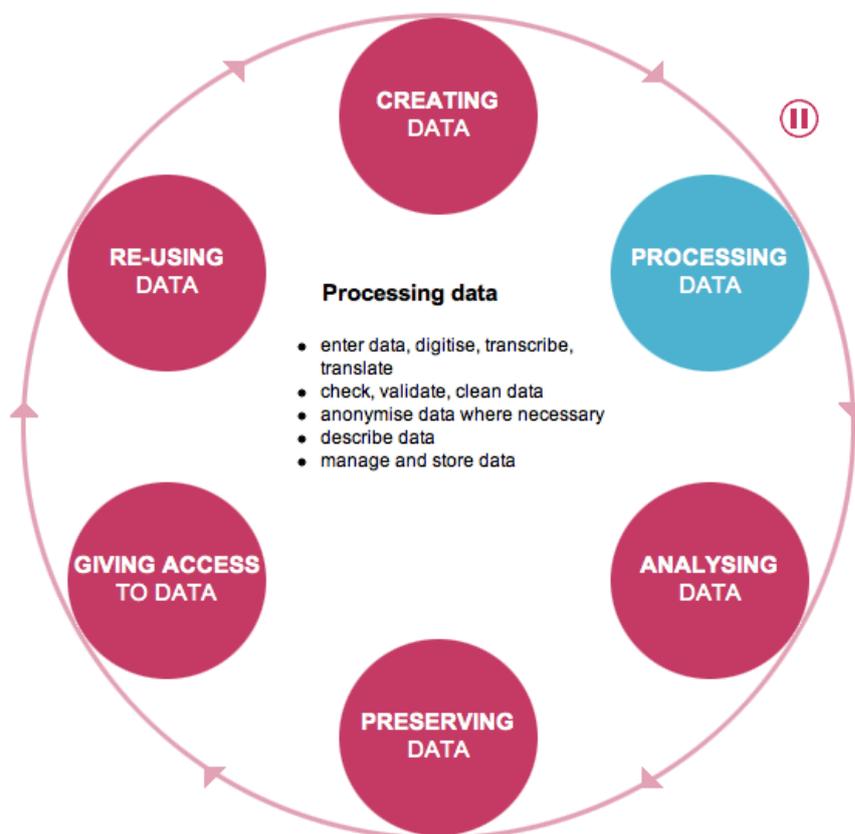
⁴⁴ GEROUDET Marie-Madeleine, *op. cit.*, 2013, p. 13.

⁴⁵ Cf. p. 52-53, p. 63.

⁴⁶ DODDS Leigh, « What is a dataset ? », février 2013, <http://blog.ldodds.com/2013/02/09/what-is-a-dataset/> (consulté le 26 décembre 2013)

⁴⁷ WHITE Ethan P. et alii, « Nine simple ways to make it easier to (re)use your data », *Ideas in Ecology and Evolution*, vol. 6, n°2, 2013, p. 1-10, p. 3, « 3. Provide an unprocessed form of the data », <http://library.queensu.ca/ojs/index.php/IEE/article/view/4608/4898> (consulté le 2 octobre 2013)

La responsabilité des producteurs de données était également soulignée dans le rapport de la *National Science Foundation* américaine en 2005 : ce sont les communautés de chercheurs, dans leurs champs disciplinaires respectifs, qui doivent « contribuer aux premières décisions sur la nature des données à archiver (notamment, données brutes *versus* données dérivées) et sur la durée de leur conservation »⁴⁸. S'il y a désormais consensus sur la nécessité de conserver (et de diffuser) certains types de données bien spécifiques, comme celles des puces à ADN ou des structures macromoléculaires, la délimitation des corpus concernés par une conservation pérenne peut varier selon les habitudes disciplinaires et, surtout, doit se négocier à l'échelle de chaque projet de recherche aboutissant à la production de données. C'est là l'une des fonctions essentielles des « plans de gestion des données » (*data management plans*) qui sont indissociables des politiques d'ouverture et qui doivent permettre de définir le cycle de vie de la donnée.⁴⁹ Il convient donc de différencier, dans notre essai de définition, les données « pérennes », dont la valeur justifie une conservation à long-terme, des données « intermédiaires », produites au cours du processus de recherche mais dont la conservation pérenne ne s'impose pas.⁵⁰



Ill. 1. Le cycle de vie de la donnée de recherche ⁵¹

⁴⁸ *Idem*, p. 28.

⁴⁹ Cf. *infra* p. 38.

⁵⁰ INSTITUT NATIONAL DE LA RECHERCHE AGRONOMIQUE [INRA], *Rapport du groupe de travail sur la gestion et le partage des données*, Paris, juin 2012, 62 p., p. 10, http://www.pfl-cepia.inra.fr/uploads/gdp_docs/Rapport-GestionDonnees-web.pdf (consulté le 7 octobre 2013)

⁵¹ UK DATA ARCHIVE, « Research Data Lifecycle », <http://data-archive.ac.uk/create-manage/life-cycle> (consulté le 9 décembre 2013)

Au regard de la problématique de l'ouverture, toutes les données n'ont donc pas la même « valeur » : selon les disciplines, selon la manière dont a été planifiée la collecte, suivant les objectifs pour lesquels elles ont été produites (intégration à un corpus plus large, projet limité à l'échelle d'un laboratoire, etc.), l'intérêt que peut présenter leur conservation et leur diffusion varie. « Certaines données ont une valeur immédiate et durable, certaines prennent de la valeur au fil du temps, d'autres n'ont qu'une valeur transitoire, certaines enfin sont plus faciles à recréer qu'à préserver »⁵²... De cette valeur, que seul le producteur des données a la capacité de déterminer,⁵³ dépend finalement la nécessité et la faisabilité de leur ouverture.

L'OUVERTURE DES DONNEES, UN ENGAGEMENT EUROPEEN

L'apparition d'une préoccupation

Dès la déclaration fondatrice de Berlin, en 2003, les données scientifiques entrent dans le champ du débat relatif au libre-accès, puisque avec les résultats de la recherche « les données brutes » font partie des « contributions » souhaitées au mouvement *Open Access*.⁵⁴ L'année suivante, le Comité de la politique scientifique et technologique (CPST) de l'OCDE réuni à Paris rappelle que « des efforts coordonnés aux niveaux national et international sont nécessaires pour *élargir l'accès aux données de la recherche financée sur fonds publics* et contribuer à faire progresser la recherche scientifique et l'innovation »⁵⁵ et une *Déclaration sur l'accès aux données de la recherche financée par des fonds publics* est publiée. Le principe de l'« ouverture » (*openness*) y est adopté, même si les gouvernements le tempèrent en évoquant « la nécessité de restreindre l'accès dans certains cas pour protéger des intérêts sociaux, scientifiques et commerciaux ». La rédaction de lignes directrices pour l'accès aux données de la recherche est en tout cas décidée. Elles sont approuvées par le CPST en octobre 2006, adossées à une recommandation de l'OCDE et entérinées par son Conseil à la fin de la même année.

⁵² BORGMAN L. Christine, « Research Data : Who will share what, with whom, when and why ? », *Fifth China-North America Library Conference 2010, 8-12 septembre 2010, Beijing*, 21 p., p. 3, <http://works.bepress.com/cgi/viewcontent.cgi?article=1237&context=borgman> (consulté le 9 décembre 2013)

⁵³ Cf. infra p. 39.

⁵⁴ RECODE, *Deliverable D1 : Stakeholder Values and Ecosystems*, septembre 2013, 101 p., p. 7, http://recodeproject.eu/wp-content/uploads/2013/10/RECODE_D1-Stakeholder-values-and-ecosystems_Sept2013.pdf (consulté le 23 octobre 2013)

⁵⁵ ORGANISATION DE COOPERATION ET DE DEVELOPPEMENT ECONOMIQUES (OCDE), « Déclaration sur l'accès aux données de la recherche financée par des fonds publics adoptée le 30 janvier 2004 à Paris », <http://www.oecd.org/fr/science/sci-tech/sciencetechnologieetinnovationpourle21emesieclereunionducomitedelapolitiquescientifiqueettechnologiquedelocdeaniveauministeriel29-30janvier2004-communiquefinal.htm> (consulté le 14 octobre 2013)

Ces *Principes* (et la recommandation à laquelle ils sont joints), bien que n'étant pas « juridiquement contraignant[s] », ⁵⁶ ont constitué un jalon important dans l'histoire « politique » de l'ouverture des données. Leur objectif est avant tout de « donner des orientations aux institutions cherchant à définir une ligne de conduite » en matière de gestion, de conservation et de mise à disposition des données. Parmi ces recommandations, la principale est bien sûr l'ouverture, par défaut, des données numériques de la recherche financée sur fonds publics ⁵⁷. Le document insiste par ailleurs déjà sur des points essentiels, comme la nécessité de privilégier une approche disciplinaire, de mettre en place des mécanismes incitatifs pour les chercheurs et de penser la pérennité de l'accès aux données avant même que ces données ne soient produites. ⁵⁸ S'ils n'ont pas eu d'effets immédiats dans l'ensemble des pays membres, les *Principes* de l'OCDE ont eu un impact particulièrement important au Royaume-Uni, où ils sont à la racine de l'engagement des organismes de financement de la recherche en matière d'*open research data*.

L'engagement progressif de l'Union européenne en faveur de l'ouverture des données

Parallèlement, l'Union européenne structure sa politique en matière de libre accès aux résultats de la recherche. Le 17 décembre 2007, le Conseil scientifique du Conseil européen de la recherche (CER) ⁵⁹ publie des recommandations demandant la mise en accès libre des résultats de recherches financées par le CER, dans un délai de six mois suivant leur publication, dans des archives ouvertes disciplinaires appropriées (*research repositories*) ou des archives ouvertes institutionnelles (*institutional repository*) ⁶⁰. Le document, par ailleurs, accorde une large place à la question de l'accès aux données « brutes » :

« 2. Le CER pense qu'il est essentiel que les données brutes ⁶¹ – comme par exemple, en sciences de la vie : les séquences de nucléotides / protéines, les coordonnées atomiques macromoléculaires et les données épidémiologiques anonymisées – soient déposées dans les bases de données adéquates le plus rapidement possible, de préférence dès publication et au plus tard dans les 6 mois. » ⁶²

⁵⁶ 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., p. 8, <http://www.oecd.org/fr/science/sci-tech/38500823.pdf> (consulté le 29 mai 2013)

⁵⁷ Ce que rappellera la communication de la Commission Européenne du 14 février 2007 « sur l'information scientifique à l'ère numérique ».

⁵⁸ ORGANISATION DE COOPERATION ET DE DEVELOPPEMENT ECONOMIQUES (OCDE), *Principes et lignes directrices*, *op. cit.*, p. 26-27.

⁵⁹ Organisme de financement paneuropéen de la recherche, mis en place dans le cadre du 7^e PCRD (FP7, 2007-2013). <http://erc.europa.eu/about-erc/mission> (consulté le 15 octobre 2013). Sur les PCRD, cf. infra p. 25.

⁶⁰ <http://www.openaire.eu/fr/component/content/article/48-erc-guidelines> (consulté le 15 octobre 2013); Marine... p. 16-17s

⁶¹ « Primary data », dans le document non traduit. Sur ces questions lexicales, cf. supra.

⁶² 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> (consulté le 15 octobre 2013);

Le CER citait notamment le cas de grands entrepôts de données thématiques, comme la *DNA DataBank of Japan* (DDBJ), l'*European Molecular Biology Laboratory* (EMBL) *Nucleotide Sequence Database* et *GenBank*, la base de séquences ADN du grand organisme de financement de la recherche américain, les *National Institutes of Health* (NIH). Ces bases de données de référence, lancées dans les années 1980, connectées et fédérées au sein de l'*International Nucleotide Sequence Database Collaboration*, qui assurent la collecte et l'archivage de l'ensemble des séquences nucléotidiques primaires connues, sont donc ultra-spécialisées⁶³. À l'instar de la *Worldwide Protein Data Bank* (wPDB), collection mondiale de données sur la structure 3D de macromolécules biologiques, ces archives constituent presque des exceptions : d'abord parce que dans leur périmètre disciplinaire, elles incarnent un modèle d'archives centralisées vers lesquelles convergent naturellement les données produites dans ces disciplines, ensuite parce qu'elles ne se concentrent que sur un type de données bien spécifique, des données « hautement standardisées »⁶⁴. Aussi, dès 2007, le Conseil européen de la recherche, tout en considérant comme prioritaires les données pour lesquelles des infrastructures thématiques dédiées existent, juge indispensable que l'ensemble des données sur lesquelles s'appuient les résultats des recherches qu'il finance soient accessibles en libreaccès après la publication de ces résultats.

À la suite des recommandations du CER, la Commission Européenne s'est d'abord focalisée sur les résultats de la recherche, en entérinant l'obligation de dépôt dans une archive ouverte, après une période d'embargo possible de 6 à 12 mois, des publications issues de 20 % des projets financés par le 7^{ème} Programme-cadre pour la recherche et le développement technologique (PCRD). L'infrastructure OpenAIRE (*Open Access Infrastructure for Research in Europe*) a été mise en place à la fin de l'année 2009 pour accompagner la réussite de cet *Open Access Pilot*,⁶⁵ qui ne concernait que les articles – et pas les données sous-jacentes⁶⁶.

Les données de la recherche se sont cependant retrouvées au cœur des préoccupations européennes après la publication en octobre 2010 du rapport *Riding the wave. How Europe can gain from the rising tide of scientific data*⁶⁷, rédigé à l'attention de la Commission Européenne par son « Groupe d'experts sur la gestion des données scientifiques ». Cette étude, qui dresse le portrait idéal de ce que devrait être, à l'horizon 2030, une infrastructure collaborative des données de la

⁶³ NICOL Aurore, CARUSO Julie & ARCHAMBAULT Éric, *Open Data Access Policies and Strategies in the European Research Area and Beyond*, Science-Metrix, août 2013, 16 p., p. 5, http://www.science-metrix.com/pdf/SM_EC_OA_Data.pdf (consulté le 27 septembre 2013). Voir également « GenBank Overview », <http://www.ncbi.nlm.nih.gov/genbank/> (consulté le 15 octobre 2013) ; *Science as an open...*, p. 83

⁶⁴ WHITE Ethan P. et alii, « Nine simple ways to make it easier to (re)use your data », *Ideas in Ecology and Evolution*, vol. 6, n°2, 2013, p. 1-10,

<http://library.queensu.ca/ojs/index.php/IEE/article/view/4608/4898> (consulté le 2 octobre 2013)

⁶⁵ COMMISSION EUROPEENNE, *Open Access Pilot in FP7*, 2008, 3 p., http://ec.europa.eu/research/science-society/document_library/pdf_06/open-access-pilot_en.pdf (consulté le 21 octobre 2013)

⁶⁶ COMMISSION EUROPEENNE, « Communication de la Commission [...]. Pour un meilleur accès aux informations scientifiques : dynamiser les avantages des investissements publics dans le domaine de la recherche », Bruxelles, Commission Européenne, 17 juillet 2012, p. 10, <http://eur-lex.europa.eu/LexUriServ/LexUriServ.do?uri=COM:2012:0401:FIN:FR:PDF> (consulté le 14 octobre 2013).

⁶⁷ [COMMISSION EUROPEENNE] HIGH LEVEL-EXPERT GROUP ON SCIENTIFIC DATA, *Riding the wave : how Europe can gain from the rising tide of scientific data*, UE, octobre 2010, 36 p., <http://cordis.europa.eu/fp7/ict/e-infrastructure/docs/hlg-sdi-report.pdf> (consulté le 21 octobre 2013)

recherche, garantissant « leur accessibilité, leur utilisation, leur réutilisation et leur fiabilité »⁶⁸, délivre un certain nombre de recommandations à destination de la Commission Européenne. L'ambition principale, pour 2030, est ainsi décrite :

*« Tous les acteurs, des autorités scientifiques et étatiques à l'ensemble du public, sont conscients de l'importance critique de la conservation et du partage des données [...] produites au cours du processus de recherche. Les chercheurs et les spécialistes de toutes les disciplines peuvent trouver les données dont ils ont besoin, y accéder et les traiter. Ils peuvent avoir confiance en leur capacité à utiliser et comprendre ces données, et peuvent évaluer leur degré de fiabilité. Les producteurs de données tirent bénéfice du fait de les ouvrir le plus largement possible [opening it to broad access], et préfèrent entreposer leurs données en toute confiance dans des dépôts fiables »*⁶⁹

Pour remplir ces objectifs, le rapport préconise notamment que « tous les États-membres publient leur politique et leur feuille de route concernant la conservation et le partage des données scientifiques », que « les agences de financement tiennent compte de la publication des données [...] dans l'avancement de carrière des chercheurs » et que « les agences européennes et nationales imposent la rédaction de plans de gestion des données ».

La Commission Européenne s'est appropriée les conclusions de ce rapport le 17 juillet 2012, dans une communication adressée au Parlement européen⁷⁰, puis dans une recommandation aux États-membres « relative à l'accès aux informations scientifiques et à leur conservation »⁷¹ suivie, le même jour, d'une autre communication sur l'Espace européen de la recherche (EER)⁷².

La première communication reconnaît qu'« il devient de plus en plus important d'améliorer l'accès aux données de la recherche » et de ne plus cantonner le débat sur le libreaccès aux seules publications scientifiques.⁷³ Elle pose comme principal objectif qu'avant 2014, « des politiques de libreaccès aux articles et données scientifiques [soient] établies dans tous les États membres à tous les niveaux pertinents ». C'est naturellement la principale préconisation de la recommandation du 17 juillet 2012, qui invite les États européens à « définir des politiques claires en matière de diffusion des données de la recherche [...] et de libre accès à ces dernières » et à garantir qu'elles deviennent « accessibles, utilisables et réutilisables par le public au moyen d'infrastructures électroniques »⁷⁴. Cette préconisation « politique » s'accompagne de remarques

⁶⁸ *Idem*, p. 4.

⁶⁹ [COMMISSION EUROPEENNE] HIGH LEVEL-EXPERT GROUP ON SCIENTIFIC DATA, *op. cit.*, p. 4.

⁷⁰ COMMISSION EUROPEENNE, « Communication de la Commission [...]. Pour un meilleur accès aux informations scientifiques [...] », *op. cit.* Cf. en particulier les points 4.2, 5.2.

⁷¹ COMMISSION EUROPEENNE, « Recommandation de la Commission relative à l'accès aux informations scientifiques et à leur conservation », Bruxelles, Commission Européenne, 17 juillet 2012, http://ec.europa.eu/research/science-society/document_library/pdf_06/recommendation-access-and-preservation-scientific-information_fr.pdf (consulté le 21 octobre 2013)

⁷² COMMISSION EUROPEENNE, « Communication de la Commission [...]. Un partenariat renforcé pour l'excellence et la croissance dans l'Espace européen de la recherche », Bruxelles, Commission Européenne, 17 juillet 2012, http://ec.europa.eu/research/era/pdf/era-communication/era-communication_fr.pdf (consulté le 21 octobre 2013).

⁷³ COMMISSION EUROPEENNE, « Communication de la Commission [...]. Pour un meilleur accès aux informations scientifiques [...] », 17 juillet 2012, *op. cit.*, p. 4

⁷⁴ COMMISSION EUROPEENNE, « Recommandation de la Commission [...] », 17 juillet 2012, *op. cit.*, p. 7.

techniques, la recommandation de la Commission abordant également, succinctement, la question des métadonnées, des identifiants pérennes ou de l'interopérabilité des dépôts.

Ces prises de position récentes des instances européennes sont indissociables de leur objectif de structurer, avant 2014, un « Espace européen de la recherche » (*European Research Area*), défini comme « un espace de recherche unifié [...] dans lequel chercheurs, connaissances scientifiques et technologies circulent librement et grâce auquel l'Union et ses États membres renforceront leurs bases scientifiques et technologiques »⁷⁵. Dans sa communication sur l'EER du 17 juillet, la Commission réaffirme son souci de « promouvoir une innovation ouverte (*Open Innovation*) ». Elle incite à nouveau les États membres à définir leurs politiques d'accès aux informations scientifiques et responsabilise également les organisations d'acteurs de la recherche (« chercheurs, universités, organismes de financement et d'exécution »), les invitant à « adopter et mettre en œuvre des mesures en faveur de la liberté d'accès aux publications et données issues de la recherche financée par des fonds publics ». Dans ce même document, enfin, la Commission s'engage à :

« [mettre] en place un accès libre aux publications scientifiques, en tant que principe général applicable à tous les projets financés au titre du programme Horizon 2020 ; pour les données de la recherche, elle développera une méthode souple tenant compte des différents domaines scientifiques et des intérêts des entreprises »⁷⁶

Les données dans *Horizon 2020* et l'élaboration d'un *Pilot for Open Research Data* ⁷⁷

Si elles constituent bien des marqueurs politiques importants, les recommandations de l'Union européenne n'ont qu'une valeur incitative et ne sont pas légalement contraignantes pour les États membres. Cependant, l'Union européenne est aussi un organisme de financement de la recherche et a vocation, à ce titre, à déterminer les règles d'accès et de dissémination de l'information scientifique issue de ses crédits. Les « Programmes-cadres pour la recherche et le développement technologique » (PCRD) constituent le principal instrument européen de financement de la recherche. Le prochain PCRD, baptisé *Horizon 2020* (2014-2020), comporte de forts engagements en faveur de l'*open access* qui prolongent et étendent les expérimentations du PCRD précédent (2007-2013).

À partir de 2014, en effet, *toutes les publications* issues de recherches menées avec l'aide de fonds du programme *Horizon 2020* devront être mises à

⁷⁵ COMMISSION EUROPEENNE, « Communication de la Commission [...]. Un partenariat renforcé », 17 juillet 2012, *op. cit.*, p. 3

⁷⁶ *Idem*, p. 16.

⁷⁷ La plupart des éléments qui suivent sont tirés de l'intervention de Kirsti Ala-Mutka, membre de l'unité « Digital Science » à la DG Connect (« Directorate General for Communication Networks, Content and Technology ») de la Commission Européenne, lors de la journée d'étude ADBU du 19 septembre 2013. ALA-MUTKA Kirsti, « Open Access and Digital Science in Horizon 2020 », *Données de la recherche : quel rôle pour la documentation ?*, journée d'étude du congrès annuel de l'ADBU, Le Havre, 19 septembre 2013

disposition en libre-accès, selon les voies verte ou dorée – les coûts de publication engagés dans ce dernier cas pouvant être remboursés par la Commission Européenne.⁷⁸

Par ailleurs, le libre-accès aux données scientifiques devient un des objectifs stratégiques du 8^{ème} PCRD : après l'*Open Access Pilot* de 2008, la Commission, comme elle l'avait annoncé dans sa communication du 17 juillet 2012⁷⁹, développe actuellement un *Open Research Data Pilot* afin de promouvoir l'ouverture et la réutilisation des données (et des métadonnées associées) générées au cours de recherches financées par *Horizon 2020*. Il n'est pas envisagé, pour l'instant, d'obliger les chercheurs à ouvrir leurs données, mais la Commission, ainsi que l'a annoncé Kirsti Ala-Mukta lors du congrès de l'ADBU du Havre, réfléchit à la possibilité d'appliquer à l'ouverture des données le principe de l'*opt-out* (« *opting out of openness* »). Le pilote devrait avant tout insister sur les conditions préliminaires indispensables à la réussite d'une politique d'ouverture, en incitant par exemple les chercheurs à rédiger, en amont, des plans de gestion des données (*data management plans*). Des consultations publiques ont en tout cas été menées, notamment en février et juillet 2013, associant chercheurs, financeurs, industriels, éditeurs et bibliothécaires.

À l'issue de la consultation du 2 juillet 2013,⁸⁰ l'infrastructure OpenAIREplus (qui a succédé en décembre 2011 à OpenAIRE), la Ligue des bibliothèques européennes de recherche (LIBER) et la *Confederation of Open Access Repositories* (COAR) ont formulé des recommandations sur le périmètre à donner à l'*Open Research Data Pilot*. Ces acteurs particulièrement impliqués dans la promotion de l'ouverture des données préconisent notamment :

- qu'une campagne de lobbying sur les bienfaits du partage des données soit organisée en Europe à l'attention des chercheurs ;
- que le pilote couvre l'ensemble des données de recherche et métadonnées associées produites dans le cadre d'un projet financé par l'Europe, dès lors que ce projet a fait l'objet d'une publication (rapport, compte-rendu, livrable, article, etc.) ;
- que toutes ces données soient déposées dans au moins une archive ouverte de données (*open data repository*) librement accessible, garantissant si possible la conservation des données à long terme ;
- d'ouvrir ces données grâce aux licences appropriées (CC0, ODbL) ;

⁷⁸ COMMISSION EUROPEENNE, « Données scientifiques : le libre accès aux résultats de recherche stimulera la capacité d'innovation en Europe », 17 juillet 2012, http://europa.eu/rapid/press-release_IP-12-790_fr.htm (consulté le 22 octobre 2013).

⁷⁹ COMMISSION EUROPEENNE, « Communication de la Commission [...]. Pour un meilleur accès aux informations scientifiques [...] », 17 juillet 2012, *op. cit.*, p. 11.

⁸⁰ REILLY Susan, « European Commission Public Consultation on Open Research Data : some impressions from the day », 3 juillet 2013, <http://www.libereurope.eu/blog/european-commission-public-consultation-on-open-research-data-some-impressions-from-the-day> (consulté le 22 octobre 2013) ; « EC Public consultation on open research data on 2 July in Brussels », 13 juin 2013, <http://ec.europa.eu/digital-agenda/en/news/ec-public-consultation-open-research-data-2-july-brussels> (consulté le 22 octobre 2013).

- que les données sur lesquelles reposent les conclusions d'un article (*research publication*) soient déposées dans une archive ouverte de données dès sa publication ;
- que les données générées au cours d'une recherche faisant l'objet d'une publication, mais qui ne sont pas nécessairement liées aux conclusions de cette publication, soient quand même déposées et ouvertes, « aussitôt que possible ».⁸¹

En février 2013, des acteurs du monde de la recherche ont également donné leur avis sur le contenu du pilote, recommandant que les recherches financées dans le cadre du premier pilier d'*Horizon 2020*, l'« Excellence scientifique », soient obligatoirement concernées par le pilote et que celui-ci contienne une liste des archives où pourraient être déposées les données. Ils insistent surtout sur l'absolue nécessité que la Commission, avant de contraindre à l'ouverture, oblige les chercheurs à produire, systématiquement, des plans de gestion des données.⁸²

L'élaboration de ce pilote est la dernière illustration de l'intérêt que porte la Commission Européenne aux données de la recherche. Dès 2007, avant même les recommandations du CER, elle relevait, dans sa « Communication sur l'information scientifique à l'ère numérique »⁸³ la nécessité d'appuyer des initiatives « qui conduisent à un accès plus large et à une meilleure diffusion de l'information scientifique [...], en particulier en ce qui concerne les publications et les données brutes générées par les activités de recherche soutenues par un financement public »⁸⁴. Ce positionnement s'inscrit dans une dynamique européenne et internationale⁸⁵ plus large d'engagement en faveur de l'*Open data*, marqué par de nombreuses directives (directive PSI n°2003/98/CE, directive Inspire n°2007/2/CE), mais l'intérêt de la Commission Européenne pour les données de recherche ne se limite pas à la rédaction de recommandations à destination des États membres.

Elle finance en effet depuis de nombreuses années des infrastructures de recherche (IR), disciplinaires ou non, consacrées à la valorisation et à la

⁸¹ OPENAIRE, LIBER & COAR, « Horizon 2020 – Outline of a Pilot for Open Research Data. Joint statement by OpenAIRE, LIBER and COAR », 3 juillet 2013, 4 p., p. 2-3, http://www.coar-repositories.org/files/Horizon_2020_Open_Data_Pilot_20130703_final.pdf (consulté le 21 octobre 2013)

⁸² « Recommendations of the participants of the workshop on Open Access to Research Data to the European Commission », 21 février 2013, disponible en ligne (consulté le 24 octobre 2013). Ils ont été entendus, puisque le 16 décembre 2013 la Commission a annoncé que certains projets bénéficiant de crédits dans le cadre d'*Horizon 2020* devraient obligatoirement comporter un plan de gestion des données : cf. COMMISSION EUROPEENNE, *Guidelines on Data Management in Horizon 2020*, v. 1.0, 16 décembre 2013, 6 p., http://ec.europa.eu/research/participants/data/ref/h2020/grants_manual/hi/oa_pilot/h2020-hi-oa-data-mgt_en.pdf (consulté le 20 décembre 2013). Voir la conclusion de ce mémoire.

⁸³ COMMISSION EUROPEENNE, « Communication de la Commission [...] sur l'information scientifique à l'ère numérique : accès, diffusion et préservation », Bruxelles, Commission Européenne, 14 février 2007, <http://eur-lex.europa.eu/LexUriServ/LexUriServ.do?uri=COM:2007:0056:FIN:FR:PDF> (consulté le 21 octobre 2013)

⁸⁴ *Idem*, p. 9.

⁸⁵ Pour ne citer que cet exemple récent, on se reportera à la *Charte pour l'ouverture des données publiques* signée par le G8 en juin 2013. Elle encourage d'ailleurs l'ouverture des « données relatives au génome humain » et des « résultats d'expérience ». Cf. <https://www.gov.uk/government/publications/open-data-charter/g8-open-data-charter-and-technical-annex> (consulté le 24 octobre 2013)

préservation des données scientifiques. En 2012 la Commission prévoyait encore de consacrer, grâce au 7^{ème} PCRD, « 45 millions d'euros à des dépenses dans le domaine des infrastructures permettant le libre accès à des articles et données de recherche » et dans celui de la conservation numérique.⁸⁶ Cet effort sera poursuivi dans le cadre du prochain PCRD. Parmi ces infrastructures qui envisagent plus ou moins directement la question de l'accessibilité des données, et sur lesquelles nous reviendrons, l'on peut citer le projet EUDAT (*European Data*, 2011-2014), dédié à la construction d'une infrastructure de préservation des données, la « Collaborative Data Infrastructure » (CDI),⁸⁷ et OpenAIREplus (2011-2014), dont l'objectif principal est de donner plus de visibilité aux jeux de données en les reliant aux publications et à des informations sur les projets de recherche dont ils sont issus. D'autres infrastructures abordent ces questions sous un angle plus disciplinaire, comme DARIAH (*Digital Research Infrastructure for the Arts and Humanities*), qui cherche à faciliter l'accessibilité et l'utilisation à long terme de l'ensemble des données de recherche produites en Europe dans toutes les disciplines des sciences humaines et sociales.⁸⁸ Dans le domaine des sciences du vivant, l'infrastructure distribuée ELIXIR a vocation à fédérer les initiatives en matière de collecte, de gestion, de conservation et de diffusion de données biologiques,⁸⁹ et le projet ViBRANT (*Virtual Biodiversity Research and Access Network for Taxonomy*), financé par le 7^{ème} PCRD, vise à la création d'un écosystème collaboratif unifié où les chercheurs travaillant sur la biodiversité puissent partager leurs données⁹⁰. ViBRANT a ainsi contribué au lancement récent d'un journal en libre-accès, le *Biodiversity Data Journal*, destiné à la publication de jeux de données taxonomiques⁹¹.

L'Europe agit donc dans de très nombreuses directions et à travers de nombreuses initiatives pour faciliter l'accessibilité et la diffusion des données de la recherche. Ses incitations à ouvrir les données ne revêtent cependant aucune dimension contraignante et l'opposition de certains États membres font encore obstacle à l'enracinement de cette dynamique d'ouverture⁹². Pourtant, certains organismes étatiques de financement de la recherche, à l'échelle internationale, placent aujourd'hui l'ouverture des données au cœur de leurs préoccupations.

⁸⁶ COMMISSION EUROPEENNE, « Communication de la Commission [...]. Pour un meilleur accès aux informations scientifiques [...] », 17 juillet 2012, *op. cit.*, p. 14.

⁸⁷ « Le projet européen EUDAT », <http://www.cines.fr/spip.php?rubrique385> (consulté le 23 octobre 2013) ; MASSOL Marion, « L'archivage à l'heure de l'Europe », *La Gazette du Cines. Archivage numérique pérenne*, CINES, février 2013, 59 p., p. 53-56, http://www.cines.fr/IMG/pdf/CINES_GAZETTE_Special_Archivage-2.pdf (consulté le 23 octobre 2013)

⁸⁸ Cf. <http://dariah.eu/about/our-partners.html> (consulté le 23 octobre 2013). L'antenne nationale de DARIAH est la TGIR (« Très Grande Infrastructure de Recherche ») Huma-Num, issue de la fusion, le 1^{er} mars 2013, du TGE Adonis et du Corpus-IR.

⁸⁹ ELIXIR, *ELIXIR. Data for life*, 14 p., http://www.elixir-europe.org/sites/elixir-europe.org/files/documents/elixir_square_brochure_final_printed.pdf (consulté le 24 octobre 2013)

⁹⁰ Cf. <http://vibrant.eu/node/1> (consulté le 24 octobre 2013)

⁹¹ SMITH Vincent & alii, « Beyond Dead Trees : Integrating the Scientific Process in the *Biodiversity Data Journal* », *Biodiversity Data Journal*, septembre 2013, 8 p., <http://biodiversitydatajournal.com/articles.php?id=995> (consulté le 24 octobre 2013)

⁹² GEROUDET Marie-Madeleine, *Étude prospective sur les données de la recherche*, rapport sous la dir. d'Isabelle Le Bescond, mai 2013, SCD Univ. Lille 1, 62 p., p. 35-36.

LES POLITIQUES DES AGENCES DE FINANCEMENT : UN PHENOMENE INTERNATIONAL

NOTA : on évoquera ici la prise en compte « politique », par des organismes de financement de la recherche à l'échelle internationale, de la nécessité d'ouvrir les données de recherche. Les retombées de ces politiques sur les établissements de recherche, ou leur aspect plus ou moins contraignant pour les chercheurs, seront évoquées en II.B.

Si l'Europe peut prendre des décisions normatives à travers ses règlements et ses directives, elle agit également comme agence de financement de la recherche et, à ce titre, peut déterminer les modalités de diffusion des résultats de la recherche qu'elle subventionne – c'est l'esprit des pilotes sur l'*Open Access* et sur l'*Open Research Data*. D'autres organismes de financement de la recherche se sont saisis, ces dernières années, de cette question, et participent de l'impulsion générale donnée à l'ouverture, cette fois-ci à l'échelle globale. Un imposant rapport du JISC, en 2008,⁹³ avait fait la synthèse des initiatives internationales relatives au partage des données de la recherche : il prouve que cette préoccupation n'est pas toute récente et que certains pays s'attellent à la problématique de l'ouverture depuis de nombreuses années.

Les cas américain, britannique et canadien

C'est notamment le cas de l'Australie, des États-Unis et du Royaume-Uni. Les agences américaines se sont très tôt engagées sur la voie du partage des données. Dès 2003, les *National Institutes of Health* (NIH), première agence fédérale en termes de financement dans le domaine biomédical et dont on connaît l'engagement en faveur du libreaccès, ont publié une *Déclaration sur le partage des données de recherche*, qui reste encore d'actualité. Les données doivent être rendues accessibles au plus tard au moment de la publication des résultats basés sur ces données. Les études d'association pangénomique (*Genome-Wide Association Studies*) qu'elles financent font l'objet d'une attention particulière : l'accessibilité des données y est régulée par une politique particulière (2008), et une archive, en partie ouverte, a été mise en place pour les accueillir (*dbGaP*).⁹⁴ Une nouvelle politique, dont l'ébauche a été publiée le 20 septembre 2013, vise à élargir l'accessibilité des données de toutes les recherches en génomique.⁹⁵ En 2010, c'est la *National Science Foundation* (NSF) qui élabore sa politique de partage des données et depuis 2011 les chercheurs qu'elle subventionne doivent rédiger en amont un plan de gestion de

⁹³ JISC, *A Comparative Study of International Approaches to Enabling the Sharing of Research Data*, 2008, 108 p., http://www.jisc.ac.uk/media/documents/programmes/preservation/national_data_sharing_report_final.pdf (consulté le 25 octobre 2013)

⁹⁴ Cf. « NIH Grants Policy Statement », http://grants.nih.gov/grants/policy/nihgps_2012/nihgps_ch8.htm#_Toc271264950 (consulté le 25 octobre 2013); « Policy for Sharing of Data Obtained in NIH Supported or Conducted Genome-Wide Association Studies (GWAS) », <http://grants.nih.gov/grants/guide/notice-files/NOT-OD-07-088.html> (consulté le 25 octobre 2013)

⁹⁵ Cf. <https://www.federalregister.gov/articles/2013/09/20/2013-22941/draft-nih-genomic-data-sharing-policy-request-for-public-comments> (consulté le 25 octobre 2013). La consultation publique sur le contenu de cette *Politique* a été close le 25 novembre 2013.

leurs données, détaillant la manière dont ils entendent se conformer aux exigences de la *NSF Data Sharing Policy*⁹⁶. Ces décisions ne sont pas sans lien avec la création en 2006, par le *National Science and Technology Council* (NSTC) de la Maison-Blanche, de l'« Interagency Working Group on Digital Data » (IWGdd), réunissant des représentants de nombreuses agences fédérales. Le rapport de l'IWGdd, publié en 2009, invitait l'ensemble des agences à se doter de *data policies* et à inciter leurs chercheurs à rédiger des plans de gestion des données.⁹⁷ Le gouvernement fédéral américain a réitéré en février 2013 son souhait d'améliorer l'accessibilité des résultats de la recherche financée sur fonds publics et comprenant « les publications évaluées par les pairs et les données scientifiques numériques ». Dans un mémo adressé aux agences fédérales, l'*Office of Science and Technology Policy* (OSTP), qui conseille le président américain dans le domaine des sciences et des technologies, appelle chacune d'entre elles à mettre sur pied un plan (*Agency Public Access Plan*) qui permette d'optimiser l'impact et la traçabilité (*accountability*) de l'investissement fédéral dans la recherche. Les articles produits sur crédits fédéraux devront être disponibles en libre-accès dans les douze mois suivant leur publication. Concernant les données, le mémo précise que celles issues d'une recherche financée, intégralement ou en partie, par le gouvernement fédéral, devront « être archivées et *rendues accessibles au public* afin d'en permettre la recherche, la récupération et l'analyse »⁹⁸. S'inscrivant dans la lignée des conclusions de l'IWGdd et des initiatives de la NSF, l'OSTP demande que chacun de ces plans « encourage le dépôt des données dans des archives ouvertes (*publicly accessible databases*) » et oblige les chercheurs à systématiquement rédiger des plans de gestion des données. Cette impulsion fédérale est étroitement liée à la proposition de loi présentée en février 2013 devant le 113^{ème} Congrès des Etats-Unis, le *Fair Access to Science and Technology Research* (FASTR) *Act*. Ce texte, version remaniée d'un ancien projet de loi, le *Federal Research Public Access Act* (FRPAA), prévoit le dépôt en libre-accès de l'ensemble des articles de recherche financé sur fonds public, dans les six mois suivant leur publication⁹⁹. La prise en compte de la question des données de la recherche dans le mémo de l'OSTP, concurremment à l'examen du *FASTR Act*, prouve que les débats sur l'*Open Research Data* sont désormais indissociables de ceux sur l'*Open Access*. Les agences fédérales devaient communiquer leurs plans à l'OSTP avant le mois de septembre 2013.

Le Royaume-Uni fait également partie des pays où l'ouverture des données scientifiques fait l'objet depuis plusieurs années d'un vrai volontarisme politique. Les déclarations de l'OCDE de 2004-2007 ont joué un rôle majeur dans ce processus, même si certains organismes anglais de financement de la

⁹⁶ Cf. « Dissemination and Sharing of Research Results », <http://www.nsf.gov/bfa/dias/policy/dmp.jsp> (consulté le 25 octobre 2013) ; DIETRICH Dianne, ADAMUS Trisha, MINER Alison, STEINHART Gail, « De-Mystifying the Data Management Requirements of Research Funders », *Issues in Science and Technology Librarianship*, n° 70 (2012), <http://www.istl.org/12-summer/refereed1.html> (consulté le 16 octobre 2013)

⁹⁷ DIETRICH Dianne & alii, *op. cit.* ; JISC, *op. cit.*, p. 18.

⁹⁸ OFFICE OF SCIENCE AND TECHNOLOGY POLICY, « Increasing Access to the Results of Federally Funded Scientific Research », *Memorandum for the Heads of Executive Departments and Agencies*, 22 février 2013, http://www.whitehouse.gov/sites/default/files/microsites/ostp/ostp_public_access_memo_2013.pdf (consulté le 30 octobre 2013)

⁹⁹ « FASTR introduced in U.S. Congress to drastically expand public access to federally funded research », 14 février 2013, <http://creativecommons.org/weblog/entry/36699> (consulté le 30 octobre 2013)

recherche, comme le *Natural Environment Research Council* (NERC), s'étaient déjà emparés de cette question auparavant. À partir du milieu des années 2000, l'ensemble de ces organismes, *Research Councils* et *Wellcome Trust*, se sont efforcés de définir leur politique de données. La politique définie par le *Biotechnology and Biological Sciences Research Council* (BBSRC) en avril 2007 et révisée en juin 2010, notamment, a fait date¹⁰⁰. Un effort d'harmonisation a ensuite conduit à l'adoption en avril 2011, par l'ensemble des conseils de recherche anglais, de *Principes communs* en matière de politique de données¹⁰¹. Déclinés en sept grands points, ils prennent acte du fait que les données de la recherche, en tant que « bien public produit dans l'intérêt du public », doivent être « ouvertes » (*made openly available*) avec le moins de restrictions possible. Comme le note Sarah Jones, « la gestion des données et leur conservation y sont décrites comme *les moyens essentiels pour parvenir à cette fin* »¹⁰² : aussi, comme aux États-Unis, la plupart des organismes anglais de financement de la recherche conditionnent l'octroi de crédits à la rédaction, en amont, de plans de gestion des données. L'une des spécificités du cas britannique est que ces impulsions politiques des agences de financement de la recherche ont été accompagnées d'investissements substantiels dédiés à la mise en place des infrastructures et des services nécessaires à la gestion des données produites par les chercheurs. Ces investissements ont été portés par le JISC (*Joint Information Systems Committee*), organisme public financé par les *Research Councils* et chargé de promouvoir auprès des universités britanniques l'utilisation des technologies de l'information et de la communication. S'appuyant en particulier sur son *Digital Curation Center* (DCC)¹⁰³, fondé en 2004, le JISC pilote depuis 2009 le programme de financement « Managing Research Data ». Entré dans sa seconde phase en 2011, il a notamment permis de soutenir, nous y reviendrons, 17 projets d'infrastructures et de services autour de la gestion des données de la recherche à l'échelle des universités britanniques¹⁰⁴. Le rapport de la *Royal Society* de Londres, *Science as an open enterprise*, a récemment préconisé que ce programme soit prolongé et étendu au-delà des 17 premières institutions-pilotes durant les cinq prochaines années, afin de favoriser l'adoption à l'échelle nationale de politiques institutionnelles de gestion des données¹⁰⁵.

Si les États-Unis et le Royaume-Uni, à travers l'engagement de leurs agences de financement de la recherche, ont incontestablement plusieurs longueurs d'avance, d'autres pays se saisissent de la question. On ne retiendra

¹⁰⁰ BBSRC *Data Sharing Policy*, juin 2010, <http://www.bbsrc.ac.uk/web/FILES/Policies/data-sharing-policy.pdf> (consulté le 12 novembre 2013)

¹⁰¹ RCUK *Common Principles on Data Policy*, <http://www.rcuk.ac.uk/research/Pages/DataPolicy.aspx> (consulté le 30 octobre 2013)

¹⁰² JONES Sarah, « Developments in Research Funder Data Policy », *The International Journal of Digital Curation*, vol. 7, n° 1, 2012, p. 114-125, p. 118 <http://ijdc.net/index.php/ijdc/article/view/209/278> (consulté le 12 novembre 2013)

¹⁰³ <http://www.dcc.ac.uk/>

¹⁰⁴ Pour des informations plus détaillées sur la seconde phase (2011-2013) du programme MRD du JISC, cf. http://www.jisc.ac.uk/whatwedo/programmes/di_researchmanagement/managingresearchdata.aspx (consulté le 13 novembre 2013).

¹⁰⁵ THE ROYAL SOCIETY, *Science as an open enterprise : summary report*, The Royal Society, juin 2012, p. 73 http://royalsociety.org/uploadedFiles/Royal_Society_Content/policy/projects/sape/2012-06-20-SAOE.pdf (consulté le 27 septembre 2013)

que l'exemple du Canada, dont les Instituts de recherche en santé (IRSC), qui financent la recherche médicale canadienne, ont élaboré des politiques de libre accès depuis de nombreuses années,¹⁰⁶ mais cherchent désormais à étendre le périmètre de leurs exigences relativement aux données de la recherche. Si cette politique reste pour l'instant plutôt frileuse et recoupe ce que réclament déjà la plupart des revues médicales, à savoir le dépôt dans une archive ouverte disciplinaire de référence¹⁰⁷ des « données de recherche biomédicale relatives à une publication », « les IRSC sont déterminés à améliorer l'accès aux résultats de la recherche et vont examiner la possibilité d'étendre la politique afin d'y inclure éventuellement le matériel de recherche et d'autres données de recherche ». Du reste, depuis le 1^{er} janvier 2013, les chercheurs « titulaires d'une subvention [doivent] conserver les ensembles de données originaux découlant d'une recherche financée par les IRSC pour une période minimale de cinq ans suivant la fin de la période de subvention. Cela s'applique à toutes les données, publiées ou non. »¹⁰⁸ Cet élargissement du spectre des données concernées, combiné à une obligation (temporaire) de conservation de la totalité des données produites, constitue sans doute une première étape avant l'adoption de politiques plus exigeantes en matière de gestion et de partage des données, inspirées des initiatives américaines et britanniques.

Du chercheur à l'institution de recherche, vers une responsabilité partagée

L'ensemble de ces initiatives internationales témoignent en tout cas de l'impulsion décisive que donnent les organismes de financement de la recherche aux processus d'ouverture des données, selon des modalités et à des rythmes divers. Cette logique descendante (*top-down*)¹⁰⁹ suivant laquelle l'ouverture est la résultante finale des impulsions « venues du haut », responsabilise avant tout les chercheurs, si bien que l'*Open research Data* pourrait n'être perçu que comme une affaire ne regardant que les financeurs et les scientifiques. Citant le cas du Royaume-Uni, Sarah Jones a ainsi écrit en 2012 que « c'est aux chercheurs que les politiques des agences de financement confient la responsabilité de prendre des dispositions pour la gestion de leurs données de recherche, en rédigeant des plans de gestion de leurs données ». Pourtant, l'une de ces agences, en mai 2011, a privilégié une autre approche, en mettant l'accent non plus seulement sur les responsabilités des chercheurs, récipiendaires des subventions, mais sur celles des organismes de recherche auxquels ces chercheurs sont rattachés. L'EPSRC, l'agence qui finance la recherche britannique dans le domaine de l'ingénierie et des sciences

¹⁰⁶ COUTURE Marc, DUBE Marcel & MALISSARD Pierrick, *Propriété intellectuelle et université. Entre la libre circulation des idées et la privatisation des savoirs*, 2010, Presses de l'université du Québec, 384 p., p. 198.

¹⁰⁷ La « Politique de libre accès des IRSC » canadiens propose une typologie des données produites dans le cadre de recherches biomédicales, et indique l'archive ouverte la plus à même d'accueillir chaque type de données. Les chercheurs sont invités à déposer les données les plus générales dans DRYAD. Cf. <http://www.cihr-irsc.gc.ca/f/46068.html#5.1.2> (consulté le 13 novembre 2013)

¹⁰⁸ *Idem*, cf. 5.1., « 5.1.2. Données de recherche relatives à une publication ».

¹⁰⁹ J'emprunte la notion à HODSON Simon, « Les politiques d'accompagnement des données : une discussion du paysage », *Journées FRÉDOC 2013, Problématiques et enjeux autour des données de la recherche, 7-10 octobre 2013, Aussois*, http://renatis.cnrs.fr/IMG/pdf/Hodson_FRéDOC_Presentation_FR.pdf (consulté le 14 novembre 2013)

physiques¹¹⁰, demande ainsi à chacun des organismes de recherche bénéficiant de ses financements de définir une politique et des procédures de gestion et de partage des données, en conformité avec ses propres attentes¹¹¹ : aussi l'EPSRC est-il le seul *Research Council* à ne pas soumettre les chercheurs à l'obligation de rédiger des plans de gestion lors de leur demande de subvention, confiant à l'organisme de recherche le soin de définir sa politique en la matière et de la faire entièrement appliquer – au plus tard au 1^{er} mai 2015.¹¹² Même si certaines universités britanniques, comme celle d'Edimbourg, avaient déjà défini leur politique institutionnelle sur la gestion des données en amont, le volontarisme de l'EPSRC a incité les établissements universitaires à agir.¹¹³

Posé ce contexte, il convient maintenant, dans la partie qui suit, d'étudier en détail *la manière dont on peut rendre, au-delà des positions de principe, les données accessibles et réutilisables* : en un mot, *ouvertes*. Nous reviendrons d'abord, en nous positionnant du côté des chercheurs, sur les différents acteurs qui participent de ce processus et sur les obligations faites, selon les cas (éditeurs, organismes de recherche, communautés scientifiques), aux producteurs de données (du simple dépôt d'un jeu de données dans une archive ouverte à la préparation de la gestion de ce jeu de données tout au long de son cycle de vie ; de la diffusion des données sur lesquelles repose une publication à celle de l'ensemble des données produites au cours du processus de recherche). Nous envisagerons ensuite les conditions juridiques et techniques qui permettent la réutilisation des données de la recherche.

¹¹⁰ Pour un aperçu de l'importance de l'EPSRC dans le financement de la recherche britannique, cf. JONES Sarah, *op. cit.*, p. 120-121.

¹¹¹ <http://www.epsrc.ac.uk/about/standards/researchdata/Pages/expectations.aspx> (consulté le 14 novembre 2013)

¹¹² <http://www.dcc.ac.uk/resources/policy-and-legal/research-funding-policies/epsrc> (consulté le 14 novembre 2013)

¹¹³ Cf. *infra* p. 57 et suivantes.

OUVRIR LES DONNEES : JEU D'ACTEURS, SOLUTIONS JURIDIQUES ET TECHNIQUES

LES CHERCHEURS FACE A LEURS DONNEES : IMPULSIONS EDITORIALES, INSTITUTIONNELLES ET DISCIPLINAIRES

Incitations éditoriales

Malgré l'implication récente, à l'échelle internationale, des organismes de financement de la recherche, ce ne sont pas les seuls acteurs à l'œuvre dans les processus d'ouverture, les éditeurs de revues scientifiques ayant également mis en place des mécanismes contraignants pour la diffusion des données de recherche. Ils restent cependant limités, et parfois insatisfaisants. Comme nous l'avons vu dans notre essai de définition des *research data*, et comme on le voit dans la *Data Publication Pyramid* tirée du rapport du projet européen ODE (*Opportunities for Data Exchange*) sur l'intégration des données de recherche dans les publications,¹¹⁴ les éditeurs contribuent à leur diffusion (outre celles qui, *de facto*, sont intégrées aux articles) en établissant parfois des règles sur l'accessibilité des données de recherche sur lesquelles reposent les conclusions des articles qu'ils publient. Il est important de noter que ces politiques ne concernent qu'une certaine catégorie de données, celles qui servent de preuve et étaient la démonstration scientifique. L'ensemble des données produites au cours du processus de la recherche, celles qui constituent la base de la *Data Pyramid*, ne sont pas concernées.

Beaucoup d'éditeurs et de revues se sont dotés de politiques de données. BioMed Central, notamment, indique que « la proposition d'un manuscrit à [l'une de ses revues] implique que les données facilement reproductibles décrites dans ce manuscrit, *incluant toutes les données brutes pertinentes*, soient rendues librement accessibles à n'importe quel scientifique souhaitant les utiliser à des fins non commerciales »¹¹⁵. Certaines revues BioMed Central arrêtent donc progressivement, depuis 2011, leur politique d'accessibilité des données justificatives, même si dans la plupart des cas les auteurs ne sont pas contraints de déposer ces données dans une archive ouverte et sont seulement encouragés à le faire. Cet engagement, encore timide, en faveur de l'accessibilité des données, s'est accompagné d'une consultation sur l'*Open data* lancée en 2012 à l'issue de laquelle, en septembre 2013, l'éditeur a décidé que les données *intégrées* aux articles (ou aux *supplementary files*) seraient

¹¹⁴ REILLY Susan & alii, *Report on Integration of Data and Publications*, ODE [Opportunities for Data Exchange] project, octobre 2011, 87 p., p. 36. Cf. annexe I, p. 94.

<http://www.libereurope.eu/sites/default/files/ODE-ReportOnIntegrationOfDataAndPublication.pdf> (consulté le 7 octobre 2013)

¹¹⁵ « Availability of supporting data », <http://www.biomedcentral.com/about/supportingdata> (consulté le 18 novembre 2013)

placées par défaut sous licence *Creative Commons* CC0, les plaçant donc dans le domaine public.¹¹⁶

Cette initiative éditoriale n'est pas isolée : un article publié en 2011 sur « La disponibilité publique des données de recherche publiées dans les revues à facteur élevé » avait démontré la diversité, mais aussi l'hétérogénéité des politiques d'éditeur concernant l'accessibilité des données.¹¹⁷ Surtout, ses auteurs avaient conclu que même lorsqu'une revue fait de la disponibilité des données justificatives une condition de la publication (comme les revues du groupe *Nature*), cette obligation n'est pas forcément respectée, du fait des réticences des chercheurs à les libérer. Beaucoup d'éditeurs, du reste, n'établissent de politiques réellement contraignantes que pour certains types de données, celles pour lesquelles des archives ouvertes de référence existent. Ainsi, *Nature* n'a défini de mandat de dépôt (dans des archives de données disciplinaires) que pour quelques types de données très standardisés : séquences protéiques, séquences ADN, structures macromoléculaires, données cristallographiques, etc. Pour les autres jeux de données, l'éditeur ne fait que « fortement recommander » leur dépôt dans une archive ouverte.¹¹⁸

Quand bien même : « on demande rarement aux pairs (*referees*) d'évaluer si les données adéquates sont diffusées, et on ne leur donne accès à ces données au cours de la procédure d'évaluation qu'encore plus rarement »¹¹⁹. Enfin, alors que la diffusion des données justificatives répond à des impératifs de transparence et de reproductibilité, « même lorsque ces données sont rendues publiquement accessibles par leurs producteurs, il est souvent impossible pour d'autres chercheurs de reproduire les résultats publiés « du fait d'une mauvaise annotation ou description de la manière dont ont été traitées et analysées les données ».¹²⁰ Les directives fournies sur ce point par les éditeurs sont très variables et même lorsque des standards de description des données existent et que les revues demandent aux auteurs de les suivre, la vérification des résultats présentés dans les conclusions d'un article n'est pas garantie, comme l'a montré un article de 2008 sur la difficulté de reproduire les analyses des données d'expression issues des puces à ADN publiées dans *Nature Genetics*, faute de métadonnées solides.¹²¹ Les lignes bougent, cependant, et le groupe *Nature* a par

¹¹⁶ HRYNASZKIEWICZ Iain, BUSCH Stefan & COCKERILL Matthew J., « Licensing the future : report on BioMed Central's public consultation on open data in peer-reviewed journals », *BMC Research Notes*, août 2013, <http://www.biomedcentral.com/content/pdf/1756-0500-6-318.pdf> (consulté le 19 novembre 2013)

¹¹⁷ ALSHEIKH-ALI Alawi A., QURESHI Waqas, AL-MALLAH Mouaz H. & IOANNIDIS John P. A., « Public Availability of Published Research Data in High-Impact Journals », *PLoS ONE*, 2011, 6 (9), <http://www.plosone.org/article/info%3Adoi%2F10.1371%2Fjournal.pone.0024357> (consulté le 15 octobre 2013)

¹¹⁸ « Availability of data and materials », <http://www.nature.com/authors/policies/availability.html> (consulté le 19 novembre 2013) ; INSTITUT NATIONAL DE LA RECHERCHE AGRONOMIQUE [INRA], *Rapport du groupe de travail sur la gestion et le partage des données*, Paris, juin 2012, 62 p., p. 19, http://www.pfl-cepia.inra.fr/uploads/gdp_docs/Rapport-GestionDonnees-web.pdf (consulté le 7 octobre 2013)

¹¹⁹ POISOT Timothée, MOUNCE Ross & GRAVEL Dominique, « Moving toward a sustainable ecological science : don't let data go to waste ! », nov. 2012, 14 p., p. 8, http://figshare.com/articles/Moving_toward_a_sustainable_ecological_science_don_t_let_data_go_to_waste_/693745

¹²⁰ ALSHEIKH-ALI Alawi A., QURESHI Waqas, AL-MALLAH Mouaz H. & IOANNIDIS John P. A., *op. cit.*, <http://www.plosone.org/article/info%3Adoi%2F10.1371%2Fjournal.pone.0024357> (consulté le 15 octobre 2013)

¹²¹ En anglais, « microarray gene expression analyses ». Cf. IOANNIDIS John P., ALLISON David B., BALL Catherine A., COULIBALY Issa, CUI Xiangqin *et al.*, « Repeatability of published microarray gene expression analyses », *Nature Genetics*, n° 41, 2009, p. 149-155.

exemple adopté en mai 2013 des mesures éditoriales pour faciliter l'interprétation et améliorer la fiabilité des résultats publiés : les chercheurs proposant des articles dans le domaine des sciences doivent désormais s'assurer, par l'intermédiaire d'une *checklist*,¹²² qu'ils donnent suffisamment de détails méthodologiques sur leurs analyses et qu'ils précisent les modalités d'accès à leurs données.¹²³

Les impulsions éditoriales sont donc bien réelles, de plus en plus nombreuses, mais comme a permis de le montrer le projet JorD (*Journal Research Data Policy Bank*) du JISC, les politiques de données sont souvent insuffisantes, peu claires et pas assez précises sur les archives où déposer les données, les standards à respecter, etc.¹²⁴ Le problème essentiel, soulevé par Jane H. Smith lors de la conférence LIBER 2013 de Munich, est que « les chercheurs ne savent tout simplement pas comment partager leurs données »¹²⁵ et que les éditeurs, au-delà de l'impératif de diffusion et d'ouverture qu'ils s'efforcent de promouvoir, ne prennent pas part à la gestion (standardisation, archivage, etc.) des données des chercheurs qu'ils publient¹²⁶. Certes, « on pourrait considérer comme infaisable que les revues soient directement impliquées dans la conservation (*storage*) et la publication des données, étant donné les coûts induits et leur manque d'expertise dans la gestion des données. Pourtant, on ne comprend pas pourquoi les éditeurs ne devraient pas être responsables de la maintenance de bases de données, dans lesquelles les données essentielles utilisées comme preuves dans les articles publiés seraient archivées et rendues accessibles »¹²⁷.

Une offre éditoriale pour le partage des données de recherche est entrain d'émerger, incarnée par les *data journals* ou par des expérimentations, comme l'« Article du Futur » d'Elsevier ou la plate-forme *Scientific Data*, que devrait lancer, au printemps 2014, le groupe *Nature*¹²⁸. Toutefois, il s'agit là de tentatives de création d'un *publisher-driven model for data sharing*,¹²⁹ plus que d'un effort d'amélioration et de rationalisation des modalités de diffusion des

¹²² *Nature Reporting Checklist For Life Sciences Articles*, <http://www.nature.com/authors/policies/checklist.pdf> (consulté le 20 novembre 2013)

¹²³ NATURE PUBLISHING GROUP, « Raising standards », *Nature Immunology*, n° 14, 2013, p. 415, http://www.nature.com/ni/journal/v14/n5/full/ni.2603.html?WT.ec_id=NI-201305 (consulté le 20 novembre 2013)

¹²⁴ Sur ce projet de 2012, cf. <http://jordproject.wordpress.com> (consulté le 20 novembre 2013) ; cf. HODSON Simon, *op. cit.*, 2013, http://renatis.cnrs.fr/IMG/pdf/Hodson_FReDOC_Presentation_FR.pdf (consulté le 14 novembre 2013)

¹²⁵ SMITH Jane H., HUSSAIN Azhar, BAMKIN Marianne & STURGES Paul, « Journals and their policies on data sharing : lessons from the JoRD Project », *LIBER 42nd Annual Conference, 26-29 June 2013, Munich*, <http://www.liber2013.de/index.php?id=68#c181> (consulté le 20 novembre 2013)

¹²⁶ RECODE, *Deliverable D1 : Stakeholder Values and Ecosystems*, septembre 2013, 101 p., p. 52, http://recodeproject.eu/wp-content/uploads/2013/10/RECODE_D1-Stakeholder-values-and-ecosystems_Sept2013.pdf (consulté le 23 octobre 2013)

¹²⁷ LEONELLI Sabina, SMIRNOFF Nicholas, MOORE Jonathan, COOK Charis & BASTOW Ruth, « Making open data work for plant scientists », *Journal of Experimental Botany*, septembre 2013, p. 4113-4114, <http://jxb.oxfordjournals.org/content/early/2013/09/15/jxb.ert273.full> (consulté le 20 novembre 2013)

¹²⁸ GEROUDET Marie-Madeleine, *Étude prospective sur les données de la recherche*, rapport sous la dir. d'Isabelle Le Bescond, mai 2013, SCD Univ. Lille 1, 62 p., p. 52. Cf. le communiqué de presse du groupe *Nature*, « NPG to launch *Scientific Data* to help scientists publish and reuse research data », 4 avril 2013, http://www.nature.com/press_releases/scientificdata.html (consulté le 20 novembre 2013)

¹²⁹ THESSEN Anne E., PATTERSON David J., « Data Issues in the Life Sciences », *ZooKeys*, novembre 2011, n° 150, p. 15-51, p. 22, <http://www.ncbi.nlm.nih.gov/pmc/articles/PMC3234430/> (consulté le 28 octobre 2013)

données justificatives associées aux articles publiés. Pour cela, les éditeurs se reposent encore très largement sur les chercheurs et les archives de données thématiques (*community-endorsed public repositories*). Dans un premier temps, dans tous les cas, les éditeurs pourraient accentuer leur rôle dans la diffusion des données « en s’assurant que les mandats de dépôt des données dans des archives ouvertes soient appliqués et pas seulement “rhétoriques”, soit par l’intermédiaire d’un contrôle éditorial accru, soit en demandant aux pairs d’évaluer aussi les plans de partage des données. »¹³⁰

Les mandats institutionnels : l’importance de la gestion des données et des DMP

Ainsi que le note Odile Hologne, directrice de l’IST à l’Institut national de la recherche agronomique (INRA), l’ouverture ne saurait constituer un point de départ : au contraire, elle ne doit être envisagée que comme le résultat nécessaire d’une bonne politique de gestion des données, qui en constitue le préalable indispensable. « La planification de [cette] gestion fournit un mécanisme pour définir les activités de gestion clés qui sont nécessaires pour assurer l’intégrité [...] des jeux de données générés au cours du processus de recherche »¹³¹ et donc leur réutilisation : c’est la raison pour laquelle les politiques institutionnelles des organismes de financement et des universités placent les plans de gestion des données au centre de leurs exigences.

Comme nous l’avons vu, la rédaction de DMP (*data management plans*)¹³² par les chercheurs fait partie des conditions imposées par beaucoup d’agences de financement de la recherche anglo-saxonnes pour toute demande de subvention : la question du partage et de la diffusion des données n’est qu’une composante parmi d’autres d’un plan de gestion (et l’importance de cette section est très variable selon les plans), mais il concerne l’ensemble des données produites au cours du projet de recherche. Un DMP sous sa forme classique, tel que décrit dans le *Guide de demande de subventions* de la NSF, doit rassembler des informations sur les différents types de données qui seront produits au cours du processus de recherche, et sur leur cycle de vie (création, curation, diffusion, conservation).¹³³ Le plan de gestion constitue le pendant, concret, d’une politique de données, puisqu’il est censé expliciter la manière dont le chercheur entend se conformer à la *Data policy* de son agence de financement ou de son université et garantir, à terme, la diffusion des données produites. Il est donc un outil indispensable à l’ouverture.

¹³⁰ POISOT Timothée, MOUNCE Ross & GRAVEL Dominique, *op. cit.*, 2012, p. 8

¹³¹ THORLEY Mark, *NERC Data Policy – Guidance Notes*, 10 p., p. 5, <http://www.nerc.ac.uk/research/sites/data/documents/datapolicy-guidance.pdf> (consulté le 21 novembre 2013)

¹³² Pour faciliter la lecture, nous utiliserons ce sigle (*Data Management Plans*) pour désigner « les plans de gestion des données de la recherche ».

¹³³ *NSF Grant Proposal Guide*, http://www.nsf.gov/pubs/policydocs/pappguide/nsf13001/gpg_2.jsp#dmp (consulté le 20 novembre 2013) ; cf. « Scientists Seeking NSF Funding Will Soon Be Required to Submit Data Management Plans », http://www.nsf.gov/news/news_summ.jsp?cntn_id=116928 (consulté le 21 novembre 2013)

La politique de données du *Natural Environment Research Council* (NERC) britannique, par exemple, révisée en janvier 2011, contient désormais une obligation formelle (effective depuis le 1^{er} août 2012) faite à tout demandeur de subvention d'inclure une esquisse de DMP à sa candidature. Si elle est validée, un plan de gestion complet doit être rédigé et présenté par le responsable du projet (PI : *principal investigator*), dans les trois mois qui suivent la mise à la disposition des fonds, à l'un des centres de données du NERC pour validation.¹³⁴ Il doit clairement préciser la répartition des rôles, au sein de l'équipe de recherche, en matière de collecte et d'enrichissement des données ; le format des données et les standards utilisés pour les métadonnées ; les modalités d'archivage, de diffusion et de conservation des données produites ou collectées au cours du projet de recherche.¹³⁵ En somme : poser les bases du futur partage en définissant quelles données seront diffusées, quand et comment. La prise en compte dès le début du projet de la question des métadonnées est également essentielle, puisqu'elle constitue la principale condition de la compréhension et de la réutilisation future des données¹³⁶. Comme le montre le cas du NERC, les plans de gestion des données ont une importance décisive sur au moins trois points : ils établissent contractuellement une obligation de dépôt (à des fins de conservation et de diffusion), définissent le cycle de vie des données collectées (en statuant sur leur « valeur », et donc sur la nécessité, ou non, d'une conservation pérenne), et jouent un rôle essentiel dans la chronologie de l'ouverture en précisant dans quels délais les données pourront être diffusées. En ce qui concerne les données collectées sur fonds du NERC, elles bénéficient de l'importante infrastructure mise à la disposition des chercheurs par ce *Research Council*. Ils doivent en effet déposer leurs données dans l'un des sept centres de données (*data centres*) que le NERC finance, comme le *British Oceanographic Data Centre* (BODC) ou le *Polar Data Center* (PDC), et chaque plan de gestion des données doit intégrer cette nécessité¹³⁷. Ils n'ont alors plus à se soucier de l'accessibilité des données, puisque le NERC gère également un *Data Catalogue Service* (DCS), disponible en ligne, rassemblant l'ensemble des métadonnées associées à des jeux de données conservés par le NERC¹³⁸. Seules les données qui auront été identifiées par l'équipe de recherche, en concertation avec le centre de données, comme ayant de la valeur à long terme, seront archivées de manière pérenne : afin d'aider les chercheurs à se positionner sur cette notion de « valeur », le NERC a produit une « Data Value Checklist » qui leur permet de statuer sur les jeux de données concernés par l'impératif de conservation à long terme.¹³⁹ Parmi ces critères, le fait que la donnée collectée est irremplaçable (données d'observation), qu'elle a un fort potentiel de réutilisation, que son coût de génération est important, qu'elle a des chances

¹³⁴ NERC Data Policy, <http://www.nerc.ac.uk/research/sites/data/policy.asp?cookieConsent=A> (consulté le 21 novembre 2013).

¹³⁵ NERC Data Management Plan Template, <http://www.nerc.ac.uk/research/sites/data/documents/dmp-template.pdf> (consulté le 21 novembre 2013)

¹³⁶ Cf. infra p. 69-72.

¹³⁷ THORLEY Mark, *NERC Data Policy – Guidance Notes*, p. 5 & p. 9.

¹³⁸ THORLEY Mark, « Data in the research process : a funder [NERC]'s perspective », *OpenAIREplus workshop "Linking Open Access publications to data – policy development and implementation"*, 11 juin 2012, http://www.openaire.eu/fr/about-openaire/publications-presentations/presentations/doc_details/393-data-in-the-research-process-a-funders-perspective--mark-thorley-national-environment-research-c (consulté le 21 novembre 2013). Cf. le *Data Catalogue Service* : <http://data-search.nerc.ac.uk/> (consulté le 21 novembre 2013).

¹³⁹ NERC Data Value Checklist, <http://www.nerc.ac.uk/research/sites/data/documents/data-value-checklist.pdf> (consulté le 21 novembre 2013)

d'être citée dans une publication ou qu'elle est tout simplement « brute », constituent des arguments de poids en faveur de sa conservation¹⁴⁰.

Le plan a également pour objet de déterminer l'étendue de la période d'exclusivité accordée à l'équipe de recherche pour l'exploitation de ses données avant leur diffusion. Cette possibilité de placer temporairement les données sous embargo, reconnue quasi-unaniment par les organismes de financement, permet de concilier les intérêts scientifiques ou industriels des chercheurs et l'ouverture de leurs données. En effet, « beaucoup de chercheurs souhaitent conserver l'usage exclusif de leurs données jusqu'à qu'ils aient pu les exploiter un maximum pour leurs publications ».¹⁴¹ Les *Research Councils* britanniques, et beaucoup d'autres, reconnaissent ce droit à l'embargo et les intègrent à leurs mandats :

« Pour s'assurer que les équipes de recherche bénéficient de la reconnaissance appropriée pour l'effort qu'implique la collecte et l'analyse de données, ceux qui entreprennent un travail financé par un conseil de recherche peuvent se voir accorder une période limitée d'usage privilégié des données qu'ils ont collectés afin de leur permettre de publier les résultats de leur recherche. »¹⁴²

Cette période d'embargo peut varier selon les disciplines et les politiques de données des agences de financement. Le NERC, même s'il encourage les chercheurs à ne pas soumettre leurs données à un embargo, les autorise à le faire pendant 2 ans « à compter de la collecte des données, [...] c'est-à-dire le moment où elles sont obtenues depuis un instrument, une expérience ou une analyse », et non à compter de la clôture du projet financé par l'agence. Cette durée, cependant, est négociable.¹⁴³ En Autriche, l'*Austrian Science Fund* (FWF) a également arrêté à 2 ans la période d'embargo possible avant l'ouverture des données de ses chercheurs, mais cette fois-ci « à compter de la clôture du projet ».¹⁴⁴ À l'échelle du DMP, la chronologie de la diffusion des données peut être déterminée selon le type de données. L'*Australian Antarctic Division*, l'agence qui administre les stations australiennes en Antarctique et pilote la recherche sur ces territoires, à travers notamment l'*Australian Antarctic program* (AAP), place la rédaction d'un DMP au centre des obligations faites à ses chercheurs et décline dans sa politique de données les périodes d'embargo dont ces dernières, selon leurs caractéristiques, peuvent bénéficier.¹⁴⁵

¹⁴⁰ Pour un aperçu plus large de ces critères, se reporter à la *checklist*, mais également à UNIVERSITY OF BRISTOL RESEARCH DATA SERVICE, *Data Management Planning. NERC funding applicants*, 1^{er} août 2013, 8 p., p. 5-6, <http://data.bris.ac.uk/files/2013/09/data.bris-NEC-guide-v31.pdf> (consulté le 21 novembre 2013)

¹⁴¹ GRIFFITHS Aaron, « The Publication of Research Data : Researcher Attitudes and Behaviour », *The International Journal of Data Curation*, n°1, vol. 4, 2009, p. 46-56, <http://www.ijdc.net/index.php/ijdc/article/view/101/76> (consulté le 7 octobre 2013)

¹⁴² RCUK *Common Principles on Data Policy*, <http://www.rcuk.ac.uk/research/Pages/DataPolicy.aspx> (consulté le 30 octobre 2013)

¹⁴³ THORLEY Mark, *NERC Data Policy – Guidance Notes*, p. 4 & 6.

¹⁴⁴ « Open Access Policy for FWF-funded projects », http://www.fwf.ac.at/en/public_relations/oai/ (consulté le 22 novembre 2013)

¹⁴⁵ « The Australian Antarctic program Data Policy, v 2.0 », septembre 2013, https://data.aad.gov.au/aadc/about/data_policy.cfm#WhoOwnsData (consulté le 22 novembre 2013)

Le plan de gestion des données, dont on peut avoir un aperçu idéal dans la « Checklist for a Data Management Plan » élaborée par le *Digital Curation Center* et révisée en 2013,¹⁴⁶ est donc devenu un levier indispensable des politiques de données et un incontournable des mandats institutionnels. Les universités britanniques, dans le sillage des demandes de l'EPSRC déjà évoquées,¹⁴⁷ se dotent progressivement de politiques contraignantes dans lesquelles la gestion des données, qui s'incarne d'abord dans la rédaction d'un DMP, est perçue comme une priorité, condition préalable de l'ouverture. Avec le soutien du JISC, plusieurs projets, comme *DaMaRO* (« Data Management Rollout », « Déploiement de la gestion des données ») à Oxford,¹⁴⁸ ou *Open Exeter* dans l'université du même nom, ont permis à ces institutions de développer des mandats qui définissent très clairement les responsabilités qui incombent aux chercheurs pour la gestion et le partage de leurs données.¹⁴⁹ La rédaction d'un plan de gestion en fait systématiquement partie. C'est le cas à l'université d'Exeter où, depuis le 1^{er} octobre 2013, les chercheurs doivent se conformer à cette règle :

« [...] 4. Tout projet de recherche doit inclure un plan de gestion des données de recherche qui aborde les questions de collecte, de gestion, d'intégrité, de confidentialité, de sécurité, de sélection, d'archivage et d'élimination, de commercialisation, de coût, du partage et de la publication des données de recherche, ainsi que la production des métadonnées descriptives permettant leur visibilité et leur réutilisation [...]. »¹⁵⁰

Ces mandats, qui se résument donc en une obligation de gestion et en une obligation de dépôt, qu'ils émanent d'agences de financement ou d'institutions de recherche, réclament des infrastructures et des services, que nous aborderons par la suite. Ils participent d'une logique descendante selon laquelle des politiques contraignantes obligent les chercheurs à agir, alors que pour beaucoup la gestion des données produites ou collectées ne fait pas partie de leur pratique scientifique.¹⁵¹ Lors de la journée d'étude de l'ADBU du 19 septembre 2013, au Havre, Alex Ball, personnel détaché du DCC à l'université de Bath, a ainsi relevé qu'à l'issue de l'audit qui avait été réalisé sur place sur les pratiques de gestion, par les chercheurs, de leurs données, 81 % d'entre eux ne rédigeaient pas de DMP, entraînant souvent des pertes de données irrémédiables... Pourtant, à ces impulsions institutionnelles répondent parfois, de la part de certaines communautés de chercheurs, une vraie prise de

¹⁴⁶ DCC, *Checklist for a Data Management Plan, v 4.0*, 2013, http://www.dcc.ac.uk/sites/default/files/documents/resource/DMP_Checklist_2013.pdf (consulté le 21 novembre 2013). Voir un extrait en annexe II, p. 95.

¹⁴⁷ Cf. supra p. 32-33.

¹⁴⁸ Cf. <http://blogs.it.ox.ac.uk/damaro/> (consulté le 22 novembre 2013)

¹⁴⁹ Cf. <http://www.dcc.ac.uk/resources/policy-and-legal/institutional-data-policies/uk-institutional-data-policies> (consulté le 22 novembre 2013).

¹⁵⁰ *Exeter University Open Access Research and Research Data Management Policy*, avril 2013, https://ore.exeter.ac.uk/repository/bitstream/handle/10036/4280/OA_RDM_Policy_Final.pdf?sequence=4 (consulté le 22 novembre 2013)

¹⁵¹ GRIFFITHS Aaron, « The Publication of Research Data : Researcher Attitudes and Behaviour », *The International Journal of Data Curation*, n°1, vol. 4, 2009, p. 46-56, p. 51 <http://www.ijdc.net/index.php/ijdc/article/view/101/76> (consulté le 7 octobre 2013)

conscience de la nécessité d'une bonne gestion des données et de l'importance de les mettre à la disposition de tous.

Logique *bottom-up* et culture du partage

Ce panorama des différents acteurs qui définissent des politiques de données et influent sur leur ouverture ne saurait être complet si nous n'évoquons les impulsions qui émanent des communautés scientifiques elles-mêmes. Certaines disciplines reposent déjà sur une vraie « culture du partage » qui, au-delà des mandats institutionnels, contribuent à la diffusion de bonnes pratiques, à la définition d'outils techniques (standards de métadonnées, archive ouverte de données disciplinaire) et, partant, à l'ouverture des données. Cette logique ascendante (*bottom-up*) est bien sûr complémentaire des incitations institutionnelles évoquées ci-dessus, comme le notait en 2008 le rapport du *Research Information Network*, « Partager ou ne pas partager » :

*« Beaucoup d'agences de financement mettent des politiques en place afin d'assurer que les jeux de données potentiellement utiles à d'autres soit gérés de manière à permettre leur visibilité, leur accessibilité et leur réutilisation. Mais les exigences des agences de financement ne s'accordent pas parfaitement aux normes culturelles en vigueur dans certaines disciplines de recherche. Quelques disciplines ont largement devancé les agences de financement en ayant, depuis longtemps, une culture du partage, et en développant les infrastructures et les méthodes pour qu'elle s'épanouisse. Dans d'autres disciplines, le partage des données n'est pas la norme et dans ce cas, en effet, les politiques des financeurs peuvent influencer de manière significative sur les attitudes et le comportement des chercheurs. »*¹⁵²

La génomique, la cristallographie ou l'astronomie comptent parmi ces disciplines qui ont déjà intégré l'impératif d'ouverture. En astronomie, le partage des données fait consensus¹⁵³ et les données produites dans les observatoires sont généralement rendues publiques après un an d'embargo¹⁵⁴ : c'est par exemple le cas, en France, des données collectées par le satellite CoRoT du CNES (qui a cessé d'émettre en novembre 2012) qui étaient mises en libreaccès dans une archive dédiée après un an d'usage exclusif par les chercheurs impliqués dans le projet.¹⁵⁵ Par ailleurs, la communauté internationale des astrophysiciens, unifiée à travers l'*International Virtual Observatory Alliance* (IVOA),¹⁵⁶ développe depuis des années des standards techniques permettant l'interopérabilité des données collectées à l'échelle

¹⁵² RESEARCH INFORMATION NETWORK (RIN), *To Share or not to Share : Publication and Quality Assurance of Research Data Outputs*, juin 2008, 56 p., <http://www.rin.ac.uk/system/files/attachments/To-share-data-outputs-report.pdf> (consulté le 22 novembre 2013).

¹⁵³ RESEARCH INFORMATION NETWORK (RIN), *op. cit.*, p. 32 ;

¹⁵⁴ THE ROYAL SOCIETY, *Science as an open enterprise : summary report*, The Royal Society, juin 2012, 104 p., p. 86 ; LIEFKE Carolin, « The Astronomical Importance of Discoverability », in ODE, *Ten Tales of Drivers & Barriers in Data Sharing*, 25 p., p. 19 ;

¹⁵⁵ <http://idoc-corot.ias.u-psud.fr/index.jsp> (consulté le 22 novembre 2013)

¹⁵⁶ Initiative citée en exemple dans le récent rapport de THE ROYAL SOCIETY, *op. cit.*, p. 86 ; cf. aussi <http://www.ivoa.net/> (consulté le 22 novembre 2013)

globale.¹⁵⁷ L'IVOA et l'« Observatoire virtuel » qu'elle contribue à mettre en place fait d'ailleurs partie des initiatives étudiées par la *Research Data Alliance* comme exemple réussi d'infrastructure de données internationales.¹⁵⁸ La communauté des chercheurs en cristallographie est également très soudée. Sous l'impulsion de l'*International Union of Crystallography* (IUCr), elle a développé un format-modèle pour les données qu'elle génère, le format CIF (*Crystallographic Information Framework*),¹⁵⁹ et dispose d'un outil permettant de vérifier l'intégrité d'un fichier CIF¹⁶⁰ et de plusieurs archives de données de référence, comme *The Cambridge Structural Database* (CSD)¹⁶¹ ou *The Crystallography Open Database* (COD)¹⁶². Depuis quelques années, enfin, elle s'efforce d'élargir le champ des données devant être ouvertes en ne se limitant plus seulement au partage des données « dérivées » mais en y intégrant aussi les données « brutes ». L'IUCr a mis en place en août 2012 un groupe de travail sur le dépôt des données de diffraction (DDDWG) qui prépare des recommandations sur la diffusion des données brutes afin d'éviter les fraudes et permettre la vérification des données, le développement de logiciels et la validation des résultats de la recherche¹⁶³.

Dans d'autres disciplines, la multiplication d'articles sur les vertus de l'ouverture et les modalités techniques et juridiques du partage prouve qu'à l'heure actuelle de nombreux chercheurs s'intéressent à la question de l'*Open Research Data* et cherchent à acculturer leur communauté. C'est le cas dans le domaine de l'écologie et de la biologie où des articles récents, publiés dans des revues comme *The Journal of Experimental Botany* ou *Ideas in Ecology and Evolution*, intitulés « *Making open data work for plant scientists* » ou « *Nine simple ways to make it easier to (re)use your data* », contribuent à la diffusion des bonnes pratiques en matière d'ouverture¹⁶⁴. Ils prouvent, incontestablement, que l'ouverture des données ne peut se situer qu'à l'intersection de politiques institutionnelles et d'initiatives qui, quand elles n'émanent pas des communautés elles-mêmes, doivent être assimilées par ces communautés et intégrées, progressivement, par leurs représentants.

¹⁵⁷ <http://www.dcc.ac.uk/resources/metadata-standards/international-virtual-observatory-alliance-technical-specifications> (consulté le 22 novembre 2013)

¹⁵⁸ « Astronomy : connecting the IVOA and RDA », <https://europe.rd-alliance.org/Content/RDAEuropeecosystem.aspx?Cat=0!1!0!0> (consulté le 23 novembre 2013) ;

¹⁵⁹ THE ROYAL SOCIETY, *op. cit.*, p. 90 ; « CIF », <http://www.iucr.org/resources/cif> (consulté le 29 novembre 2013) ;

¹⁶⁰ Cf. <http://checkcif.iucr.org/> (consulté le 29 novembre 2013)

¹⁶¹ CODATA-ICSTI TASK GROUP ON DATA CITATION STANDARDS, « Out of Cite, Out of Mind: The Current State of Practice, Policy, and Technology for the Citation of Data », *Data Science Journal*, vol. 12, 13 septembre 2013, p. 26.

¹⁶² GRAZULIS et alii, « Crystallography Open Database (COD) : an open-access collection of crystal structures and platform for world-wide collaboration », *Nucleic Acids Research*, 2012, vol. 40, p. 420-427, <http://nar.oxfordjournals.org/content/40/D1/D420.full.pdf+html> (consulté le 29 novembre 2013)

¹⁶³ RESEARCH INFORMATION NETWORK (RIN), *op. cit.*, p. 15 ; « IUCr Diffraction Data Deposition Working Group », <http://www.iucr.org/resources/data/dddwg> (consulté le 29 novembre 2013). Sur cette dynamique en faveur du partage des données brutes en cristallographie, cf. HELLIWELL John R., « Deposition and use of raw diffraction images », *Crystallographic Information and Data Management : a Satellite Symposium to the 28th European Crystallographic Meeting, University of Warwick, 25 août 2013*, http://www.iucr.org/_data/assets/pdf_file/0004/80266/JRHelliwel_ECM28.pdf (consulté le 29 novembre 2013)

¹⁶⁴ LEONELLI Sabina, SMIRNOFF Nicholas, MOORE Jonathan, COOK Charis & BASTOW Ruth, *op. cit.*, 2013, & WHITE Ethan P. et alii, *op. cit.*, 2013.

Quoi qu'il en soit, ouvrir une donnée ne consiste pas seulement à la déposer dans une archive ouverte. Il convient aussi d'en définir les conditions de réutilisation et de faire sauter les verrous juridiques qui pourraient l'empêcher. Mais quel est, notamment en France, le statut de la donnée ? Et quels mécanismes permettent sa « libération » ?

LES ENJEUX JURIDIQUES DE L'OUVERTURE

Données de recherche et propriété intellectuelle...

L'ambition de l'ouverture des données de la recherche se heurte à un obstacle de taille, relevé par Aaron Griffiths en 2009 : les chercheurs ne comprennent pas toujours à *qui appartiennent* les jeux de données produits au cours de leur recherche et donc s'ils ont le droit de les rendre publiquement accessibles.¹⁶⁵ Cette difficulté avait également été soulevée lors de la journée d'étude sur les données de la recherche organisée par le CNRS le 29 mai 2012 et elle est évoquée dans le *Rapport* du groupe de travail de l'INRA sur la gestion et le partage des données. La problématique se heurte en effet :

« [*•* à] *l'indifférence, la méconnaissance, voire la défiance de la plupart des chercheurs relativement aux problèmes juridiques.*

[*•* au] *développement de la recherche au sein de consortiums internationaux regroupant des organismes n'ayant pas les mêmes environnements juridiques ni les mêmes positions stratégiques.* »¹⁶⁶

Le statut des données dans quatre pays européens a fait l'objet d'un rapport du *Knowledge Exchange* en 2011, mais celui-ci ne s'est pas arrêté sur le cas de la France et ne s'est intéressé qu'aux données les moins « travaillées », « dans leur forme la plus primaire », celles qui peuvent donc être assimilées à de purs enregistrements factuels (mesures, observations).¹⁶⁷ Il a conclu que dans ces quatre pays, l'Allemagne, la Finlande, les Pays-Bas et le Royaume-Uni, ces données de recherche étaient potentiellement protégées au titre de la propriété intellectuelle, mais que les critères applicables variaient d'un pays à l'autre, tout comme variaient, le cas échéant, le détenteur des droits d'auteur. Quel que soit le pays, du reste, les données peuvent être protégées par le droit *sui generis* de bases de données¹⁶⁸. Il convient cependant de noter que le droit européen des bases de données, notamment dans sa transposition en droit français, n'a pas pour objectif de protéger « un investissement substantiel » placé dans la

¹⁶⁵ GRIFFITHS Aaron, *op. cit.*, 2009, p. 51.

¹⁶⁶ INSTITUT NATIONAL DE LA RECHERCHE AGRONOMIQUE [INRA], *op. cit.*, 2012, p. 13, http://www.pfl-cepia.inra.fr/uploads/gdp_docs/Rapport-GestionDonnees-web.pdf (consulté le 7 octobre 2013)

¹⁶⁷ KNOWLEDGE EXCHANGE, *The legal status of the research data in the Knowledge Exchange partner countries (Netherlands, Denmark, Germany, United Kingdom)*, Centre for Intellectual Property Law (CIER), octobre 2011, 59 p., p. 10-11.

¹⁶⁸ Il est l'émanation de la directive européenne 96/9/CE du 11 mars 1996, cf. <http://eur-lex.europa.eu/LexUriServ/LexUriServ.do?uri=CELEX:31996L0009:FR:NOT> (consulté le 29 novembre 2013)

création de données, mais bien dans leur *agrégation* et leur présentation. Aussi, si de nombreuses collections de données correspondent à la définition d'une base de données, elles ne bénéficient pas nécessairement d'une protection au titre du droit *sui generis*. Concernant la donnée elle-même, en résumé, sa protection juridique est fonction de la manière dont sont interprétés les termes de la loi : si en général, pour être éligible à cette protection, la donnée doit atteindre un certain degré d'« originalité » et de « créativité » (voire de subjectivité) dans la manière dont elle a été produite, des pays donnent une définition plus large des données concernées par le droit d'auteur. Au Royaume-Uni, dans le contexte du monde de la recherche, une donnée peut être protégée au titre de la propriété intellectuelle si sa « création » a demandé « du savoir-faire, du jugement et du travail » : « le droit d'auteur (*copyright*) peut s'y appliquer aux données de recherche qui ont été collectées d'une manière créative (*recorded in a creative way*) avec un minimum d'effort – par opposition à un pur enregistrement des faits ». ¹⁶⁹ Le rapport du *Knowledge Exchange* concluait à la nécessité, idéale, de remettre à plat dans le droit européen l'ensemble des règles juridiques appliquées à la protection des données de recherche, en particulier des données brutes :

« Pour annihiler les effets potentiels de la protection au titre de la propriété intellectuelle des informations factuelles comme les données de recherche brutes, il est recommandable d'introduire une exception ou une limitation qui garantirait que les données de recherche puissent être librement utilisées à des fins académiques¹⁷⁰. Par ce biais, les chercheurs seraient autorisés à copier, modifier, distribuer et/ou publier des données de recherche sans avoir à en demander la permission au détenteur des droits d'auteur, [...] les rendant ainsi véritablement disponibles en libre-accès. »¹⁷¹

Dans l'attente d'une telle mutation, le rapport insistait sur l'importance des « arrangements contractuels », c'est-à-dire des outils ponctuels permettant la réutilisation des données, licences et *waivers*. ¹⁷² En tout cas, même quand la donnée n'est *a priori* pas éligible à la protection, elle peut être à tort, *de facto*, soumise à une sorte de propriété intellectuelle, un « monopole de fait » qui en entraverait l'ouverture :

« Même si aucune protection au titre de la propriété intellectuelle ne s'applique, les données de recherche peuvent être sujettes à des monopoles de fait. Le fait que des données relèvent du "domaine public" et ne soient pas protégées par les droits de la propriété intellectuelle ne signifie pas que leurs créateurs, ou les personnes qui les possèdent, aient à les communiquer à un tiers. »¹⁷³

Dans le droit français, comme ailleurs, les données brutes ne sont pas protégées par le droit d'auteur, mais leur mise en collection « originale » peut justifier la protection (de cet agencement particulier) par le droit *sui generis* des

¹⁶⁹ KNOWLEDGE EXCHANGE, *op. cit.*, 2011, p. 20.

¹⁷⁰ Comme on le constate, cette proposition contredit en partie ce que l'on entend généralement par « données ouvertes », puisqu'elles doivent aussi pouvoir être réutilisées à des fins commerciales ou industrielles.

¹⁷¹ KNOWLEDGE EXCHANGE, *op. cit.*, 2011, p. 52.

¹⁷² KNOWLEDGE EXCHANGE, *op. cit.*, 2011, p. 53-56 ; cf. infra p. 48.

¹⁷³ *Idem*, p. 45.

producteurs de bases de données.¹⁷⁴ Pourtant, si l'enregistrement factuel du réel ne s'assimile pas, en droit, à une création, le fait que la donnée soit systématiquement le résultat d'une opération scientifique et, partant, d'une méthode, mais également de processus de validation ou de sélection,¹⁷⁵ rend périlleuse l'identification définitive de données libres de droit¹⁷⁶. Cet obstacle juridique a récemment été mis en avant dans la presse par Serge Bauin, ancien directeur de l'IST au CNRS :

« Les législations anglo-saxonnes tendent à considérer que la protection des données de recherche commence très tôt, tandis que les législations continentales européennes estiment que les données brutes ne sont pas "protégeables". "Mais une donnée n'est jamais brute : pour la produire, il faut toujours une mesure, considère Serge Bauin. Et quand cette donnée est le fruit d'une collaboration internationale, on entre dans un cauchemar juridique !" »¹⁷⁷

... « un cauchemar juridique »

Un effort de clarification, dans le droit français, du statut juridique des données de recherche est donc encore à mener, la priorité restant de définir ce qu'est une donnée brute libre de droits. En l'état actuel, comme le remarque Marc Leobet, on peut globalement considérer que les données des chercheurs « ne sont pas légalement protégées »¹⁷⁸ et que les chercheurs ont toute latitude pour décider des modalités de leur divulgation.¹⁷⁹ La réflexion rejoint ici celle, plus large, de l'accès aux résultats de la recherche. Partant du postulat que les enseignants-chercheurs ou les chercheurs des EPST sont des agents publics,¹⁸⁰ ne sont-ils pas soumis au régime du droit d'auteur des agents publics et donc contraints de céder partie de leurs droits moraux et patrimoniaux à l'État ?¹⁸¹ Non, car les dispositions prévues par le *Code de la propriété intellectuelle* relativement au droit d'auteur des agents publics « ne s'appliquent pas aux agents auteurs d'œuvres dont la divulgation n'est pas soumise, en vertu de leur statut ou des règles qui régissent leurs fonctions, à un contrôle préalable de

¹⁷⁴ DANTANT Martin, « Droit d'auteur des chercheurs, logiciels, bases de données et archives ouvertes », support de formation du réseau Mistral-Doc, 30 mai 2013, 128 p., http://mistral.cnrs.fr/IMG/pdf/Droit_d_auteur_des_rechercheurs_Logiciels_Bases_de_Donnees_et_Archives_Ouvertes_-_Marseille.pdf (consulté le 1^{er} octobre 2013)

¹⁷⁵ FAYET Sylvie, « "Données" de la recherche, les mal-nommés », 15 novembre 2013, <http://urfistinfo.hypotheses.org/2581> (consulté le 2 décembre 2013)

¹⁷⁶ Sur ce point, voir une contribution récente sur la notion de « donnée brute » : GITELMAN Lisa éd., *Raw Data is an Oxymoron*, 2013, MIT Press.

¹⁷⁷ ROSIER Florence, « Surfer sur le tsunami des données de la recherche », *Le Monde*, 7 juin 2012, http://abonnes.lemonde.fr/sciences/article/2012/06/07/surfer-sur-le-tsunami-de-donnees-de-recherche_1714722_1650684.html (consulté le 23 octobre 2013)

¹⁷⁸ Marc Leobet est chargé de mission INSPIRE au MEDDE, <http://www.rencontres-sig-la-lettre.fr/wp-content/uploads/2010/05/TR-INSPIRE-BD-V3.pdf> (consulté le 3 décembre 2013)

¹⁷⁹ GEROUDET Marie-Madeleine, *Étude prospective sur les données de la recherche*, rapport sous la dir. d'Isabelle Le Bescond, mai 2013, SCD Univ. Lille 1, 62 p., p. 26-27.

¹⁸⁰ On se référera, comme Michèle Battisti et Martin Dantant, au décret n°84-431 du 6 juin 1984 (<http://www.legifrance.gouv.fr/affichTexte.do?cidTexte=LEGITEXT000006064492>) et à celui n°83-1260 du 30 décembre 1983 (<http://www.dsi.cnrs.fr/rmlr/textesintegaux/volume5/5112-d83-1260.htm>) (consultés le 2 décembre 2013)

¹⁸¹ Sur cette question, voir la synthèse de la direction des affaires juridiques du CNRS : <http://www.dgdr.cnrs.fr/daj/propriete/droits/droits5.htm> (consulté le 2 décembre 2013)

l'autorité hiérarchique » (article L. 111-1). C'est bien le cas des enseignants-chercheurs, qui « jouissent d'une pleine indépendance et d'une entière liberté d'expression dans l'exercice de leurs [...] activités de recherche », ¹⁸² mais également des chercheurs des EPST. Le chercheur public-auteur bénéficie donc d'un régime dérogatoire : comme l'a confirmé la loi du 1^{er} août 2006, il jouit de la totalité des droits de propriété intellectuelle sur ses « productions » ¹⁸³, donc sur ses publications et, *a priori*, sur les données qu'il collecte, assemble, traite, etc. ¹⁸⁴

Ajoutons par ailleurs que les données, même numériques, produites dans le cadre d'une activité de recherche par des agents publics pourraient être considérées, au regard de la loi, comme des archives publiques – sans contradiction avec les mécanismes de la propriété intellectuelle ¹⁸⁵. Les « données brutes » sont ainsi clairement intégrées au *Référentiel de gestion des archives de la recherche* produit par le réseau AURORE en septembre 2012. ¹⁸⁶

Cette indétermination juridique, encore compliquée par l'extrême diversité des types de données, les clauses de confidentialité intégrées à certains contrats de recherche, le régime spécifique associé à certaines données sensibles (confidentialité des données personnelles, notamment) ¹⁸⁷, n'est toutefois pas inconciliable avec la dynamique d'ouverture des données. Dans l'attente de définitions plus précises sur le cadre légal dans lequel, en Europe notamment, certaines données de recherche pourraient être rendues accessibles, réutilisables, modifiables, à la faveur d'une harmonisation des droits d'auteur ou d'une législation sur la « fouille de textes et de données » (*text and data mining*) à des fins de recherche scientifique, ¹⁸⁸ l'utilisation de licences demeure le mécanisme le plus adapté. Ainsi que l'a révélé très récemment un rapport d'OpenAIRE sur le droit des données, beaucoup d'éléments du cadre légal européen doivent encore être clarifiés. Même lorsque des données de recherche sont accessibles par l'intermédiaire de bases de données, et quand bien même ces données seraient libres de droits, la protection juridique *sui generis* des bases de données entre potentiellement en contradiction avec le paradigme de l'« ouverture » qui, on le sait, ne se limite pas à la simple accessibilité des données. En effet, si la

¹⁸² Article L. 952-2 du *Code de l'éducation*, cf. DANTANT Martin, *op. cit.*, 2013, p. 26 ; cf. également le décret n°84-431 du 6 juin 1984.

¹⁸³ LATREILLE Antoine, « L'appropriation des connaissances scientifiques à l'ère numérique », *Éthique publique*, 2012, vol. 14, n° 2, <http://ethiquepublique.revues.org/1008> (consulté le 2 décembre 2013)

¹⁸⁴ INRA, *op. cit.*, 2012, p. 14 ; DANTANT Martin, *op. cit.*, 2013, p. 26 ; LEOBET Marc, « Éléments juridiques des données des chercheurs », avril 2012, <http://georezo.net/blog/inspire/2012/04/06/2563/> (consulté le 3 décembre 2013)

¹⁸⁵ Sur ce débat, voir DELAUNAY Guillaume, *Les archives scientifiques en sciences humaines et sociales : état de l'art*, mémoire DCB sous la dir. de J.-F. Foucaud, 2012, 79 p., p. 20-23, http://memsic.ccsd.cnrs.fr/docs/00/68/64/99/PDF/MA_moire_Archives_scientifiques.pdf (consulté le 3 décembre 2013)

¹⁸⁶ *Référentiel de gestion* pour le « traitement et [la] conservation des archives des laboratoires de recherche, des chercheurs et des enseignants-chercheurs dans les universités, les organismes de recherche et les agences sanitaires », septembre 2012, 16 p.,

<http://f.hypotheses.org/wp-content/blogs.dir/1176/files/2013/02/tableaugestionarchivesaurores.pdf> (consulté le 3 décembre 2013)

¹⁸⁷ INRA, *op. cit.*, 2012, p. 15.

¹⁸⁸ Un groupe de travail de la Commission européenne a été mis en place pour réfléchir à cette question : <http://ec.europa.eu/licences-for-europe-dialogue/en/content/text-and-data-mining-working-group-wg4-0> (consulté le 3 novembre 2013). Cf., à ce même sujet, BATTISTI Michèle, « Quel droit pour faire du *text* et du *data mining* ? », http://www.a-brest.net/article13036.html#_ftn1 (consulté le 3 décembre 2013)

directive 96/9/CE prévoit des exceptions à la protection du contenu d'une base de données, invitant les États membres à autoriser « l'extraction » et « la réutilisation »¹⁸⁹ d'« une partie substantielle » de ce contenu « à des fins d'illustration de l'enseignement ou de recherche scientifique », cette possibilité n'a fait l'objet d'aucune harmonisation à l'échelle européenne.¹⁹⁰ En France, cette exception a été assimilée, mais seulement « à des fins exclusives d'illustration ».¹⁹¹ C'est la raison pour laquelle OpenAIRE, qui travaille à la construction d'une infrastructure européenne des données, a publié des « Recommandations au législateur européen » qui rappellent, pour partie, celles formulées par *Knowledge Exchange* en 2011 :

« L'exception à des fins de recherche scientifique est, en l'état actuel, inappropriée. Nous recommandons l'introduction d'une nouvelle exception pour la recherche, contraignante, à l'échelle européenne. [Cette exception] devrait être élargie, incluant explicitement l'analyse, la réutilisation et la reproduction d'une base de données dans son ensemble [...] à des fins scientifiques et non-commerciales [...]. Enfin, les exceptions contenues en l'article 9 de la directive 96/9/CE devraient être rendues obligatoires et reformulées de manière à aboutir à l'harmonisation du cadre légal européen »¹⁹²

Dans ce contexte en mouvement et, juridiquement, incertain, le meilleur moyen de définir les obligations juridiques attachées à une donnée est d'avoir recours à une licence ou une *waiver*.¹⁹³

Permettre la réutilisation : licences & *waivers*

Il n'est pas ici question de présenter de manière exhaustive les licences ouvertes à disposition des chercheurs pour définir le degré d'ouverture des données qu'ils produisent. L'adoption de telle ou telle licence est laissée à la responsabilité du chercheur, même si certaines archives de données n'acceptent qu'un seul type de licences.

¹⁸⁹ Ces notions sont définies en détails dans la directive, chap. III, article 7 : <http://eur-lex.europa.eu/LexUriServ/LexUriServ.do?uri=OJ:L:1996:077:0020:0028:FR:PDF> (consulté le 3 décembre 2013)

¹⁹⁰ DIETRICH Nils, GUIBAUT Lucie, MARGONI Thomas & alii, *OpenAIRE. Study on licensing of publications and research data : summary of findings*, novembre 2013, 9 p., p. 5, http://www.openaire.eu/en/about-openaire/publications-presentations/public-project-documents/doc_download/621-openairelicensingstudysummary (consulté le 3 décembre 2013). La version complète de ce rapport doit désormais être disponible en ligne. Cf. également BALL Alex, *How to License Research Data*, DCC, 2012, p. 3, http://www.dcc.ac.uk/sites/default/files/documents/publications/reports/guides/How_To_License_Research_Data.pdf (consulté le 3 décembre 2013)

¹⁹¹ Article L. 342-3 du Code de la propriété intellectuelle : <http://www.legifrance.gouv.fr/affichCodeArticle.do?cidTexte=LEGITEXT000006069414&idArticle=LEGIARTI000006279251&dateTexte=&categorieLien=cid> (consulté le 3 décembre 2013)

¹⁹² DIETRICH Nils, GUIBAUT Lucie, MARGONI Thomas & alii, *op. cit.*, 2013, p. 8. Des recommandations similaires ont été formulées par un groupe de travail du projet européen *Pro-iBiosphere* (1^{er} sept. 2012-31 août 2014) qui s'attache à la création d'une infrastructure européenne des données dans le domaine de la biodiversité. Ses experts ont proposé de réviser la directive 96/9/CE et militent pour l'harmonisation des droits européens en matière de réutilisation du contenu des bases de données pour la recherche scientifique. Cf. « *Pro-iBiosphere Draft policy on Open Access for data and information* », août 2013, 24 p., p. 20, http://wiki.pro-ibiosphere.eu/w/media/7/7f/Pro-iBiosphere_WP2_Plazi_D2.4.1_VFF_30082013.pdf (consulté le 3 décembre 2013)

¹⁹³ BATTISTI Michèle, « Libérons les données ! De quelques aspects juridiques », mars 2010, <http://www.paralipomenes.net/archives/622> (consulté le 3 décembre 2013)

La *waiver*, ou « renonciation », n'est pas à proprement parler une licence, mais un outil juridique permettant au détenteur des droits de propriété intellectuelle sur une donnée de les abandonner, faisant basculer cette donnée dans le domaine public. Les partisans de l'ouverture des données privilégient presque unanimement l'utilisation de la CC0, développée par l'organisation *Creative Commons*, qui offre un potentiel de réutilisation illimité à l'utilisateur de la donnée ainsi libérée¹⁹⁴. Certaines archives ouvertes de données parmi les plus connues, comme Dryad, sur laquelle nous reviendrons, placent automatiquement les données déposées sur sa base sous CC0 : si elle n'oblige pas formellement le (ré)utilisateur à mentionner l'origine et le créateur de la donnée, les normes internationales de la recherche scientifique et l'évaluation par les pairs suffisent à concilier abandon des droits sur la donnée et mention légitime de son producteur.¹⁹⁵ Des éditeurs scientifiques plébiscitent également l'attribution de la CC0, comme Pensoft¹⁹⁶, ou BioMed Central¹⁹⁷.

L'utilisation des autres licences *Creative Commons* était jusqu'à récemment globalement déconseillée, étant jugées trop restrictives, ou tout simplement inappropriées aux données, notamment parce que, centrées sur le droit d'auteur, elles négligeaient par exemple le droit *sui generis* des bases de données¹⁹⁸. Ce n'est désormais plus le cas puisqu'à la fin du mois de novembre 2013, *Creative Commons* a publié une version 4.0 de ses licences qui prend en compte le droit des bases de données¹⁹⁹. Cette nouvelle mouture des licences CC est d'ailleurs privilégiée par OpenAIRE.²⁰⁰ Une autre licence, développée par l'*Open Knowledge Foundation* et utilisée dans des projets *Open Data*, reste néanmoins utilisable : l'*Open Database License* (ODbL). Néanmoins, ses clauses d'attribution et de « partage dans les mêmes conditions » (*Attribution-ShareAlike*), qui en font l'équivalent d'une licence CC BY-SA, peuvent compliquer son utilisation dans le cadre de projets d'ouverture de données de recherche. L'écueil de « l'empilage des

¹⁹⁴ Parmi de nombreuses références : POISOT Timothée & alii, *op. cit.*, 2012, p. 7-8 ; WHITE Ethan P. et alii, *op. cit.*, 2013, p. 7 ; HAGEDORN Gregor & alii, « Creative Commons licenses and the non-commercial condition : Implications for the re-use of biodiversity information », *ZooKeys*, novembre 2011, n° 150, p. 127-149, <http://www.ncbi.nlm.nih.gov/pmc/articles/PMC3234435/> (consulté le 30 octobre 2013) ; COMMISSION EUROPEENNE, « Report of the European Commission Public Consultation on Open Research Data », octobre 2013, 15 p., https://ec.europa.eu/digital-agenda/sites/digital-agenda/files/Report_2013-07-OpenResearchData-Consultation-FINAL1.pdf (consulté le 23 octobre 2013).

¹⁹⁵ « Why does Dryad use CC0 ? », octobre 2011, <http://blog.datadryad.org/2011/10/05/why-does-dryad-use-cc0/> (consulté le 4 décembre 2013)

¹⁹⁶ LYUBOMIR Penev & alii, *Pensoft Data Publishing Policies and Guidelines for Biodiversity Data*, 2011, 34 p., p. 5-7, http://www.pensoft.net/J_FILES/Pensoft_Data_Publishing_Policies_and_Guidelines.pdf (consulté le 4 décembre 2013)

¹⁹⁷ « Opening up the data – an update to BioMed Central's Copyright and License Agreement », août 2013, <http://blogs.biomedcentral.com/bmcblog/2013/08/21/opening-up-the-data-an-update-to-biomed-centrals-copyright-and-license-agreement/> (consulté le 4 décembre 2013). Cf. supra (...).

¹⁹⁸ CARTIER Aurore, *Bibliothèques et Open data. Et si on ouvrait les bibliothèques sur l'avenir ?*, mémoire DCB sous la dir. de Frédérique Schlosser, 2013, 107 p., p. 31, <http://www.enssib.fr/bibliotheque-numerique/documents/60401-bibliotheque-et-open-data-et-si-on-ouvrait-les-bibliotheques-sur-l-avenir.pdf> (consulté le 15 octobre 2013) ; BALL Alex, *op. cit.*, 2012, p. 6 ;

¹⁹⁹ MAUREL Lionel, « Deux enjeux stratégiques dans la nouvelle version 4.0 des *Creative Commons* : *Open Data* et clause non-commerciale », 30 novembre 2013, <http://scinfolex.com/2013/11/30/deux-enjeux-strategiques-dans-la-nouvelle-version-4-0-des-creative-commons-open-data-et-clause-non-commerciale> (consulté le 4 décembre 2013) ; BOUVIER Stéphanie, « Quelles licences de diffusion pour une archive ouverte ? », 27 novembre 2013, <http://blog.univ-angers.fr/projetao/2013/11/27/quelles-licences-de-diffusion-pour-une-archive-ouverte-institutionnelle/#.Up9EICiwqQt> (consulté le 4 décembre 2013)

²⁰⁰ DIETRICH Nils, GUIBAUT Lucie, MARGONI Thomas & alii, *op. cit.*, 2013, p. 9.

attributions » (*attribution stacking* ou *cascading attribution*), notamment, est souvent mis en avant, et constitue l'un des principaux arguments en faveur de l'utilisation de licences du domaine public :

« La clause d'attribution n'est pas problématique si les données [réutilisées] ne sont combinées qu'avec des données issues d'un petit nombre de "datasets". [Mais] agréger des données issues d'un grand nombre de "datasets" peut se révéler infaisable s'il faut s'atteler au fardeau administratif que représente le fait de créditer individuellement chaque contributeur de ce super-jeu de données [...]. C'est ce problème qui est souvent connu sous le nom d'"attribution stacking" »²⁰¹

Cette difficulté, dans un contexte où la réutilisation de données ouvertes deviendrait la norme, pourrait vite devenir insurmontable, d'autant plus si la combinaison et l'agrégation des données, à la faveur de l'essor des *big data*, est automatisée.²⁰² Aussi, comme on l'a déjà vu avec le cas de Dryad, un vrai courant d'opinion est apparu parmi les partisans de l'*Open Research Data* pour que l'idéal de réutilisation ne soit pas parasité par les clauses d'attribution des licences et que chaque communauté scientifique, s'auto-régulant, suive ses propres normes en la matière tout en adoptant des licences du domaine public.²⁰³ Ces appels n'ont pas été entendus dans le récent « Manifeste d'Amsterdam sur les principes de citation des données » qui réclame que *l'ensemble* des contributeurs à la création d'un jeu de données soient systématiquement crédités.²⁰⁴ La place importante accordée au producteur de données partagées dans certains cas, comme dans celui des données sur la maladie d'Alzheimer diffusées sous l'impulsion de l'*Alzheimer's Disease Neuroimaging Initiative* (ADNI), a pu faire débat : les critiques formulées à l'encontre de l'ADNI,²⁰⁵ qui doit être créditée comme (co)auteur de l'ensemble des articles publiés à partir de données diffusées dans sa base, prouvent que la manière dont on crédite(nt) le(s) producteur(s) d'une donnée ouverte peut se révéler problématique et que les impératifs d'attribution entrent parfois en contradiction avec « l'esprit » de la libre réutilisation des données.

Aussi, comme on le retrouve dans les *Panton Principles for Open Data in Science* de l'OKF,²⁰⁶ le mécanisme idéal pour dissiper toute ambiguïté sur la possibilité de réutilisation de données reste le fait de les placer explicitement dans le domaine public, à travers la *Public Domain Dedication and License* (PDDL) ou la renonciation *Creative Commons Zero* (CC0). En France, Thierry Chanier, directeur du Laboratoire de Recherche sur le Langage de l'université Blaise-Pascal de Clermont-Ferrand, militant de l'ouverture, est partisan de cette solution. Si la

²⁰¹ BALL Alex, *op. cit.*, 2012, p. 7.

²⁰² DODDS Leigh, « How do we attribute data ? », avril 2013, <http://blog.ldodds.com/2013/04/30/how-do-we-attribute-data/> (consulté le 4 décembre 2013)

²⁰³ MILLER Paul, « Getting it right with data attribution », mai 2013, <http://cloudofdata.com/2013/05/getting-it-right-with-data-attribution> (consulté le 4 décembre 2013)

²⁰⁴ https://docs.google.com/document/d/1ON0vy2_jT2VxL_Cdm03HgMSNnN1A6VzvDQrTi577-ig/edit?pli=1 (consulté le 4 décembre 2013).

²⁰⁵ ROHLFING Torsten, POLINE Jean-Baptiste, « Why shared data should not be acknowledged on the author byline », *NeuroImage*, vol. 59, n° 4, février 2012, p. 4189-4195, <http://www.sciencedirect.com/science/article/pii/S1053811911011554> (consulté le 4 décembre 2013); *Alzheimer's Disease Neuroimaging Initiative (ADNI) Data Use Agreement*, rév. octobre 2013, http://adni.loni.usc.edu/wp-content/uploads/how_to_apply/ADNI_Data_Use_Agreement.pdf (consulté le 4 décembre 2013).

²⁰⁶ <http://pantonprinciples.org/> (consulté le 4 décembre 2013)

licence CC-BY reste un moindre mal, elle complique les possibilités de réutilisation et il est donc préférable de ne pas gréver les données de clauses d'attribution. Là encore, de toute façon, la libération de la donnée ne se fait pas nécessairement au détriment du chercheur-producteur : en rendant possible la simple citation d'un jeu de données comme s'il s'agissait d'une référence bibliographique,²⁰⁷ et en se fiant aux conventions académiques, un producteur de données peut légitimement supposer que son travail ne sera pas pillé.²⁰⁸

DES INFRASTRUCTURES COMPLEMENTAIRES ? ÉTUDES DE CAS

La majorité des initiatives en matière de données de recherche se jouent à l'heure actuelle sur le terrain des *liens entre publications et données* : la dynamique d'ouverture étant justifiée par des impératifs de validation et de reproductibilité des résultats de la recherche,²⁰⁹ l'objectif est d'abord de donner accès aux données sur lesquelles ces résultats sont bâtis, sous une forme moins contraignante (du point de vue de la quantité des données présentées) et plus satisfaisante (meilleur potentiel de réutilisation) que les traditionnels « matériaux supplémentaires » (*supplementary files*) associés aux articles scientifiques.²¹⁰ Puisqu'il a été démontré, du reste, que la création de liens réciproques entre articles et données avaient des effets bénéfiques sur le « succès » des publications et leur citation,²¹¹ de nombreuses solutions techniques ont été élaborées pour permettre la construction d'un écosystème associant publications et données de recherche : il repose sur l'identification pérenne de chaque jeu de données, permettant ainsi leur citation,²¹² et sur des infrastructures garantissant leur archivage.

²⁰⁷ Cf. note 212.

²⁰⁸ CHANIER Thierry, « A viewpoint on the place of CALL within the Digital Humanities : considering CALL journals, research data and the sharing of research results », *EUROCALL 2013. Learning from the Past, Looking to the Future*, université d'Évora, Portugal, 11-14 septembre 2013, <http://edutice.archives-ouvertes.fr/docs/00/86/20/24/PDF/eurocall2013-chanier-130915.pdf> (consulté le 30 octobre 2013)

²⁰⁹ NATURE, « Must try harder », mars 2012, vol. 483, p. 509, <http://www.nature.com/nature/journal/v483/n7391/pdf/483509a.pdf> (consulté le 26 décembre 2013)

²¹⁰ Sur les modalités « classiques » d'intégration des données de recherche aux articles scientifiques, cf. REILLY Susan & alii, *Report on Integration of Data and Publications*, ODE [Opportunities for Data Exchange] project, octobre 2011, 87 p., p. 36-49,

<http://www.libereurope.eu/sites/default/files/ODE-ReportOnIntegrationOfDataAndPublication.pdf> (consulté le 7 octobre 2013)

²¹¹ On consultera notamment HENNEKEN Edwin A., ACCOMAZZI Alberto, « Linking to Data – Effect on Citation Rates in Astronomy », novembre 2011, <http://arxiv.org/abs/1111.3618> (consulté le 26 décembre 2013) ; HOLOGNE Odile, « How could forest science journals contribute to the development of open data ? », *Meeting of the Editors of Forest Science Journals*, Nancy, 24 septembre 2013, <http://prodira.inra.fr/?locale=fr#!ConsultNotice:208058> (consulté le 26 décembre 2013)

²¹² Un rapport très complet a été publié en septembre 2013 par le groupe opérationnel du CODATA (*Committee on Data for Science and Technology*) qui se consacre aux « Pratiques et aux normes de citation des données » : CODATA-ICSTI TASK GROUP ON DATA CITATION STANDARDS, « Out of Cite, Out of Mind: The Current State of Practice, Policy, and Technology for the Citation of Data », *Data Science Journal*, vol. 12, 13 septembre 2013, 75 p., https://www.jstage.jst.go.jp/article/dsj/12/0/12_OSOM13-043/_pdf (consulté le 26 décembre 2013). On peut également se reporter à BALL Alex, DUKE Monica, *How to Cite Datasets and Link to Publications*, DCC, juin 2012, 12 p.

http://www.dcc.ac.uk/sites/default/files/documents/publications/reports/guides/How_to_Cite_Link.pdf (consulté le 26 décembre 2013)

La diversité des solutions disciplinaires

Parmi ces infrastructures, l'on citera le cas de Dryad et Pangaea. Dryad est un entrepôt de données américain spécialisé dans la biologie de l'évolution et l'écologie, destiné à accueillir « la longue traîne » des données ne pouvant être archivées dans de grands entrepôts thématiques comme GenBank ou TreeBASE.²¹³ L'entrepôt est administré par une association à but non lucratif réunissant des éditeurs, des sociétés savantes et une bibliothèque, la *German National Library of Medicine*. Il est en grande partie financé par la *National Science Foundation* (NSF) américaine, mais aussi depuis 2013 par les frais d'adhésion des membres de l'association, les frais de publication (*Data Publishing Charges*) assumés par les organismes souhaitant faire bénéficier leurs chercheurs du dépôt dans Dryad et les frais de publication assumés parfois par les chercheurs eux-mêmes – 80 dollars par « paquet » de données associées à un article.²¹⁴ Dryad cible particulièrement les revues, les éditeurs, les sociétés savantes éditrices, en leur proposant de souscrire pour une période de deux ans un contrat forfaitaire leur permettant de déposer autant de jeux de données qu'ils le souhaitent. Ces revues peuvent alors intégrer à leur procédure de soumission en ligne des manuscrits le dépôt des données associées dans Dryad. Quelles que soient les modalités de dépôt, les données doivent obligatoirement être associées à un article. Chaque jeu de données déposé se voit assigner un identifiant pérenne, en l'occurrence un DOI (*Digital Object Identifier*). Ce type d'identifiant repose sur la suite logicielle « Handle »²¹⁵ et ne peut être attribué que par une agence officielle membre du consortium DataCite.²¹⁶ Créé en 2009, le consortium DataCite réunit, sous l'égide de la *German National Library of Science and Technology* (TIB), des bibliothèques, des instituts de recherche et, depuis 2010, pour la France, l'Institut de l'information scientifique et technique (INIST),²¹⁷ qui œuvrent pour favoriser l'identification pérenne des données. L'identifiant peut ensuite être intégré à l'article publié, qui est lui-même cité dans Dryad : par l'intermédiaire du DOI, le jeu de données est définitivement accessible depuis l'article et vice versa.²¹⁸ Dryad hébergeait en décembre 2013 près de 4500 « paquets » de données placées (après embargo ou non) sous licence CC0. Dans son domaine de spécialité, cet entrepôt reste l'une des infrastructures les plus efficaces au service de l'ouverture et de la diffusion des données de recherche.

C'est également le cas de Pangaea, un entrepôt de données administré par des instituts de recherche allemands spécialisés dans les sciences marines.²¹⁹ Tout type de données dans le domaine des sciences de la vie et de la terre peut y être déposé (tant que ces données peuvent être géoréférencées) et Pangaea collabore depuis plusieurs années avec des éditeurs scientifiques pour assurer l'accessibilité et la

²¹³ CODATA-ICSTI TASK GROUP ON DATA CITATION STANDARDS, *op. cit.*, 2013, p. 71.

²¹⁴ <http://datadryad.org/pages/pricing> (consulté le 26 décembre 2013)

²¹⁵ BRASE Jan, FARQUHAR Adam, « Access to Research Data », *D-Lib Magazine*, vol. 17, n° 1/2, 2011, <http://www.dlib.org/dlib/january11/brase/01brase.html> (consulté le 14 décembre 2013)

²¹⁶ BALL Alex, DUKE Monica, *op. cit.*, 2012, p. 3-4.

²¹⁷ <http://www.inist.fr/?DataCite> (consulté le 14 décembre 2013) ; « Faciliter la citation et l'accès à la recherche : attribuer un DOI », <http://doccitantist.lirmm.fr/spip.php?article236> (consulté le 28 décembre 2013)

²¹⁸ Pour un exemple de jeu de données entreposé dans Dryad : <http://datadryad.org/resource/doi:10.5061/dryad.7bn22> (consulté le 26 décembre 2013)

²¹⁹ <http://www.pangaea.de/> (consulté le 26 décembre 2013)

visibilité des jeux de données associés à des articles publiés – et en cours de publication. Comme dans Dryad, chaque jeu de données qui y est archivé reçoit un identifiant DOI et il est relié à l'article qui le documente, lui-même étant accessible depuis Pangaea.²²⁰

Solutions éditoriales et *data journals*

L'importance du lien entre publications et données, rappelé en mars 2013 par le collectif Force 11 lors de la conférence « Beyond the PDF 2 » à Amsterdam,²²¹ contribue également à bouleverser le monde de l'édition scientifique qui voit dans l'intégration des données à son modèle d'affaires « un nouvel eldorado »²²². La plus connue des initiatives éditoriales en la matière est sans doute l'« Article du futur », projet lancé en 2009 par Elsevier pour renouveler la forme traditionnelle de l'article scientifique.²²³ Les données n'y sont plus seulement citées ou reliées par l'intermédiaire d'un DOI, mais intégrées grâce à des outils de visualisation, contribuant ainsi à l'enrichissement du contenu de l'article. Elsevier a ainsi conclu un partenariat avec Pangaea pour inclure dans ScienceDirect une carte interactive de localisation des données associées à un article déposées dans Pangaea.²²⁴ Plus généralement, certains entrepôts de données (comme Dryad) sont représentés dans ScienceDirect sous la forme d'une bannière permettant d'accéder au jeu de données auquel l'article fait référence.²²⁵ Dernière illustration de son intérêt pour ces questions, Elsevier a mis en place en 2013 un service « Données de recherche » afin d'encourager les chercheurs à publier leurs données.²²⁶ Si elle soulève encore des problèmes « éthiques » - jusqu'à quel point un article peut-il être enrichi après avoir été revu par les pairs ?, la publication augmentée est l'une des réponses éditoriales apportée à la question de l'ouverture des données.

Les *data journals* font également partie des initiatives des éditeurs scientifiques pour renouveler le lien entre publications et données et contribuer à leur diffusion. Contrairement à un article classique, le *data paper* est seulement dédié à la description d'un ou plusieurs jeux de données, et à la publication de

²²⁰ BALL Alex, DUKE Monica, *op. cit.*, 2012, p. 9. Pour un exemple de jeu de données entreposé dans Pangaea : <http://doi.pangaea.de/10.1594/PANGAEA.824434> (consulté le 26 décembre 2013)

²²¹ FORCE 11, « Declaration of Data Citation Principles. Draft », mars 2013, <http://www.force11.org/datacitation> (consulté le 27 décembre 2013). La version définitive de ce manifeste devrait être adoptée au début de l'année 2014.

²²² HOLOGNE Odile, « Professionnels IST et données de la recherche : des discours incantatoires aux actions concrètes », *Documentaliste*, vol. 50, n° 3, octobre 2013, p. 31.

²²³ Cf. supra p. 37-38. GEROUDET Marie-Madeleine, *op. cit.*, 2013, p. 49-50.

²²⁴ « How Elsevier is connecting data and research articles on ScienceDirect », <http://www.elsevier.com/about/content-innovation/database-linking> (consulté le 28 décembre 2013) ; REILLY Susan & alii, *Report on Integration of Data and Publications*, ODE [Opportunities for Data Exchange] project, octobre 2011, 87 p., p. 51.

²²⁵ Pour un exemple : <http://www.sciencedirect.com/science/article/pii/S105579031100279X> (consulté le 28 décembre 2013). Pour une présentation détaillée de l'« Article du futur » : AALBERSBERG Jan IJsbrand & alii, « Bringing Digital Science Deep Inside the Scientific Article : the Elsevier Article of the Future Project », *LIBER Quarterly*, vol. 23 (3) à paraître, <http://liber.library.uu.nl/index.php/lq/article/view/8446> (consulté le 28 décembre 2013)

²²⁶ « Elsevier Research Data Services », <http://researchdata.elsevier.com/> (consulté le 28 décembre 2013) ; MARQUES David, « Research Data Driving New Services », *Library Connect Newsletter*, vol. 11, n° 1, 2013, <http://libraryconnect.elsevier.com/articles/best-practices/2013-02/research-data-driving-new-services> (consulté le 28 décembre 2013)

leurs métadonnées. Les données de recherche associées, décrites dans l'article, sont déposées dans un entrepôt disciplinaire et reliées à l'article. L'objectif du *data journal* est avant tout de compenser le manque de visibilité des données produites dans le cadre de petits projets scientifiques et de valoriser la rédaction de métadonnées au même titre que celle d'un article classique.²²⁷ Le groupe *Nature* lancera ainsi en mai 2014 *Scientific Data*, une revue en libre-accès, évaluée par les pairs, décrivant des jeux de données.²²⁸ Comme d'autres *data journals*, elle permettra aux chercheurs de n'être plus seulement crédités pour leurs articles, mais également pour les données qu'ils produisent.²²⁹ Afin de faciliter leur réutilisation, les métadonnées associées aux articles de *Scientific Data* seront placées sous licence CC0.²³⁰ D'autres formes de publication reposant sur la description de données de recherche voient actuellement le jour comme, dans le domaine des sciences environnementales, le *Biodiversity Data Journal* : les données n'y sont pas seulement décrites, mais intégrées au maximum au corps même de l'article.²³¹

Gestionnaires d'entrepôts de données et éditeurs ne sont pas les seuls acteurs à s'attacher à la consolidation des liens entre publications et données. D'autres initiatives s'inscrivent dans le mouvement de l'*Open research data* et, à une autre échelle, permettent le développement d'infrastructures propices à la visibilité et à la circulation des données de recherche. C'est le cas d'OpenAIREplus.

Fédérer des entrepôts interopérables : l'ambition d'OpenAIRE

L'initiative européenne OpenAIRE, dans son prolongement « OpenAIREplus » (décembre 2011-mai 2014), mise sur une solution distribuée et transdisciplinaire pour améliorer la visibilité des données produites à l'échelle européenne. OpenAIREplus poursuit sur le versant « données » l'effort engagé dès 2009 pour les publications. Comme pour les publications, il s'agit d'abord de mettre à la disposition des chercheurs ne disposant pas d'un entrepôt institutionnel ou disciplinaire une archive orpheline, Zenodo,²³² créée en collaboration avec le CERN et lancée au printemps 2013.

²²⁷ CHAVAN Vishwas, PENEV Lyubomir, « The data paper : a mechanism to incentivize data publishing in biodiversity science », *BMC Bioinformatics*, vol. 12 suppl. 15, 2011, <http://www.biomedcentral.com/1471-2105/12/S15/S2> (consulté le 28 décembre 2013)

²²⁸ Pour un exemple de *data paper* : <http://www.nature.com/sdata/2013/samplearticle2.html> (consulté le 28 décembre 2013)

²²⁹ LEONELLI Sabina & alii, « Making open data work for plant scientists », *Journal of Experimental Botany*, septembre 2013, p. 4114.

²³⁰ <http://blogs.nature.com/scientificdata/2013/12/09/metadata-associated-with-data-descriptor-articles-to-be-released-under-cc0-waiver/> (consulté le 28 décembre 2013)

²³¹ SMITH Vincent & alii, *op. cit.*, 2013. Pour un exemple d'article du *BDJ* : <http://biodiversitydatajournal.com/articles.php?id=1013> (consulté le 28 décembre 2013)

²³² <http://zenodo.org/> (consulté le 14 décembre 2013)

L'un des principaux objectifs d'OpenAIREplus est de promouvoir l'interopérabilité des entrepôts²³³ et, à ce titre, OpenAIREplus milite pour la conformité des archives de données avec le protocole OAI-PMH. Outre la gestion de l'archive orpheline Zenodo, compatible OAI, le projet a en effet vocation à créer un point d'accès unique à des jeux de données distribués, en assurant la mise en réseau d'entrepôts institutionnels ou thématiques. OpenAIREplus a ainsi travaillé à la modélisation d'un schéma-type de métadonnées,²³⁴ processus qui a abouti au printemps 2013 avec la publication en ligne de *guidelines*. Les responsables du projet se sont inscrits dans le sillage des initiatives de l'agence DataCite en adoptant le *DataCite Metadata Schema*, sous une forme adaptée, pour permettre le moissonnage et l'importation de métadonnées depuis les entrepôts de données.²³⁵

À terme, l'objectif d'OpenAIREplus est de moissonner suivant le protocole OAI-PMH les métadonnées (décrites au format *DataCite*), exposées selon ses recommandations, relatives à des jeux de données. Seules les données issues d'un projet de recherche financé par l'Europe et/ou reliées à une publication – l'une et l'autre informations étant intégrées au schéma de métadonnées, seront présentées sur le portail OpenAIREplus.²³⁶ Le volet « données de recherche » n'avait pas encore, en décembre 2013, été incorporé au portail : de la même manière qu'à l'heure actuelle il présente la liste des archives ouvertes (de publications) compatibles OpenAIREplus,²³⁷ le portail devrait bientôt donner la liste des entrepôts de données moissonnables. OpenAIREplus a signé en octobre 2013 une convention avec *re3data.org*, un répertoire international d'entrepôts de données financé par la *German Research Foundation* (DFG),²³⁸ afin de faciliter l'intégration à OpenAIRE des entrepôts recensés par *re3data.org*.²³⁹

²³³ RETTBERG Najla, SCHMIDT Birgit, « Building a Collaborative Open Access Infrastructure for European Researchers », *LIBER Quarterly*, vol. 22, n° 3, 2012, p. 160-175, <http://liber.library.uu.nl/index.php/lq/article/view/URN%3ANBN%3ANL%3AUI%3A10-1-113938/8514> (consulté le 14 décembre 2013)

²³⁴ GEROUDET Marie-Madeleine, *op. cit.*, 2013, p. 20 ; LUTZ Jean-François, « Le projet OpenAIREplus », *Données de la recherche : enjeux, perspectives, politique(s), journée d'étude organisée par le CNRS, Paris, 29 mai 2012*.

²³⁵ OPENAIREPLUS, *Data Guidelines : Use of DataCite Metadata Schema*, juin 2013, https://guidelines.openaire.eu/wiki/Data_Guidelines:_Use_of_DataCite_Metadata_Schema (consulté le 14 décembre 2013)

²³⁶ OPENAIREPLUS, *OpenAIRE Guidelines : For Data Archives*, juin 2013, https://guidelines.openaire.eu/wiki/OpenAIRE_Guidelines:_For_Data_Archives (consulté le 24 décembre 2013)

²³⁷ <https://www.openaire.eu/fr/component/openaire/compliantrepos/default/620> (consulté le 24 décembre 2013)

²³⁸ PAMPAL Heinz & alii, « Making Research Data Repositories Visible : The re3data.org Registry », *PLoS ONE*, 8 (11), novembre 2013, <http://www.plosone.org/article/info%3Adoi%2F10.1371%2Fjournal.pone.0078080> (consulté le 26 décembre 2013)

²³⁹ <http://www.re3data.org/2013/10/memorandum-of-understanding-between-openaire-and-re3data-org/> (consulté le 26 décembre 2013)

DONNEES, ETABLISSEMENTS DE RECHERCHE ET BIBLIOTHEQUES

On l'a vu, la problématique de l'ouverture est à la croisée de nombreuses impulsions. Les initiatives reposent sur l'engagement de plusieurs acteurs complémentaires parmi lesquels les chercheurs, producteurs des données, occupent bien sûr une place essentielle. Ce dernier chapitre, en partie prospectif, a pour but d'étudier la position qu'occupent, dans cet écosystème, les bibliothèques, et la manière dont elles peuvent accompagner, à l'échelle des établissements universitaires, les politiques d'ouverture.

LA PRISE EN MAIN DES DONNEES PAR LES ETABLISSEMENTS DE RECHERCHE : METHODOLOGIE

Des responsabilités des institutions en matière de gestion et de conservation

Nous avons déjà rapidement mentionné le fait que dans certains pays, comme au Royaume-Uni, des établissements universitaires se dotaient de politiques institutionnelles de données²⁴⁰ et définissaient leurs responsabilités (et celles de leurs chercheurs) en matière de gestion, de conservation et de partage. Nous allons nous arrêter en détails sur ces politiques institutionnelles, sur les missions que choisissent d'assumer, aux États-Unis, en Australie ou en Grande-Bretagne, les institutions de recherche, mais également, puisque les universités françaises devront bientôt s'atteler au sujet des *research data*, sur la méthodologie mise en place par les établissements anglo-saxons pour définir et mettre en œuvre leurs politiques. Le *Code de la recherche*, modifié en juillet 2013, intègre désormais « l'organisation de l'accès libre aux données scientifiques » comme faisant partie des objectifs de la recherche publique française et donc des missions des établissements de recherche français.²⁴¹

Si la gestion des données est d'abord une affaire de chercheur, beaucoup d'universités américaines ou australiennes intègrent cet objectif à leur mission. Le cas de l'université Monash, en Australie, est particulièrement bien documenté et fait partie des exemples mis en avant par le JISC²⁴² : l'université elle-même, d'ailleurs, cherche à s'affirmer comme « *leader* régional, national et international

²⁴⁰ Cf. supra p. 32.

²⁴¹ *Code de la recherche*, « chap. II Objectifs et moyens institutionnels de la recherche publique », article L112-1 modifié par la loi n° 2013-660 du 22 juillet 2013, article 16. http://www.legifrance.gouv.fr/affichCode.do?sessionId=1B91C4435EFC974490B5CC8CE2295188.tpdjo14v_1?idSectionTA=LEGISCTA000006151273&cidTexte=LEGITEXT000006071190&dateTexte=20131113 (consulté le 10 décembre 2013)

²⁴² JONES Sarah, « Bringing it all together : a case study on the improvement of research data management at Monash University », Edinburgh, DCC, 2013, http://www.dcc.ac.uk/resources/developing-rdm-services/improving-rdm-monash#_ftn2 (consulté le 10 décembre 2013)

en matière de gestion des données de recherche ». ²⁴³ Le chantier de la gestion des données y a été ouvert au milieu des années 2000 et l'actuelle politique institutionnelle ne fut adoptée qu'en 2010, après plusieurs années de travail avec les chercheurs qui permirent, simultanément, de les sensibiliser à l'importance du *data management*. Cette politique, dont on retrouve un écho dans celle de l'université d'Edimbourg, ²⁴⁴ prend acte du fait que la gestion des données est « une responsabilité partagée » entre les chercheurs et l'université, laquelle s'engage en matière de conservation, d'accompagnement des producteurs dans la gestion de leurs données et de formation ²⁴⁵. La stratégie de l'université est donc avant tout une stratégie d'accompagnement. Si la gestion des données est placée au cœur des priorités de l'institution, cela ne se traduit jamais en termes contraignants : contrairement au Royaume-Uni où, du fait des exigences des agences de financement, les institutions de recherche font de la rédaction d'un plan de gestion une formalité *obligatoire*, ²⁴⁶ c'est une approche partenariale qui est ici privilégiée, l'idée étant d'accompagner progressivement les chercheurs dans l'amélioration de leurs compétences de gestion (*data management skills*) ²⁴⁷. Pour le reste, les engagements pris par l'université australienne font écho à la majorité des politiques institutionnelles britanniques, comme celle d'Oxford où :

« 10. L'université a la responsabilité :

- de fournir des services et des infrastructures pour l'archivage, la sauvegarde, le dépôt et la conservation des données de recherche [...] ;
- de faire bénéficier les chercheurs de formations, de soutien et de conseils sur la gestion des données de recherche [...] » ²⁴⁸

D'un point de vue méthodologique, on peut avoir un aperçu des démarches et des acteurs engagés à travers les cas des universités britanniques de Bath, Leeds ou Northampton. Certaines de ces institutions ont d'abord pu bénéficier du soutien logistique du JISC qui, à travers son programme « Managing Research Data » (octobre 2011-juillet 2013), a financé près d'une vingtaine de projets institutionnels autour de la gestion des données de recherche ²⁴⁹. Cette initiative est indissociable de l'impulsion donnée au printemps 2011 par la plus importante agence de financement britannique pour que les établissements de recherche bénéficiant de ses crédits définissent, en conformité avec ses propres attentes, une politique et des procédures de gestion des données. ²⁵⁰ À Bath et à Leeds, un comité

²⁴³ MONASH UNIVERSITY, « Research Data Management Strategy and Strategic Plan 2012-2015 », avril 2012, 17 p., p. 15.

²⁴⁴ « Edinburgh University Research Data Management Policy », mai 2011, <http://www.ed.ac.uk/schools-departments/information-services/about/policies-and-regulations/research-data-policy> (consulté le 10 décembre 2013)

²⁴⁵ « Monash University Research Data Management Policy », 2010, <http://www.policy.monash.edu/policy-bank/academic/research/research-data-management-policy.html> (consulté le 10 décembre 2013)

²⁴⁶ Cf. supra p. 32.

²⁴⁷ JONES Sarah, *op. cit.*, 2013.

²⁴⁸ OXFORD UNIVERSITY, « Policy on the Management of Research Data and Records », http://www.admin.ox.ac.uk/media/global/wwwadminoxacuk/localsites/researchdatamanagement/documents/Policy_on_the_Management_of_Research_Data_and_Records.pdf (consulté le 10 décembre 2013). Cf. annexe III, p. 97.

²⁴⁹ Jusqu'à sa clôture à l'été 2013, ce programme a été piloté par Simon Hodson, désormais directeur exécutif de CODATA. Pour consulter la liste complète des projets institutionnels financés par le JISC : <http://researchdata.jiscinvolve.org/wp/> (consulté le 10 décembre 2013)

²⁵⁰ Cf. supra p. 32-33.

de pilotage chargé de statuer les questions de gestion des données avaient déjà été mis en place en amont.²⁵¹ Ils ont candidaté avec succès auprès du JISC et ont joué un rôle important dans la coordination de leurs projets respectifs, baptisés *Research 360* à Bath et *RoaDMap* à Leeds. De la méthodologie mise en œuvre, on retiendra qu'elle a conduit à associer étroitement au projet la gouvernance de l'université, représentée dans les comités de pilotage par le *Pro-Vice-Chancellor (PVC) for Research* qui, à la tête du *Research Committee* du Sénat de l'université, supervise la stratégie universitaire de recherche²⁵². À Bath, sa participation a été décisive, qui a permis d'anticiper les difficultés que pourrait soulever la feuille de route stratégique au cours du processus de validation par le *Vice-Chancellor's Group (VCG)* qui pilote le Sénat de l'université²⁵³. La bibliothèque était bien sûr représentée dans ces projets, notamment en la personne, à Bath, du *Research Publication Librarian*.²⁵⁴

L'évaluation de *l'existant* constitue une étape essentielle qui permet, préalablement à la définition d'une politique et d'une feuille de route institutionnelles, de faire le point sur les pratiques et les besoins des communautés de chercheurs en matière de gestion des données. À Northampton et à Bath, comme dans la plupart des cas au Royaume-Uni, ce sont les bibliothèques qui se sont chargées du déploiement de cet audit, utilisant des outils mis en place par le *Digital Curation Center*. L'université de Northampton, durant huit semaines de mai à juin 2010, a utilisé la méthodologie « DAF » (*Data Asset Framework*) afin d'identifier les types de données produites dans les laboratoires, la manière dont les chercheurs envisagent leur cycle de vie, leurs besoins en termes d'infrastructure et de soutien de la part de l'université. Cette méthodologie repose généralement sur des entretiens ciblés avec des chercheurs représentatifs des disciplines pratiquées, une enquête en ligne puis des entretiens approfondis avec des chercheurs s'étant portés volontaires à l'issue de cette enquête.²⁵⁵ Les résultats de cette enquête permettent d'affiner, à l'échelle d'une université et des différentes disciplines concernées, le périmètre et les objectifs d'une politique institutionnelle de gestion des données.²⁵⁶ Dans le cas de Northampton, l'équipe-projet en charge de cet audit était pilotée par le bibliothécaire responsable des services à la recherche (*Research Support Librarian*).²⁵⁷ Il a donné lieu à un rapport qui à l'origine de l'adoption de la politique universitaire de gestion des données.²⁵⁸ Les établissements

²⁵¹ À Leeds : http://library.leeds.ac.uk/info/422/policies/219/research_data_steering_and_working_groups/1 (consulté le 10 décembre 2013) ; à Bath, cf. PINK Catherine, « Meeting the Data Management Compliance Challenge : Funder Expectations and Institutional Reality », *The International Journal of Digital Curation*, vol. 2, n° 8, 2013, p. 157-171, p. 158, <http://www.ijdc.net/index.php/ijdc/article/viewFile/8.2.157/325> (consulté le 10 décembre 2013)

²⁵² Cf. l'organigramme de l'université de Bath : <http://www.bath.ac.uk/statutory-bodies-committees/bodies-and-committees-senate/index.html> (consulté le 10 décembre 2013)

²⁵³ PINK Catherine, *op. cit.*, 2013, p. 160.

²⁵⁴ Cf. la composition de l'équipe du projet *Research 360* : <http://blogs.bath.ac.uk/research360/team/> (consulté le 10 décembre 2013)

²⁵⁵ PICKTON Miggie, JONES Sarah, MARIEKE Guy, « Leading from the library : data management initiatives at the University of Northampton », *The role of libraries in data curation, access and preservation : an international perspective. Conference Session 116, 78th IFLA General Conference and Assembly, 11-17 août 2012, Helsinki*, 10 p., p. 4-5, <http://conference.ifla.org/past/2012/116-pickton-en.pdf> (consulté le 10 décembre 2013).

²⁵⁶ Sur la méthode *Data Asset Framework*, voir sa présentation par le *Digital Curation Center* : <http://www.dcc.ac.uk/resources/repository-audit-and-assessment/data-asset-framework> (consulté le 10 décembre 2013). Se reporter également à : <http://www.data-audit.eu/> (consulté le 5 décembre 2013).

²⁵⁷ PICKTON Miggie & alii, *op. cit.*, 2012, p. 4.

²⁵⁸ « The University of Northampton Research Data Policy », <http://www.northampton.ac.uk/staff/academic-research/research-data-policy> (consulté le 10 décembre 2013)

universitaires peuvent avoir aussi recours à l'outil CARDIO (*Collaborative assessment of research data infrastructure and objectives*) du DCC, comme l'a fait l'université de Bath en complément de la méthode « DAF ».²⁵⁹

Une politique formalisée de gestion des données et/ou une feuille de route stratégique détaillant les objectifs de l'université en termes d'infrastructure et les actions à mettre en œuvre pour y parvenir font partie des livrables de ces projets institutionnels²⁶⁰. Si l'adoption d'une *data management policy* est toujours considérée comme indispensable, l'élaboration de son contenu, qui doit recueillir l'aval de la gouvernance, est un processus long. Trouver des définitions qui font consensus,²⁶¹ englober la totalité des disciplines, trancher entre contrainte et incitation,²⁶² font partie des difficultés qui peuvent retarder les choses. Par ailleurs, son insertion dans la chronologie du projet peut varier, comme le rappelle Sarah Jones :

*« Ce qui constitue le meilleur moment pour publier une politique institutionnelle fait l'objet de controverses. Des craintes ont été exprimées sur le fait qu'adopter une politique avant d'avoir développé une infrastructure et des services de soutien pourrait creuser un fossé entre les aspirations et la réalité d'une mise en œuvre effective et adaptée. D'autres considèrent une politique comme un point de départ utile, un moyen de gagner du terrain et de proposer une vision ambitieuse qui motivera et guidera le développement de services. [...] Sans infrastructure une politique peut être difficile à mettre en œuvre, mais sans politique, encourager l'investissement dans le développement d'une infrastructure peut se révéler compliqué »*²⁶³

Faut-il un entrepôt institutionnel ?

Parmi les engagements essentiels pris par les universités qui se saisissent de la question de la gestion des données, il y a en effet celui, comme à Edimbourg ou à Monash, de participer à leur conservation en mettant à la disposition des chercheurs un entrepôt de données. La création d'un tel entrepôt est ainsi inscrit à la feuille de route de l'université de Bath.²⁶⁴ Suivant l'exemple américain, de nombreuses universités britanniques se sont dotées d'un tel entrepôt ou sont sur le point de le faire. Dans la plupart des cas, il s'agit d'adapter l'archive ouverte institutionnelle déjà en place, mais dédiée aux publications, au stockage des

²⁵⁹ PINK Catherine, *op. cit.*, 2013, p. 159.

²⁶⁰ Pour le cas de l'université de Bath, dont la feuille de route est inspirée de celle de l'université Monash, voir *University of Bath Roadmap for EPSRC. Compliance with Research Data Management Expectations*, avril 2012, 17 p., <http://www.bath.ac.uk/rds/assets/pdf/University-of-Bath-Roadmap-for-EPSRC.pdf> (consulté le 11 décembre 2013)

²⁶¹ JONES Sarah, *op. cit.*, 2013, p. 3.

²⁶² PICKTON Miggie & alii, *op. cit.*, 2012, p. 6.

²⁶³ JONES Sarah, PRYOR Graham & WHYTE Angus, « How to Develop Research Data Management Services. A guide for HEIs », Edinburgh, DCC, 2013, http://www.dcc.ac.uk/resources/how-guides/how-develop-rdm-services#_ftn2 (consulté le 11 décembre 2013) ; JONES Sarah, « Bringing it all together : a case study on the improvement of research data management at Monash University », Edinburgh, DCC, 2013, http://www.dcc.ac.uk/resources/developing-rdm-services/improving-rdm-monash#_ftn2 (consulté le 10 décembre 2013)

²⁶⁴ Objectif 7.1 de la feuille de route déjà citée, p. 13, <http://www.bath.ac.uk/rds/assets/pdf/University-of-Bath-Roadmap-for-EPSRC.pdf> (consulté le 12 décembre 2013)

données, sur le modèle de ce qui a été fait à l'université Monash.²⁶⁵ À travers le projet *DataPool* (novembre 2011-mars 2013), financé par le JISC, une réflexion similaire a été menée à l'université de Southampton pour la création d'une infrastructure numérique de données à partir des solutions logicielles SharePoint et EPrints.²⁶⁶ L'université connaît bien ce dernier logiciel d'archive ouverte, utilisée notamment par l'Institut national polytechnique de Toulouse,²⁶⁷ puisqu'elle est à l'origine de sa création. Une communauté d'utilisateurs potentiels ayant déjà mis en place un entrepôt de publications EPrints a été identifiée (en la personne des universités de Leeds et Glasgow) et associée à l'étude menée par Southampton.²⁶⁸ Les universités de Cambridge et d'Edimbourg, quant à elles, ont créé leur entrepôt à partir du logiciel DSpace, là encore l'un des principaux logiciels d'archives ouvertes utilisés dans le monde.²⁶⁹ Il est aussi utilisé par Dryad.

Ces entrepôts de données n'en sont, pour l'essentiel, qu'au stade de l'expérimentation. L'impératif d'interopérabilité et les difficultés inhérentes à l'élaboration de schémas de métadonnées satisfaisants, adaptés à la diversité des disciplines, font partie des obstacles techniques qui inscrivent la structuration et l'évaluation de ces entrepôts dans le temps long. Par ailleurs, se pose la question de l'intérêt de ces entrepôts institutionnels dans un paysage international très morcelé, mais structuré, dans certaines disciplines, autour de bases de données thématiques. Certaines communautés de chercheurs s'inquiètent de l'incapacité des solutions institutionnelles à assimiler l'ensemble des normes disciplinaires :

*« Le fait qu'une seule institution doive avoir la capacité et la possibilité d'archiver la large gamme des types de données produites par ses chercheurs fait partie des inconvénients possibles qu'il y a à faire des universités les principaux responsables de l'archivage et de la diffusion des données. Ce serait une tâche monstrueuse et pénible, et il semble impossible qu'une seule université ait l'expertise et le budget suffisant pour conserver et héberger toutes les données qu'elle produit : il y a beaucoup trop de variables selon les disciplines et les types de données. Une solution plus raisonnable, pour les universités, consiste à demander que les données, selon leur type, soient déposées dans des bases de données internationales de référence, quand c'est possible. [...] Dans le cas où cette base de donnée n'existerait pas, l'entrepôt institutionnel peut provisoirement servir de dépôt. »*²⁷⁰

²⁶⁵ JONES Sarah, « Bringing it all together : a case study on the improvement of research data management at Monash University », Edinburgh, DCC, 2013, http://www.dcc.ac.uk/resources/developing-rdm-services/improving-rdm-monash#_ftn2 (consulté le 10 décembre 2013). Cf. *Monash University Research Repository* : <http://arrow.monash.edu.au/vital/access/manager/Index>.

²⁶⁶ JONES Sarah, PRYOR Graham & WHYTE Angus, *op. cit.*, 2013. Ce projet est très bien documenté, notamment par l'intermédiaire d'un blog (<http://datapool.soton.ac.uk/datapool/>). Cf. également

²⁶⁷ RIGEADE Marine, *Les archives ouvertes institutionnelles en France. État des lieux et perspectives*, mémoire DCB sous la dir. de Sandrine Malotiaux, 2012, 100 p., <http://www.enssib.fr/bibliotheque-numerique/documents/56708-les-archives-ouvertes-institutionnelles-en-france-etat-des-lieux-et-perspectives.pdf> (consulté le 15 octobre 2013)

²⁶⁸ HITCHCOCK Steve & WHITE Wendy, *Towards research data cataloguing at Southampton using Microsoft SharePoint and EPrints : a progress report*, University of Southampton, mai 2013, 15 p., p. 13, <http://eprints.soton.ac.uk/352813/3/eprints-sharepoint-report-final10.pdf> (consulté le 12 décembre 2013)

²⁶⁹ RICE Robin, EKMEKIOGLU Cuna, HAYWOOD Jeff & alii, « Implementing the Research Data Management Policy : University of Edinburgh Roadmap », *The International Journal of Digital Curation*, vol. 2, n° 8, 2013, p. 194-204, <http://www.ijdc.net/index.php/ijdc/article/view/8.2.194/327> (consulté le 10 décembre 2013)

²⁷⁰ LEONELLI Sabina, SMIRNOFF Nicholas, MOORE Jonathan, COOK Charis & BASTOW Ruth, « Making open data work for plant scientists », *Journal of Experimental Botany*, septembre 2013, p. 4114, <http://jxb.oxfordjournals.org/content/early/2013/09/15/jxb.ert273.full> (consulté le 20 novembre 2013)

La solution institutionnelle présente des avantages considérables, déjà relevés dans le cas d'archives ouvertes de publications²⁷¹. Elle offre notamment une garantie de pérennité que des dépôts disciplinaires, qui pourraient n'être plus financés ou arriver à saturation, comme c'est actuellement le cas de certaines collections spécialisées²⁷², n'offrent pas. Il semble cependant pertinent, dans un contexte où l'on privilégie un modèle « distribué » d'archivage des données, que les établissements prennent en compte les solutions autres qu'institutionnelles dans leur stratégie, comme les y invite le *Digital Curation Center* britannique.²⁷³ L'essentiel reste que les solutions (complémentaires) adoptées soient en phase avec la politique définie par l'institution en matière de conservation et de partage des données, puisque certains établissements s'engagent sur ces questions : l'université de Bath, d'ici à la fin de l'année 2013, devrait ainsi avoir publié sa « Stratégie de conservation des données de recherche » afin de satisfaire, entre autres, aux attentes de l'EPSRC. À la demande de cette agence, les institutions doivent en effet préserver, à l'horizon 2015, les données produites sur ses financements au minimum 10 années après publication des premiers résultats qui en sont tirés.²⁷⁴ L'ensemble des politiques institutionnelles imposent des périodes minimales de conservation (trois ans à Oxford),²⁷⁵ tandis que les plans de gestion des données doivent permettre de déterminer les données ayant de la valeur à long-terme et devant donc bénéficier d'une conservation pérenne dans un entrepôt. Dans la poursuite de cet objectif de pérennité, l'université d'Edimbourg, qui s'est engagée à « garantir une continuité d'accès à son patrimoine de données ayant une valeur à long-terme »,²⁷⁶ va tenter d'obtenir pour son entrepôt de données *DataShare* la certification « Data Seal of Approval » (DSA), développée par le *Data Archiving and Networked Services* (DANS)²⁷⁷. Cet institut néerlandais qui a également créé l'archive de données EASY,²⁷⁸ a mis en place une certification qui récompense les entrepôts engagés dans l'archivage numérique à long terme des données répondant à certains critères inspirés de la norme OAIS (*Open archival information system*). Le CINES, qui l'a obtenu en 2010,²⁷⁹ aurait en France un rôle essentiel à jouer pour accompagner les universités souhaitant s'engager dans la conservation à long terme des données produites par leurs chercheurs.

Une autre solution, complémentaire ou non de celle de l'entrepôt institutionnel, est envisagée dans certaines politiques de données et consiste à

²⁷¹ RIGEADE Marine, *op. cit.*, 2012, p. 47.

²⁷² INRA, *op. cit.*, 2012, p. 19.

²⁷³ JONES Sarah, PRYOR Graham & WHYTE Angus, *op. cit.*, 2013.

²⁷⁴ *University of Bath Roadmap for EPSRC. Compliance with Research Data Management Expectations*, avril 2012, 17 p., p. 13, <http://www.bath.ac.uk/rdso/assets/pdf/University-of-Bath-Roadmap-for-EPSRC.pdf> (consulté le 11 décembre 2013)

²⁷⁵ OXFORD UNIVERSITY, *op. cit.*, 201?.

²⁷⁶ Cette feuille de route est présentée, pour l'exemple, en annexe. UNIVERSITY OF EDINBURGH, *Research Data Management Roadmap (August 2012-August 2014)*, v. 1.1 juin 2013, 9 p., http://www.ed.ac.uk/polopoly_fs/1.101179!/fileManager/UoE-RDM-Roadmap130722.pdf (consulté le 12 décembre 2013)

²⁷⁷ RICE Robin, EKMEKCIOGLU Cuna, HAYWOOD Jeff & alii, *op. cit.*, 2013, p. 201 ; RECODE, *Deliverable D1 : Stakeholder Values and Ecosystems*, septembre 2013, 101 p., p. 56, http://recodeproject.eu/wp-content/uploads/2013/10/RECODE_D1-Stakeholder-values-and-ecosystems_Sept2013.pdf (consulté le 23 octobre 2013)

²⁷⁸ GEROUDET Marie-Madeleine, *op. cit.*, 2013, p. 48.

²⁷⁹ « Les services de conservation de données proposés par le TGE Adonis », *Dossier thématique Huma-Num*, février-mars 2013, 14 p., <http://www.huma-num.fr/sites/default/files/ressourcesdoc/la-lettre-fev-mars2013.pdf> (consulté le 12 décembre 2013)

mettre en place un catalogue de métadonnées documentant les données produites au sein de l'université. L'on voit ici que l'enjeu de la gestion des données dépasse largement celui de leur « ouverture », l'ensemble des données étant ici concernées, et même si la mise en place d'un tel outil participe de leur accessibilité. La création d'un tel catalogue est à l'étude à l'université d'Edimbourg²⁸⁰ et est en voie d'aboutir à Oxford, dans le cadre du projet *DaMaRO*. Dans ce dernier cas, le catalogue, baptisé *DataFinder*, est complémentaire de l'entrepôt de données institutionnel, *DataBank*. *DataBank* est censé accueillir « des données qui ne peuvent pas être placées dans un entrepôt thématique plus approprié » mais *DataFinder*, *a contrario*, recensera les références « de l'ensemble des données de recherche produites à l'université, qu'elles soient archivées localement ou en dehors de l'institution ». ²⁸¹ Un schéma de métadonnées générique a été élaboré à partir du schéma proposé par le consortium *DataCite* (dont la dernière version a été publiée en juillet 2013), et peut-être complété par des métadonnées spécifiques à la discipline concernée. ²⁸² Il est également possible d'établir un lien avec une publication et/ou un autre jeu de données, par l'intermédiaire de leur identifiant pérenne DOI notamment. Enfin, et surtout, *DataFinder* pourra moissonner suivant le protocole OAI-PMH les métadonnées associées à des données produites par les chercheurs d'Oxford mais déposées dans des entrepôts extérieurs. ²⁸³

L'on manque encore de recul sur ces politiques et ces infrastructures de données, sur la manière dont les chercheurs se les approprient et sur leur viabilité technique. L'on retiendra, quelles que soient les solutions envisagées, que l'enjeu de l'ouverture est toujours envisagé, dans ces universités, à travers le prisme du *data management*, à l'intersection de la définition d'une politique institutionnelle, de la création d'infrastructures et de la mise en place de services. C'est sur ce dernier point que nous allons maintenant insister, en évoquant le rôle que jouent (et sont amenées à jouer) les bibliothèques dans ce processus. Auparavant, nous aborderons le cas de l'INRA qui, en France, fait partie des institutions pionnières à se saisir de la question de la gestion des données et donne ainsi quelques pistes méthodologiques.

Le cas de l'INRA, développé à de nombreuses reprises par Odile Hologne, directrice de l'IST de cet établissement, prouve que la définition d'une politique de données doit nécessairement impliquer la gouvernance. C'est en effet à l'issue d'une visite, en 2009, au BBSRC, l'une des agences de financement de la recherche britannique, ²⁸⁴ que la présidente de l'institut a demandé au conseil scientifique de se saisir du dossier et de piloter une étude sur la gestion et le partage des données de recherche. Un groupe de travail, piloté par Dominique

²⁸⁰ RICE Robin, EKMEKCIOGLU Cuna, HAYWOOD Jeff & alii, *op. cit.*, 2013, p. 200.

²⁸¹ WILSON James A. J., JEFFREYS Paul W., « Towards a Unified University Infrastructure : The Data Management Roll-Out at the University of Oxford », *The International Journal of Digital Curation*, vol. 2, n° 8, 2013, p. 235-246, p. 238, <http://www.ijdc.net/index.php/ijdc/article/view/8.2.235/331> (consulté le 13 décembre 2013)

²⁸² RUMSEY Sally, JEFFERIES Neil, « Challenges in Building an Institutional Research Data Catalogue », *The International Journal of Digital Curation*, vol. 8, n° 2, 2013, <http://www.ijdc.net/index.php/ijdc/article/view/8.2.205/328> (consulté le 12 décembre 2013) ; JONES Sarah, PRYOR Graham & WHYTE Angus, *op. cit.*, 2013. Pour une description plus technique, RUMSEY Sally, JEFFERIES Neil, « DataFinder : a Research Data Catalogue for Oxford », *Ariadne*, n° 71, juin 2013, <http://www.ariadne.ac.uk/issue71/rumsey-jefferies> (consulté le 13 décembre 2013). Cf. supra (...) sur l'appropriation de ce même schéma de métadonnées par le projet OpenAIREplus.

²⁸³ Pour une description technique complète de *DataFinder*, cf. RUMSEY Sally, JEFFERIES Neil, *op. cit.*, 2013.

²⁸⁴ Cf. supra p. 31.

Pontier, professeur à l'université Lyon 1, a été mis en place et a rendu ses conclusions en juin 2012.²⁸⁵ Parmi ses recommandations : la définition d'une politique institutionnelle, mais également la sensibilisation des chercheurs et la prise en compte du cycle de vie des données dès l'élaboration des projets de recherche. Ce dernier point est essentiel et constitue le fil rouge de la totalité des politiques institutionnelles : si elles hésitent entre incitation et contrainte, toutes intègrent la rédaction, en amont des projets de recherche, de plans de gestion des données (*data management plans*).²⁸⁶ C'est ce qu'a par exemple inscrit l'université de Bath à sa feuille de route.²⁸⁷ Afin de coordonner la mise en œuvre de cette politique, l'INRA a mis en place au printemps 2013 des groupes de travail thématiques mobilisant des compétences scientifiques, informatiques et documentaires (compétences IST)²⁸⁸. Un groupe IST est plus particulièrement chargé de réfléchir à la sensibilisation des chercheurs, d'évaluer la pérennité et la fiabilité d'entrepôts non-institutionnels comme Dryad ou Pangaea,²⁸⁹ d'étudier comment créer un entrepôt ou un annuaire des données de l'INRA. L'INRA s'engage également dans la définition de standards disciplinaires, gages d'interopérabilité des données, en ayant intégré les groupes de travail de la *Research Data Alliance*, initiative internationale lancée en mars 2013 pour favoriser la circulation des données de recherche. Elle est notamment impliquée dans le groupe de travail pour l'interopérabilité des données sur le blé (« Wheat Data Interoperability WG »)²⁹⁰. Ainsi que le note Odile Hologne, le cas de l'INRA « montre la diversité des possibilités d'action des professionnels IST, principalement sur le volet « partage » des données, du politique à l'opérationnel »²⁹¹. C'est cette implication des professionnels de l'information scientifique et technique, et donc des bibliothécaires, que nous allons désormais évoquer.

BIBLIOTHEQUES ET DONNEES : CE QU'ELLES FONT ET COMMENT ELLES DOIVENT S'Y PREPARER

Dans cet écosystème où interagissent agences de financement, gouvernance des établissements de recherche, éditeurs, créateurs d'entrepôts disciplinaires, producteurs de données, quelle est la place des bibliothèques ? La question n'est pas neuve, mais connaît une vraie actualité : le rôle des bibliothèques dans la

²⁸⁵ INRA, *op. cit.*, 2012.

²⁸⁶ Cf. supra p. 38 et suivantes.

²⁸⁷ *University of Bath Roadmap for EPSRC. Compliance with Research Data Management Expectations*, avril 2012, 17 p., p. 6, <http://www.bath.ac.uk/rds/assets/pdf/University-of-Bath-Roadmap-for-EPSRC.pdf> (consulté le 11 décembre 2013). C'est l'objectif 1.6, « Embed research data management planning (DMP) within each research project lifecycle ».

²⁸⁸ AVENTURIER Pascal, « Données ouvertes de la recherche : nouvelles pratiques de publication et de partage », support de cours ENSSIB, novembre 2013, <http://prodirna.inra.fr/?locale=fr#!ConsultNotice:217160> (consulté le 12 décembre 2013)

²⁸⁹ HOLOGNE Odile, « Données de la recherche : rôle des professionnels IST à l'INRA », *Données de la recherche : quel rôle pour la documentation ?*, journée d'étude du congrès annuel de l'ADBU, Le Havre, 19 septembre 2013, <http://prodirna.inra.fr/?locale=fr#!ConsultNotice:207958> (consulté le 12 décembre 2013)

²⁹⁰ <https://rd-alliance.org/working-groups/wheat-data-interoperability-wg.html> (consulté le 12 décembre 2013)

²⁹¹ HOLOGNE Odile, « Professionnels IST et données de la recherche : des discours incantatoires aux actions concrètes », *Documentaliste*, vol. 50, n° 3, octobre 2013, p. 30-31.

gestion des données de recherche a été l'un des principaux thèmes de la dernière conférence annuelle de l'association LIBER, organisée en juin 2013 à Munich, et une journée d'étude était consacrée à ce sujet lors du 43^{ème} congrès de l'ADBU, le 19 septembre 2013, au Havre.²⁹² Ces événements ont permis de constater que de nombreuses bibliothèques étrangères se saisissent de la problématique mais que la France accuse un retard considérable, exception faite d'initiatives pionnières auxquelles participent des professionnels de l'IST, comme à l'INRA ou au Centre de données astronomiques de Strasbourg.²⁹³

L'implication des bibliothèques

L'expérience prouve pourtant que dans la majorité des cas étudiés, au Royaume-Uni par exemple, les bibliothèques sont en première ligne : elles sont étroitement associées à la définition des politiques institutionnelles, participent aux comités de pilotage et groupes de travail et contribuent à défricher le terrain en étant responsables des audits préalables, comme à Bath ou Monash.²⁹⁴ L'association LIBER, qui a publié en juillet 2012 des recommandations sur la manière dont les bibliothèques pouvaient s'impliquer dans la gestion des données de recherche les encouragent à « participer activement au développement de politiques institutionnelles de données ».²⁹⁵ Les autres grandes missions qu'identifie l'association comme relevant des responsabilités des bibliothèques dans la mise en œuvre concrète de ces politiques sont, entre autres, la formation à la gestion des données, la création d'un entrepôt ou d'un catalogue institutionnels et la contribution « technique » à l'ouverture des données (interopérabilité, identification pérenne des données, développement de standards de métadonnées, etc.).

Ces recommandations ont été, dans bon nombre de cas, entendues. Dans de nombreuses institutions américaines, australiennes ou britanniques, les bibliothèques s'affirment comme des structures de sensibilisation et de soutien à la gestion des données de recherche. Au Royaume-Uni, les feuilles de route qui sont actuellement rédigées dans les universités pour organiser la mise en place de politiques institutionnelles de données définissent clairement la responsabilité des bibliothèques à cet égard. À Exeter, le groupe de travail créé dans le cadre du projet *Open Exeter* (octobre 2011-mars 2013) propose, dans ses recommandations, que la bibliothèque soit chargée de superviser les futurs services de formation et

²⁹² <http://adbu.fr/lehavre2013/files/2013/06/Programme-Journée-détude-ADBU.pdf> (consulté le 13 décembre 2013)

²⁹³ GENOVA Françoise, « Le CDS : des données au service de la communauté scientifique », *Documentaliste*, vol. 50, n° 3, octobre 2013, p. 47-49.

²⁹⁴ JONES Kara, « Assessing institutional data storage and management using the Data Asset Framework (DAF) methodology at the University of Bath », University of Bath, mai 2011, 19 p., http://opus.bath.ac.uk/24960/1/DAF_report_May_2011.pdf (consulté le 13 décembre 2013) ; INRA, *op. cit.*, 2012, p. 21.

²⁹⁵ *Ten recommendations for libraries to get started with research data management. Final report of the LIBER working group on E-Science/Data Management*, juillet 2012, 3 p., <http://www.libereurope.eu/sites/default/files/The%20research%20data%20group%202012%20v7%20final.pdf> (consulté le 13 décembre 2013)

d'assistance à la gestion des données.²⁹⁶ C'est aussi le cas à Bath où depuis août 2013 et la clôture du projet *Research 360*, c'est désormais la bibliothèque, à travers son département des services à la recherche (*Research Services*), qui assiste les chercheurs dans l'élaboration de plans de gestion des données.²⁹⁷ La boucle est bouclée, puisque c'est aussi ce département qui avait piloté l'audit sur la gestion des données en 2011. Le cas britannique est, certes, un cas particulier, car les bibliothèques universitaires peuvent compter, dans la structuration de services autour des données de recherche, sur l'aide du *Digital Curation Center*, bras armé du JISC pour la création d'outils communs et l'accompagnement des initiatives institutionnelles. Le DCC fournit en effet aux bibliothèques des solutions « clé en main » pour assister leurs chercheurs, comme *DMPonline*, outil permettant la création en ligne de plans de gestion des données.²⁹⁸ Les bibliothèques n'ont ainsi plus qu'à s'approprier l'outil, former les chercheurs à son utilisation et, pourquoi pas, l'adapter aux besoins spécifiques de leur institution. C'est en effet l'un des grands avantages de *DMPonline*, qui peut être « customisé » par les universités britanniques : elles peuvent notamment définir un « modèle » (*template*) institutionnel de plan de gestion, intégré à *DMPonline* mais adapté à leur politique de données.²⁹⁹ Le logiciel étant *open source*, les universités peuvent aussi l'intégrer à leur système d'information, ou « l'institutionnaliser » en l'adaptant à leur charte graphique, en y greffant leur logo ou en créant leur propre URL (*ex* : dmponline.southampton.ac.uk).³⁰⁰ À Northampton, la bibliothèque a travaillé avec le DCC pour créer un modèle institutionnel de plan de gestion³⁰¹ et celle d'Oxford proposera bientôt aux chercheurs de l'université une version intégrée de *DMPonline*.³⁰²

Qu'elles s'approprient ou non ce type de solutions techniques, les bibliothèques impliquées sur le sujet des données de recherche agissent donc prioritairement comme soutiens dans la rédaction de plans de gestion des données. Même au sein d'institutions pourtant très en avance, comme la Monash University, l'amélioration de ce service, sous la responsabilité des bibliothèques, apparaît encore comme une priorité dans le déploiement des politiques institutionnelles. La directrice de la bibliothèque de Monash est fortement responsabilisée sur ces questions : selon la feuille de route stratégique de l'université,³⁰³ elle doit veiller à ce qu'une méthodologie et des outils pour la planification de la gestion des données soient en place avant 2014. Outre la mise à disposition des chercheurs

²⁹⁶ OPEN EXETER, *Recommendation report for the implementation of research data management policy and for a sustainable research data management service at the University of Exeter*, février 2013, 16 p., https://ore.exeter.ac.uk/repository/bitstream/handle/10871/11682/T%26FRecommendation_report_Final.pdf?sequence=1 (consulté le 15 décembre 2013)

²⁹⁷ <http://www.bath.ac.uk/research/data/policy/index.html> (consulté le 15 décembre 2013)

²⁹⁸ https://dmponline.dcc.ac.uk/about_us (consulté le 15 décembre 2013)

²⁹⁹ ASHLEY Kevin, « DMPonline – current status », mars 2013, <http://www.dcc.ac.uk/blog/dmponline-current-status> (consulté le 15 décembre 2013)

³⁰⁰ DONNELLY Martin, « Bringing it all back home : Tailoring DMPonline for your institution », mai 2012, <http://www.dcc.ac.uk/blog/tailoring-dmp-online-for-your-institution> (consulté le 15 décembre 2013)

³⁰¹ PICKTON Miggie, JONES Sarah, MARIEKE Guy, *op. cit.*, 2012, p. 6.

³⁰² Pour en savoir plus sur ce projet, cf. SHOTTON David, « The Oxford DMPonline Project. Creating customized data management plans », juin 2013, http://damaro.oucs.ox.ac.uk/docs/Shotton-DMPonline_DaMaRO-Workshop_28-06-2013.pdf (consulté le 15 décembre 2013)

³⁰³ « 4.2 Increased focus on data planning », MONASH UNIVERSITY, « Research Data Management Strategy and Strategic Plan 2012-2015 », avril 2012, 17 p., p. 14. Cf. annexe IV, p. 99.

d'un outil dédié, comme *DMPonline* ou son équivalent américain, *DMPTool*,³⁰⁴ cette mission d'assistance peut prendre plusieurs formes : documentation en ligne, modèles de plans de gestion, *checklist*, entretiens personnalisés, etc.

Ce soutien est indissociable de la sensibilisation des chercheurs et du développement de leurs compétences en matière de gestion et de partage des données de recherche, et ce volet « formation » fait généralement partie des attributions des bibliothèques universitaires au sein des dispositifs institutionnels. C'est le cas à l'université Monash,³⁰⁵ à Exeter ou à Bristol, comme dans la plupart des bibliothèques américaines. Cette mission s'inscrit de fait dans le prolongement des services traditionnellement proposés par les BU, en France et à l'étranger, pour la formation des usagers (recherche documentaire, bibliographie, bibliométrie, etc.). La sensibilisation des communautés universitaires commence dès le lancement des projets institutionnels et les équipes mises en place à cette occasion en assument dans un premier temps la responsabilité. À Bristol, par exemple, un « Research Data Service » a été créé dans le cadre du projet *Data.bris* (octobre 2011-avril 2013),³⁰⁶ qui est chargé de gérer l'entrepôt institutionnel et d'« assister les chercheurs sur tous les aspects du *research data management* ».³⁰⁷ Ce service est rattaché à la bibliothèque, piloté par la bibliothèque, mais associe également la DSI (*IT services*) et le département *Research and Enterprise Development* (RED), l'instance universitaire chargée de la valorisation de la recherche. Le service a produit « Research Data Bootcamp », tutoriel d'une trentaine de minutes sur le concept de « données de la recherche » et sur l'importance des bonnes pratiques de gestion.³⁰⁸ Le « Research Data Service » de la bibliothèque de Bristol a recruté en novembre 2013 une « Senior Research Data Librarian » dont la principale mission consiste à développer des outils d'aide à la rédaction de plans de gestion et à mettre en place des formations et des actions de sensibilisation à destination des chercheurs.³⁰⁹ Un dispositif similaire est en place à Exeter où la bibliothèque est désormais dotée d'une « Open Access and Data Curation Team » qui a pris le relais de l'équipe-projet d'*Open Exeter* (octobre 2011-mars 2013) : elle gère l'entrepôt institutionnel et fournit assistance et formations sur les questions de *data management*. Le cas d'Exeter est particulièrement représentatif de l'implication des bibliothèques. L'*Open Access and Data Curation Team* est dirigée par l'ancienne chef de projet *Open Exeter*, laquelle faisait partie du service de « Research Support » de la bibliothèque.³¹⁰ Le comité de pilotage du projet, par ailleurs, était dirigé par la directrice de la bibliothèque.³¹¹ On notera, enfin, que le JISC a apporté son soutien logistique à des projets spécifiquement consacré à la création d'outils de formation : MANTRA, créé en 2011 à l'université

³⁰⁴ <https://dmp.cdlib.org/> (consulté le 15 décembre 2013)

³⁰⁵ « 3.1 Develop the data management skills and knowledge of Monash researchers », MONASH UNIVERSITY, « Research Data Management Strategy and Strategic Plan 2012-2015 », avril 2012, 17 p., p. 13.

³⁰⁶ <http://data.bris.ac.uk/2013/08/20/launch-of-data-bris-pilot-research-data-service/> (consulté le 16 décembre 2013)

³⁰⁷ <http://data.bris.ac.uk/> (consulté le 16 décembre 2013)

³⁰⁸ <http://data.bris.ac.uk/research/bootcamp/> (consulté le 16 décembre 2013)

³⁰⁹ <http://data.bris.ac.uk/2013/11/15/welcome-to-the-team/> (consulté le 16 décembre 2013). Nous avons également pu consulter sa fiche de poste, mise en ligne en août 2013.

³¹⁰ « Open Access and Data Curation Team », http://as.exeter.ac.uk/divisions/crs/ae/open_access_&_data/ (consulté le 16 décembre 2013)

³¹¹ <http://as.exeter.ac.uk/library/resources/openaccess/openexeter/projectinformation/projectgovernance/>

d'Edimbourg à partir du logiciel *open source* Xerte Online Toolkits,³¹² est cependant accessible à tous et est utilisé dans d'autres universités. À Bath, un module de formation en ligne sera développé à partir du logiciel Xerte (favorisant ainsi l'échange de contenus avec MANTRA) et intégré à Moodle, l'ENT de l'université.³¹³

Les bibliothécaires sont-ils les bons interlocuteurs ? Compétences et formations.

L'on a vu que les bibliothèques jouaient un rôle important dans la préparation, la définition et la mise en œuvre des politiques institutionnelles de gestion des données. Mais quelles compétences sont mobilisées ? La bibliothèque est-elle le bon interlocuteur ? Cette implication est-elle transposable en France ? Un article particulièrement intéressant a été publié sur cette question il y a quelques mois,³¹⁴ présentant un tableau de synthèse sur le rôle des bibliothécaires dans la gestion des données et la manière dont il fait écho à certaines de leurs missions « traditionnelles », tout en mobilisant de nouvelles compétences :

³¹² <http://datalib.edina.ac.uk/mantra/> (consulté le 16 décembre 2013). RICE Robin, EKMEKCIOGLU Cuna, HAYWOOD Jeff & alii, *op. cit.*, 2013, p. 202 ; JONES Sarah, PRYOR Graham & WHYTE Angus, *op. cit.*, 2013.

³¹³ PINK Catherine, *Research 360 : JISC Final Report*, University of Bath, juillet 2013, 26 p., http://opus.bath.ac.uk/36362/2/R360_JiscFinalReport_v1.0.pdf (consulté le 16 décembre 2013)

³¹⁴ COX Andrew, VERBAAN Eddy, SEN Barbara, « Upskilling Liaison Librarians for Research Data Management », *Ariadne*, décembre 2012, n° 70, <http://www.ariadne.ac.uk/issue70/cox-et-al> (consulté le 16 décembre 2013)

ROLE	ALIGNMENT WITH EXISTING ROLES	COMPETENCIES REQUIRED
POLICY AND ADVOCACY		
Lead on institutional data policy	Advocacy role e.g. in the area of open access	Strategic understanding and influencing skills
SUPPORT AND TRAINING		
Bring data into undergraduate research-based learning, promoting data information literacy	Information literacy training	Understanding of RDM best practices as they apply to relevant disciplines; pedagogic skills
Teach data literacy to postgraduate students		
Develop researcher data awareness		
Provide an advice service to researchers (and research administrators) Eg on writing Data Management plans or advice on RDM within a project. Advice on licensing data. Advice on data citation. Perhaps measurement of impact of data sharing.	Reference and enquiry roles; producing print and Web-based guides; copyright advice.	Reference interview, knowledge of RDM principles
Provide advice as above through a Web portal	Library Web site	Knowledge of institutional and extra-institutional resources
Signpost who in the institution should be consulted in relation to a particular question	Role of library as point of enquiry and the reference interview	Knowledge of institution
Promote data reuse by making known what is available internally and externally; explaining data citation	Marketing of library resources	Knowledge of researchers' needs, knowledge of available material
AUDITING AND REPOSITORY MANAGEMENT		
Audit to identify data sets for archiving, create a catalogue of materials or to identify RDM needs	Metadata skills	
Develop and manage access to data collections	Collection development, digital library management and metadata management	Audit interviews, knowledge of RDM principles, metadata, licensing
Develop local data curation capacity	Open access role. Preservation role.	Knowledge of RDM principles, relevant technologies and processes, metadata

*Ill. 2. Le rôle des bibliothécaires en matière de gestion des données : compétences requises et compétences actuelles*³¹⁵

Il montre que l'implication des bibliothécaires dans la gestion des données de recherche s'inscrit dans le prolongement naturel de certaines de leurs compétences, par exemple en matière de métadonnées. D'autres auteurs ont aussi récemment insisté sur la légitimité des professionnels de la documentation et de l'information scientifique et technique à intervenir sur le sujet des données de la recherche, comme Odile Hologne, qui met en avant les points d'ancrage du métier comme autant d'arguments justifiant leur implication. Parmi ces points d'ancrage, l'on retiendra la maîtrise des problématiques inhérentes au mouvement du libre-accès,

³¹⁵ *Idem.*

qui fournit « un cadre politique, des outils techniques [et] juridiques » qui préparent les professionnels à l'enjeu du partage des données.³¹⁶ Les compétences développées par les SCD dans leur effort de valorisation de la production scientifique, par l'intermédiaire d'archives ouvertes institutionnelles,³¹⁷ sont réinjectables dans tout projet relatif à la gestion des données de recherche. Les SCD déploient déjà, pour les publications, le même type d'infrastructures que certains entrepôts de données. Interrogée sur la répartition possible des tâches à l'échelle des établissements universitaires français, Charlotte Maday, responsable du Bureau des archives de l'université Paris Diderot et présidente de la section AURORE, tout en insistant sur la nécessité de prendre en compte « la dimension archivistique » des données de recherche,³¹⁸ et en appelant, sur cette question, à une collaboration SCD/archives, rappelait que « les SCD sont bien plus armés que les archivistes sur les questions de conservation de l'électronique », notamment parce qu'ils « disposent de bien plus de moyens ». Magalie Moysan, adjointe du Bureau des archives de cette même université, notait par ailleurs que les chercheurs, *a priori*, se tourneraient « beaucoup plus facilement vers les bibliothèques en raison de leur rôle dans la conservation des publications ».

Par ailleurs, la participation des bibliothécaires à la gestion des données de recherche leur permet de capitaliser sur les connaissances acquises en matière de gestion des métadonnées³¹⁹. Cette connaissance, cependant, est avant tout technique, et générale :

*« Les documentalistes disposent d'un socle de compétences et d'outils acquis dans la gestion des publications qui est transposable à la gestion des données, mais il est important de souligner qu'ils n'ont généralement pas une connaissance intime de la nature des recherches, et des jeux de données associés. Ils ne sont donc pas en mesure de définir les métadonnées correspondantes »*³²⁰

Plusieurs pistes s'offrent alors, que les bibliothécaires sont à mêmes de suivre : contribution à la modélisation ou à l'adaptation d'un schéma générique de métadonnées et aux modalités techniques de son « enrichissement » (selon les disciplines),³²¹ formation à la compréhension de schémas disciplinaires (comme l'*Ecological Metadata Language*, le *Darwin Core*, etc.),³²² collaboration avec les

³¹⁶ HOLOGNE Odile, *op. cit.*, 2013, p. 31.

³¹⁷ RIGEADE Marine, *Les archives ouvertes institutionnelles en France. État des lieux et perspectives*, mémoire DCB sous la dir. de Sandrine Malotiaux, 2012, 100 p., p. 52-55, <http://www.enssib.fr/bibliotheque-numerique/documents/56708-les-archives-ouvertes-institutionnelles-en-france-etat-des-lieux-et-perspectives.pdf> (consulté le 15 octobre 2013)

³¹⁸ Elle l'a encore rappelé dans un récent billet intitulé « Les données de la recherche sont des archives comme les autres », 24 septembre 2013, <http://lotteauxfraises.wordpress.com/2013/09/24/les-donnees-de-la-recherche-sont-des-archives-comme-les-autres-2/> (consulté le 16 septembre 2013)

³¹⁹ RIGEADE Marine, *op. cit.*, p. 54-55 (consulté le 16 septembre 2013)

³²⁰ INRA, *op. cit.*, 2012, p. 38.

³²¹ Cf. ce que nous avons évoqué pour *DataFinder* à Oxford, voir RUMSEY Sally, JEFFERIES Neil, « Challenges in Building an Institutional Research Data Catalogue », *The International Journal of Digital Curation*, vol. 8, n° 2, 2013, p. 205-214, p. 209, <http://www.ijdc.net/index.php/ijdc/article/view/8.2.205/328> (consulté le 12 décembre 2013)

³²² Pour un exemple développé de schéma de métadonnées disciplinaires, on peut se reporter à WIECZOREK John, BLOOM David, GURALNICK Robert & alii, « Darwin Core : An Evolving Community-Developed Biodiversity Data Standard », *PLoS ONE*, 7 (1), janvier 2012,

<http://www.plosone.org/article/info%3Adoi%2F10.1371%2Fjournal.pone.0029715> (consulté le 16 décembre 2013)

chercheurs pour l'écriture des métadonnées. C'est sans doute là l'une des principales difficultés et l'un des principaux enjeux de l'implication des bibliothécaires dans les processus de gestion et de partage des données, les SCD ayant beaucoup de mal à rentrer dans les laboratoires. Du reste :

« Les bibliothécaires manquent généralement d'une expérience directe de la recherche, et peuvent donc avoir du mal à comprendre les motifs et les pratiques des chercheurs. Comprendre la diversité des "données de la recherche" elles-mêmes, selon les cultures disciplinaires et sous-disciplinaires et les pratiques variables de production des données, est [...] indispensable. Les bibliothécaires ont besoin d'adopter la perspective des chercheurs dans leur relation à la gestion des données de recherche. [...] Transposer l'expertise et les pratiques actuelles des bibliothèques, notamment en matière de création de métadonnées, dans le contexte des données de recherche, n'est pas simple. »³²³

L'exemple du Centre de données de Strasbourg (CDS), récemment développé par Françoise Genova, s'il n'est pas directement emprunté au monde des SCD, prouve que le défi repose en effet sur une meilleure collaboration et une meilleure intégration des professionnels de la documentation dans l'univers des producteurs de données.³²⁴ Cette UMR rattachée au CNRS et à l'université de Strasbourg et vouée à la diffusion de données astronomiques publiées associe étroitement, dans son fonctionnement, des chercheurs, des informaticiens et des documentalistes, constituant ensemble une équipe « intégrée ». Le CDS gère plusieurs bases de données, dont la base de référence Simbad pour la nomenclature et la bibliographie des objets astronomiques,³²⁵ qui comptait début juillet 2013 plus de 7 millions d'objets. Elle répertorie tous les noms des objets ainsi que les références des articles dans lesquels ces objets ont été cités. Les documentalistes sont au cœur de cette collecte d'informations et sont chargés de l'ingestion des données dans Simbad : lecture des articles, identification des objets, ajout, le cas échéant, d'un objet dans la nomenclature, etc. Le succès de ces tâches repose sur une forte interaction avec les chercheurs : « les contacts entre documentalistes et astronomes [y] sont constants, avec des réunions et [...] des échanges sur des cas complexes, avec des astronomes "référents" qui répondent plus spécifiquement aux questions de l'un des documentalistes ».³²⁶ Les documentalistes contribuent ainsi à l'élaboration de données « à forte valeur ajoutée », en les mettant en forme et en leur associant des métadonnées :

« Il s'agit d'identifier, de sélectionner, de vérifier, d'homogénéiser les données, et de créer des métadonnées permettant qu'elles soient réutilisées. C'est un métier qui a un très fort impact sur la recherche : la possibilité d'accéder à des données bien décrites, sous forme numérique, permet de les exploiter facilement. »³²⁷

L'INRA aspire à développer de telles collaborations, en intégrant plus directement les professionnels de l'IST au sein des équipes de recherche.³²⁸ Dans

³²³ COX Andrew, VERBAAN Eddy, SEN Barbara, *op. cit.*, 2012.

³²⁴ GENOVA Françoise, « Le CDS : des données au service de la communauté scientifique », *Documentaliste*, vol. 50, n° 3, octobre 2013, p. 47-49.

³²⁵ <http://simbad.u-strasbg.fr/simbad/> (consulté le 17 décembre 2013)

³²⁶ GENOVA Françoise, *op. cit.*, 2013, p. 49.

³²⁷ *Idem.*

³²⁸ INRA, *op. cit.*, 2012, p. 38-39.

tous les cas, il apparaît clairement qu'« une implication forte des chercheurs, producteurs de données, [...] est nécessaire », en particulier « dans tous les processus relatifs à la construction des métadonnées »³²⁹.

Si la contribution des bibliothécaires à la gestion des données peut donc être perçue comme un prolongement de compétences-métier déjà bien affirmées, notamment en ce qui concerne l'indexation et la « curation » de métadonnées,³³⁰ elle dépend encore de conditions qui ne sont pas à l'heure actuelle totalement réunies. D'abord parce que, comme nous l'a écrit la directrice du SCD d'une grande université parisienne, sur ce sujet, « nous ne pouvons nous passer de la collaboration des chercheurs », et que ces derniers sont encore dans la majorité des cas très peu sensibles au travail fourni par les SCD pour la valorisation de leurs publications par l'intermédiaire d'une archive ouverte institutionnelle. Ensuite, parce que les bibliothécaires ne sont pas suffisamment au fait des problématiques de gestion et de partage des données, que leur connaissance du milieu de la recherche est souvent limitée et que leur formation ne les prépare pas à ce défi.³³¹ Même au Royaume-Uni où, nous l'avons vu, les institutions universitaires sont en pleine mutation, intégrant la question des données à leurs politiques d'établissement, la question de la formation des bibliothécaires reste un enjeu décisif. Des outils particulièrement innovants ont été mis en place comme, au début de l'année 2013, *Immersive Informatics*, un programme pilote anglo-australien élaboré par les universités de Melbourne et de Bath³³² et dont on devrait connaître le bilan dans les mois qui viennent. Cette formation comprend des ateliers thématiques, mais également des « immersions » en laboratoire, chaque participant étant appelé à adopter un jeu de données qu'il apprendra à gérer en vue de sa conservation et de sa diffusion futures.³³³

³²⁹ *Idem.*

³³⁰ HOLOGNE Odile, *op. cit.*, 2013, p. 31.

³³¹ DELHAYE Marlène, « Liber 2013 », septembre 2013, <http://marlenescorner.net/2013/09/01/liber-2013/> (consulté le 17 décembre 2013)

³³² <http://immersivinformatics.org/> (consulté le 17 décembre 2013)

³³³ <http://immersivinformatics.org/index.php/pilot> (consulté le 17 décembre 2013)

BILAN : QUI FAUT-IL MOBILISER, EN FRANCE, POUR PARTICIPER A L'ELABORATION DE POLITIQUES DE DONNEES ?

À l'échelle des établissements, le levier de la gouvernance et de la coordination

On l'a vu dans le cas des universités britanniques, on l'a dit dans le cas de l'INRA, l'implication de la gouvernance est absolument nécessaire à la réussite des politiques institutionnelles de données. La Ligue européenne des universités de recherche (LERU), vient tout juste de publier sa *Feuille de route pour les données de recherche*,³³⁴ à la rédaction de laquelle a été associé, pour la France, Serge Fdida, professeur d'informatique à l'UPMC, ancien vice-président en charge des relations internationales et désormais vice-président Europe de cette université. La publication de ce document et l'implication d'un représentant d'une des plus importantes universités françaises dans sa préparation, constituera peut-être un tournant dans le développement de politiques institutionnelles de données. Parmi ses recommandations, la LERU invite chacun de ses membres à mettre en place un comité de pilotage sur les données de recherche (« RDM Steering Group ») et à définir, si ce n'est pas déjà fait, une politique institutionnelle et une feuille de route stratégique.³³⁵

Lors de la journée d'étude du 19 septembre au congrès annuel de l'ADBU 2013, Grégory Colcanap, coordinateur du consortium Couperin, a insisté sur la nécessité que les universités se mobilisent, afin que les organismes de recherche, CNRS et INRA en tête, ne soient pas les seuls à intégrer à leur politique d'établissement la gestion des données. Rien ne pourra se faire sans investissement de la gouvernance. Les conseils scientifiques, les vice-présidents Recherche et la Conférence des présidents d'université (CPU), à travers notamment sa « Commission de la recherche et de l'innovation », font partie des acteurs et des instances qui, dans les mois qui viennent, pourraient créer une dynamique.

La possibilité de confier, à l'échelle des établissements, le travail de « défrichage » à un coordinateur, chargé de mission réunissant autour de lui l'ensemble des acteurs concernés par l'élaboration d'une politique de données (chercheurs, bibliothèque, gouvernance, DSI, etc.) semble ensuite particulièrement séduisante. C'est la démarche qui a été adoptée à l'université de Bath, où a été créé au début du projet *Research 360* un poste de « Data Scientist » autonome servant

³³⁴ LERU RESEARCH DATA WORKING GROUP, *LERU Roadmap for Research Data*, 18 décembre 2013, 34 p., http://www.leru.org/files/publications/AP14_LERU_Roadmap_for_Research_data_final.pdf (consulté le 20 décembre 2013)

³³⁵ *Idem*, p. 31.

d'interface entre les facultés, la bibliothèque, la DSI et la Direction de la recherche (et notamment le service de valorisation).³³⁶ Cette fonction de médiation, tout au long du projet *Research 360*, s'est appuyée sur une solide expérience de la recherche : Catherine Pink, la « Data Scientist » en question, était spécialiste en génétique de l'évolution et avait travaillé quatre ans dans la recherche privée au développement de fongicides. Cette expérience s'est révélée décisive pour convaincre et mobiliser les chercheurs.³³⁷ Si ce scénario n'est pas facilement transposable, la mise en place d'un comité de pilotage associant le SCD, la DSI, la Direction de la recherche et certaines unités de recherche pilotes constituerait, à l'échelle des établissements, une solution efficace.

Capitaliser sur l'expérience acquise : le CINES

Ainsi que l'a récemment rappelé Jean-Pierre Theron, membre du projet ISAAC au Centre informatique national de l'enseignement supérieur (CINES), la position du CINES comme l'un des plus grands centres de calcul intensif français et européen a naturellement contribué à placer l'archivage électronique des données scientifiques au cœur de ses préoccupations.³³⁸ Le CINES a donc mis en place, dans le cadre du projet ISAAC (« Information scientifique archivée au Cines »), un service d'archivage *intermédiaire* des données scientifiques à l'attention de la communauté des chercheurs. Le CINES a profité de sa participation au projet européen EUDAT,³³⁹ lancé en 2011 et dédié à la construction d'une infrastructure européenne de préservation des données numériques, et d'une enquête réalisée par ses soins auprès de 150 laboratoires de recherche, pour structurer ce service encore en phase de développement. Il comporte un important volet « conservation », avec la préservation à moyen terme (3 à 5 ans) des données versées et la possibilité de les constituer en archive définitive.³⁴⁰ Surtout, il répond au besoin des communautés scientifiques en matière de description des données :

« Les communautés dans lesquelles se regroupent les chercheurs sont très différentes, les laboratoires sont très spécialisés dans leur domaine d'activités et la description des données n'est pas forcément leur priorité. [...] L'objectif des métadonnées est d'ajouter un niveau descriptif suffisamment pertinent pour permettre l'exploitation et le

³³⁶ PINK Catherine, *Research 360 : JISC Final Report*, University of Bath, juillet 2013, 26 p., p. 9, http://opus.bath.ac.uk/36362/2/R360_JiscFinalReport_v1.0.pdf (consulté le 16 décembre 2013)

³³⁷ PINK Catherine, « Meeting the Data Management Compliance Challenge : Funder Expectations and Institutional Reality », *The International Journal of Digital Curation*, vol. 2, n° 8, 2013, p. 157-171, p. 162-163, <http://www.ijdc.net/index.php/ijdc/article/viewFile/8.2.157/325> (consulté le 10 décembre 2013)

³³⁸ THERON Jean-Pierre, « L'archivage des données scientifiques : ISAAC, une solution du CINES », *La Gazette du Cines. Archivage numérique pérenne*, CINES, février 2013, 59 p., p. 46-48, http://www.cines.fr/IMG/pdf/CINES_GAZETTE_Special_Archivage-2.pdf (consulté le 23 octobre 2013)

³³⁹ LECARPENTIER Damien & alii, « EUDAT : A New Cross-Disciplinary Data Infrastructure for Science », *The International Journal of Digital Curation*, vol. 8, n° 1, 2013, 9 p., <http://www.ijdc.net/index.php/ijdc/article/view/8.1.279/312> (consulté le 23 octobre 2013)

³⁴⁰ « Archivage intermédiaire des données scientifiques », <http://www.cines.fr/spip.php?rubrique273> (consulté le 20 décembre 2013)

partage des données dans les meilleures conditions. [Pourtant], la plupart des laboratoires ne possèdent pas de standard concernant les jeux de métadonnées »³⁴¹

Le service ISAAC accompagnera donc les chercheurs dans la création de métadonnées descriptives, adaptées aux données concernées, afin d'en faciliter la compréhension et le partage au sein d'un groupe d'utilisateurs autorisés. D'un point de vue organisationnel, ISAAC repose sur la constitution de « comités thématiques d'archivage » (CTA) spécialisés, associant un référent CINES à un groupe d'experts dans la discipline, chargés de définir les modalités du dépôt (format des données et métadonnées).³⁴²

Ce service, qui ouvrira prochainement, ne présente pas de solution pour l'« ouverture » des données, mais explore les conditions techniques indispensables à leur conservation et à leur partage (dans un cercle restreint) à moyen terme. Il confirme que les solutions *ad hoc* doivent être élaborées au plus près des communautés de chercheurs et que la documentation des données, par l'intermédiaire de descripteurs idoines, est le terrain principal qu'il faut investir pour épauler les laboratoires face au défi de la gestion des données de recherche. Il prouve également que le CINES, dans le paysage français, doit être un interlocuteur essentiel sur toutes les questions relatives au stockage et à l'archivage des données. À ce titre, il a vocation à être associé à la mise en place du segment BSN 10.³⁴³

Au-delà du projet ISAAC, le CINES a en effet déjà démontré son efficacité dans le cadre du « Très Grand Équipement » (TGE) Adonis, qui a fusionné en mars 2013 avec Corpus-IR pour former la « Très Grande Infrastructure de Recherche » (TGIR) Huma-Num.³⁴⁴ Lancé en 2004 par le CNRS afin de répondre à « la croissance de la production de données », à « leur diversité » et à « la fragilité des dispositifs supports » dans le secteur des SHS,³⁴⁵ le TGE Adonis s'est consolidé, notamment à partir de 2007, comme « un ensemble de services pour le traitement, le stockage, la diffusion et la conservation de données numériques » à l'attention des chercheurs en SHS.³⁴⁶ Huma-Num, portée par l'UMS 3598 placée sous la tutelle du CNRS, d'Aix-Marseille Université et du Campus Condorcet, est ainsi la grande infrastructure pour les données de recherche française.³⁴⁷ Le dispositif associe trois composantes : une grille de services d'aide à la gestion des données

³⁴¹ « Focus sur les données scientifiques et leur problématique d'archivage », <http://www.cines.fr/spip.php?rubrique383> (consulté le 20 décembre 2013)

³⁴² « Le projet ISAAC », <http://www.cines.fr/spip.php?rubrique369> (consulté le 20 décembre 2013)

³⁴³ Cf. infra « Une impulsion nationale ? », p. 76.

³⁴⁴ « Human-Num, le TGIR des humanités numériques », <http://www.huma-num.fr/> (consulté le 20 décembre 2013)

³⁴⁵ THIBAUT François & JOUVE Bertrand, *Les Infrastructures de recherche en sciences humaines et sociales. Rapport du groupe "Infrastructures" de l'Alliance Athéna* [Alliance nationale des sciences humaines et sociales], septembre 2012, 40 p., p. 14-15, <http://www.cnrs.fr/fr/pdf/allianceathena/121012-infrastructure-sh/> (consulté le 3 octobre 2013) ; <http://www.cnrs.fr/inshs/recherche/TGE-ADONIS.php> (consulté le 3 octobre 2013)

³⁴⁶ TGE ADONIS, « Les services de conservation de données proposés par le TGE Adonis », *La Lettre de la Grille Adonis*, n° 5, février-mars 2013, 14 p., <http://www.huma-num.fr/sites/default/files/ressourcesdoc/la-lettre-fev-mars2013.pdf>

³⁴⁷ Elle a été présentée par son directeur, Marc Renneville, lors de la journée d'étude du 19 septembre 2013 au congrès de l'ADBU.

(traitement et stockage, en partenariat avec le centre de calcul de l'IN2P3),³⁴⁸ un point d'accès unique pour la diffusion et la valorisation des données, la plateforme *Isidore*,³⁴⁹ et un service d'archivage à long terme des données pérennes, associant le CINES et le Service interministériel des Archives de France (SIAF). Le principal support de cette conservation pérenne est le service d'archivage PAC (« Plateforme d'Archivage au Cines »), bien connu de certains SCD et SICD comme la BUPMC ou la BIUS.³⁵⁰ Dans le cadre du TGE Adonis, un projet pilote a été mené pour l'archivage des données orales du CRDO (Centre de ressources pour la description de l'oral), dont l'antenne d'Aix est devenue en 2011 le *Speech and Language Data Repository* (SLDR, « Banque de données parole et langage »).³⁵¹ Le CINES assure l'archivage numérique pérenne des données du SLDR selon le modèle OAIS.³⁵² Il vient de renouveler avec la TGIR Huma-Num, pour une période de cinq ans (2013-2018), la convention qui le liait au TGE Adonis. La communauté des producteurs de données en SHS, par l'intermédiaire d'Huma-Num, pourra donc bénéficier des compétences du CINES pour l'archivage pérenne de ses données.³⁵³

Une impulsion nationale ?

Les contours et les orientations d'une politique nationale des données de recherche demeurent, pour l'instant, la principale inconnue du problème. Alain Colas, responsable de la Mission de l'information scientifique et technique et du réseau documentaire (MISTRD) au Ministère de l'Enseignement supérieur et de la Recherche (MESR), a rappelé lors des journées FRÉDoc 2013 que le 10^{ème} segment du projet « Bibliothèque Scientifique Numérique », consacré aux données de recherche, était sur le point d'être mis en place.³⁵⁴ La création de BSN 10 avait été annoncée lors des journées *Open Access* organisées par Couperin, mais son lancement a été retardé à plusieurs reprises.³⁵⁵ Aucun groupe de travail n'ayant encore été mis en place, difficile de savoir quel sera le contenu de ce segment, et ce qu'il supposera en termes d'infrastructure. L'ambition de la MISTRD, semble-t-il, est d'associer étroitement à l'élaboration de ce segment les chercheurs, en en confiant le pilotage à « un scientifique reconnu et ayant des responsabilités

³⁴⁸ TGE ADONIS, *op. cit.*, 2013.

³⁴⁹ <http://www.rechercheisidore.fr/> (consulté le 20 décembre 2013)

³⁵⁰ Cf. l'ensemble du numéro spécial de *La Gazette du CINES* consacré à « L'archivage numérique pérenne », février 2013, 59 p., http://www.cines.fr/IMG/pdf/CINES_GAZETTE_Special_Archivage-3-pdfA3b-2.pdf (consulté le 20 décembre 2013)

³⁵¹ <http://crdo.up.univ-aix.fr/> (consulté le 20 décembre 2013)

³⁵² BEL Bernard, « Mutualisation et archivage pérenne des données orales : un nouveau cadre technique et juridique au service de la recherche en linguistique », *Les archives de la recherche : problèmes et enjeux de la construction du savoir scientifique, 18-20 janvier 2012, Université Paris-Sud 11 & CNAM, Paris, Paris, 2012*, 16 p., <http://aune.lpl.univ-aix.fr/~fulltext/4871.pdf> (consulté le 1^{er} octobre 2013)

³⁵³ Pour une présentation exhaustive : TGE ADONIS, *op. cit.*, 2013.

³⁵⁴ Au 18 décembre 2013, ce segment n'apparaît toujours pas sur le portail BSN : <http://www.bibliothequescientifiquenumerique.fr/> (consulté le 18 décembre 2013). COLAS Alain, « Quelle politique nationale pour les données de la recherche ? », *Journées FRÉDOC 2013, Problématiques et enjeux autour des données de la recherche, 7-10 octobre 2013, Aussois*, http://renatis.cnrs.fr/IMG/pdf/Colas_FREDOC_2013_Presentation.pdf (consulté le 18 décembre 2013)

³⁵⁵ GEROUDET Marie-Madeleine, *op. cit.*, 2013, p. 37.

politiques », ³⁵⁶ et la gouvernance des universités, par l'intermédiaire de la Conférence des présidents d'université (CPU). ³⁵⁷ La mobilisation de ces acteurs est, en effet, indispensable. Le CNRS devrait participer au déploiement de ce segment qui, dans un premier temps, aura pour objectif d'esquisser les contours d'une future stratégie nationale.

Il ne semble pas pertinent de créer une infrastructure nationale de stockage et de diffusion (une sorte de « HAL » des données), et il conviendrait donc plutôt de miser sur une solution distribuée. L'hypothèse la plus viable pourrait être la création d'un point d'accès unique, national, à ces données distribuées, mais se pose alors la question de sa visibilité, et de son rapport à d'autres infrastructures, comme OpenAIREplus. ³⁵⁸ S'il y a architecture distribuée, dans tous les cas, il sera nécessaire de déterminer de quelles manières les établissements de recherche et les universités y participent.

Par ailleurs, parallèlement à la définition d'une politique nationale des données de recherche, il est nécessaire, comme l'ont récemment relevé des chercheurs de *Deuxième labo* dans une tribune publiée dans *Le Monde*, ³⁵⁹ que l'État, dans son engagement en faveur de l'*Open data*, se positionne sur l'ouverture des données sur la recherche. En effet, alors que l'Agence nationale de la recherche (ANR) est le principal interlocuteur pour le financement de la recherche sur projets, ³⁶⁰ les données sur les projets financés par l'agence manquent pour l'instant totalement de visibilité. ³⁶¹ L'ouverture de ces données, sous l'impulsion de la mission Etalab, ³⁶² constituerait pourtant un progrès essentiel qui rejoint, par bien des aspects, les justifications apportées au mouvement de l'*Open research data*. À terme, pourquoi ne pas rêver d'une base de données nationale, sur le modèle d'OpenAIREplus, associant informations sur les projets de recherche (laboratoires, montants, chercheurs, durée, etc.), données et résultats de recherche correspondants ?

Comme le notait le rapport de l'INRA, ³⁶³ le retard de la France sur ces thématiques est en partie lié à l'absence d'« une instance de coordination forte », ³⁶⁴ au service des établissements de recherche et dotée de moyens importants, comme le *Digital Curation Center* britannique. Pourtant, comme au Royaume-Uni, le

³⁵⁶ Intervention d'Alain Colas lors de la journée d'étude « Données de la recherche : quel rôle pour la documentation ? », organisée lors du congrès annuel de l'ADBU au Havre, le 19 septembre 2013.

³⁵⁷ Alain Colas nous l'a confirmé de vive voix.

³⁵⁸ Cf. supra p. 54-55.

³⁵⁹ BLANCHARD Antoine, SABUNCU Elifsu, « Pour un accès aux données sur la recherche française », *Le Monde*, 2 décembre 2013, http://lemonde.fr/sciences/article/2013/12/02/pour-un-acces-aux-donnees-sur-la-recherche-francaise_3523991_1650684.html (consulté le 18 décembre 2013).

³⁶⁰ Sur le financement de la recherche publique française, on se reportera au rapport de la Cour des comptes, *Le Financement public de la recherche, un enjeu national*, juin 2013, 283 p., <http://www.ccomptes.fr/Publications/Publications/Le-financement-public-de-la-recherche-un-enjeu-national> (consulté le 18 décembre 2013)

³⁶¹ L'ANR ne fait pas partie des nouveaux producteurs de données ajoutés à la dernière version du portail *Data.gouv.fr*, lancée en décembre 2013.

³⁶² CARTIER Aurore, *op. cit.*, 2013, p. 26-27.

³⁶³ INRA, *op. cit.*, 2012, p. 48.

³⁶⁴ GEROUDET Marie-Madeleine, *op. cit.*, 2013, p. 37.

volontarisme est la clé. Au-delà de BSN 10, conditionner en France toute mise à disposition de fonds publics à la rédaction, en amont et au cours des projets de recherche, d'un plan de gestion des données, comme le fait la Commission européenne, constituerait une avancée décisive. L'ANR vient de renouveler son soutien aux archives ouvertes,³⁶⁵ elle pourrait désormais s'engager, avec ce type de disposition, en faveur d'une gestion réfléchie des données de recherche.

³⁶⁵ L'ANR a signé cette convention le 2 avril 2013 : http://www.bibliothequescientifiquenumerique.fr/IMG/pdf/Convention_HAL_AO_26-03-2013.pdf (consulté le 18 décembre 2013)

CONCLUSION

Comme l'a récemment rappelé Ghislaine Chartron,³⁶⁶ les métiers de l'information sont en pleine évolution, et la figure du *data librarian*, spécialisé dans la gestion et la documentation des données, s'impose progressivement dans le paysage bibliothéconomique anglo-saxon. Pierre Carbone, dans un rapport de mars 2013 de l'IGB, a également eu l'occasion de rappeler que les bibliothèques étaient entrées « dans une décennie de mutation radicale » avec, notamment, l'épanouissement de professionnels « au service de la formation et de la recherche » :

« Il convient d'observer que, même si le phénomène est encore marginal dans les BU, on assiste à l'émergence de nouvelles compétences, notamment autour de la bibliométrie, de la valorisation des archives ouvertes ou de la gestion des données de la recherche, compétences beaucoup plus développées à l'heure actuelle dans les services d'information scientifique et technique des EPST »³⁶⁷

L'INRA et le CNRS, aiguillonnés par leur direction de l'information scientifique et technique (DIST), ont en effet pris les choses en main,³⁶⁸ mais les universités françaises restent pour l'instant à l'écart de la dynamique d'ouverture. Pourtant, l'annonce par la Commission Européenne, le 16 décembre 2013, que certains chercheurs bénéficiant de crédits européens dans le cadre du 8^{ème} PCRD devront obligatoirement rédiger des plans de gestion des données, pourrait avoir des conséquences décisives. Ces plans décriront « quelles données un projet de recherche générera, si et comment elles seront exploitées ou rendues accessibles pour être vérifiées et réutilisées, comment elles seront traitées (*curated*) et préservées ». ³⁶⁹ La rédaction systématique de plans de gestion des données, que le CNRS s'apprête également à plébisciter,³⁷⁰ constitue bien un levier essentiel : la mise en place de politiques de partage et d'ouverture repose avant tout sur la bonne planification de leur gestion, tout au long de leur cycle de vie.³⁷¹

Le cas britannique le prouve : c'est aussi à l'échelle des établissements universitaires que doit se préparer l'ouverture des données. Et si c'est d'abord à la gouvernance de se saisir de cette question, les SCD ont un rôle majeur à jouer. Le sujet des données de recherche, dans les années qui viennent, peut constituer une opportunité de taille pour consolider les liens, trop souvent distendus, entre

³⁶⁶ CHARTRON Ghislaine, « Changement de périmètre pour l'IST », *Journée "Documentation Mathématique"*, 21 mai 2013, CIRM, <http://www.rnbn.org/cirm2013/perimetre-IST.pdf> (consulté le 15 novembre 2013)

³⁶⁷ CARBONE Pierre & alii, *Quels emplois dans les bibliothèques ? État des lieux et perspectives*, Inspection générale des bibliothèques, rapport n° 2012-020, mars 2013, 82 p., http://cache.media.enseignementsup-recherche.gouv.fr/file/2013/20/3/Quels_emplois_dans_les_bibliotheques_Etat_des_lieux_et_perspectives_247203.pdf (consulté le 15 octobre 2013)

³⁶⁸ Cf. INRA, *op. cit.*, 2012 ; CNRS, *op. cit.*, 2013.

³⁶⁹ COMMISSION EUROPEENNE, *Guidelines on Data Management in Horizon 2020*, v. 1.0, 16 décembre 2013, 6 p., http://ec.europa.eu/research/participants/data/ref/h2020/grants_manual/hi/oa_pilot/h2020-hi-oa-data-mgt_en.pdf (consulté le 20 décembre 2013)

³⁷⁰ CNRS, *op. cit.*, 2013, p. 37.

³⁷¹ GUINDON Alex, « La gestion des données de recherche en bibliothèque universitaire », *Documentation et bibliothèques*, vol. 59, n° 4, octobre-décembre 2013, p. 189-200.

chercheurs et bibliothécaires. Des actions peuvent être d'ores-et-déjà menées en ce sens par les professionnels de la documentation, comme la réalisation d'enquêtes auprès de laboratoires pilotes afin de comprendre les modalités de production et de gestion des données, ou la mise en place de groupes de travail collaboratifs sur les pratiques existantes en matière de création de métadonnées. L'enjeu principal, pour les bibliothécaires, consiste à « comprendre les méthodes et les besoins des chercheurs » :

*« Avec la question de la documentation électronique, les bibliothèques se sont positionnées sur des questions techniques, administratives et juridiques liées à l'acquisition et à la mise en œuvre des accès ; elles ne sont pas rentrées dans les contenus. Concernant les données de la recherche en revanche, les questions d'acquisition, d'accès et d'exploitation sont indissociables des impératifs scientifiques et ne peuvent être envisagés qu'avec la communauté des chercheurs. Le cadre de l'établissement n'est pas forcément pertinent, pas plus que les standards documentaires traditionnels. **Entre recherche et documentation, il y a un espace à investir pour les bibliothécaires et les chercheurs qui ont envie de travailler ensemble à la diffusion des données** ».*³⁷²

Si ce rapprochement peut encore sembler utopique, il n'y a pas de raison, comme on a pu le voir avec l'exemple du Centre de données astronomiques de Strasbourg,³⁷³ de le considérer comme impossible à moyen terme. Il reposera sur un double changement de culture : celle des bibliothécaires, dont la formation doit permettre de les sensibiliser à cet enjeu, celle des chercheurs, dont les carrières doivent être plus directement conditionnées par le respect de bonnes pratiques de gestion des données.³⁷⁴ Espérons que le vœu formulé par Robert Genet, directeur général pour la recherche et l'innovation (DGRI) au MESR, de « mobiliser, grâce à BSN 10, les chercheurs, de susciter leur désir et de modifier leurs pratiques actuelles »,³⁷⁵ se réalise. En attendant, les SCD doivent « pousser la porte des laboratoires »³⁷⁶ et commencer à défricher, à l'échelle des établissements, le gigantesque terrain des données de recherche.

³⁷² FAYET Sylvie, « « Données » de la recherche, les mal-nommées », 15 novembre 2013, <http://urfistinfo.hypotheses.org/2581> (consulté le 4 décembre 2013)

³⁷³ Cf. supra p. 71-72. GENOVA Françoise, « Le CDS : des données au service de la communauté scientifique », *Documentaliste*, vol. 50, n° 3, octobre 2013, p. 47-49.

³⁷⁴ Grégory Miura, directeur adjoint du SCD de l'université Bordeaux III, a comparé cette nécessaire acculturation des chercheurs à la gestion des données à l'obligation de dépôt électronique des thèses, lors de la journée d'étude du 19 septembre au Havre.

³⁷⁵ Intervention de Robert Genet lors de la journée d'étude « Données de la recherche : quel rôle pour la documentation ? », organisée lors du congrès annuel de l'ADBU au Havre, le 19 septembre 2013.

³⁷⁶ Pour reprendre l'expression de Stéphane Pouyllau, directeur-adjoint technique d'Huma-Num, lors des journées FRÉDOC 2013 : <http://www.donneesdelarecherche.fr/spip.php?article402> (consulté le 1^{er} janvier 2014)

Bibliographie

GENERALITES

AVENTURIER Pascal, « Données ouvertes de la recherche : nouvelles pratiques de publication et de partage », support de cours ENSSIB (M2 PN-AN), novembre 2013, <http://prodinra.inra.fr/?locale=fr#!ConsultNotice:217160> (consulté le 12 décembre 2013)

BLANCHARD Antoine, SABUNCU Elifsu & STROPPA Yvan, « Pour une recherche reproductible, publiez vos codes et données », *Le Monde*, 15 juillet 2013, http://lemonde.fr/sciences/article/2013/07/15/pour-une-recherche-reproductible-publiez-vos-codes-et-donnees_3447825_1650684.html (consulté le 30 décembre 2013)

BLANCHARD Antoine, SABUNCU Elifsu, « Pour un accès aux données sur la recherche française », *Le Monde*, 2 décembre 2013, http://lemonde.fr/sciences/article/2013/12/02/pour-un-acces-aux-donnees-sur-la-recherche-francaise_3523991_1650684.html (consulté le 18 décembre 2013)

BORGMAN L. Christine, « Research Data : Who will share what, with whom, when and why ? », *Fifth China-North America Library Conference 2010, 8-12 septembre 2010, Beijing*, 21 p.,

<http://works.bepress.com/cgi/viewcontent.cgi?article=1237&context=borgman>
(consulté le 9 décembre 2013)

CARBONE Pierre & alii, *Quels emplois dans les bibliothèques ? État des lieux et perspectives*, Inspection générale des bibliothèques, rapport n° 2012-020, mars 2013, 82 p.,

http://cache.media.enseignementsup-recherche.gouv.fr/file/2013/20/3/Quels_emplois_dans_les_bibliotheques_Etat_des_lieux_et_perspectives_247203.pdf (consulté le 15 octobre 2013)

CARTIER Aurore, *Bibliothèques et Open data. Et si on ouvrait les bibliothèques sur l'avenir ?*, mémoire DCB sous la dir. de Frédérique Schlosser, 2013, 107 p., <http://www.enssib.fr/bibliotheque-numerique/documents/60401-bibliotheque-et-open-data-et-si-on-ouvrait-les-bibliotheques-sur-l-avenir.pdf> (consulté le 15 octobre 2013)

CHARTRON Ghislaine, « La valeur des services documentaires en prise avec le numérique », *BBF*, vol. 57, n° 5, 2012, p. 14-18, <http://bbf.enssib.fr/consulter/bbf-2012-05-0014-003> (consulté le 12 novembre 2013)

CHARTRON Ghislaine, « Changement de périmètre pour l'IST », *Journée 'Documentation Mathématique'*, 21 mai 2013, *CIRM*, <http://www.rnbnm.org/cirm2013/perimetre-IST.pdf> (consulté le 15 novembre 2013)

« Compte-rendu des journées FRéDOC 2013 », novembre 2013, <http://www.donneesdelarecherche.fr/spip.php?article402> (consulté le 31 décembre 2013)

COUTURE Marc, DUBE Marcel & MALISSARD Pierrick, *Propriété intellectuelle et université. Entre la libre circulation des idées et la privatisation des savoirs*, 2010, Presses de l'université du Québec, 384 p.

DOORN Peter, DILLO Ingrid, VAN HORIK René, « Lies, Damned Lies and Research Data : Can Data Sharing Prevent Data Fraud ? », *The International Journal of Digital Curation*, vol. 8, n° 1, 2013, p. 229-243, <http://www.ijdc.net/index.php/ijdc/article/view/8.1.229/308> (consulté le 31 décembre 2013)

FAYET Sylvie, « « Données » de la recherche, les mal-nommées », 15 novembre 2013, <http://urfistinfo.hypotheses.org/2581> (consulté le 4 décembre 2013)

HODSON Simon, « Les politiques d'accompagnement des données : une discussion du paysage », *Journées FRéDOC 2013, Problématiques et enjeux autour des données de la recherche, 7-10 octobre 2013, Aussois*, http://renatis.cnrs.fr/IMG/pdf/Hodson_FReDOC_Presentation_FR.pdf (consulté le 14 novembre 2013)

LACOMBE Romain, BERTIN Pierre-Henri, VAUGLIN François & VIEILLEFOSSE Alice, *Pour une politique ambitieuse des données publiques. Rapport remis au ministre de l'Industrie le 13 juillet 2011*, 2011, 111 p., p. 25, <http://www.ladocumentationfrancaise.fr/var/storage/rapports-publics/114000407/0000.pdf> (consulté le 30 décembre 2013)

MOLLOY Laura, « Research Data Management in France : DCC and Jisc MRD programme suggest possible approaches », 24 septembre 2013, <http://www.dcc.ac.uk/blog/research-data-management-france-dcc-and-jisc-mrd-programme-suggest-possible-approaches> (consulté le 22 novembre 2013)

NATURE, « Must try harder », mars 2012, vol. 483, p. 509, <http://www.nature.com/nature/journal/v483/n7391/pdf/483509a.pdf> (consulté le 26 décembre 2013)

RIGEADE Marine, *Les archives ouvertes institutionnelles en France. État des lieux et perspectives*, mémoire DCB sous la dir. de Sandrine Malotaux, 2012, 100 p., <http://www.enssib.fr/bibliotheque-numerique/documents/56708-les-archives-ouvertes-institutionnelles-en-france-etat-des-lieux-et-perspectives.pdf> (consulté le 15 octobre 2013)

ROSIER Florence, « Surfer sur le tsunami des données de la recherche », *Le Monde*, 7 juin 2012, http://abonnes.lemonde.fr/sciences/article/2012/06/07/surfer-sur-le-tsunami-de-donnees-de-recherche_1714722_1650684.html (consulté le 23 octobre 2013)

STAMBOLIYSKA Rayna, « Politiques publiques pour l'ouverture des données scientifiques », *Science ouverte*, 10 décembre 2013, <http://donnees.hypotheses.org/71> (consulté le 16 décembre 2013)

JALONS POLITIQUES

ALA-MUTKA Kirsti, « Open Access and Digital Science in Horizon 2020 », *Données de la recherche : quel rôle pour la documentation ?*, journée d'étude du congrès annuel de l'ADBU, Le Havre, 19 septembre 2013

COMMISSION EUROPEENNE, « Communication de la Commission [...] sur l'information scientifique à l'ère numérique : accès, diffusion et préservation », Bruxelles, Commission Européenne, 14 février 2007, <http://eur-lex.europa.eu/LexUriServ/LexUriServ.do?uri=COM:2007:0056:FIN:FR:PDF> (consulté le 21 octobre 2013)

COMMISSION EUROPEENNE, « Données scientifiques : le libre accès aux résultats de recherche stimulera la capacité d'innovation en Europe », 17 juillet 2012, http://europa.eu/rapid/press-release_IP-12-790_fr.htm (consulté le 22 octobre 2013)

COMMISSION EUROPEENNE, « Communication de la Commission [...]. Pour un meilleur accès aux informations scientifiques : dynamiser les avantages des investissements publics dans le domaine de la recherche », Bruxelles, Commission Européenne, 17 juillet 2012, <http://eur-lex.europa.eu/LexUriServ/LexUriServ.do?uri=COM:2012:0401:FIN:FR:PDF> (consulté le 14 octobre 2013)

COMMISSION EUROPEENNE, « Recommandation de la Commission relative à l'accès aux informations scientifiques et à leur conservation », Bruxelles, Commission Européenne, 17 juillet 2012, http://ec.europa.eu/research/science-society/document_library/pdf_06/recommendation-access-and-preservation-scientific-information_fr.pdf (consulté le 21 octobre 2013)

COMMISSION EUROPEENNE, « Communication de la Commission [...]. Un partenariat renforcé pour l'excellence et la croissance dans l'Espace européen de la recherche », Bruxelles, Commission Européenne, 17 juillet 2012, http://ec.europa.eu/research/era/pdf/era-communication/era-communication_fr.pdf (consulté le 21 octobre 2013)

COMMISSION EUROPEENNE, « Report of the European Commission Public Consultation on Open Research Data », octobre 2013, 15 p., https://ec.europa.eu/digital-agenda/sites/digital-agenda/files/Report_2013-07-OpenResearchData-Consultation-FINAL1.pdf (consulté le 23 octobre 2013)

COMMISSION EUROPEENNE, « La Commission lance un projet pilote de libre accès aux données issues de la recherche financée sur fonds publics », 16 décembre 2013, http://europa.eu/rapid/press-release_IP-13-1257_fr.htm (consulté le 28 décembre 2013)

COMMISSION EUROPEENNE, *Guidelines on Data Management in Horizon 2020*, v. 1.0, 16 décembre 2013, 6 p., http://ec.europa.eu/research/participants/data/ref/h2020/grants_manual/hi/oa_pilot/h2020-hi-oa-data-mgt_en.pdf (consulté le 20 décembre 2013)

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> (consulté le 15 octobre 2013)

[COMMISSION EUROPEENNE] HIGH LEVEL-EXPERT GROUP ON SCIENTIFIC DATA, *Riding the wave : how Europe can gain from the rising tide of scientific data*, UE, octobre 2010, 36 p., <http://cordis.europa.eu/fp7/ict/e-infrastructure/docs/hlg-sdi-report.pdf> (consulté le 21 octobre 2013)

OFFICE OF SCIENCE AND TECHNOLOGY POLICY, « Increasing Access to the Results of Federally Funded Scientific Research », *Memorandum for the Heads of Executive Departments and Agencies*, Washington, 22 février 2013, http://www.whitehouse.gov/sites/default/files/microsites/ostp/ostp_public_access_memo_2013.pdf (consulté le 30 octobre 2013)

OPENAIRE, LIBER & COAR, « Horizon 2020 – Outline of a Pilot for Open Research Data. Joint statement by OpenAIRE, LIBER and COAR », 3 juillet 2013, 4 p.,

http://www.coar-repositories.org/files/Horizon_2020_Open_Data_Pilot_20130703_final.pdf
(consulté le 21 octobre 2013)

ORGANISATION DE COOPERATION ET DE DEVELOPPEMENT ECONOMIQUES (OCDE), « Déclaration sur l'accès aux données de la recherche financée par des fonds publics adoptée le 30 janvier 2004 à Paris », Comité de la politique scientifique et technologique de l'OCDE, <http://www.oecd.org/fr/science/sci-tech/sciencetechnologieetinnovationpourle21emesieclereunionducomitedelapolitiquescientifiqueettechnologiquedelocdeaniveauministeriel29-30janvier2004-communiquefinal.htm> (consulté le 14 octobre 2013)

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> (consulté le 29 mai 2013)

« Recommendations of the participants of the workshop on Open Access to Research Data to the European Commission », 21 février 2013, http://recodeproject.eu/wp-content/uploads/2013/04/2013_OARD_Recommendations_public.pdf (consulté le 24 octobre 2013).

SYNTHESES ET RAPPORTS

COMMISSION EUROPEENNE, *National Open Access and Preservation Policies in Europe. Analysis of a questionnaire to the European Research Area Committee*, 2011, Luxembourg, Publications office of the European Union, 69 p., http://ec.europa.eu/research/science-society/document_library/pdf_06/open-access-report-2011_en.pdf (consulté le 21 octobre 2013)

GEROUDET Marie-Madeleine, *Étude prospective sur les données de la recherche*, rapport sous la dir. d'Isabelle Le Bescond, mai 2013, SCD Univ. Lille 1, 62 p. [document non public, communiqué par l'auteur expurgé des éléments qui ne pouvaient être diffusés qu'en interne]³⁷⁷

JISC, *A Comparative Study of International Approaches to Enabling the Sharing of Research Data*, 2008, 108 p., http://www.jisc.ac.uk/media/documents/programmes/preservation/national_data_sharing_report_final.pdf (consulté le 25 octobre 2013)

MINISTERE DE L'ENSEIGNEMENT SUPERIEUR ET DE LA RECHERCHE, *Stratégie nationale. Infrastructures de recherche 2012-2020*, MESR, octobre 2012, 53 p., http://www.donneesdelarecherche.fr/IMG/pdf/MESR_Strategie-Nationale-Infrastructures_22_10_12.pdf (consulté le 2 octobre 2013)

NATIONAL SCIENCE FOUNDATION, *Long-Lived Digital Data Collections : Enabling Research and Education in the 21st Century*, septembre 2005, 87 p.,

³⁷⁷ Je remercie à nouveau vivement l'auteur d'avoir bien voulu me communiquer ce rapport.

<http://www.nsf.gov/pubs/2005/nsb0540/nsb0540.pdf> (consulté le 5 décembre 2013)

NICOL Aurore, CARUSO Julie & ARCHAMBAULT Éric, *Open Data Access Policies and Strategies in the European Research Area and Beyond*, Science-Metrix, août 2013, 16 p., http://www.science-metrix.com/pdf/SM_EC_OA_Data.pdf (consulté le 27 septembre 2013)

RECODE, *Deliverable D1 : Stakeholder Values and Ecosystems*, septembre 2013, 101 p., http://recodeproject.eu/wp-content/uploads/2013/10/RECODE_D1-Stakeholder-values-and-ecosystems_Sept2013.pdf (consulté le 23 octobre 2013)

RESEARCH INFORMATION NETWORK (RIN), *To Share or not to Share : Publication and Quality Assurance of Research Data Outputs*, juin 2008, 56 p., <http://www.rin.ac.uk/system/files/attachments/To-share-data-outputs-report.pdf> (consulté le 22 novembre 2013)

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 (consulté le 27 septembre 2013)

SHEARER KATHLEEN, *Les données de la recherche : un potentiel insoupçonné. Une trousse de sensibilisation réalisée pour l'Association des bibliothèques de recherche du Canada (ABRC)*, ABRC, 2009, 15 p., http://www.carl-abrc.ca/uploads/pdfs/data_mgt_toolkit-f.pdf (consulté le 15 octobre 2013)

SURF, *European Landscape Study of Research Data Management*, Utrecht, SURF, mai 2013, 55 p., http://www.sim4rdm.eu/sites/default/files/uploads/documents/SIM4RDM%20landscape%20report%20vs1%204_14.08.13.pdf (consulté le 1^{er} octobre 2013)

LES POLITIQUES DE DONNEES : INITIATIVES EDITORIALES ET DISCIPLINAIRES

AALBERSBERG Jan IJsbrand & alii, « Bringing Digital Science Deep Inside the Scientific Article : the Elsevier Article of the Future Project », *LIBER Quarterly*, vol. 23 (3) à paraître, <http://liber.library.uu.nl/index.php/lq/article/view/8446> (consulté le 28 décembre 2013)

ALSHEIKH-ALI Alawi A., QURESHI Waqas, AL-MALLAH Mouaz H. & IOANNIDIS John P. A., « Public Availability of Published Research Data in High-Impact Journals », *PLoS ONE*, 2011, 6 (9), <http://www.plosone.org/article/info%3Adoi%2F10.1371%2Fjournal.pone.0024357> (consulté le 15 octobre 2013)

BEL Bernard, « Mutualisation et archivage pérenne des données orales : un nouveau cadre technique et juridique au service de la recherche en linguistique », *Les archives de la recherche : problèmes et enjeux de la construction du savoir scientifique, 18-20 janvier 2012, Université Paris-Sud 11 & CNAM, Paris*, Paris, 2012, 16 p., <http://aune.lpl.univ-aix.fr/~fulltext/4871.pdf> (consulté le 1^{er} octobre 2013)

CHAVAN Vishwas, PENEV Lyubomir, « The data paper : a mechanism to incentivize data publishing in biodiversity science », *BMC Bioinformatics*, vol. 12

suppl. 15, 2011, <http://www.biomedcentral.com/1471-2105/12/S15/S2> (consulté le 28 décembre 2013)

COMMISSION EUROPEENNE, *Livre vert. Connaissance du milieu marin 2020 [Marine Knowledge 2020] : de la cartographie des fonds marins à la prévision océanographique*, Union Européenne, 2012, 23 p., http://ec.europa.eu/maritimeaffairs/documentation/publications/documents/marine-knowledge-2020-green-paper_fr.pdf (consulté le 7 octobre 2013)

DIETRICH Dianne, ADAMUS Trisha, MINER Alison, STEINHART Gail, « De-Mystifying the Data Management Requirements of Research Funders », *Issues in Science and Technology Librarianship*, n° 70 (2012), <http://www.istl.org/12-summer/refereed1.html> (consulté le 16 octobre 2013)

DIGITAL CURATION CENTER (DCC), « Overview of funders' data policies », <http://www.dcc.ac.uk/resources/policy-and-legal/overview-funders-data-policies> (consulté le 21 octobre 2013)

ELIXIR, *ELIXIR. Data for life*, 14 p., http://www.elixir-europe.org/sites/elixir-europe.org/files/documents/elixir_square_brochure_final_printed.pdf (consulté le 24 octobre 2013)

GRIFFITHS Aaron, « The Publication of Research Data : Researcher Attitudes and Behaviour », *The International Journal of Data Curation*, n°1, vol. 4, 2009, p. 46-56, <http://www.ijdc.net/index.php/ijdc/article/view/101/76> (consulté le 7 octobre 2013)

HOLOGNE Odile, « How could forest science journals contribute to the development of open data ? », *Meeting of the Editors of Forest Science Journals*, Nancy, 24 septembre 2013, <http://prodinra.inra.fr/?locale=fr#!ConsultNotice:208058> (consulté le 26 décembre 2013)

HRYNASZKIEWICZ Iain, BUSCH Stefan & COCKERILL Matthew J., « Licensing the future : report on BioMed Central's public consultation on open data in peer-reviewed journals », *BMC Research Notes*, août 2013, <http://www.biomedcentral.com/content/pdf/1756-0500-6-318.pdf> (consulté le 19 novembre 2013)

HUMA-NUM, « Huma-Num, la nouvelle très grande infrastructure de recherche pour les humanités numériques », *La Lettre de l'INSHS*, septembre 2013, http://www.huma-num.fr/sites/default/files/lettre_info25_tribune.pdf (consulté le 20 décembre 2013)

JONES Sarah, « Developments in Research Funder Data Policy », *The International Journal of Digital Curation*, vol. 7, n° 1, 2012, <http://ijdc.net/index.php/ijdc/article/view/209/278> (consulté le 12 novembre 2013)

LAGERSTROM Jill, « Measuring the Impact of the Hubble Space Telescope : open data as a catalyst for science », *76th IFLA General Conference and Assembly, 10-15 August 2010, Gothenburg, Sweden*, 2010, 11 p., <http://conference.ifla.org/past/2010/155-lagerstrom-en.pdf> (consulté le 30 décembre 2013)

LECARPENTIER Damien & alii, « EUDAT : A New Cross-Disciplinary Data Infrastructure for Science », *The International Journal of Digital Curation*, vol. 8, n° 1, 2013, 9 p., <http://www.ijdc.net/index.php/ijdc/article/view/8.1.279/312> (consulté le 23 octobre 2013)

LEONELLI Sabina, SMIRNOFF Nicholas, MOORE Jonathan, COOK Charis & BASTOW Ruth, « Making open data work for plant scientists », *Journal of Experimental Botany*, septembre 2013, p. 4109-4117, <http://jxb.oxfordjournals.org/content/early/2013/09/15/jxb.ert273.full> (consulté le 20 novembre 2013)

MARQUES David, « Research Data Driving New Services », *Library Connect Newsletter*, vol. 11, n° 1, 2013, <http://libraryconnect.elsevier.com/articles/best-practices/2013-02/research-data-driving-new-services> (consulté le 28 décembre 2013)

POISOT Timothée, MOUNCE Ross & GRAVEL Dominique, « Moving toward a sustainable ecological science : don't let data go to waste ! », nov. 2012, 14 p., à paraître, http://figshare.com/articles/Moving_toward_a_sustainable_ecological_science_don_t_let_data_go_to_waste_/693745

SMITH Vincent & alii, « Beyond Dead Trees : Integrating the Scientific Process in the *Biodiversity Data Journal* », *Biodiversity Data Journal*, septembre 2013, 8 p., <http://biodiversitydatajournal.com/articles.php?id=995> (consulté le 24 octobre 2013)

TGE ADONIS, « Les services de conservation de données proposés par le TGE Adonis », *La Lettre de la Grille Adonis*, n° 5, février-mars 2013, 14 p., <http://www.huma-num.fr/sites/default/files/ressourcesdoc/la-lettre-fev-mars2013.pdf>

THIBAUT François & JOUVE Bertrand, *Les Infrastructures de recherche en sciences humaines et sociales. Rapport du groupe "Infrastructures" de l'Alliance Athéna* [Alliance nationale des sciences humaines et sociales], septembre 2012, 40 p., p. 14-15, <http://www.cnrs.fr/fr/pdf/allianceathena/121012-infrastructure-sh/> (consulté le 3 octobre 2013)

THORLEY Mark, « Data in the research process : a funder [NERC]'s perspective », *OpenAIREplus workshop "Linking Open Access publications to data – policy development and implementation"*, 11 juin 2012, http://www.openaire.eu/fr/about-openaire/publications-presentations/presentations/doc_details/393-data-in-the-research-process-a-funders-perspective--mark-thorley-national-environment-research-c (consulté le 21 novembre 2013)

LES DONNEES, ASPECTS JURIDIQUES ET TECHNIQUES

NOTA : pour une bibliographie complète et régulièrement mise à jour sur la question de la « curation » des données de la recherche dans les publications anglo-saxonnes, notamment américaines, consulter BAILEY Charles W. éd., *Research Data Curation Bibliography*, Houston, Digital Scholarship, juin 2013, <http://digital-scholarship.org/rdcb/rdcb.htm> (consulté le 15 octobre 2013)

AALBERSBERG Jan IJsbrand, KÄHLER Ove, « Supporting Science through the Interoperability of Data and Articles », *The Magazine of Digital Library Research*, janvier-février 2011, vol. 17, n° 1-2, <http://www.dlib.org/dlib/january11/aalbersberg/01aalbersberg.html> (consulté le 28 octobre 2013)

BALL Alex, *How to License Research Data*, Digital Curation Center (DCC), 2012, 16 p.,

http://www.dcc.ac.uk/sites/default/files/documents/publications/reports/guides/How_To_License_Research_Data.pdf (consulté le 30 septembre 2013)

BALL Alex, DUKE Monica, *How to Cite Datasets and Link to Publications*, DCC, juin 2012, 12 p.,

http://www.dcc.ac.uk/sites/default/files/documents/publications/reports/guides/How_to_Cite_Link.pdf (consulté le 26 décembre 2013)

BERMES Emmanuelle, « Des identifiants pérennes pour les ressources numériques. L'expérience de la BnF », *International Preservation News*, Paris, n° 40, décembre 2006, p. 16-26, <http://archive.ifla.org/VI/4/news/ipnn40.pdf> (consulté le 1^{er} octobre 2013)

BRASE Jan, FARQUHAR Adam, « Access to Research Data », *D-Lib Magazine*, vol. 17, n° 1/2, 2011, <http://www.dlib.org/dlib/january11/brase/01brase.html> (consulté le 14 décembre 2013)

CHANIER Thierry, « A viewpoint on the place of CALL within the Digital Humanities : considering CALL journals, research data and the sharing of research results », *EUROCALL 2013. Learning from the Past, Looking to the Future, université d'Évora, Portugal, 11-14 septembre 2013*, <http://edutice.archives-ouvertes.fr/docs/00/86/20/24/PDF/eurocall2013-chanier-130915.pdf> (consulté le 30 octobre 2013)

CHANIER Thierry, « Results of the survey “CALL within the Digital Humanities : considering CALL journals, research data and the sharing of research results” », 21 octobre 2013, <http://edutice.archives-ouvertes.fr/docs/00/87/64/30/PDF/Chanier-Eurocall2013-survey-researchdata.pdf> (consulté le 30 octobre 2013)

CODATA-ICSTI TASK GROUP ON DATA CITATION STANDARDS, « Out of Cite, Out of Mind: The Current State of Practice, Policy, and Technology for the Citation of Data », *Data Science Journal*, vol. 12, 13 septembre 2013, https://www.jstage.jst.go.jp/article/dsj/12/0/12_OSOM13-043/_pdf (consulté le 30 octobre 2013)

DANTANT Martin, « Droit d'auteur des chercheurs, logiciels, bases de données et archives ouvertes », support de formation du réseau Mistral-Doc, 30 mai 2013, 128 p.,

http://mistral.cnrs.fr/IMG/pdf/Droit_d_auteur_des_rechercheurs_Logiciels_Bases_de_Donnees_et_Archives_Ouvertes_-_Marseille.pdf (consulté le 1^{er} octobre 2013)

DIETRICH Nils, GUIBAUT Lucie, MARGONI Thomas & alii, *OpenAIRE. Study on licensing of publications and research data : summary of findings*, novembre 2013, 9 p., http://www.openaire.eu/en/about-openaire/publications-presentations/public-project-documents/doc_download/621-openairelicensingstudysummary (consulté le 3 décembre 2013)³⁷⁸

FORCE 11, « Declaration of Data Citation Principles. Draft », mars 2013, <http://www.force11.org/datacitation> (consulté le 27 décembre 2013)

GRÜTTEMEIER Herbert, « DataCite au service des données : identifier pour valoriser », support d'intervention à l'université d'été GFII [Groupement français

³⁷⁸ La version complète de ce rapport doit désormais être disponible en ligne.

de l'industrie de l'information], 12 septembre 2013, <http://www.gfii.fr/uploads/docs/DataCite.pdf> (consulté le 5 octobre 2013)

HAGEDORN Gregor & alii, « Creative Commons licenses and the non-commercial condition : Implications for the re-use of biodiversity information », *ZooKeys*, novembre 2011, n° 150, p. 127-149, <http://www.ncbi.nlm.nih.gov/pmc/articles/PMC3234435/> (consulté le 30 octobre 2013)

HENNEKEN Edwin A., ACCOMAZZI Alberto, « Linking to Data – Effect on Citation Rates in Astronomy », novembre 2011, <http://arxiv.org/abs/1111.3618> (consulté le 26 décembre 2013)

KNOWLEDGE EXCHANGE, *The legal status of the research data in the Knowledge Exchange partner countries (Netherlands, Denmark, Germany, United Kingdom)*, Centre for Intellectual Property Law (CIER), octobre 2011, 59 p., http://www.knowledge-exchange.info/Files/Filer/downloads/Primary%20Research%20Data/Legal%20Stat%20us%20Research%20Data/KE-CIER_Report_legal_status_of_research_data_Final.pdf (consulté le 30 septembre 2013)

MASSOL Marion, « L'archivage à l'heure de l'Europe », *La Gazette du Cines. Archivage numérique pérenne*, CINES, février 2013, 59 p., p. 53-56, http://www.cines.fr/IMG/pdf/CINES_GAZETTE_Special_Archivage-2.pdf (consulté le 23 octobre 2013)

OPENAIREPLUS, *Data Guidelines : Use of DataCite Metadata Schema*, juin 2013, https://guidelines.openaire.eu/wiki/Data_Guidelines:_Use_of_DataCite_Metadata_Schema (consulté le 14 décembre 2013)

PAMPPEL Heinz & alii, « Making Research Data Repositories Visible : The re3data.org Registry », *PLoS ONE*, 8 (11), novembre 2013, <http://www.plosone.org/article/info%3Adoi%2F10.1371%2Fjournal.pone.0078080> (consulté le 26 décembre 2013)

REILLY Susan & alii, *Report on Integration of Data and Publications*, ODE [Opportunities for Data Exchange] project, octobre 2011, 87 p., <http://www.libereurope.eu/sites/default/files/ODE-ReportOnIntegrationOfDataAndPublication.pdf> (consulté le 7 octobre 2013)

RETTBERG Najla, SCHMIDT Birgit, « Building a Collaborative Open Access Infrastructure for European Researchers », *LIBER Quarterly*, vol. 22, n° 3, 2012, p. 160-175, <http://liber.library.uu.nl/index.php/lq/article/view/URN%3ANBN%3ANL%3AUI%3A10-1-113938/8514> (consulté le 14 décembre 2013)

THESSSEN Anne E., PATTERSON David J., « Data Issues in the Life Sciences », *ZooKeys*, novembre 2011, n° 150, p. 15-51, <http://www.ncbi.nlm.nih.gov/pmc/articles/PMC3234430/> (consulté le 28 octobre 2013)

WHITE Ethan P. et alii, « Nine simple ways to make it easier to (re)use your data », *Ideas in Ecology and Evolution*, vol. 6, n°2, 2013, p. 1-10, <http://library.queensu.ca/ojs/index.php/IEE/article/view/4608/4898> (consulté le 2 octobre 2013)

LES UNIVERSITES ET LA GESTION DES DONNEES DE RECHERCHE

CENTRE NATIONAL DE LA RECHERCHE SCIENTIFIQUE, *Schéma d'orientation stratégique de l'information scientifique et technique (IST)*. « Mieux partager les connaissances », novembre 2013, 49 p., <http://www.cnrs.fr/dist/docs/131119-orientation.pdf> (consulté le 31 décembre 2013)

HITCHCOCK Steve & WHITE Wendy, *Towards research data cataloguing at Southampton using Microsoft SharePoint and EPrints: a progress report*, University of Southampton, mai 2013, 15 p., <http://eprints.soton.ac.uk/352813/3/eprints-sharepoint-report-final10.pdf> (consulté le 12 décembre 2013)

HODSON Simon, « Seven rules of successful research data management in universities », *The Guardian Higher Education Network blog*, juillet 2013, <http://www.jisc.ac.uk/blog/seven-rules-of-successful-research-data-management-in-universities-16-jul-2013> (consulté le 30 décembre 2013)

INSTITUT NATIONAL DE LA RECHERCHE AGRONOMIQUE [INRA], *Rapport du groupe de travail sur la gestion et le partage des données*, Paris, juin 2012, 62 p., http://www.pfl-cepia.inra.fr/uploads/gdp_docs/Rapport-GestionDonnees-web.pdf (consulté le 7 octobre 2013)

JONES Sarah, « Bringing it all together : a case study on the improvement of research data management at Monash University », Edinburgh, Digital Curation Center, 2013, http://www.dcc.ac.uk/resources/developing-rdm-services/improving-rdm-monash#_ftn2 (consulté le 10 décembre 2013)

JONES Sarah, PRYOR Graham & WHYTE Angus, « How to Develop Research Data Management Services. A guide for HEIs' », Edinburgh, Digital Curation Center, 2013, http://www.dcc.ac.uk/resources/how-guides/how-develop-rdm-services#_ftn2 (consulté le 11 décembre 2013)

LERU RESEARCH DATA WORKING GROUP, *LERU Roadmap for Research Data*, 18 décembre 2013, 34 p., http://www.leru.org/files/publications/AP14_LERU_Roadmap_for_Research_data_final.pdf (consulté le 20 décembre 2013)

PINK Catherine, « Meeting the Data Management Compliance Challenge : Funder Expectations and Institutional Reality », *The International Journal of Digital Curation*, vol. 2, n° 8, 2013, p. 157-171, <http://www.ijdc.net/index.php/ijdc/article/viewFile/8.2.157/325> (consulté le 10 décembre 2013)

PINK Catherine, *Research 360 : JISC Final Report*, University of Bath, juillet 2013, 26 p., http://opus.bath.ac.uk/36362/2/R360_JiscFinalReport_v1.0.pdf (consulté le 16 décembre 2013)

RICE Robin, EKMEKCIOGLU Cuna, HAYWOOD Jeff & alii, « Implementing the Research Data Management Policy : University of Edinburgh Roadmap », *The International Journal of Digital Curation*, vol. 2, n° 8, 2013, p. 194-204, <http://www.ijdc.net/index.php/ijdc/article/view/8.2.194/327> (consulté le 10 décembre 2013)

RUMSEY Sally, JEFFERIES Neil, « Challenges in Building an Institutional Research Data Catalogue », *The International Journal of Digital Curation*, vol. 8, n° 2,

2013, p. 205-214, <http://www.ijdc.net/index.php/ijdc/article/view/8.2.205/328> (consulté le 12 décembre 2013)

RUMSEY Sally, JEFFERIES Neil, « DataFinder : a Research Data Catalogue for Oxford », *Ariadne*, n° 71, juin 2013, <http://www.ariadne.ac.uk/issue71/rumsey-jefferies> (consulté le 13 décembre 2013).

WILSON James A. J., JEFFREYS Paul W., « Towards a Unified University Infrastructure : The Data Management Roll-Out at the University of Oxford », *The International Journal of Digital Curation*, vol. 2, n° 8, 2013, p. 235-246, <http://www.ijdc.net/index.php/ijdc/article/view/8.2.235/331> (consulté le 13 décembre 2013)

BIBLIOTHEQUES ET DONNEES DE LA RECHERCHE

AYRIS Paul, « LIBER and LERU participate in EC consultation on Open Research Data », 3 juillet 2013, <http://www.libereurope.eu/news/liber-and-leru-participate-in-ec-consultation-on-open-research-data> (consulté le 22 octobre 2013)

CORRALL Sheila, KENNAN Mary Anne, AFZAL Waseem, « Bibliometrics and Research Data Management Services : Emerging Trends in Library Support for Research », *Library Trends*, 61 (3), 2013, p. 636-674, http://d-scholarship.pitt.edu/18948/1/08_61_3_corrall_636%2D674.pdf

COX Andrew, VERBAAN Eddy, SEN Barbara, « Upskilling Liaison Librarians for Research Data Management », *Ariadne*, n° 70, novembre 2012, <http://www.ariadne.ac.uk/issue70/cox-et-al> (consulté le 16 décembre 2013)

GENOVA Françoise, « Le CDS : des données au service de la communauté scientifique », *Documentaliste*, vol. 50, n° 3, octobre 2013, p. 47-49.

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

HOLOGNE Odile, « Données de la recherche : rôle des professionnels IST à l'INRA », *Données de la recherche : quel rôle pour la documentation ?*, journée d'étude du congrès annuel de l'ADBU, Le Havre, 19 septembre 2013, <http://prodinra.inra.fr/?locale=fr#!ConsultNotice:207958> (consulté le 12 décembre 2013)

HOLOGNE Odile, « Professionnels IST et données de la recherche : des discours incantatoires aux actions concrètes », *Documentaliste*, vol. 50, n° 3, octobre 2013, p. 30-31.

PICKTON Miggie, JONES Sarah, MARIEKE Guy, « Leading from the library : data management initiatives at the University of Northampton », *The role of libraries in data curation, access and preservation : an international perspective. Conference Session 116, 78th IFLA General Conference and Assembly, 11-17 août 2012, Helsinki*, 10 p., <http://conference.ifla.org/past/2012/116-pickton-en.pdf> (consulté le 10 décembre 2013)

REILLY Susan, « The role of libraries in supporting data exchange », *The role of libraries in data curation, access and preservation : an international perspective. Conference Session 116, 78th IFLA General Conference and Assembly, 11-17 août 2012, Helsinki*, 7 p., <http://conference.ifla.org/past/2012/116-reilly-en.pdf> (consulté le 5 décembre 2013)

REILLY Susan, « European Commission Public Consultation on Open Research Data : some impressions from the day », 3 juillet 2013, <http://www.libereurope.eu/blog/european-commission-public-consultation-on-open-research-data-some-impressions-from-the-day> (consulté le 22 octobre 2013)

RESEARCH LIBRARIES UK (RLUK), *Re-skilling for Research. An investigation into the role and skills of subject and liaison librarians required to effectively support the evolving information needs of researchers*, janvier 2012, 112 p., <http://www.rluk.ac.uk/files/RLUK%20Re-skilling.pdf> (consulté le 30 décembre 2013)

Ten recommendations for libraries to get started with research data management. Final report of the LIBER working group on E-Science/Data Management, juillet 2012, 3 p., <http://www.libereurope.eu/sites/default/files/The%20research%20data%20group%202012%20v7%20final.pdf> (consulté le 13 décembre 2013)

Table des annexes

LA PUBLICATION DES DONNEES : DES DONNES BRUTES AUX DONNEES PUBLIEES..... 94

Source : REILLY Susan & alii, *Report on Integration of Data and Publications*, ODE [Opportunities for Data Exchange] project, octobre 2011, 87 p., p. 36
<http://www.libereurope.eu/sites/default/files/ODE-ReportOnIntegrationOfDataAndPublication.pdf>

LE CONTENU D’UN PLAN DE GESTION : LA CHECKLIST DU DIGITAL CURATION CENTER..... 95

Source : DIGITAL CURATION CENTER, *Checklist for a Data Management Plan, v 4.0*, 2013,
http://www.dcc.ac.uk/sites/default/files/documents/resource/DMP_Checklist_2013.pdf

UNE POLITIQUE INSTITUTIONNELLE DE DONNEES : LE CAS D’OXFORD 97

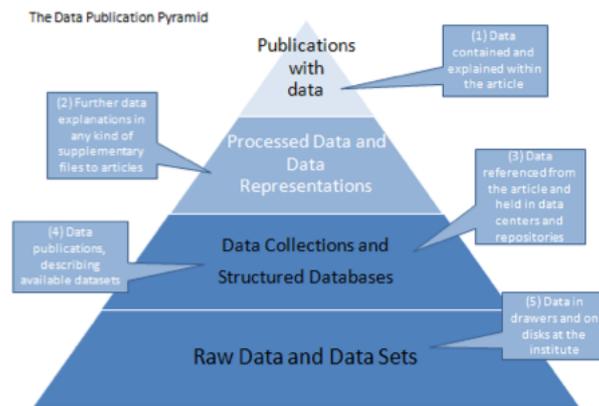
Source : OXFORD UNIVERSITY, « Policy on the Management of Research Data and Records »,
http://www.admin.ox.ac.uk/media/global/wwwadminoxacuk/localsites/researchdatamanagement/documents/Policy_on_the_Management_of_Research_Data_and_Records.pdf

LA FEUILLE DE ROUTE STRATEGIQUE D’UNE UNIVERSITE : LE CAS DE MONASH..... 99

Source : MONASH UNIVERSITY, « Research Data Management Strategy and Strategic Plan 2012-2015 », avril 2012, 17 p.,

<https://confluence-vre.its.monash.edu.au/download/attachments/39752006/Monash+University+Research+Data+Management+Strategy-publicrelease.pdf?version=1&modificationDate=1334289180000>

LA PUBLICATION DES DONNEES : DES DONNES BRUTES AUX DONNEES PUBLIEES



Source : REILLY Susan & alii, *Report on Integration of Data and Publications*, ODE [Opportunities for Data Exchange] project, octobre 2011, 87 p., p. 36
<http://www.libereurope.eu/sites/default/files/ODE-ReportOnIntegrationOfDataAndPublication.pdf>

LE CONTENU D'UN PLAN DE GESTION : LA *CHECKLIST* DU *DIGITAL CURATION CENTER*



Checklist for a Data Management Plan, v4.0

Please cite as: DCC. (2013). *Checklist for a Data Management Plan*. v.4.0. Edinburgh: Digital Curation Centre. Available online: <http://www.dcc.ac.uk/resources/data-management-plans>

DCC Checklist	DCC Guidance and questions to consider
Administrative Data	
ID	A pertinent ID as determined by the funder and/or institution.
Funder	State research funder if relevant
Grant Reference Number	Enter grant reference number if applicable [POST-AWARD DMPs ONLY]
Project Name	If applying for funding, state the name exactly as in the grant proposal.
Project Description	<p>Questions to consider:</p> <ul style="list-style-type: none"> - What is the nature of your research project? - What research questions are you addressing? - For what purpose are the data being collected or created? <p>Guidance:</p> <p>Briefly summarise the type of study (or studies) to help others understand the purposes for which the data are being collected or created.</p>
PI / Researcher	Name of Principal Investigator(s) or main researcher(s) on the project.
PI / Researcher ID	E.g ORCID http://orcid.org/
Project Data Contact	Name (if different to above), telephone and email contact details
Date of First Version	Date the first version of the DMP was completed
Date of Last Update	Date the DMP was last changed
Related Policies	<p>Questions to consider:</p> <ul style="list-style-type: none"> - Are there any existing procedures that you will base your approach on? - Does your department/group have data management guidelines? - Does your institution have a data protection or security policy that you will follow? - Does your institution have a Research Data Management (RDM) policy? - Does your funder have a Research Data Management policy? - Are there any formal standards that you will adopt? <p>Guidance:</p> <p>List any other relevant funder, institutional, departmental or group policies on data management, data sharing and data security. Some of the information you give in the remainder of the DMP will be determined by the content of other policies. If so, point/link to them here.</p>
Data Collection	
What data will you collect or create?	<p>Questions to consider:</p> <ul style="list-style-type: none"> - What type, format and volume of data? - Do your chosen formats and software enable sharing and long-term access to the data? - Are there any existing data that you can reuse? <p>Guidance:</p> <p>Give a brief description of the data, including any existing data or third-party sources that will be used, in each case noting its content, type and coverage. Outline and justify your choice of format and consider the implications of data format and data volumes in terms of storage, backup and access.</p>
How will the data be collected or created?	<p>Questions to Consider:</p> <ul style="list-style-type: none"> - What standards or methodologies will you use? - How will you structure and name your folders and files? - How will you handle versioning? - What quality assurance processes will you adopt? <p>Guidance:</p> <p>Outline how the data will be collected/created and which community data standards (if any) will be used. Consider how the data will be organised during the project, mentioning</p>

	for example naming conventions, version control and folder structures. Explain how the consistency and quality of data collection will be controlled and documented. This may include processes such as calibration, repeat samples or measurements, standardised data capture or recording, data entry validation, peer review of data or representation with controlled vocabularies.
Documentation and Metadata	
What documentation and metadata will accompany the data?	<p>Questions to consider:</p> <ul style="list-style-type: none"> - What information is needed for the data to be read and interpreted in the future? - How will you capture / create this documentation and metadata? - What metadata standards will you use and why? <p>Guidance:</p> <p>Describe the types of documentation that will accompany the data to help secondary users to understand and reuse it. This should at least include basic details that will help people to find the data, including who created or contributed to the data, its title, date of creation and under what conditions it can be accessed.</p> <p>Documentation may also include details on the methodology used, analytical and procedural information, definitions of variables, vocabularies, units of measurement, any assumptions made, and the format and file type of the data. Consider how you will capture this information and where it will be recorded. Wherever possible you should identify and use existing community standards.</p>
Ethics and Legal Compliance	
How will you manage any ethical issues?	<p>Questions to consider:</p> <ul style="list-style-type: none"> - Have you gained consent for data preservation and sharing? - How will you protect the identity of participants if required? e.g. via anonymisation - How will sensitive data be handled to ensure it is stored and transferred securely? <p>Guidance:</p> <p>Ethical issues affect how you store data, who can see/use it and how long it is kept. Managing ethical concerns may include: anonymisation of data; referral to departmental or institutional ethics committees; and formal consent agreements. You should show that you are aware of any issues and have planned accordingly. If you are carrying out research involving human participants, you must also ensure that consent is requested to allow data to be shared and reused.</p>
How will you manage copyright and Intellectual Property Rights (IPR) issues?	<p>Questions to consider:</p> <ul style="list-style-type: none"> - Who owns the data? - How will the data be licensed for reuse? - Are there any restrictions on the reuse of third-party data? - Will data sharing be postponed / restricted e.g. to publish or seek patents? <p>Guidance:</p> <p>State who will own the copyright and IPR of any data that you will collect or create, along with the licence(s) for its use and reuse. For multi-partner projects, IPR ownership may be worth covering in a consortium agreement. Consider any relevant funder, institutional, departmental or group policies on copyright or IPR. Also consider permissions to reuse third-party data and any restrictions needed on data sharing.</p>
Storage and Backup	
How will the data be stored and backed up during the research?	<p>Questions to consider:</p> <ul style="list-style-type: none"> - Do you have sufficient storage or will you need to include charges for additional services? - How will the data be backed up? - Who will be responsible for backup and recovery? - How will the data be recovered in the event of an incident? <p>Guidance:</p> <p>State how often the data will be backed up and to which locations. How many copies are being made? Storing data on laptops, computer hard drives or external storage devices alone is very risky. The use of robust, managed storage provided by university IT teams is preferable. Similarly, it is normally better to use automatic backup services provided by IT Services than rely on manual processes. If you choose to use a third-party service, you</p>

[Extrait]. Source :

http://www.dcc.ac.uk/sites/default/files/documents/resource/DMP_Checklist_2013.pdf

UNE POLITIQUE INSTITUTIONNELLE DE DONNEES : LE CAS D'OXFORD



Policy on the Management of Research Data and Records

1. The University of Oxford seeks to promote the highest standards in the management of research data and records¹ as fundamental to both high quality research² and academic integrity.
2. The University recognises that accurate and retrievable research data are an essential component of any research project and necessary to verify and defend, when required, the process and outcomes of research. Research data are valuable to researchers for the duration of their research, and may well have long-term value for research, teaching and for wider exploitation for the public good, by individuals, government, business and other organisations, as a project develops and after research results have been published.
3. The University acknowledges its obligations under research funders' data-related policy statements³ and codes of practice to ensure that sound systems are in place to promote best practice, including through clear policy, guidance, supervision, training and support.
4. Researchers⁴, departments/faculties, divisions, central administrative units and service providers and, where appropriate, research sponsors and external collaborators, need to work in partnership to implement good practice and meet relevant legislative, research funder and regulatory requirements.
5. Research data and records should be:
 - a. Accurate, complete, authentic and reliable;
 - b. Identifiable, retrievable, and available when needed;
 - c. Secure and safe;
 - d. Kept in a manner that is compliant with legal obligations and, where applicable, the requirements of funding bodies and project-specific protocols approved under the University Policy on the Ethical Conduct of Research Involving Human Participants and Personal Data⁵.
 - e. Able to be made available to others in line with appropriate ethical, data sharing and open access principles.
6. Research data and records should be retained for as long as they are of continuing value to the researcher and the wider research community, and as long as specified by research funder, patent law, legislative and other regulatory requirements. The minimum retention period for research data and records is three (3) years after publication or public release of the work of the research. In many instances, researchers will resolve to retain research data and records for a longer period than the minimum requirement.

¹ Research data and records are defined as the recorded information (regardless of the form or the media in which they may exist) necessary to support or validate a research project's observations, findings or outputs.

² Research is defined as per the Frascati manual, i.e. creative work undertaken on a systematic basis in order to increase the stock of knowledge, including knowledge of man, culture and society, and the use of this stock of knowledge to devise new applications.

³ An overview of the major research funders' data policies is available at <http://www.dcc.ac.uk/resources/policy-and-legal/overview-funders-data-policies>. See also the RCUK Common Principles on Data Policy (2011) at <http://www.rcuk.ac.uk/research/Pages/DataPolicy.aspx>

⁴ Researchers are defined as members of the University including staff and doctoral students, and those who are not members of the University but who are conducting research on University premises or using University facilities.

⁵ <http://www.admin.ox.ac.uk/curec/policystatement/>

7. Where research is supported by a contract with or a grant to the University that includes specific provisions regarding ownership, retention of and access to data, the provisions of that agreement will take precedence.
8. If research data and records are to be deleted or destroyed, either because the agreed period of retention has expired or for legal or ethical reasons, this should be done so in accordance with all legal, ethical, research funder and collaborator requirements and with particular concern for confidentiality and security.
9. Researchers are responsible for:
 - Managing research data and records in accordance with the principles and requirements in 5-8 above;
 - Developing and documenting clear procedures for the collection, storage, use, re-use, access and retention or destruction of the research data and records associated with their research. This shall include, where appropriate, defining protocols and responsibilities in a joint or multi-institution collaborative research project. This information should be incorporated, where appropriate, in a research data management plan;
 - Planning for the ongoing custodianship (at the University or using third-party services) of their data after the completion of the research or, in the event of their departure or retirement from the University, reaching agreement with the head of department/faculty (or his/her nominee) as to where such data will be located and how this will be stored;
 - Ensuring that any requirements in relation to research data and records management placed on their research by funding bodies or regulatory agencies or under the terms of a research contract with the University are also met.
10. The University is responsible for:
 - Providing access to services and facilities for the storage, backup, deposit and retention of research data and records that allow researchers to meet their requirements under this policy and those of the funders of their research;
 - Providing researchers with access to training, support and advice in research data and records management;
 - Providing the necessary resources to those operational units charged with the provision of these services, facilities and training.
11. The University's Research and Information Sub-Committee, a sub-committee of the University Research Committee, is responsible for guiding the development and updating of this policy.

Relationship with existing policies

12. This policy will operate in conjunction with other University policies such as:
 - Academic Integrity in Research (<http://www.admin.ox.ac.uk/personnel/cops/researchintegrity/>)
 - Policy on the ethical conduct of research involving human participants and personal data (<http://www.admin.ox.ac.uk/curec/policystatement/>)
 - Intellectual property policy (http://www.admin.ox.ac.uk/statutes/790-121.shtml#_Toc28143157)
 - Data protection policy (<http://www.admin.ox.ac.uk/dataprotection/policy/>)
 - Freedom of Information (<http://www.admin.ox.ac.uk/foi/>)
 - Information Security Policy (<http://www.it.ox.ac.uk/infosec/ispolicy/>)

Source :

http://www.admin.ox.ac.uk/media/global/wwwadminoxacuk/localsites/researchdatamanagement/documents/Policy_on_the_Management_of_Research_Data_and_Records.pdf

LA FEUILLE DE ROUTE STRATEGIQUE D'UNE UNIVERSITE : LE CAS DE MONASH

3. Skills and knowledge

By 2015 Monash University's researchers – including Higher Degree by Research students and, over time, Honours and undergraduate researchers – will have well-developed data management awareness, knowledge and skills. Data management skills will be seen as essential graduate skills that are necessary in the research sector and transferable to a range of other workplaces. The University will offer a range of professional development opportunities – both stand-alone and embedded in the curriculum, delivered through as many channels as possible – that meet the needs of researchers in different disciplines and at different career stages. The development needs and career paths of data management professionals in central units will also be addressed, and the University will be able to recruit, retain and develop high quality staff that contribute both to the University and national capability.

Goal	Initiatives	Key measures	Responsibility/key stakeholders
3.1 Develop the data management skills and knowledge of Monash researchers	<ul style="list-style-type: none"> Project to re-develop, expand and better coordinate the program of research data skills development opportunities available to Monash researchers Work with faculties and the new Monash Institute of Graduate Research to explore embedding research data management skills development in coursework curricula and professional development offerings 	<ul style="list-style-type: none"> Continuous improvement in program of learning opportunities, in terms of number and range of opportunities Increased attendance (e.g. enrolment / attendance statistics) High quality training (e.g. participant feedback, impact surveys) 	<ul style="list-style-type: none"> University Librarian MeRC eSolutions Director, Monash Institute of Graduate Research Associate Deans of Research and Research Training
3.2 Develop the skills and knowledge of professional staff <ul style="list-style-type: none"> Library staff Technical staff – MeRC and eSolutions 	<ul style="list-style-type: none"> Provide targeted development opportunities for professional staff Support opportunities for professional staff to network and participate in communities of practice – Monash, national, international Proactively address staff retention and career planning Promote research data management career paths at entry level through guest lectures, placements, and participation in curriculum review 	<ul style="list-style-type: none"> Continuous improvement in program of opportunities to gain skills and knowledge and to network with other data management professionals Retention of key staff Feedback from staff 	<ul style="list-style-type: none"> University Librarian CIO Director, MeRC
3.3 Integrate research data skills as part of research-led teaching	<ul style="list-style-type: none"> Integrate data management concepts and services into the Graduate Certificate of Academic Practice (GCAP) Map data management knowledge, skills and attributes to educational frameworks such as the Research Skill Development (RSD) framework, the Monash graduate attributes, Australian Qualifications Framework, professional accreditation schemes and other relevant frameworks. Develop action plan for increasing access to, and re-use of, research data by Honours and undergraduate researchers. 	<ul style="list-style-type: none"> GCAP contribution made in first semester 2012 and evaluated for inclusion in future years Skills mapping work completed by June 2012 Action plan developed by December 2012 	<ul style="list-style-type: none"> University Librarian Library Office of the PVC Teaching and Learning Monash Institute of Graduate Research MeRC

13

[Extrait]. Source : <https://confluence-vre.its.monash.edu.au/download/attachments/39752006/Monash+University+Research+Data+Management+Strategy-publicrelease.pdf?version=1&modificationDate=133428918000>

Table des illustrations

- p. 20. Ill. 1. Le cycle de vie de la donnée de recherche
- p. 69. Ill. 2. Le rôle des bibliothécaires en matière de gestion des données : compétences requises et compétences actuelles

Table des matières

SIGLES ET ABREVIATIONS	9
INTRODUCTION.....	11
DEFINITIONS & JALONS POLITIQUES	15
<i>Data et datasets : essai de définition</i>	15
<i>Qu'est-ce qu'une donnée ?</i>	15
<i>De la donnée brute à la donnée dérivée</i>	17
<i>La donnée et son cycle de vie.....</i>	19
L'ouverture des données, un engagement européen	21
<i>L'apparition d'une préoccupation.....</i>	21
<i>L'engagement progressif de l'Union européenne en faveur de l'ouverture des données</i>	22
Les politiques des agences de financement : un phénomène international	29
<i>Les cas américain, britannique et canadien.....</i>	29
<i>Du chercheur à l'institution de recherche, vers une responsabilité partagée</i>	32
OUVRIR LES DONNEES : JEU D'ACTEURS, SOLUTIONS JURIDIQUES ET TECHNIQUES	35
Les chercheurs face à leurs données : impulsions éditoriales, institutionnelles et disciplinaires	35
<i>Incitations éditoriales.....</i>	35
<i>Les mandats institutionnels : l'importance de la gestion des données et des DMP.....</i>	38
<i>Logique bottom-up et culture du partage.....</i>	42
Les enjeux juridiques de l'ouverture	44
<i>Données de recherche et propriété intellectuelle.....</i>	44
<i>... « un cauchemar juridique »</i>	46
<i>Permettre la réutilisation : licences & waivers.....</i>	48
Des infrastructures complémentaires ? Études de cas.....	51
<i>La diversité des solutions disciplinaires.....</i>	52
<i>Solutions éditoriales et data journals.....</i>	53
<i>Fédérer des entrepôts interopérables : l'ambition d'OpenAIRE.....</i>	54
DONNEES, ETABLISSEMENTS DE RECHERCHE ET BIBLIOTHEQUES.....	57
La prise en main des données par les établissements de recherche : méthodologie.....	57
<i>Des responsabilités des institutions en matière de gestion et de conservation</i>	57

<i>Faut-il un entrepôt institutionnel ?</i>	60
Bibliothèques et données : ce qu’elles font et comment elles doivent s’y préparer	64
<i>L’implication des bibliothèques</i>	65
<i>Les bibliothécaires sont-ils les bons interlocuteurs ? Compétences et formations.</i>	68
Bilan : qui faut-il mobiliser, en France, pour participer à l’élaboration de politiques de données ?	73
<i>À l’échelle des établissements, le levier de la gouvernance et de la coordination</i>	73
<i>Capitaliser sur l’expérience acquise : le CINES</i>	74
<i>Une impulsion nationale ?</i>	76
CONCLUSION	79
BIBLIOGRAPHIE	81
Généralités	81
Jalons politiques	82
Synthèses et rapports	84
Les politiques de données : initiatives éditoriales et disciplinaires	85
Les données, aspects juridiques et techniques	87
Les universités et la gestion des données de recherche	90
Bibliothèques et données de la recherche	91
TABLE DES ANNEXES	93
TABLE DES ILLUSTRATIONS	101
TABLE DES MATIERES	103