Mémoire de recherche / septembre 2018

Diplôme national de master

Domaine - sciences humaines et sociales

Mention – sciences de l'information et des bibliothèques

Parcours – archives numériques

Créer, préserver, communiquer.

Une étude comparative du *records management* et des plans de gestion des données.

JACQUES Stephanie

Sous la direction de Céline Guyon
Maître de conférences associé – Enssib

Remerciements

Merci à Céline Guyon pour ses remarques pertinentes et son appui durant la rédaction de ce mémoire.

Merci à Hélène Chambefort, Anne Cornilleau, Alina Dancieu, Christine Hadrossek, Madeleine Hubert, Magalie Moysan et Cynthia Pedroja pour leur disponibilité et nos échanges qui ont permis d'alimenter mes réflexions.

Merci à Elodie Bouillon, Clara Henry et Emmy Toussirot pour le soutien moral et les sessions de travail.

Merci à Odile Rompré-Brodeur et Karine Jacques pour leur relecture.

Finalement, merci à Jonathan Berthier pour son soutien et ses encouragements.

(cc) BY-NC-ND

Résumé : *Le* records management *et les plans de gestion des données sont tous deux des outils servant à identifier et gérer les informations à caractère probant parmi celles qui sont créées dans le cadre de l'activité d'une entreprise ou d'un projet de recherche. Par conséquent, ceux-ci partagent des concepts qui sont leur sont similaires. Ce texte s'interroge donc à reconnaître comment ces outils se ressemblent, et comment leurs différences peuvent possiblement se complémenter.*

Descripteurs : records management, *plan de gestion des données, données de la recherche, preuve, outils de gestion.*

Abstract : *Both records management and data management plan are tools used to identify and manage the information with probative value amongst that created by the activities of an organisation or a research project. Because of that, they share similar concepts. Therefore, this essay will examine their shared concepts as well as explore their differences and if they can complemente each other.*

Keywords : records management, data management plan, research data, life cycle, management tools, proof.

Droits d'auteurs

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

Sommaire

SIGLES ET ABREVIATIONS 7

INTRODUCTION 9

COMPRENDRE LE *RECORDS MANAGEMENT* ET LES PLANS DE GESTION DE DONNEES 11

Le *records management* 12

Plus que de la gestion de documents d'activité 12

Historique du records management 14

Quelques clés pour mieux comprendre le records management 16

Gérer les documents probants et leurs métadonnées 20

Mettre en place un système de *records management* 24

La norme ISO 15489-2 25

Le mode d'emploi de l'ADBS 26

Les plans de gestion de données 29

L'objet d'une longue réflexion 29

Donner sens aux « données de la recherche » 33

Le contenu d'un plan de gestion des données 35

La pratique concernant les PGD 37

Mettre en place les plans de gestion de données 42

Bâtir une offre de service concernant les données de la recherche 42

Data curation profiles 44

LE RECORDS MANAGEMENT ET LES PLANS DE GESTION DE DONNEES, COMPLEMENTAIRES ? 47

Les points communs 47

Revalorisation des métiers des sciences de l'information 47

Démarche d'évaluation 49

Tout n'est pas records, tout n'est pas données de la recherche 50

Contexte de création du record / des données de la recherche 52

Les métadonnées 54

Définir les responsabilités 55

Le cycle de vie 56

Établir une durée de conservation 62

La question du droit d'accès 63

Les points de divergence 64

Des finalités différentes 64

Une mise en place permanente, un projet ponctuel 66

Mettre en place un système ou rédiger un document 67

(cc) BY-NC-ND

Des données authentiques ? Des *records* accessibles ? 68
Des records FAIR ? 69
Des données de la recherche authentiques, fiables, intègres et exploitables ? 74
Une démarche pouvant en aider une autre ? 78
Faciliter la rédaction d'un plan de gestion des données 78
Un plan de gestion de données pour identifier et gérer nos records 82
CONCLUSION 85
SOURCES 89
BIBLIOGRAPHIE 93
ANNEXES 101
TABLE DES ILLUSTRATIONS 123
TABLE DES MATIERES 125

(cc) BY-NC-ND

Sigles et abréviations

AAF	Association des archivistes français
ADBS	Association des professionnels de l'information et de la documentation
AFNOR	Association française de normalisation
ANR	Association nationale de la Recherche
APROGED	Association des professionnels de la gestion électronique des documents
BNF	Bibliothèque nationale de France
CCSD	Centre pour la Communication Scientifique Directe
CDSP	Centre de données socio-politiques
CNRS	Centre national de recherche scientifique
DCP	*Data curation profile*
DDR	Données de la recherche
DINSIC	Direction interministérielle du numérique et du système d'information et de communication de l'Etat
DIRKS	*Designing and Implementing Record Keeping System*
DMP	*Data management plan*
DOI	*Digital object identifier*
DSI	Direction des systèmes informatiques
DUA	Durée d'utilité administrative
ELIPSS	Étude longitudinale par internet pour les sciences sociales
FAIR	*Findable, accessible, interoperable, reusable*
INIST	Institut de l'information scientifique et technique
ISO	*International Organization for Standardization*
NEH	*National Endowment for the Humanities*
NIH	*National Institutes of Health*
NSF	*National Science Foundation*
OCDE	Organisation de Coopération et de Développement Économiques
P.R.O.	*Public Records Office*
PGD	Plan de gestion des données
RIATE	Réseau interdisciplinaire pour l'aménagement et la cohésion des territoires de l'Europe et de ses voisinages
RM	*Records management*
SGDA	Système de gestion des documents d'activités

(cc) BY-NC-ND

INTRODUCTION

Le dernier siècle a vu croître la quantité d'informations qui circule à une vitesse exponentielle. Plusieurs facteurs ont contribué à cette croissance incroyable ; l'accroissement de la population en est sûrement un exemple. Mais le développement de nouveaux modes de communication, facilitant le partage de l'information dans l'espace-temps, y a très certainement aussi contribué. Dans plusieurs cas, ces nouveaux modes de communications ont aussi soulevé un autre problème : celui de la fragilité des supports où est conservée cette information. On se retrouve donc face à une montagne d'information, potentiellement difficile à conserver à long terme. Face à des moyens (humains, matériels, financiers) souvent limités, différentes organisations doivent bien souvent faire des choix. Que garder ? Comment conserver l'information de façon à ce qu'elle reflète le mieux possible l'activité d'où elle découle ? Comment s'assurer que cette information sera toujours lisible dans le temps ?

Dans le domaine des archives, et plus particulièrement des archives administratives, le *records management* (RM) fût originalement mis en place dans les pays anglo-saxons au milieu du 20^{e} siècle afin de gérer la quantité de plus en plus importante de papiers qui s'accumulaient dans les bureaux des agences fédérales. En France, cette méthode de gestion des documents a pris de l'ampleur alors que l'administration numérique se développait, afin de mieux gérer des documents aux supports souvent instables. Le *records management* se concentre sur les documents essentiels sur la gestion des documents essentiels aux organisations, soit ceux qui les engagent. Puisque ces documents servent à prouver la conformité des activités des entreprises et des administrations à la réglementation et aux exigences les affectant, il est important de pouvoir les conserver dans la forme où ils ont été validés par les autorités pertinentes. En bref, le *records management* permet, entre autres, de s'assurer que la forme et le contenu des documents corresponds aux exigences qui sont demandées par leur contexte de création.

Ces questions, ayant été soulevées dans le domaine des sciences de l'information depuis maintenant plusieurs années, préoccupent aujourd'hui plus que les archivistes et documentalistes de ce monde. L'omniprésence des documents et

(cc) BY-NC-ND

des données au sein de nos sociétés font que des acteurs de plusieurs domaines doivent faire face à cette surabondance d'information.

Aujourd'hui, face au déferlement des données et face aux différents mouvements « open » (*open access, open data, open science*) qui poussent à la publication des données de la recherche, le domaine scientifique commence à mettre en place les plans de gestion de données (PGD), des documents servant à détailler comment, dans un projet de recherche, les données seront obtenues, organisées, conservées et diffusées. Le but étant d'identifier les données à conserver et les informations qui doivent les accompagner. L'apparition de ce type de document est relativement récente, moins d'une quinzaine d'années. Les équipes de recherche en sont encore, pour la plupart, à tenter de comprendre comment s'articule cet outil dans les différents projets de recherche ainsi que la façon optimale de les utiliser.

Ces deux outils, qui partagent plusieurs concepts, seront donc l'objet de ce mémoire. L'objectif sera de comprendre et d'analyser les démarches d'écriture d'un plan de gestion des données et de mise en place du *records management* ainsi que de les comparer afin de voir à quels points elles se rapprochent véritablement. Concrètement, notre questionnement est le suivant : est-il possible de rapprocher les concepts de *records management* et ceux des plans de gestion des données et, si oui, est-ce que ces deux outils peuvent se compléménter ?

Pour répondre à cette problématique, nous nous intéresserons dans une première partie à ces deux outils, soit le RM et les PGD, afin de bien les comprendre. Après avoir mis au clair leurs objectifs, nous expliciterons leur historique et ce qui a mené à leur création. Ensuite, nous détaillerons et préciserons les concepts en lien avec ces outils, en plus d'expliquer les différents éléments qui les constituent. Puis, nous observerons différentes façons de mettre en place le RM et les PGD. Cette première partie nous permettra de mettre en place les bases pour, par la suite, comparer ces outils.

La deuxième partie s'attardera donc à comparer le *records management* et les plans de gestion de données. Quels sont les concepts qui les rapprochent ? En quoi sont-ils différents ? Peut-on dire que les objets auxquels ils s'intéressent (les *records* dans un cas et les données de la recherche dans l'autre) se ressemblent ? Finalement, cela nous permettra de réfléchir aux façons dont ces deux outils peuvent potentiellement se compléter.

(cc) BY-NC-ND

COMPRENDRE LE *RECORDS MANAGEMENT* ET LES PLANS DE GESTION DE DONNEES

Afin d'arriver à comparer et comprendre comment deux outils peuvent se complémenter, il est d'abord important de les comprendre, de connaître d'où ils proviennent et comment ils sont mis en œuvre aujourd'hui. Pour ce faire, je me suis d'abord penché sur l'importante littérature existante au niveau du *records management*. Cette démarche (comme nous le verrons plus bas) ayant été introduite en France il y a presque 20 ans et existant depuis encore plus longtemps dans les pays anglo-saxons a fait l'objet de nombreux textes, tant sur sa théorie que sur sa mise en pratique. C'est donc ces textes qui auront servi de base afin de mieux comprendre cette démarche.

Toutefois, la pratique de rédiger des plans de gestion de données est relativement nouvelle, puisqu'elle ne date que d'une dizaine d'années dans le monde et qu'elle est encore balbutiante en France. Il est donc difficile d'avoir accès à un corpus de textes important concernant les PGD et la façon dont les gens se les sont appropriés. C'est pourquoi j'ai choisi de faire des entretiens afin d'obtenir, d'une part, des retours d'expériences de gens ayant mis des plans de gestion de données dans le cadre d'un projet de recherche et, d'autre part, la vision de personnes travaillant étroitement avec les données de la recherche, qui sont au courant de cette nouvelle tendance sans pour autant l'avoir déjà expérimentée. J'ai d'abord contacté Magalie Moysan (responsable du bureau des archives à la direction du pilotage et des études à l'université Paris-Diderot) et Hélène Chambefort (responsable du service des archives à l'Institut national de santé et de recherche médicale) pour des discussions informelles, afin de commencer à m'aiguiller sur les PGD. Ces premières discussions auront orienté mes lectures et m'auront permis de contacter d'autres personnes travaillant sur la question.

Par la suite, quatre entretiens semi-dirigés ont été réalisés dans le cadre de ce mémoire. Deux avec une grille de questions cherchant à connaître leur démarche lors de la rédaction d'un plan de gestion des données et leurs impressions. Puis deux autres en adaptant cette grille pour des personnes n'ayant pas, pour l'instant, rédigé de PGD[1]. D'abord, j'ai interrogé Alina Dancieu et Anne Cornilleau, du Centre de

[1] Vous trouverez ces grilles et la synthèse de ces entretiens en annexe.

(cc) BY-NC-ND

données socio-politiques (CDSP) de SciencesPo, qui ont testé le plan de gestion de données dans le cadre du projet d'Étude longitudinale par internet pour les sciences sociales (ELIPSS). Par la suite, j'ai pu discuter avec Christine Hadrossek, chargée de l'information scientifique et technique pour l'équipe du Réseau interdisciplinaire pour l'aménagement et la cohésion des territoires de l'Europe et de ses voisinages (RIATE), qui a testé le plan de gestion de données dans le cadre d'un projet d'étude sur la typologie socio-économique des régions frontalières. J'ai pu contacter ces personnes grâce à Magalie Moysan, qui m'a mise en contact avec ces dernières. Puis, j'ai réalisé deux autres entretiens, un avec Cynthia Pedroja, chef de projet Gestion des données de la recherche – Direction des ressources et de l'information scientifique à SciencesPo Paris et un autre avec Madeleine Hubert, *data librarian*, chargée de la gestion, du signalement, de l'archivage et de la diffusion des données de la recherche à l'Université de Lorraine. J'ai eu la chance de rencontrer Mme Pedroja dans le cadre du cours Le web : collecte, signalement, accès aux archives du master 2 ARN de l'Enssib ; c'est elle qui par la suite m'a mise en contact avec Mme Hubert.

Ces entretiens m'auront permis d'avoir un retour d'expérience sur la pratique de rédaction d'un plan de gestion de données, mais aussi d'avoir une meilleure compréhension de ce qu'implique le travail de gestion des données de la recherche. Bien que ces entretiens ne sont en aucun cas un portrait exhaustif du travail sur les données de la recherche en France, ils m'auront permis de nourrir ma réflexion ainsi que de m'aider à mieux comprendre les pratiques d'aujourd'hui.

Ce sont donc mes lectures et ces entretiens qui me permettent cette plongée au cœur du *records management* et des plans de gestion de données. Maintenant, il convient de s'attarder à définir ces deux grands concepts.

LE *RECORDS MANAGEMENT*

Plus que de la gestion de documents d'activité

Le *records management*, appelé gestion des documents d'activité en français[2], est une démarche ayant pour but d'identifier les documents probants (ou engageants)

[2] Les termes *records* et *records management* ont été traduit dans la nouvelle version de l'ISO 15489 par « document d'activité » et « gestion des documents d'activités ». Toutefois, comme ces concepts se basent sur une longue tradition anglo-saxonne, ont déjà fait l'objet d'autres traductions plus ou moins formelles (au Québec, par exemple, *records*

(cc) BY-NC-ND

au sein d'une activité, puis d'assurer la conservation de leurs fiabilité, authenticité, intégrité et exploitabilité (ces termes seront définis ci-dessous) tout au long de la période où ceux-ci restent engageants, soit leur durée de conservation. Cette période est souvent nommée durée d'utilité administrative (DUA) dans le secteur public. C'est une démarche de gestion qui débute en amont du processus de création et de gestion documentaire. Il y a plusieurs objectifs :

- Rendre compte de ses activités – les entreprises s'engageant dans une démarche de *records management* le font en partie afin de conserver les preuves de bonne tenue de leurs activités et chemin faisant, cette démarche permet d'identifier là où l'entreprise ne répond pas à ses obligations, si c'était le cas. Ceci est particulièrement important en cas de litige (notamment en cas de procédures judiciaires) afin de pouvoir présenter des documents capables d'appuyer la position d'une entreprise dans le conflit ;
- Capitaliser sur les connaissances de l'entreprise – contrairement à certaines idées reçues, le numérique n'a pas simplifié la production et la gestion documentaire d'une entreprise. Il n'est pas rare qu'une entreprise produise beaucoup d'informations, au point d'en perdre le fil. Le *records management* permet d'identifier les documents clés dans les processus de l'entreprise et de les gérer tout au long de leur cycle de vie dans l'optique d'arriver à une bonne gouvernance de l'information. Tout cela permet de prendre des décisions éclairées ;
- Évaluer et simplifier les activités – la première étape d'une démarche de *records management* est d'observer, comprendre et éventuellement améliorer les processus faisant partie des activités d'une entreprise, dans le but de mieux rendre compte des activités. Ceci permet au *records manager* de reconnaître où les processus peuvent être lacunaire, intervenir, et éventuellement clarifier les responsabilités envers chaque tâche d'un processus, tout en gardant en tête l'importance de produire des documents conformes à la réglementation, tant dans leur contenu que dans leur forme.
- Pouvoir obtenir le bon document, par la bonne personne, au bon moment – un bon système de *records management* en est un où les processus sont

management était traduit par « gestion des archives courantes et intermédiaires ») et que la traduction de l'ISO est contesté, la forme originale, en anglais, sera privilégié pour le reste du texte.

(cc) BY-NC-ND

clairement identifiés, les responsabilités énoncées, où l'on connaît les états du document et l'on s'assure que ceux-ci ne peuvent être vus que par les personnes qui y sont autorisées.

Cette pratique vient un peu bouleverser la vision traditionnelle que l'on a de l'archive ou de l'archiviste. Alors que le document parvient à l'archiviste une fois son cycle de vie terminé, le *records manager*, lui, est plutôt dans l'anticipation du cycle de vie.

Historique du *records management*

Cette démarche et ses finalités découlent du contexte dans lequel est né le *records management*. Cette pratique est « née », en partie, pendant l'entre-deux-guerres aux États-Unis. L'administration américaine voyant sa taille augmenter rapidement – et avec celle-ci, le volume de sa production documentaire – considéra prendre des mesures afin de contrôler cette production et d'en réduire le fardeau. Dans les années 30, les archives nationales ont établi un « *records administration program* » afin de faciliter le transfert des *records* aux archives. Cette solution s'est toutefois révélée insuffisante et plus tard furent implantées des unités de gestion des *records* dans les différents services de l'administration. Puis, ces unités ont fusionné avec les archives nationales alors que la Commission Hoover recommande en 1948 que ces dernières soient incorporées dans un bureau de *records management*[3].

On retrouve une démarche semblable après la deuxième Guerre mondiale en Angleterre, où le *Public Records Office* (P.R.O.), se retrouve désemparé face au volume de production documentaire de l'administration anglaise. Une commission est mise en place et publiera ses conclusions en 1954. Ceci amènera à mettre en place un système de gestion des documents tout au long de leur cycle de vie. Chaque document produit devra être évalué au moins deux fois. Une fois près du moment de sa création, par les ministères eux-mêmes, afin de déterminer si le document possède un intérêt administratif quelconque. Puis, une deuxième fois, 20 à 25 après sa création, et cette fois avec l'aide d'un représentant du P.R.O. qui pourra alors déterminer si le document représente vraiment un intérêt historique ou non. C'est

[3] Cox, Richard J. *Closing an era : Historical perspectives on Modern Archives and Records Management*. Westport et Londres : Greenwood Press, 2000, p. 2-3

(cc) BY-NC-ND

seulement alors que ces documents pourront être transmis au P.R.O.[4] Comme aux États-Unis donc, on tente d'appréhender les documents avant leur dépôt aux archives.

Au fil du temps, le *records management* s'est bien implanté dans la pratique administrative anglo-saxonne et est, pour la première fois au monde, décrit sous forme de norme en Australie en 1996 : la norme AS 4390-1996. Un an plus tard, ce même pays proposera à l'*International Organization for Standardization* (ISO) d'en faire une norme internationale. C'est alors que l'Association française de normalisation (AFNOR) créera un comité rassemblant des experts de la direction des Archives de France, de la Bibliothèque nationale de France (BNF), de l'Association des archivistes français (AAF) ainsi que de l'Association des professionnels de l'information et de la documentation (ADBS) pour travailler à ce sujet au sein du sous-comité 11, rattaché à la commission générale 46 « Information et documentation »[5] de l'ISO.

En France, la pratique se développe plutôt lors de la fin des années 90 et le début des années 2000. En plus des travaux du CN11, qui mèneront à la publication en 2001 de la traduction française de la norme australienne, le moment en lui-même était très opportun :

> « Le ‘’records management’’ est apparu dès la fin des années 1990 comme une opportunité de rapprocher les pratiques de ces deux métiers [documentaliste et archiviste] et de redonner une légitimité à l'intervention des professionnels sur l'ensemble du cycle de vie documentaire, y compris sur l'organisation des documents produits en fonction des activités de l'entité concernée et des processus métiers mis en œuvre[6]. »

C'est alors vu comme une véritable opportunité[7] pour les travailleurs de l'information de démontrer concrètement leur savoir-faire ainsi que de le partager avec d'autres métiers.

---

[4] Pérotin, Yves « Le *records management* et l'administration anglaise des archives » **In :** *La Gazette des archives* **[en ligne]**, 1964 (n°44), p.6-7. Disponible sur : http://www.persee.fr/doc/gazar_0016-5522_1964_num_44_1_1753?q=records+management+et+l%27administration+anglaise (consulté le 24 mars 2018)

[5] Cottin Michel, Dessolin Baumann Sylvie. « La famille des normes ISO sur le records management » **In :** *La Gazette des archives* **[en ligne]**, 2012/4 (n°228). Normalisation et gestion des documents d'activité (records management) : enjeux et nouvelles pratiques pour notre profession. pp. 119-133. Disponible sur : www.persee.fr/doc/gazar_0016-5522_2012_num_228_4_4989 (consulté le 07 février 2018)

[6] « Présentation » **In :** *Documentaliste-Sciences de l'Information* **[en ligne]**, 2009/2 (Vol. 46), p. 28. Disponible sur : https://www-cairn-info.docelec.insa-lyon.fr/revue-documentaliste-sciences-de-l-information-2009-2-page-28.htm (consulté le 07 février 2018)

[7] Dessolin-Baumann Sylvie, Wiegandt Caroline, « Un métier en devenir » **In :** *Documentaliste-Sciences de l'Information* **[en ligne]**, 2009/2 (Vol. 46), pp. 30-31. Disponible sur : https://www-cairn-info.docelec.insa-lyon.fr/revue-documentaliste-sciences-de-l-information-2009-2-page-30.htm (consulté le 07 février 2018)

(cc) BY-NC-ND

De plus, il est possible d'avancer que l'avènement du numérique ainsi que le développement de l'administration électronique aient contribué à faire développer la pratique en France. La loi du 13 mars 2000[8] indique qu'il est possible de reconnaître la valeur juridique des documents électroniques, à condition que l'on puisse bien identifier la personne dont émanent ces documents et que ceux-ci soient conservés de façon à en garantir leur intégrité. Bien que l'on ne reconnaîtra la valeur probante de la copie électronique qu'avec l'ordonnance du 10 février 2016, cela permet, dès le début des années 2000, d'envisager de conserver les documents nés numériques tels quels, sans les imprimer, afin d'économiser sur les coûts de stockage des documents papier. Toutefois, les documents numériques étant plus fragiles à préserver et plus facilement modifiables par autrui, il devient de plus en plus important pour l'archiviste (et par extension le service producteur du *records*) de contrôler le document lors de tout son cycle de vie, et notamment de se rapprocher du moment de la création du document afin d'assurer des conditions de conservation optimale. La pratique du *records management* consistant à prendre en compte le document en amont de sa création était donc une méthode intéressante afin de faire face à ces nouveaux défis.

Quelques clés pour mieux comprendre le *records management*

Afin de mieux comprendre cette démarche, cette prochaine partie s'appuiera sur la norme *ISO 15489 Information et documentation – Gestion des documents d'activités*. Dans sa première version de 2001, celle-ci comprenait deux parties. Une première afin de poser les concepts et les principes de la démarche, révisée en 2016 et dont nous nous servirons ici pour définir les concepts clés du *records management,* et une deuxième partie, abandonnée en 2016, qui définissait les lignes directrices pour la mise en place d'une telle démarche, dont nous parlerons un peu plus loin.

Qu'est-ce qu'un record *?*

Historiquement, le *records management* vient de la volonté de contrôler en amont le cycle de vie des documents. Cette différence avec la gestion des archives

[8] Loi no 2000-230 du 13 mars 2000 portant adaptation du droit de la preuve aux technologies de l'information et relative à la signature électronique **[en ligne]**. JORF n°62 du 14 mars 2000 page 3968. Disponible sur : https://www.legifrance.gouv.fr/affichTexte.do?cidTexte=JORFTEXT000000399095&categorieLien=id (consulté le 13 août 2018)

(cc) BY-NC-ND

telle qu'elle a été envisagée en France jusqu'à récemment est aussi bien illustrée avec la définition que l'on fait des archives, versus celle que l'on a des *records*. Alors que les premières sont définies dans la loi française comme suit : « Les archives sont l'ensemble des documents, y compris les données, quels que soient leur date, leur lieu de conservation, leur forme et leur support, produits ou reçus par toute personne physique ou morale et par tout service ou organisme public ou privé dans l'exercice de leur activité.[9] » La norme *ISO 15489 Information et documentation – Gestion des documents d'activités*, elle, définit les *records* comme suit : « informations créées, reçues et préservées comme preuve et actif par une personne physique ou morale dans l'exercice de ses obligations légales ou la conduite des opérations liées à son activité[10] ». Bien que les deux définitions évoquent la production documentaire dans le cadre d'une activité, on constate que la définition de *records* sous-entend qu'un tri sera fait parmi l'ensemble des documents, pour ne gérer que ceux qui ont une nature de preuve ou d'actif informationnel, afin de définir en amont quels seront les documents à archiver.

Toutefois, ces définitions laissent une certaine liberté quant à la forme que peuvent prendre les documents, tant qu'elle est conforme aux réglementations applicables au *record* en particulier. Ceux-ci peuvent donc avoir une apparence multiple, ils peuvent être lettre, mail, formulaire, compte-rendu, facture, décision, données, etc. Cette forme changera selon le contexte : « Glossaries show that a 'record' to a computer scientist is a group of fields within a table and database ; to a lawyer it is a financial document or an account of proceedings in a case ; to a telecommunications officer it is the unit of data that is transmitted from sender to receiver.[11] » C'est le contenu, les éléments s'articulant sur ce document qui lui donne toute sa valeur de preuve ou d'actif. Le *record* est donc la représentation physique ou numérique d'une information probante. Le contenu prendra ensuite la forme demandée par l'exécution de l'activité et la réglementation le touchant.

Pour être véritablement *record*, la fonction de représentation d'une information probante n'est pas la seule caractéristique qui importe. Puisqu'ils se

[9] *Article L211-1*. **[en ligne]** Code du patrimoine. Disponible sur : https://www.legifrance.gouv.fr/affichCode.do;jsessionid=986188EC1F33C71FEB56F633683887FE.tplgfr30s_3?idSectionTA=LEGISCTA000006159940&cidTexte=LEGITEXT000006074236&dateTexte=20180813 (consulté le 13 août 2018)

[10] ISO, *Information et documentation - Gestion des documents d'activité Partie 1 : concepts et principes = Information and documentation - Records management Part 1 : Concepts and principles*. ISO 15489-1. Paris : AFNOR, 2016, p. 3

[11] Williams, Caroline. « Records and archives : concepts, roles and definition. » **In :** Brown, Caroline (edited by). *Archives and Recordkeeping : Theory into practice.* Londres : Facet Publishing, 2014, p. 10

(cc) BY-NC-ND

doivent d'agir à titre de preuve d'un évènement ou d'une activité, il est essentiel que l'on puisse instiller une confiance envers ces documents. La norme ISO 15489 base cette confiance sur quatre caractéristiques fondamentales : l'authenticité, l'intégrité, la fiabilité et l'exploitabilité.

Authenticité, fiabilité, intégrité, exploitabilité

La caractéristique d'authenticité du document renvoie à ses origines. A-t-il bien été créé au moment où il prétend avoir été créé ? A-t-il vraiment été créé par la personne qui prétend l'avoir fait ? Est-il vraiment ce qu'il prétend être ? Au minimum, l'auteur, la date et le contenu du document doivent donc être clairement identifiés sur un *record* (ou dans ses métadonnées) afin que l'on puisse le qualifier d'authentique.

En plus de la présence de ces éléments, on doit pouvoir s'assurer que ceux-ci (ainsi que le *record* lui-même) n'aient pas été altérés ou perdus ; ou à tout le moins, s'il y a effectivement eu une altération, qu'elle ait été documentée et expliquée, c'est-à-dire justifiée. En un mot, on doit pouvoir s'assurer de l'intégrité du document.

De plus, le document doit aussi être une représentation complète et exacte de l'évènement auquel il se rapporte. Par conséquent, il est indispensable que celui-ci soit créé au plus près de l'évènement lui-même. Ces éléments mèneront à la création d'un document fiable.

Finalement, un *record* doit être exploitable. Dans un premier temps parce que son authenticité, son intégrité et sa fiabilité font que l'on peut vraiment s'en servir à titre de preuve, puis parce qu'il doit être facilement retrouvable et que l'on puisse le restituer dans un délai de temps raisonnable.

Autour du record : les métadonnées

Afin d'assurer ces caractéristiques d'authenticité, de fiabilité, d'intégrité et d'exploitabilité du *record,* celui-ci doit s'accompagner de métadonnées capables de nous démontrer ces caractéristiques. Par exemple, si l'on souhaite s'assurer de la bonne provenance d'un document, il convient que celui-ci soit accompagné de la signature d'une autorité compétente. Si l'on souhaite être capable de vérifier son caractère non altéré, cela peut signifier qu'il faut accompagner le document d'un

(cc) BY-NC-ND

calcul d'empreinte prouvant que le document n'a pas subi de modifications depuis son enregistrement. Il existe plusieurs types de métadonnées que l'on peut associer à un *record*, et la norme ISO 15489 en identifie principalement deux catégories.

D'abord, les métadonnées de gestion servent à connaître et comprendre le contexte dans lequel le document est créé :

- « le contexte opérationnel;
- les dépendances et les relations entre documents d'activité et systèmes documentaires;
- les relations avec les contextes légal et social;
- les relations avec les acteurs qui créent, gèrent et utilisent les documents d'activité[12]. »

Puis, la norme précise aussi quelles métadonnées doivent être associées au document, précisant au passage leur caractère crucial afin d'en faire des documents probants :

- « la description du contenu du document d'activité;
- la structure du document d'activité (par exemple, sa forme, son format et les relations entre ses composantes constitutives);
- le contexte opérationnel dans lequel le document d'activité a été créé ou reçu et utilisé;
- les relations avec d'autres documents d'activité et d'autres métadonnées;
- les identifiants et autres informations nécessaires à sa récupération et à sa communication, comme les informations de format ou de stockage;
- les actions et les évènements affectant le document d'activité tout au long de son existence (en incluant la date l'heure des actions, les changements apportés aux métadonnées et les acteurs entreprenant ces actions)[13]. »

La norme ne précise pas quelle forme exactement ces métadonnées doivent prendre ainsi que la façon précise dont elles doivent être structurées. Un *records manager* pourra donc s'appuyer sur d'autres normes ou standard afin de créer et

[12] ISO, *Information et documentation - Gestion des documents d'activité*. *Op cit.* p. 5-6

[13] *Ibid.* p. 6

(cc) BY-NC-ND

gérer ces métadonnées. Citons par exemple la norme *ISO 23081 Métadonnées pour les documents d'activités*, MoReq2, Dublin Core, EAD, etc.

Comme les métadonnées permettent de démontrer le caractère probant des documents, il est évident que ces dernières doivent être gérées selon les mêmes standards afin de conserver le caractère probant des documents.

Gérer les documents probants et leurs métadonnées

Le système de gestion des documents d'activités (SGDA)

Plus globalement encore, les documents et leurs métadonnées doivent être intégrés dans un système qui permet de garantir les caractéristiques d'authenticité, fiabilité, intégrité et exploitabilité. Attention, il est important de préciser ici que le mot système n'est pas un synonyme de logiciel ou de système d'exploitation ; le système de gestion des *records* n'a pas à être exclusivement informatique. Toutefois, celui-ci doit posséder, au minimum, un certain nombre de caractéristiques énoncées par la norme *ISO 30300 Information et documentation - Systèmes de gestion des documents d'activité – Principes essentiels et vocabulaire* :

- Fiable

Tout comme la fiabilité des *records* est indiquée par leur caractère complet et exact, le SGDA doit lui aussi traiter les documents probants de façon complète et exacte. C'est-à-dire qu'il doit prendre en compte l'ensemble des *records* relevant de son domaine d'application, mais aussi les organiser aussi de façon cohérente selon les contextes de production et ainsi garantir la confiance des utilisateurs en protégeant les documents et en intégrant les informations nécessaires à une bonne gestion du cycle de vie des documents. En somme, « les systèmes documentaires fiables permettent la continuité de l'activité et la gestion des risques[14] ».

- Sécurisé

[14] *ISO, Information et documentation - Systèmes de gestion des documents d'activité – Principes essentiels et vocabulaire = Information and documentation – Management system for records – Fundamentals and vocabulary.* ISO 30300. Paris : AFNOR, 2011, p. 4

(cc) BY-NC-ND

Le système doit pouvoir protéger les *records* des modifications abusives, par exemple en gérant les droits d'accès sur ces derniers et en conservant une trace des altérations sur les *records*.

- Conforme

Le système doit pouvoir répondre aux exigences de l'organisme en matière de standards, normes, attentes des parties prenantes, etc.

- Exhaustif

Le système ne peut laisser de côté des documents ou des informations. Le cas contraire risquerait d'entraîner de la confusion par rapport à ce qui doit, ou non, être traité.

- Caractère systématique

Tous les documents pertinents doivent automatiquement être traités par le système. Afin d'instaurer cet automatisme, le système doit être régi par des règles et politiques claires, et les responsabilités de chacun envers les documents et leur gestion doivent être bien établies.

Afin de posséder toutes ces caractéristiques, le système de gestion documentaire ne peut être qu'un outil créé et utilisé ponctuellement. La norme ISO 30300 (tout comme l'ISO 15489) est très claire sur ce point. Il est important d'avoir une volonté claire de la direction afin de créer des politiques et des méthodes de travail qui placent ce système au cœur des activités de l'entreprise. C'est aussi pourquoi la norme met en avant l'approche par processus afin de déterminer quels sont les *records*, comment ceux-ci doivent être gérés et intégrés au SGDA. Comme le montre la figure 1 ci-dessous, c'est en analysant quelles sont les activités de l'entreprise, dans quels contextes ces activités se déroulent et en tenant compte des intérêts des différentes parties prenantes que l'on peut s'assurer de se conformer aux exigences du milieu et ainsi créer des documents qui serviront de preuves pour les activités.

(cc) BY-NC-ND

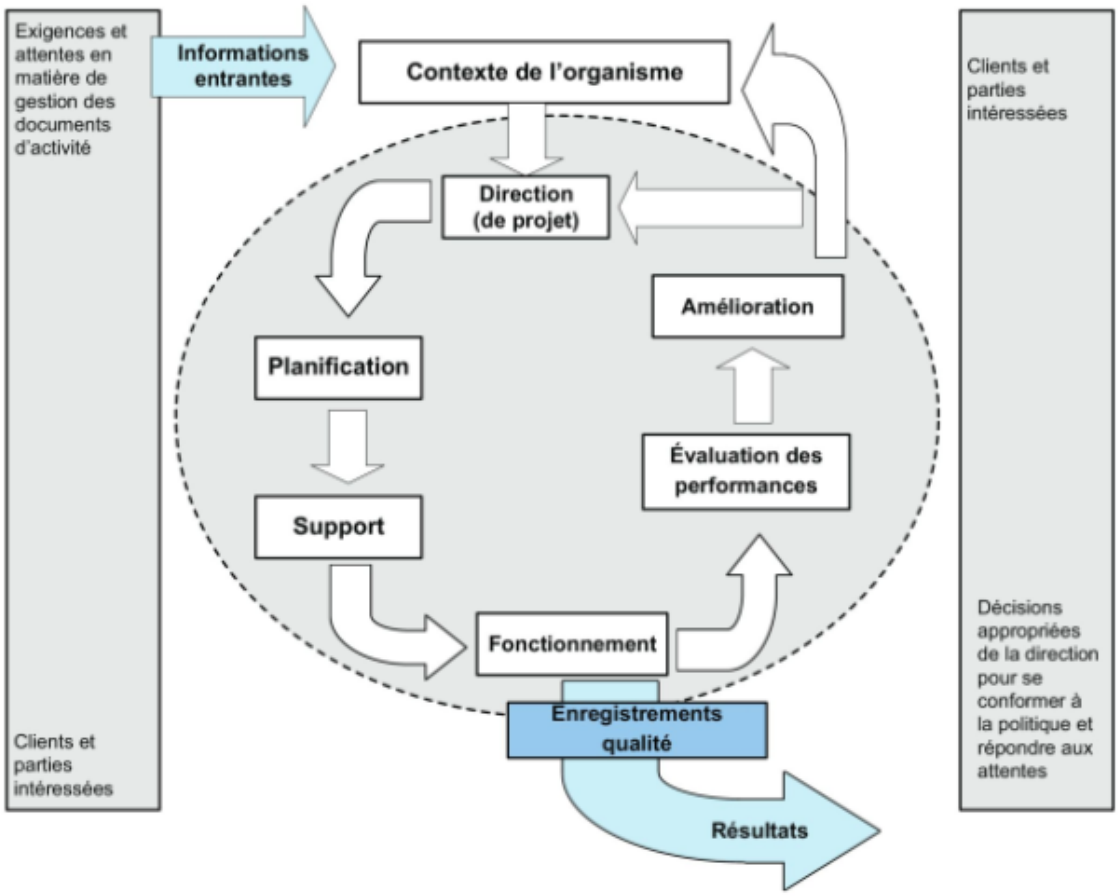

Figure 1 – Structure d'un SGDA selon la norme ISO 30300[15]

L'analyse de processus

Il existe différentes méthodologies d'analyse de processus. Dans la famille des normes d'information et de documentation, la norme ISO 26122 *Analyse des processus pour la gestion des informations et documents d'activités* propose deux méthodes afin de procéder à une analyse des processus : l'analyse fonctionnelle et l'analyse séquentielle.

L'analyse fonctionnelle consiste à identifier, en partant du plus général au plus spécifique, toutes les composantes faisant partie d'un processus. Par exemple, pour le processus « Écrire un mémoire », on pourrait identifier plusieurs activités : recherche bibliographique, rédaction, révision. Chacune de ces activités se divise ensuite en transaction : lecture, rédaction d'une fiche de lecture, sélection des références…

[15] ISO, « Structure d'un SGDA ». (2011) **[schéma]. In :** *ISO, Information et documentation - Systèmes de gestion des documents d'activité – Principes essentiels et vocabulaire*. *Op. cit.* p. 6

(cc) BY-NC-ND

L'analyse séquentielle cherche plutôt à analyser le déroulement chronologique d'une activité. Dans celle-ci, on y décrit tous les éléments composant un processus, à la suite des uns et des autres (selon le degré de précision souhaité). Pour reprendre notre exemple, dans le processus « Écrire un mémoire », on commencerait par l'activité sélection d'un sujet, puis recherche bibliographique, sélection des références, lectures, rédaction des fiches de lecture, et ainsi de suite.

Ces deux types d'analyse ne sont pas nécessairement opposables. Un *records manager* peut faire le choix d'en utiliser une ou de les combiner. Dans les deux cas, il est important de définir le périmètre de l'analyse, c'est-à-dire de déterminer si on l'applique à une seule activité, à une division de l'entreprise ou à l'organisation entière. Le but de l'analyse est d'identifier les éléments clés d'un processus, et précisément les moments où est créée l'information à conserver afin de s'assurer que les documents probants portent bien les traces de la conformité aux règlements encadrant lesdites activités. Il sera au *records manager* d'ajuster si besoin le processus ou les documents (leur forme ou leur contenu, selon les cas) qui en émanent afin de s'assurer que le *record* conservera ses caractéristiques d'authenticité, de fiabilité, d'intégrité et d'exploitabilité tout au long de son cycle de vie.

Les outils du records management

En plus du SGDA, le *records manager* doit s'aider de plusieurs outils afin de gérer les documents et leur cycle de vie :

- Référentiel de gestion des documents / tableau de gestion

Cet outil essentiel permet de recenser les différentes typologies de *records* qui, souvent, émanent de l'analyse des processus et d'indiquer leur sort final (élimination, conservation, tri). On peut aussi l'enrichir en précisant les étapes de validation du *record*, ses valeurs de conservation (historique, juridique, informationnelle), etc. Cet outil peut servir au *records manager* à réunir le détail des cycles de vie des documents en un seul endroit ; il peut aussi être utile afin de faire la promotion du *records management* dans l'entreprise ou faciliter les usages en le diffusant parmi les salariés qui peuvent ainsi prendre en main une partie du cycle de vie du document.

(cc) BY-NC-ND

- Plan de classement des activités

Cet outil, aussi basé sur l'analyse de processus, permet de décliner les groupes de processus, les processus, les activités liées à ces derniers ainsi que leurs transactions. Ce plan de classement des activités peut servir de point de références pour un organisme, afin de préciser les missions de chaque service et comment celles-ci sont exécutées. Plus encore, ce plan peut servir de base à la création d'un plan de classement dans un système informatique.

- Référentiel de métadonnées

De façon semblable au référentiel de gestion des documents qui démontre le lien entre les documents et les activités dont ils émanent, le référentiel de métadonnées effectue le lien entre les métadonnées et les documents auxquels elles se réfèrent. Ce référentiel pourra inclure des règles d'utilisation (quelles métadonnées doivent être créées lors de la création d'un document ? lors de sa capture ? comment doivent-elles être créés?).

Ces éléments font partie de la boîte à outils du *records manager*, afin de lui permettre de contrôler la façon dont seront produit les documents en amont.

Ainsi qu'on a pu le constater, le *records management* est une démarche anglo-saxonne à la base, mais qui s'adapte peu à peu à la réalité française. Celle-ci offre un avantage indéniable pour analyser la production documentaire d'un organisme, afin d'en définir les besoins de conservation et de gestion en amont. Effectivement, il s'agit ici de créer des règles de gestion claires et définies, qui pourront être mises en place et répétées aussi longtemps/souvent que nécessaire.

METTRE EN PLACE UN SYSTÈME DE *RECORDS MANAGEMENT*

Une fois que le *records management* a été identifié comme étant un système utile à mettre en place pour une organisation vient l'étape de l'application. Comment procède-t-on ? Pour ce faire, il existe plusieurs guides pratiques ou lignes directrices qui ont été publiés. Dans notre cas, nous nous appuierons sur la norme ISO 15489-2 ainsi que le mode d'emploi publié en 2000 par l'ADBS pour que l'on puisse identifier les éléments importants de la mise en place d'un système de RM.

(cc) BY-NC-ND

La norme ISO 15489-2

Jusqu'à la révision de la norme ISO 15489 en 2016, celle-ci se déclinait en deux parties : une présentant les concepts, puis l'autre présentant un guide pratique pour mettre en place le *records management*. Ce guide était basé sur la méthode australienne *Designing and Implementing Record Keeping System* (DIRKS)[16].

Ce guide pratique présente huit étapes pour mettre en place un tel système dans une organisation. Elles sont résumées ainsi[17] :

- Étape A - Enquête préliminaire : Cette première étape consiste à comprendre comment les contextes administratifs, légaux, économiques et sociaux peuvent influencer notre organisation et la façon dont on aura à gérer les *records*. Celle-ci permet aussi d'identifier les points forts et les points faibles actuels de l'entreprise par rapport à une démarche de *records management*.
- Étape B - Analyse des activités : Cette étape consiste bien souvent à réaliser une analyse de processus tel que nous en avons parlé plus haut. Cette analyse permet de bien comprendre comment un organisme réalise ses activités, d'identifier les informations créées par ces activités et surtout de connaître leur forme et leur contenu.
- Étape C - Identification des exigences archivistiques : Les exigences archivistiques sont définies par « les besoins de l'organisme en matière de création (et/ou acquisition), de gestion, d'organisation et de conservation des documents. Elle [l'étape] doit permettre d'identifier les documents qui doivent être créés, sous quelle forme, et comment ils doivent être gérés et conservés.[18] »
- Étape D - Évaluation des systèmes existants : Ici, il s'agit d'évaluer les systèmes d'archivage et de gestion documentaire existant afin de voir si ceux-ci peuvent convenir au système de *records management*.

[16] NSW State Archives. *DIRKS Methodology and Manual.* **[en ligne]**. Disponible sur : https://www.records.nsw.gov.au/recordkeeping/advice/dirks/methodology (consulté le 31 juillet 2018)

[17] ISO, *Information et documentation - "Records management" Partie 2 : guide pratique = Information and documentation - "Records management" Part 2 : guidelines.* ISO 15489-2. Paris : AFNOR, 2002, p. 4-7

[18] Couture, Cynthia., Roy, Julie. « La norme ISO 15489 : principes et application » **[en ligne]**. *Archives*, 2006-2007/2 (vol. 38) p. 159. Disponible sur : https://www.archivistes.qc.ca/revuearchives/vol38_2/38_2_Couture_Roy.htm (consulté le 10 janvier 2018)

(cc) BY-NC-ND

- Étape E - Identification de la stratégie pour la satisfaction des exigences archivistiques : C'est lors de cette étape que l'on définit la politique qui sera mise en place au sein de l'organisme ainsi que les procédures y découlant.
- Étape F - Conception d'un système d'archivage : Une fois que les stratégies et les procédures générales ont été décidées, cette étape consiste à planifier comment, concrètement, ces systèmes seront mis en œuvre.
- Étape G - Mise en œuvre d'un système d'archivage : Cette étape peut être résumée comme étant le déploiement du système de *records management*.
- Étape H - Contrôle a posteriori : L'évaluation et le contrôle est une part importante du *records management*. Une fois que le système a été déployé, il est important de contrôler que celui-ci fonctionne bien, tant pour produire des *records* de qualité que dans la façon dont il est approprié par les acteurs qui l'exécutent.

Bien que ce guide ait été abandonné par le comité ISO responsable de la norme, celui-ci semble fournir un bon outil sur lequel s'appuyer lorsque l'on met en place une méthode de *records management*. Il a, par exemple, servi de guide à l'implémentation du RM à la BnF[19] et, comme le notent Cynthia Couture et Julie Roy : « C'est une façon pour les archivistes et gestionnaires de documents québécois de s'assurer de prendre des décisions éclairées, de mesurer l'état de l'arrimage de leurs pratiques à celles de la communauté internationale, mais également de s'appuyer sur un aide-mémoire non négligeable.[20] »

Bien que cette méthode d'implémentation ait profité d'une reconnaissance internationale, ce n'est évidemment pas la seule façon de mettre en œuvre le *records management*.

Le mode d'emploi de l'ADBS

L'ADBS publiait en 2000 le livre *Records management : Mode d'emploi*[21] qui, sans être un guide étape par étape, a été structuré de façon à ce qu'on puisse l'utiliser

[19] Dhérent, Catherine. « Le pilotage et l'évaluation du système de *records management* de la Bibliothèque nationale de France » **[en ligne]**. *La Gazette des archives*, n°216, 2009-4. Chiffrer et évaluer les archives. p. 131-144. Disponible sur : http://www.persee.fr/doc/gazar_0016-5522_2009_num_216_4_4647 (consulté le 27 avril 2018)

[20] Couture, Cynthia., Roy, Julie. « La norme ISO 15489 : principes et application ». *op. cit.*, p.171

[21] Drouhet, Geneviève., Keslassy, Georges. et Morineau, Elisabeth. *Records management : mode d'emploi*. *op. cit*.

(cc) BY-NC-ND

pour s'orienter dans l'implantation d'un système RM. Sachant que la norme ISO sur le *records management* a été publiée pour la première fois en 2002, il est intéressant d'observer la façon dont on a traité la question en France, alors que le *records management* était encore peu connu. Après une introduction sur les concepts et les avantages du RM, ce mode d'emploi est divisé en deux parties : la stratégie et la mise en œuvre.

Dans la première partie, celle sur la stratégie, ce guide décrit ce qu'est un système de RM, quelles en sont les clés de réussites, les responsabilités à attribuer aux différentes parties prenantes (direction, *records manager*, collaborateur, DSI, etc.) ainsi que les métiers qui doivent intervenir dans le RM. Puis, on y décrit l'analyse de l'organisme par activités, comment organiser le plan de classement par activités, l'importance d'utiliser un vocabulaire normé au sein de l'organisme. Ensuite, on explique comment bien identifier et indexer un document… et donc comment choisir les métadonnées qui les accompagneront. La question de la conservation est abordée par la suite, et on y revoit quels sont les enjeux et les points critiques auxquels tenir compte lorsque l'on planifie la conservation de documents papiers, d'archives mixtes, de documents numériques ainsi que de courriels. Cette partie se conclut sur les systèmes informatiques et le SGDA en rappelant les caractéristiques essentielles (interopérabilité, ergonomie, sécurité, etc.) et les fonctionnalités qu'ils doivent posséder (gestion des droits, capture et traitement des documents, outils de recherche, possibilité de communiquer certains documents, stockage, récolement, destruction).

La deuxième partie sur la mise en œuvre commence par aborder les procédures de gestion, en rappelant celles faisant typiquement partie de l'archivage « classique » et en précisant ce qui est modifié lorsqu'un système de RM est mis en place (emphase sur la création et la capture du document, le versement devient moins important). Pour chacune, elle rappelle les points à prendre en compte lors de l'élaboration des procédures par un *records manager*. Par la suite, une attention particulière est portée à la capture des documents dans le système : quelles informations doit-on capturer, mais aussi comment s'assurer que notre système exécutera la capture correctement ? Cette section indique donc les spécificités techniques à prendre en compte, mais rappelle que le facteur humain a lui aussi son importance. Puis, l'ouvrage nous présente une sous-partie sur les « archives vitales » soient celles qui sont indispensables au fonctionnement d'une entreprise ou d'une

(cc) BY-NC-ND

organisation. Cette partie explique ce que sont les archives vitales et comment les reconnaître. Finalement, les deux dernières parties s'intéressent à la formation des collaborateurs et à l'audit du système.

Ce mode d'emploi du *records management* ne présente donc pas un guide étape par étape, comme le fait la norme ISO 15489-2, mais offre plutôt une vue globale du sujet avec un accent sur différents aspects de la démarche.

Des deux démarches, on retiendra d'abord qu'il est important de bien connaître le milieu dans lequel on souhaite introduire une démarche de *records management*. Il faut connaître les enjeux politiques, légaux, économiques sur lesquels le système pourrait avoir un impact et les éléments sur lesquels il est possible de s'appuyer pour faciliter l'implantation de la démarche. Cette évaluation permettra aussi de connaître quelles règles un *records manager* devra appliquer pour les documents (durée de conservation, information à inclure lors de la capture, etc.). Ensuite, les deux guides insistent sur l'importance de bien connaître les activités de l'organisation dans lequel on souhaite implanter le RM. Puis, il est important de bien planifier le système que l'on créera pour gérer les *records*, afin de s'assurer que celui-ci sera fiable, sécuritaire, et autres… Lors de sa mise en œuvre, il est important de former les collaborateurs (mise en œuvre du système) car c'est eux qui pourront assurer le succès d'une telle démarche. En dernier lieu, l'évaluation régulière du système est un élément crucial, puisqu'un système de RM ne peut être complètement figé dans le temps, mais doit pouvoir évoluer au fil des besoins et des circonstances changeantes.

(cc) BY-NC-ND

LES PLANS DE GESTION DE DONNEES

Les plans de gestion des données servent à décrire le cycle de vie des données de la recherche afin d'expliciter « la manière dont seront obtenues, documentées, analysées, disséminées et utilisées les données produites au cours et à l'issue d'un processus ou d'un projet de recherche.[22] » Le plus souvent, cet objectif implique aussi de préparer les données afin qu'elles soient plus facilement partageables et réutilisables (bien que le PGD puisse aussi ne servir qu'à planifier la préservation des données).

L'objet d'une longue réflexion

D'abord, l'accès libre

Si aujourd'hui on constate une volonté de préserver et surtout d'ouvrir les données de la recherche au partage, c'est qu'il y a d'abord eu une véritable réflexion sur la place de la production scientifique dans le domaine public. D'abord, parce que « D'après [*sic*] l'association américaine des bibliothèques de recherche, entre le milieu des années 1980 et le milieu des années 2000, le prix des revues payantes a augmenté 2,5 fois plus vite que le coût de la vie.[23] » Ensuite, parce que la généralisation de l'usage d'Internet et la facilité de partage d'informations que cet outil a rendu possible ont amené plusieurs personnes à se demander pourquoi elles devraient continuer de payer des coûts de plus en plus exorbitants pour accéder aux résultats de recherches pour lesquelles elles payaient déjà à travers l'utilisation de fonds publics pour mener ces mêmes recherches. C'est de ce constat que naît le mouvement *open access*, accès libre ou libre accès en français.

L'*open access* peut être défini de plusieurs façons, et ne signifie pas toujours exactement la même chose selon les personnes qui en font l'interprétation. Plus généralement, l'*open access* a été traduit en français soit comme accès libre ou libre accès. Bien que la différence entre ces deux traductions semble minime, celles-ci reflètent plutôt deux points de vue assez différents. D'un côté, le libre accès fait référence au fait que l'on a éliminé les barrières pour consommer une œuvre. L'œuvre est donc disponible gratuitement quelque part (le plus souvent sur internet)

[22] CARTIER, Aurore, MOYSAN, Magalie et REYMONET, Nathalie, *Réaliser un plan de gestion de données*. Paris : Université Paris Diderot, Sorbonne Paris Cité, Université Paris Descartes, 2015, p. 4.

[23] Suber Peter. « 2. Les motivations » ***In*** *Qu'est-ce que l'accès ouvert ?* **[en ligne]**. Marseille : OpenEdition Press, 2016. Disponible sur : http://books.openedition.org/oep/1605 (consulté le 21 avril 2018)

(cc) BY-NC-ND

et n'importe qui peut y accéder. On insiste donc sur l'accès à l'œuvre, car celle-ci peut toujours être protégée par le droit d'auteur, et donc il y a toujours interdiction de la reproduire, de la réutiliser, etc. L'accès libre, quant à lui, insiste plutôt sur la liberté qu'ont les usagers envers cette œuvre. Non seulement peuvent-ils y accéder librement, mais les usagers ont la possibilité d'en faire à peu près ce qu'ils en veulent[24]. C'est d'ailleurs selon cette idéologie qu'est définie l'*open access* lors des déclarations de Bethesda et de Berlin (toutes deux signées en 2003) qui définit ce concept comme la possibilité que des usagers « copient, utilisent, diffusent, transmettent et affichent l'œuvre publiquement, qu'ils créent et diffusent des œuvres dérivatives dans tout médium numérique ayant un objectif responsable, avec attribution adéquate de l'œuvre à son auteur[25] ». Ces premières définitions, donc, font part d'une volonté des institutions publiques de se réapproprier le travail scientifique.

À noter que les plans de gestion de données s'inscrivent dans le mouvement accès libre, c'est-à-dire que ce document permet d'envisager la structure des données afin que d'autres puissent les réutiliser par la suite.

Du partage des publications vers le partage des données

Puis, l'Organisation de Coopération et de Développement Économiques (OCDE) publie en 2004 une déclaration qui engage les signataires à « œuvrer pour l'établissement de régimes d'accès aux données numériques de la recherche financée sur fonds publics conformes[26] ». L'Union européenne suit le pas en 2007, en publiant une communication souhaitant « lancer un processus politique sur (a) l'accès et la diffusion de l'information scientifique et (b) les stratégies de préservation de l'information scientifique dans toute l'Union européenne[27]. »

Cette communication argumente qu'un plus grand partage des résultats de la recherche scientifique ne peut être que bénéfique à l'innovation de la société

---

[24] Dacos Marin. « Préface à l'édition française » ***In :*** SUBER Peter. *Qu'est-ce que l'accès ouvert ? op. cit.*

[25] Traduction d'une partie des déclarations de Bethesda et de Berlin provenant de : Suber Peter. « 1. Une définition » ***In*** *Qu'est-ce que l'accès ouvert ? op. cit.*

[26] *Déclaration de l'OCDE sur l'accès aux données de la recherche financée par des fonds publics* **[en ligne]**. Paris : OCDE, 2004. Disponible sur : http://openaccess.inist.fr/?Declaration-de-l-OCDE-sur-l-acces (consulté le 14 novembre 2017)

[27] Commission des communautés européennes. *Communication de la commission au parlement européen, au conseil et au comité économique et social européen sur l'information scientifique à l'ère numérique : accès, diffusion et préservation.* Bruxelles : Commission européenne, 2007, p.2 . Disponible sur : http://eur-lex.europa.eu/legal-content/FR/TXT/PDF/?uri=CELEX:52007DC0056&from=FR (consulté le 11 novembre 2017)

(cc) BY-NC-ND

européenne. Concrètement, elle encourage un plus grand partage en accès libre des publications scientifiques ainsi que la préservation et le partage des données de la recherche.

Parallèlement à ces réflexions, les pratiques sont aussi en train de changer. Le paradigme de la publication d'un article scientifique, où les chercheurs collectent les données, les analysent et publient leurs résultats change tranquillement au fil du temps. Les premiers effets se font sentir dans des disciplines travaillant avec un gros volume de données, telles que l'astronomie ou la physique où « un autre circuit se met en place, car les scientifiques vont d'abord rassembler les données éparpillées qu'ils possèdent les uns et les autres, ce qui implique de diffuser et de publier ces données, et ce n'est qu'après cette étape qu'ils entameront leur analyse.[28] » Le monde scientifique se dirige donc vers une ère du partage des données.

C'est un fait qui sera reconnu par la commission européenne, alors qu'un groupe d'experts sur les données scientifiques remet à la vice-présidente de l'agenda numérique un rapport sur le sujet. Celui-ci présente différents scénarios quant à l'avenir des données scientifiques, et recommande différentes actions à prendre afin de se saisir de cette vague déferlante de données[29]. Certaines de ces recommandations se concrétiseront en partie en 2013 lorsque le plan Horizon 2020 introduit l'obligation pour certains domaines de créer des plans de gestion des données afin de recevoir des subventions, afin d'encourager le partage et la réutilisation des données.

Les mouvements « open »

Aujourd'hui, à l'*open access* vient s'ajouter les mouvements d'*open science* et d'*open data*. Pour les résumer brièvement, le premier se définit comme suit : « Open Science is the practice of science in such a way that others can collaborate and contribute, where research data, lab notes and other research processes are freely available, under terms that enable reuse, redistribution and reproduction of the

[28] Banat-Berger Françoise., Aussant Dominique., et Chaudiron Stéphane. « Méthodes techniques et outils », **In** : *Documentaliste-Sciences de l'Information* **[en ligne]**, 2009/1 (Vol. 46), p. 12-17. Disponible sur : https://www-cairn-info.docelec.insa-lyon.fr/revue-documentaliste-sciences-de-l-information-2009-1-page-12.htm (consulté le 28 avril 2018)

[29] [Commission Européenne] High Level-Expert Group on Scientific Data, *Riding the wave : how Europe can gain from the rising tide of scientific data* **[en ligne]**. Union Européenne, octobre 2010, 36 p. Disponible sur : https://ec.europa.eu/digital-single-market/en/news/digital-agenda-unlock-full-value-scientific-data-high-level-group-presents-report (consulté le 20 octobre 2017)

(cc) BY-NC-ND

research and its underlying data and methods.[30] » C'est-à-dire que l'on cherche à donner accès non seulement aux publications, et éventuellement aux données qui accompagnent un projet de recherche, mais carrément à tous les éléments de contexte entourant ce projet.

L'*open data,* quant à lui, est une volonté de rendre disponibles publiquement toutes sortes de données : « Le concept de donnée publique est désormais connu et précis : il s'agit de données complètes, interopérables, libres de droit, dont la réutilisation est gratuite et le format non propriétaire.[31] » Ces données étant souvent produites par des organismes publics, le mouvement *open data* se base sur une réflexion semblable à l'*open access* : si ces ressources sont produites avec des fonds publics, elles devraient aussi être disponibles pour tous.

Tim Berner-Lee, le « père » du web a suggéré un format d'adoption progressif de l'*open data*, sous forme de certification à étoile avec différents niveaux d'ouverture, qui va comme suit[32] :

*	Publiez vos données sur le web (peu importe leur format) avec une licence ouverte
**	Publiez-les en tant que données structurées (par exemple, un document Excel au lieu d'une image scannée d'un tableau)
***	Publiez-les dans un format ouvert et non propriétaire (par exemple, un CSV plutôt qu'un Excel)
****	Utilisez des *uniform resource identifier* (URI) pour désigner des choses dans vos données, afin que les gens puissent faire des références à celles-ci
*****	Liez vos données à d'autres données pour y ajouter du contexte

Tableau 1 - Les niveaux d'ouverture de l'*open data*

Ces deux mouvements concernent les données de la recherche, qui font partie de la « science » et qui sont bien souvent des données produites par des projets

[30] FOSTER. *Open Science Definition* **[en ligne]**. Disponible sur : https://www.fosteropenscience.eu/foster-taxonomy/open-science-definition (consulté le 14 août 2018)

[31] Maday, Charlotte. « L'apport de la gestion des documents d'activité (*records management*) à l'ouverture des données. Réflexions basées sur les pratiques en France. » **In :** Ott, Florence (dir.). *Apport de la gestion documentaire à la gouvernance de l'information*. *Les Cahier du numérique*, 2015/2 (vol. 11), p. 152

[32] *Open Data 5** **[en ligne]**. Disponible sur : https://5stardata.info/fr/ (consulté le 14 août 2018)

(cc) BY-NC-ND

résultant de financements publics. Ainsi, on peut supposer que ces derniers viennent renforcer le principe d'ouverture des données et éventuellement appuiera la propagation de l'usage des PGD.

Donner sens aux « données de la recherche »

Cet intérêt grandissant envers les données en amène plusieurs à se poser la question : mais que sont les données de la recherche exactement ? Puisque la question du partage des données de la recherche se pose depuis maintenant environ 10 ans, on pourrait croire qu'une définition relativement précise et complète a été adoptée par la communauté. Ce n'en est pourtant pas le cas, et cela se comprend. Objets polyformes, les données de la recherche varient selon les disciplines, les sujets de recherche, les projets… Physiques ou numériques, quantitatives ou qualitatives, elles se réinventent avec chaque projet. Plus encore, est-ce que toutes données utilisées dans le cadre d'un projet de recherche sont une « donnée de la recherche » ? On peut imaginer l'exemple d'une expérience ou des données brutes sont recueillies, mais dont une première opération de « nettoyage » a été effectuée par le chercheur afin d'éliminer certaines données jugées comme non pertinentes au projet. Est-ce que ces données éliminées sont des données de la recherche[33] ?

Au fil du temps, de nombreuses personnes ont débattu de la nature des données de la recherche[34] [35] [36], en analysant plusieurs points de vue, ainsi qu'en apportant le leur au débat. Pour ma part, j'en retiendrai une définition, ou plutôt un type de définitions qui semble couramment utilisé par les instances officielles lorsque vient le temps de définir ces objets que sont les données de la recherche. Observons par exemple comment la Commission européenne définit ces dernières :

> « Les données de la recherche désignent les informations, et en particulier les faits ou chiffres collectés pour être analysés et traités pour alimenter des réflexions, discussions ou calculs. Dans un contexte de recherche, ces données sont, par exemple, des statistiques, des résultats

[33] Fayet, Sylvie. « ''Données'' de la recherche, les mal-nommées ». **In** : *Urfistinfo* **[en ligne]** (publié le 15 novembre 2013). Disponible sur : https://urfistinfo.hypotheses.org/2581 (consulté le 05 avril 2018).

[34] *Ibid.*

[35] Schöpfel, Joachim., Kergosien, Eric., et Prost, Hélène. *" Pour commencer, pourriez-vous définir 'données de la recherche' ? " Une tentative de réponse.* **[en ligne]** Atelier VADOR : Valorisation et Analyse des Données de la Recherche ; INFORSID 2017, May 2017, Toulouse, 2017. Disponible sur : http://hal.univ-lille3.fr/hal-01530937/document (consulté le 26 avril 2018)

[36] Gaillard, Rémi. *De l'Open data à l'Open research data : quelle(s) politique(s) pour les données de recherche ?*. **[en ligne]** Mémoire d'étude, diplôme de conservateur de bibliothèque. Villeurbanne : Enssib, 2014. Disponible sur : http://www.enssib.fr/bibliotheque-numerique/documents/64131-de-l-open-data-a-l-open-research-data-quelles-politiques-pour-les-donnees-de-recherche.pdf (consulté le 26 avril 2018)

(cc) BY-NC-ND

> d'expériences, des mesures, des observations sur le terrain, des résultats d'enquêtes, des enregistrements d'entretiens ou des images.[37] »

Ce paragraphe provenant des *Lignes directrices pour le libre accès aux publications scientifiques et aux données de la recherche dans Horizon 2020* reprend la définition « de base » des données (« une donnée est donc un conteneur porteur d'une information ou d'un fragment d'information[38] »), pour y ajouter quelques exemples des formes possibles que peuvent prendre les données dans le cadre de la recherche.

De même, plusieurs autres définitions offertes varient souvent sur cette thématique. Effectivement, celles-ci reprennent l'idée générale de ce qu'est une donnée (une information, un enregistrement) et s'appuient sur le processus dans lequel s'inscrivent les données (la recherche) pour expliciter davantage la forme que devrait prendre cette information. En voici quelques autres exemples :

> L'Association des archivistes français (AAF)
>
> « Les données de la recherche sont des informations, spécimens et matériaux produits, recueillis et documentés. Elles sont collectées ou exploitées à des fins de recherche et de preuves par les chercheurs et leurs équipes. À ce titre, elles constituent une partie des archives de la recherche[39] »

> OCDE
>
> « les [*sic*] ''données de la recherche'' sont définies comme des enregistrements factuels (chiffres, textes, images et sons), qui sont utilisés comme sources principales pour la recherche scientifique et sont généralement reconnus par la communauté scientifique comme nécessaires pour valider des résultats de recherche.[40] »

Ces deux dernières définitions apportent un aspect intéressant, soit les données de la recherche comme preuve des résultats de la recherche. Cette vision est d'ailleurs soutenue par Borgman, professeure des *Information Studies* à UCLA : « [research data are] entities used as evidence of phenomena for the purpose or

---

[37] Direction générale de la recherche et de l'innovation *Lignes directrice pour le libre accès aux publications scientifiques et aux données de la recherche dans Horizon 2020* **[en ligne]**. Bruxelles : Commission européenne, 2016 (v 3.1), p. 4. Disponible sur : http://www.donneesdelarecherche.fr/IMG/pdf/lignes-directrices_libre-acces_horizon2020_version3.1_tr-fr.pdf (consulté le 20 octobre 2017)

[38] Définition du glossaire OAIS dans : Banat-Berger, François. Huc, Claude. *Section 3 – Les multiples visages du document numérique*. **[en ligne]** Portail international archivistique francophone (PIAF), Cours, 2011. Disponible sur : http://www.piaf-archives.org/sites/default/files/bulk_media/m07s03/co/section3_web.html (consulté le 10 janvier 2018)

[39] Définition de l'AAF dans : CARTIER, Aurore., MOYSAN, Magalie., et REYMONET, Nathalie. *Réaliser un plan de gestion de données*. Paris : Université Paris Diderot, Sorbonne Paris Cité, Université Paris Descartes, 2015, p. 3

[40] Définition de l'OCDE dans : Serres, Alexandre (dir.) *Données de la recherche en SHS. Pratiques, représentations et attentes des chercheurs : une enquête à l'Université Rennes 2.* **[en ligne]** Rapport de recherche. Rennes : Université de Rennes 2, 2017, p. 5 Disponible sur : https://hal.archives-ouvertes.fr/hal-01635186 (consulté le 13 décembre 2017).

(cc) BY-NC-ND

research or scholarship[41] ». Les données sont donc vues comme des éléments probants de la recherche.

Il est possible d'y voir un parallèle avec les *records* évoqués plus haut... D'ailleurs, il est très facile d'adapter la définition des *records* pour en formuler convenant aux données de la recherche, et s'apparentant aux définitions de l'AAF ou de l'OCDE : « informations créées, reçues et préservées comme preuve et actif par une personne physique ou morale [lors de la conduite d'un projet de recherche][42] ». J'ajouterai même que les notions de création, réception et préservation des données de cette nouvelle définition font échos à l'objectif des PGD qui est, pour le rappel, d'expliciter « la manière dont seront obtenues, documentées, analysées, disséminées et utilisées les données produites au cours et à l'issue d'un processus ou d'un projet de recherche.[43] » En somme, ce document permet aux chercheurs de voir de façon plus claire qu'elles sont *ses* données de la recherche.

Le contenu d'un plan de gestion des données

La rédaction d'un plan de gestion des données permettra donc de mettre de l'ordre dans les données du chercheur. En premier lieu, il s'agit d'envisager en amont, soit avant la création ou la récupération des données, comment celles-ci seront traitées tout au long du projet ainsi qu'à sa suite. Le PGD est donc normalement rédigé avant le début de la recherche, mais il est fortement recommandé (parfois même exigé) de le réviser pendant le projet de recherche et encore une fois lorsque le projet est terminé.

Pour le programme Horizon 2020, l'idéal est que les données correspondent aux principes FAIR, soit qu'elles puissent être facilement retrouvées (*Findable*), qu'elles soient accessibles (*Accessible*), qu'elles répondent au besoin d'interopérabilité (*Interoperable*) et qu'elles soient réutilisables (*Reusable*)[44]. Ces principes sont nés de l'initiative GO FAIR, une initiative *bottom-up* (approche ascendante) cherchant à guider la création et la gestion des données de la recherche

[41] Borgman, Christine L. *Big Data, Little Data, No Data : Scholarship in the Networked World* **[en ligne]**. Cambridge, London : MIT Press, 2015, 411 p. Disponible sur : http://unr-ra.scholarvox.com/book/88841741 (consulté le 27 octobre 2017). p.29

[42] ISO, *Information et documentation - Gestion des documents d'activité Partie 1 : concepts et principes. Op. cit.* p. 3

[43] Cartier, Aurore., Moydsn, Magalie., et Reumonet, Nathalie. *Réaliser un plan de gestion de données*. *Op. cit* p. 4

[44] Directorate-General for Research & Innovation, *H2020 Programme: Guidelines on FAIR Data management in Horizon 2020* **[en ligne]**. Bruxelles : Commission européenne, 2016 (v 3.0), 12 p. Disponible sur : http://ec.europa.eu/research/participants/data/ref/h2020/grants_manual/hi/oa_pilot/h2020-hi-oa-data-mgt_en.pdf (consulté le 20 octobre 2017)

(cc) BY-NC-ND

afin d'en améliorer le potentiel de partage et de réutilisation. Cette initiative est une réponse aux différents mouvements émergents d'*Open Science*, mais se perçoit comme une façon de mieux gérer toutes les données, et pas seulement celles issues de la recherche[45].

Cette initiative décrit les caractéristiques FAIR comme ceci[46] :

- Retrouvable : Les données (ainsi que leurs métadonnées) doivent être lisibles à la fois par les humains et les machines. Plus spécifiquement, les métadonnées doivent permettre de retrouver les données facilement.
- Accessible : On doit pouvoir accéder aux données rapidement et les droits qui y sont rattachés doivent permettre une réutilisation.
- Interopérable : Les données doivent pouvoir s'intégrer à d'autres données, mais aussi être lues et utilisées par différents systèmes ou applications.
- Réutilisable : Les données doivent être formatées afin de permettre leur réutilisation, que ce soit dans le même contexte ou non. Cela implique de bien documenter les métadonnées.

Les PGD peuvent avoir des objectifs plus larges, et le chercheur peut en tirer divers autres bénéfices. Par exemple : répondre aux exigences de ses financeurs, améliorer l'impact de sa recherche (le PGD peut servir de base à la rédaction d'un *data paper*), économiser du temps, préserver les données, assurer une meilleure qualité des données, promouvoir l'innovation, augmenter la transparence, supporter l'accès libre (ou au contraire s'en servir pour justifier le non-partage de ses données, par exemple dans le cas de données personnelles), aider des institutions ou des pays moins riches à faire de la recherche, faire de la « bonne » science[47].

Ce plan se présente le plus souvent sous la forme d'un document écrit, contenant quelques questions auxquelles le ou les auteur(e(s)) doivent répondre. Ces auteurs peuvent être le ou les chercheur(s) responsables du projet, les documentalistes/archivistes qui accompagnent les chercheurs dans la gestion de leurs données voire même la DSI qui s'occupe des systèmes informatiques en charge

[45] GO FAIR. *GO FAIR Initiative* **[en ligne]** Disponible sur : https://www.go-fair.org/go-fair-initiative/ (consulté le 14 mars 2018)

[46] GO FAIR. *FAIR principles* **[en ligne]** Disponible sur : https://www.go-fair.org/fair-principles/ (consulté le 14 mars 2018)

[47] Puren, Marie., Riondet, Charles. *Research data management, a chance for Open Science. Methods and tutorials to create a Data Management Plan (DMP)*. **[en ligne]** Prague : DARIAH's Humanities at Scale Winter School, 2016, 81 diapositives. Disponible sur : https://hal.inria.fr/hal-01416978 (consulté le 24 octobre 2017)

(cc) BY-NC-ND

de la préservation des données. Car étant donné la nature des informations qui sont demandées dans un PGD, plusieurs expertises peuvent être impliquées. Bien que les plans de gestion de données ne se ressemblent pas tous, les différentes sections étant organisées différemment, la plupart contiennent les informations suivantes :

- Information sur le projet : Les personnes rédigeant le PGD doivent décrire leur projet, cela permet de donner du contexte aux données.
- Information sur les données : de quels types de données parle-t-on ? Données quantitatives, données qualitatives, taille des données, comment sont-elles obtenues, quels seront le ou les formats utilisés pour les données, etc.
- Responsabilité : Qui est en charge de la préservation des données pendant le projet ? Après le projet ?
- Métadonnées : De quoi les données seront-elles accompagnées ? Est-ce qu'un standard sera utilisé pour la création des métadonnées ?
- Information concernant le partage des données : Comment celles-ci seront-elles rendues accessibles ? Seront-elles déposées dans un entrepôt de données ? Si oui, lequel ? Sinon, pourquoi ? Quels sont les enjeux liés à la propriété intellectuelle ?
- Ressources : Quels sont les moyens humains, matériels et financiers pour mettre en œuvre le plan de gestion des données ?
- Archivage des données : Combien de temps seront préservées les données suite au projet ? Où seront-elles préservées ? Est-ce toutes les données qui seront préservées ou seulement une partie ?

Comme vous pouvez sûrement le constater, le contenu de certains de ces points se chevauche. Ils ne sont donc pas nécessairement adressés de façon individuelle dans les plans.

La pratique concernant les PGD

L'exemple d'un pionnier : les États-Unis

Il est possible que les mouvements que l'on a vus émerger en Europe concernant le partage des publications scientifiques et des données de la recherche proviennent de l'influence des institutions américaines, qui ont souvent été en avance sur ces questions. Pour donner un exemple parmi d'autres, c'est en 1991 qu'a

(cc) BY-NC-ND

été lancé la plateforme arXiv, la première plateforme où les chercheurs ont pu publier leurs articles scientifiques avant l'envoi à un comité de révision par les pairs, afin d'en faciliter le partage et l'accès[48].

De la même façon, les grands organismes américains qui subventionnent la recherche ont été les premiers à exiger la rédaction de plans de gestion des données comme condition à leur programme de financement. La *National Science Foundation* (NSF) exigeait déjà depuis plusieurs années le partage des données de la recherche dans un délai raisonnable à condition que les coûts résultants de cette démarche restent modestes. Elle ira plus loin en mai 2010 alors qu'elle publie un communiqué affirmant qu'elle exigera, d'ici la fin de l'année, un plan de gestion de données de deux pages qui sera soumis à un comité de révision par les pairs. Selon la NSF, cette démarche est justifiée par plusieurs facteurs : faciliter la collaboration entre chercheurs, rendre publiques les données de recherche financées par les investissements publics et de façon plus large, s'inscrire dans le mouvement souhaitant rendre le gouvernement et l'administration publique plus transparents[49].

Quelques années plus tard, cette initiative est généralisée à toutes les institutions fédérales de recherches ayant un budget annuel de plus de 100 millions de dollars à travers une législation adoptée en 2013. Cela signifie donc que les chercheurs recevant des subventions de la NSF, du *National Institutes of Health* (NIH) et du *National Endowment for the Humanities* (NEH) doivent impérativement produire des plans de gestion de données dans le cadre de leurs recherches. Au fil du temps, plusieurs universités ont implanté des programmes d'assistance aux chercheurs afin de les aider à produire ces documents. Par ailleurs, une étude réalisée en 2011 puis reprise en 2015 qui cherchait à connaître l'attitude des chercheurs face à leurs données et comment celle-ci pouvait avoir changé à la suite des obligations introduites par les organismes de subventions américains a noté que les scientifiques semblent maintenant plus engagés par rapport au partage et à la réutilisation des données qu'avant l'apparition des politiques concernant les PGD[50].

---

[48] Pepe, Alberto., Cantiello, Matteo., Nicholson, Josh. « The arXiv of the future will not look like the arXiv ». **In :** *Pré-publication Authorea* **[en ligne].** 2017, p. 1. Disponible sur : https://arxiv.org/abs/1709.07020 (consulté le 21 avril 2017)

[49] National Science Foundation, *Scientists Seeking NSF Funding Will Soon Be Required to Submit Data Management Plans* **[en ligne]**. Alexandria : National Science Foundation, 2010. Disponible sur : https://www.nsf.gov/news/news_summ.jsp?cntn_id=116928 (consulté le 09 décembre 2017)

[50] Tenopir, Carol., Dalton, Elizabeth D., Allard, Suzie., et al. « Changes in Data Sharing and Data Reuse Practices and Perceptions among Scientists Worldwide » **[en ligne]**. *PLoS One*, 2015/8 (vol. 10) 24 p. Disponible sur : https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4550246/#? (consulté le 09 décembre 2017)

(cc) BY-NC-ND

Et en France ?

Bien qu'active sur les questions de l'accès libre et du partage de données, la France semble avoir mis un peu de temps avant de s'intéresser à ces questions. Effectivement, la littérature note peu d'initiatives françaises concernant ces questions avant le début des années 2000, soit près de 10 ans après la création de la première archive ouverte américaine ! Par exemple, le Centre pour la Communication Scientifique Directe (CCSD) est créé en 2000, et les premiers dépôts de thèse ne sont possibles qu'en 2001[51]. C'est d'ailleurs de cette initiative que sera créé HAL, le portail de partage d'archives ouvertes, destiné au dépôt d'articles scientifiques.

De la même manière donc, le plan de gestion de données ne fera son apparition en France que plus tardivement, quelques années après les premières initiatives américaines. Plus précisément, leur utilisation semble encore être une pratique marginale. D'abord, parce que ceux-ci ne sont encore pas tellement connus des milieux universitaires, tels que le démontre l'enquête effectuée auprès des directeurs de laboratoire et des chercheurs et chercheuses de l'université Rennes 2 : « Le résultat des réponses données est tout à fait clair et montre une méconnaissance très partagée, puisque 89 % des répondants (97 personnes sur 109) déclarent en effet ne pas savoir ce qu'est un Plan de Gestion de Données.[52] »

Ensuite, il semblerait que même pour ceux qui connaissent le PGD et ses objectifs, son adoption n'est pas majoritaire. Des cinq personnes interrogées dans le cadre de ce mémoire, trois avaient déjà fait l'exercice de remplir un plan de gestion de données (dont deux personnes qui ont travaillées conjointement au sein du même organisme) et chaque fois, cela était fait de façon plutôt expérimentale afin de voir si ces personnes pouvaient éventuellement assister des chercheurs dans la rédaction de ces documents, comme en témoignent Alina Dancieu et Anne Cornilleau du CDSP de Sciences-po :

> « [C]'est aussi une façon de voir quelles seraient les difficultés ou les facilités, au contraire, que l'on aurait pour remplir ce type de plan de gestion

[51] CCSD. *Les temps forts* **[en ligne].** Disponible sur : https://www.ccsd.cnrs.fr/presentation/les-temps-forts/ (consulté le 26 avril 2018)

[52] SERRES, Alexandre (dir.) *Données de la recherche en SHS. Pratiques, représentations et attentes des chercheurs : une enquête à l'Université Rennes 2. op cit.* p. 107

(cc) BY-NC-ND

de données, si on est amené à l'avenir à accompagner des équipes qui auraient ce besoin-là[53]. »

Même son de cloche chez Christine Hadrossek de l'équipe du RIATE au Centre national de recherche scientifique (CNRS) :

> « Une démarche qui était, dans un premier temps, expérimentale. Il n'y avait pas particulièrement obligation de faire… il ne s'agit pas d'un projet H2020, ou… d'un projet européen qui nécessite absolument la production d'un DMP. On se retrouvait, dans mon équipe, dans une démarche de vouloir expérimenter… le plan de gestion de données, pour pouvoir le moment venu, être capable de le réaliser dans les meilleures conditions[54]. »

Cela peut s'expliquer en partie par le fait que pour l'instant, peu (voire pas) d'organismes subventionnant la recherche française n'exigent de plans de gestion de données, si ce n'est que le programme H2020.

Toutefois, on note un intérêt grandissant pour ce document. Celui-ci est de plus en plus discuté lors de conférences abordant le thème des données de la recherche. On a vu certaines universités françaises comme Rennes 2 ou Lille 3 effectuer récemment des enquêtes auprès de leurs chercheurs afin de connaître leurs besoins en termes de gestion des données de la recherche et éventuellement proposer des services d'accompagnement à la rédaction de PGD. L'institut de l'information scientifique et technique (INIST) aborde aussi ce sujet dans son module *Une introduction à la gestion et au partage des données de la recherche*. L'Association nationale de la Recherche (ANR), bien qu'elle n'exige pas (encore) de plan de gestion de données comme conditionnement pour le financement des projets, encourage toutefois les chercheurs à porter une attention particulière à la question des données lors du déroulement de leur projet[55]. Dernièrement, la *Loi pour une république numérique*, promulguée le 7 octobre 2016, encourage les politiques de l'*open data* et ce, aussi dans le cadre du partage des données de la recherche, puisqu'elle modifie le Code de la recherche pour y inclure le texte suivant :

> « Dès lors que les données issues d'une activité de recherche financée au moins pour moitié par des dotations de l'Etat, des collectivités territoriales, des établissements publics, des subventions d'agences de financement nationales ou par des fonds de l'Union européenne ne sont pas protégées par un droit spécifique ou une réglementation particulière et qu'elles ont été rendues

[53] Entretien avec Alina Dancieu, chargée d'études et Anne Cornilleau, chargée d'études, du CDSP le 15 janvier 2018

[54] Entretien avec Christine Hadrossek, Chargée de l'information scientifique et technique à l'UMS2414 Riate le 17 janvier 2018

[55] Agence nationale de recherche. *L'ANR et l'open access* **[en ligne]**. Disponible sur : http://www.agence-nationale-recherche.fr/missions-et-organisation/open-access/ (consulté le 26 avril 2018)

(cc) BY-NC-ND

publiques par le chercheur, l'établissement ou l'organisme de recherche, leur réutilisation est libre[56]. »

De plus, plusieurs initiatives gouvernementales démontrent un véritable intérêt pour un plus grand partage des données de la recherche. Nous n'avons à penser qu'à Etalab, une initiative de la Direction interministérielle du numérique et du système d'information et de communication de l'Etat (DINSIC) pour faciliter et appuyer l'ouverture des données de l'Etat français, qui spécifie vouloir, entre autres actions, « Généraliser [*sic*] progressivement via un accompagnement la mise en place de plans de gestion des données dans les appels à projets de recherche, et inciter à une ouverture des données produites par les programmes financés[57] ». Récemment, le Ministère de l'Enseignement supérieur, de la Recherche et de l'Innovation annonçait un *Plan national pour la science ouverte* avec un axe se concentrant sur « structurer et ouvrir les données de la recherche[58] » avec des mesures telles que rendre obligatoire la diffusion des données de la recherche produites dans le cadre de programmes financés par appel à projets sur fonds publics ou « promouvoir l'adoption d'une politique de données ouvertes associées aux articles publiés par les chercheurs.[59] »

Il semblerait donc que malgré des pratiques plutôt timides (pour l'instant), plusieurs éléments soient en place afin de développer les pratiques concernant le partage et la réutilisation de données, et cela passe en partie par les plans de gestion des données.

---

[56] Article 30 de la *LOI n° 2016-1321 du 7 octobre 2016 pour une République numérique (1)*. Legifrance **[en ligne].** 7 octobre 2016. Disponible sur : https://www.legifrance.gouv.fr/affichTexte.do;jsessionid=298A6C25EFA7003656E88AAFB3D19609.tplgfr25s_2?cidTexte=JORFTEXT000033202746&categorieLien=id (consulté le 16 avril 2018)

[57] Etalab. *Pour une action publique transparente et collaborative : plan d'action national pour la France 2018-2020* **[en ligne].** Paris : Direction interministérielle du numérique et du système d'information et de communication de l'Etat, 2018, p. 58. Disponible sur : https://www.etalab.gouv.fr/wp-content/uploads/2018/04/PlanOGP-FR-2018-2020-VF-FR.pdf (consulté le 13 août 2018)

[58] Ministère de l'Enseignement supérieur, de la Recherche et de l'Innovation. *Le Plan national pour la science ouverte : les résultats de la recherche scientifique ouverts à tous, sans entrave, sans délai, sans paiement* **[en ligne]** (modifié le 12 juillet 2017)**.** Disponible sur : http://www.enseignementsup-recherche.gouv.fr/cid132529/le-plan-national-pour-la-science-ouverte-les-resultats-de-la-recherche-scientifique-ouverts-a-tous-sans-entrave-sans-delai-sans-paiement.html (consulté le 13 août 2018)

[59] *Ibid.*

(cc) BY-NC-ND

METTRE EN PLACE LES PLANS DE GESTION DE DONNEES

Comme le plan de gestion de données est un document, et non tout un système, il n'existe pas vraiment de guide ou de mode d'emploi sur la façon dont ceux-ci doivent être mis en place.

Par exemple, l'Inist a participé à la création d'un outil, DMP OPIDoR, inspiré lui-même du DMPtool, un outil *open source* créé aux États-Unis en réponse à l'imposition des plans de gestion de données par la NSF et la NIH et du DMPOnline, outil développé par le Digital Curation Center au Royaume-Uni. Cet outil aide à la rédaction de plan de gestion de données. On y retrouve plusieurs modèles (soumis par les différents utilisateurs) que les gens peuvent remplir à travers cet outil ; les différentes sections des PGD sont séparées, et selon les modèles déposés, on y retrouve des indications sur le contenu nécessaire à cette section. DMP OPIDoR offre aussi la possibilité de demander de l'assistance auprès d'une personne de la même organisation que nous ou recevoir des recommandations quant à la façon de répondre à un champ de questions[60].

Toutefois, bien que cet outil semble particulièrement utile, cela ne constitue pas un guide pour véritablement procéder à une démarche en vue de la rédaction de plans de gestion de données dans un établissement… et il paraît difficile de trouver un tel document. Jusqu'à maintenant, on trouve surtout plusieurs exemples de démarches qui ont été mises en place par divers organismes afin de mieux connaître les besoins des chercheurs en termes de gestion et de conservation des données. Nous nous concentrerons donc sur cet aspect, plus particulièrement en observant dans cette section deux cas particuliers, soit les enquêtes qui sont menées dans diverses universités françaises afin de développer une offre de service en rapport aux données de la recherche et le *Data Curation Profiles*, une démarche développée par les bibliothèques de l'Université Purdue et le *Graduate School of Library and Information Science* à l'Université d'Illinois Urbana-Champaign.

Bâtir une offre de service concernant les données de la recherche

D'abord, nous avons mentionné que plusieurs universités françaises sont en train de développer leur offre envers les chercheurs et pour certaines, cela inclura

[60] INIST, *DMP OPIDoR* **[en ligne]**. Disponible sur : https://dmp.opidor.fr/ (consulté le 03 janvier 2018)

(cc) BY-NC-ND

éventuellement l'aide à la rédaction de plan de gestion de données. C'est par exemple le cas de l'université de Lorraine, qui a mené une petite enquête sur les données de la recherche. Celle-ci leur a permis de voir quels laboratoires avaient plus d'intérêts pour le sujet et tester avec eux quelques idées[61]. D'autres universités, telles que les universités Lille 3 et Rennes 2 ont toutes deux mené des enquêtes d'assez grande envergure et ont publiés le rapport de ces enquêtes sur la plateforme HAL. Une autre est en cours à l'université Lyon 1.

L'université Lille 3 a mené son enquête dans le cadre d'une démarche plus large, ayant les objectifs suivants : « faire un état des lieux, analyser la situation et préparer le terrain pour mettre en place une offre de service adaptée et utile aux doctorants et chercheurs du campus.[62] » De façon plus large, c'était aussi l'occasion de faire la promotion du sujet des données de la recherche auprès de leurs doctorants et chercheurs. L'enquête a eu lieu en 2015 sous forme de questionnaires divisés en 6 sections : Informations sur le répondant (statut, domaine), Typologie des données (sources et résultats), Pratiques de stockage et sauvegarde, Pratiques de partage, Avis par rapport au partage et aux archives de données et Besoins[63]. Les auteurs voyaient aussi cette enquête comme une première étape, une initiative dans le cadre d'une démarche plus importante incorporant les axes formation, assistance et conseil et infrastructure pour les données de la recherche.

Cette initiative a inspiré l'université Rennes 2, qui a mené son enquête en utilisant un questionnaire semblable, et en y ajoutant des entretiens qualitatifs afin de pouvoir approfondir les différents thèmes soulevés par les questions quantitatives. Leur enquête avait aussi des objectifs semblables soient : établir un état des lieux, élaborer une offre de services et alimenter la recherche.

Une fois leur enquête terminée, les auteurs de cette étude ont fait les propositions suivantes :

- Créer un groupe de travail chargé de la réflexion sur une politique des données de recherche ;

[61] Entretien avec M. Hubert, *op. cit.*

[62] Prost, Hélène., Schöpfel, Joachim. *Les données de la recherche en SHS. Une enquête à l'Université de Lille 3 : Rapport final.* **[en ligne]** Rapport de recherche. Lille : Université de Lille 3, 2015, p. 9. Disponible sur : http://hal.univ-lille3.fr/hal-01198379 (page consulté le 27 avril 2018)

[63] *Ibid.*, p.11

(cc) BY-NC-ND

- Développer des services mutualisés pour la sensibilisation et la formation ;
- Offrir un service juridique spécialisé ;
- Mieux adapter les espaces internes de stockage des données de recherche aux besoins individuels et collectifs des chercheurs ;
- Sensibiliser à la description des jeux de données et à la rédaction des Plans de Gestion de Données ;
- Sensibiliser aux solutions institutionnelles de partage et d'archivage des jeux de données[64].

Avec ces propositions, on voit déjà les traits d'une démarche qui se formalise afin d'implanter l'habitude de bien gérer les données de la recherche, travail qui ultimement devrait faciliter la rédaction de plans de gestion des données pour les chercheurs et ceux qui les assisteraient dans cette tâche.

Il ne convient donc pas de dire que ces enquêtes sont une démarche complète en vue de rédiger les plans de gestion des données, mais elles semblent être une première étape logique lorsqu'il convient de mettre en place des services d'assistance à la rédaction de PGD et plus largement d'aide à la gestion des données.

Data curation profiles

Alors que les enquêtes menées par les universités sont un moyen de connaître, de façon large, les connaissances et les habitudes des chercheurs en matière de données de la recherche, le *data curation profile* (DCP) est un moyen de connaître les besoins et les pratiques individuels en matière de données de la recherche. Les auteurs de cette méthode la présentent comme étant : « the "story" of a data set or collection, describing its origin and life cycle within a research project.[65] » Le fait d'établir un profil permettrait de commencer un dialogue avec les chercheurs à propos des données de la recherche, connaître quels sont les éventuels enjeux sur lesquels porter attention dans une démarche de gestion de données, connaître les besoins pour une bonne gestion d'un jeu de données et amorcer une démarche pour

[64] SERRES, Alexandre (dir.) *Données de la recherche en SHS. Pratiques, représentations et attentes des chercheurs : une enquête à l'Université Rennes 2. op cit***.** p. 131-136

[65] DCP. « Overview » **In** *Data Curation Profiles* **[en ligne]**. Disponible sur : http://datacurationprofiles.org/overview.php (consulté le 03 août 2018)

(cc) BY-NC-ND

une bonne gestion et préservation des données de la recherche[66]. Le but étant de connaître les données du point de vue du chercheur (le DCP n'a pas vocation à être un inventaire exhaustif des données de recherche d'un chercheur). Le profil, une fois complété, est organisé selon les sections suivantes[67] :

1. Résumé des besoins en matière de données
2. Résumé de la recherche
3. Types de données et étapes de traitement[68]
4. Propriété intellectuelle et autres informations
5. Organisation et description des données
6. Transfert des données
7. Partage et accès
8. Recherche (comment peut-on retrouver les données)
9. Outils (ceux utilisés pour produire les données et pour pouvoir les lire)
10. Liens et interopérabilité
11. Impact
12. Gestion des données
13. Préservation
14. Personnel, soient les coordonnées des personnes impliquées dans le projet de recherche.

Selon les besoins, toutes les catégories n'ont pas à être remplies mais les auteurs suggèrent d'avoir au minimum les sections 2, 3, 5 et 7.

Comme vous pouvez le constater, le contenu du DCP n'est pas très différent de celui d'un plan de gestion de données. De plus, celui-ci offre non seulement la structure générale du document, mais suggère aussi une marche à suivre. D'abord, on suggère de sélectionner les données à discuter avec le chercheur, puis de se préparer aux entretiens en faisant une recherche préliminaire sur les projets de ce dernier. Les personnes impliquées dans la création du profil devraient être identifiées dès le début, et on recommande que la personne qui conduit les entretiens

[66] DCP. « Home ». **In** *Data Curation Profiles* **[en ligne]**. Disponible sur : http://datacurationprofiles.org/index.php (consulté le 03 août 2018)

[67] Traducation libre de : CARLSON, Jake. *The Data Curation Profiles Toolkit : User Guide* **[en ligne].** West Lafayette, Purdue University, 2010. Disponible sur : https://docs.lib.purdue.edu/dcptoolkit/1/ (consulté le 03 août 2018)

[68] Le *data curation profile* propose de remplir un tableau indiquant la quantité et le format des données aux stades de données sources, données traitées, données analysées et données finalisées.

(cc) BY-NC-ND

soit aussi celle qui s'occupe de rédiger le profil. Finalement, en vue de la préparation, on recommande d'obtenir l'accord des autorités nécessaires (direction ou autre) afin de mettre en place une telle démarche. C'est aussi lors de la préparation que l'on pourra ajuster les sections du DCP si nécessaire.

La deuxième phase est celle des entretiens. Le guide d'utilisateur recommande de faire au moins deux entretiens d'une heure chacun. Pour ces entretiens, un manuel est fourni. Ce manuel propose des façons d'introduire chaque section ainsi que des exemples de questions à poser. Une feuille de travail est aussi fournie, où l'intervieweur peut prendre des notes pendant les entretiens et qui peut être transmis au chercheur en amont du processus d'entretien. On recommande d'enregistrer les entretiens ainsi que de les retranscrire.

Finalement, la troisième phase est celle de la construction du profil, où l'on met en forme les informations que l'on a pu recueillir lors des entretiens.

Une fois le profil terminé, il est possible de le déposer dans un entrepôt créé à cet effet par l'Université Purdue. Cela permet de lier le profil à un DOI (*digital object identifier*) et donc éventuellement de lier celui-ci aux publications résultantes du même projet de recherche que les données.

De ces deux démarches pour gérer les données de la recherche, on remarquera une chose : ces méthodes ne sont en fait que le début d'une démarche pour gérer les données ou rédiger un plan de gestion de données, soit des façons de réaliser des enquêtes ou des évaluations avant de prendre action pour gérer des données de la recherche. Bien que l'enquête de Rennes 2 ait proposé des initiatives pour améliorer la gestion des données de la recherche, il ne semble malheureusement avoir, pour l'instant, que peu de littérature sur comment, concrètement, effectuer une démarche de rédaction d'un plan de gestion de données autre que simplement répondre aux questions qui sont écrites sur ces derniers. Une personne pourra toutefois s'inspirer du DCP qui aborde des thèmes semblables aux PGD dans ses questions et qui pourrait éventuellement être adapté pour convenir au modèle que l'on a choisi.

(cc) BY-NC-ND

LE RECORDS MANAGEMENT ET LES PLANS DE GESTION DE DONNEES, COMPLEMENTAIRES ?

Cette première partie nous aura permis d'en connaître un peu plus sur les démarches de *records management* et de rédaction d'un plan de gestion des données de la recherche. Les lecteurs attentifs auront probablement déjà noté quelques points de ressemblances, ou au contraire, de divergences entre ces deux démarches. Ces points seront explorés plus en détail dans les pages suivantes.

LES POINTS COMMUNS

Revalorisation des métiers des sciences de l'information

Lorsque que le *records management* a commencé à se développer en France, cela a quelque peu bouleversé les pratiques archivistiques qui étaient en place jusqu'à ce jour. En effet, jusqu'à il y a une vingtaine d'années, c'était encore la théorie des trois âges[69] qui guidait les pratiques archivistiques, ce qui signifie que le rôle de l'archiviste était bien souvent relégué à la fin du cycle de vie d'un document[70]. On connaît tous bien le cliché de l'archiviste caché dans les sous-sols des entreprises dont la tâche est de s'occuper de vieux documents poussiéreux qui n'intéressent personne. Hors, le développement du *records management* en France est venu modifier la perception des métiers des sciences de l'information :

> « Les principes du « records management » et cette norme ISO 15489 publiée en 2001 représentaient donc une opportunité, pour les professionnels de l'information-documentation, de valoriser leur savoir-faire en matière de gestion de l'information, au même titre que la veille, l'intelligence économique ou le knowledge management.[71] »

Puisque que le *records management* place l'archiviste (*records manager*) en amont du cycle de vie, ceux-ci en sont venus à occuper une position différente au sien des

[69] Cette théorie avance que les archives passent par trois phases :
- les archives courantes, utilisés fréquemment pour leur valeur administrative et, par conséquent, conservé physiquement près des bureaux de l'organisme ;
- les archives intermédiaires, où la valeur administrative n'est plus très importante, mais l'archive ne possèdent pas encore totalement de valeur historique, et donc reléguées à un lieu de conservation « intermédiaire » ;
- les archives historiques.

[70] Drouhet, Geneviève., Keslassy, Georges. et Morineau, Elisabeth. *Records management : mode d'emploi*. *op. cit*., p.19

[71] Dessolin-Baumann Sylvie, Wiegandt Caroline, « Un métier en devenir ». *op. cit.*

(cc) BY-NC-ND

administrations. Ce bouleversement leur demandait aussi d'occuper un rôle plus central, plus stratégique :

> « Autrement dit, les professionnels doivent apprendre à maîtriser les techniques documentaires ou archivistiques selon leurs origines professionnelles, mais aussi à dialoguer au plus haut niveau de l'entreprise afin d'être reconnus comme partenaires.[72] »

Bien qu'il soit encore trop tôt pour l'affirmer avec certitude, il semblerait que l'engouement actuel pour les données de la recherche représente lui aussi une opportunité de revalorisation des métiers des sciences de l'information. D'abord, car les compétences demandées pour une bonne gestion des DDR sont bien souvent celles possédées par les différents métiers de l'information/documentation :

> « Des métiers liés à la gouvernance de l'information, à l'animation de la communauté, à la médiation scientifique ou encore plus spécifiques à la gestion de la donnée scientifique ou à l'architecture de l'information doivent s'imposer pour accompagner les chercheurs et être capables d'effectuer des tâches que ces derniers n'ont pas le temps de faire.[73] »

C'est d'ailleurs un sentiment qui m'a été partagé par Mme Dancieu et Mme Cornilleau du CDSP qui m'ont confié avoir trouvé cela relativement facile de compléter un plan de gestion de données, car : « on a [*sic*] l'habitude de la gestion des données[74] ».

Ensuite parce que les plans de gestion des données, ou même plus largement les pratiques de l'*open science*, sont vu comme une opportunité à saisir pour les personnes des métiers de l'information qui gravitent autour des chercheurs scientifiques : « Par contre, nous pouvons sans problème sensibiliser les chercheurs avec cette thématique [la gestion des données de la recherche] par des formations. Nous espérons qu'avec ces sensibilisations, ils viendront vers nous pour qu'on puisse les accompagner dans leurs besoins documentaires.[75] » Ces nouvelles pratiques dans le monde scientifique représentent donc une belle opportunité de revoir les rôles des bibliothécaires/documentalistes/archivistes au sein d'organismes de recherche et d'éducation. Elles représentent une occasion d'amorcer le dialogue entre ces deux

[72] *Ibid.*

[73] Besnard, Myriam., Delobelle, Grégory. « Les archives du Bureau de recherches géologiques et minières en transition » **In** « Archiver la recherche : responsabilités partagées » **In :** *Actes des journées d'études de la section Aurore - archivistes des universités, rectorats, organismes de recherche et mouvement étudiants - de l'Association des archivistes français du 23 juin 2016* **In** *La Gazette des archives*, 2017/2 (no 246), p.69

[74] Entretien avec A. Dancieu et A. Cornilleau, *op. cit.*

[75] Entretien avec M. Hubert, *op. cit.*

(cc) BY-NC-ND

corps de métiers (métiers de l'information et métiers de la recherche) et redéfinir les places que les uns occupent envers les autres.

Les métiers de l'information sont en constantes évolutions, et des occasions comme l'avènement du *records management* ou le déferlement des données que l'on connaît aujourd'hui permet de mettre en valeur les compétences qui sont possédées par ces métiers.

Démarche d'évaluation

La plupart des guides pour pratiquer le *records management* indiquent une même première étape : il faut d'abord bien comprendre l'environnement dans lequel sont produits les *records*. Ceci signifie d'être en mesure de comprendre les processus de travail des créateurs de *records*, de connaître les impératifs réglementaires et législatifs, et de s'être renseigné sur les objectifs de l'entreprise ou de la direction qui met en place cette démarche. Ainsi, la norme ISO 15489-2, qui offre un guide pratique pour la mise en place du *records management*, suggère de débuter par une étape d'enquête préliminaire, qui renseignera le *records manager* sur les éléments évoqués plus haut dans ce paragraphe. Plus encore, les auteurs de la norme ajoutent que l'enquête : « est nécessaire pour prendre des décisions efficaces en matière de système d'archivage d'un organisme. Elle aidera à cerner les problèmes d'archivage au sein d'un organisme et à évaluer la faisabilité et les risques des différentes solutions possibles.[76] » Cette première étape apparaît donc comme essentielle à une démarche de *records management*.

De même, avant de commencer la rédaction d'un plan de gestion de données, il est important de bien comprendre et d'évaluer le contexte dans lequel sera rédigé le plan de gestion de données, mais aussi la façon dont seront créées ces données, car ces informations seront celles qui guideront la rédaction du plan :

> « Pour l'instant, les défis sont des défis d'adaptations, les chercheurs travaillent avec des données qui diffèrent les uns des autres, les programmes utilisés pour traiter ces données diffèrent aussi, leurs méthodes de travail… Il faut bien connaître leur environnement et leur problématique afin d'être capable de les aider.[77] »

[76] ISO, *Information et documentation - "Records management" Partie 2 : guide pratique = Information and documentation - "Records management" Part 2 : guidelines*. ISO 15489-2. Paris : AFNOR, 2002, p.4

[77] Entretien avec M. Hubert, *op. cit.*

(cc) BY-NC-ND

Une personne cherchant à rédiger un tel plan devra donc se poser plusieurs questions avant de commencer sa rédaction : à quel type de données avons-nous affaire ? Parle-t-on de données qualitatives ? Quantitatives ? Avons-nous à faire à un gros volume de données (*big data*) ou non ? Est-ce que le projet pourrait nous amener à gérer des données personnelles/confidentielles ? Quel est le but du projet de recherche ? Est-ce qu'il existe des enjeux économiques ou politiques autour du projet de recherche en question ? Comment sont créées les données en question ? Il est évident que ces questions auront un impact sur les décisions qui seront prises pendant la rédaction du PGD. Ce sont les réponses à ces questions qui nous dicteront le lieu où l'on conservera les données, le niveau de sécurité qui les protégera, etc.

Dans les deux cas, l'évaluation est donc une étape cruciale à mettre en œuvre dès le début de la démarche. Cette évaluation est celle qui permet de comprendre le contexte dans lequel le *records manager* ou le rédacteur du plan de gestion de données travaillent, de juger les enjeux qui accompagne le déroulement d'une tellement démarche et de donner les premières orientations au projet. Plus encore, cette évaluation permettra éventuellement de choisir, parmi tous les documents et les données produits par les activités de l'entreprise ou l'activité de recherche, lesquels seront *records* et lesquels seront données de la recherche.

Tout n'est pas *records*, tout n'est pas données de la recherche

Effectivement, toutes les informations produites par une entreprise n'ont pas vocation à être *records* ou toutes les données produites lors d'un projet de recherche ne seront pas nécessairement intégrées au PGD. De façon réaliste, il faudrait beaucoup trop de moyens pour tout conserver comme il se doit et, pour certaines informations, cela ne vaut tout simplement pas la peine :

> « Appliquer la démarche de *records management* à la totalité des documents produits ou reçus par un organisme est, sauf cas particulier, inutile et mobiliserait des moyens trop importants. Aussi l'organisme doit-il faire des choix et mettre en balance le coût d'une démarche de *records management* et les risques encourus s'il n'applique que partiellement cette démarche.[78] »

Cette citation pourrait tout autant s'appliquer à un PGD. Les chercheurs ont des ressources limitées et surtout, n'ont pas nécessairement beaucoup de temps libre.

[78] Groupe métiers AAF – ADBS « Records management », « Comprendre et pratiquer le records management, Analyse de la norme ISO 15489 au regard des pratiques archivistiques françaises ». *Documentaliste-Sciences de l'Information*, 2/2005 (Vol.42). Disponible sur : https://www-cairn-info.docelec.insa-lyon.fr/revue-documentaliste-sciences-de-l-information-2005-2-page-106.htm (consulté le 10 janvier 2018)

(cc) BY-NC-ND

Exiger de devoir traiter toutes leurs données dans un PGD serait une demande irréaliste qui pousserait probablement plutôt à l'abandon de la démarche qu'à l'augmentation de la pratique. Il faut donc faire des choix et concentrer nos ressources sur les *records* ou les données pertinentes.

Rappelons d'abord que les *records* sont définis comme les informations produites ou reçues et préservées par un organisme à titre de preuve ou d'actifs organisationnels. Naturellement, dans une organisation quelconque, tous les documents ne correspondent pas à cette définition ; c'est d'ailleurs l'évaluation qui nous permettra de reconnaître lesquels le sont, ou non. Par la suite, les documents existent bien souvent en plusieurs versions et il est souvent inutile de garder toutes les différentes versions. Pensons par exemple à un compte-rendu de réunion de comité d'administration : il est tapé pendant ou juste après la réunion, puis doit être approuvé par une autorité quelconque au sein de l'organisme, probablement le président dudit comité qui y apposera sa signature par exemple. Cela ne vaut pas la peine de conserver la première version du document, mais plutôt celle qui a été validée par une signature, ce qui confère de l'authenticité et de la fiabilité au document.

Malheureusement, au niveau des données de la recherche, l'étape de validation des données n'est pas nécessairement aussi formelle. D'autant plus que les différentes étapes ou les différentes étapes de la vie des données ont chacun leur pertinence et pourraient être amenées à être partagées. Si l'on reprend la classification des données d'un projet selon le *data curation profile* (données sources, données traitées, données analysées et données finalisées), on peut voir un intérêt dans chacun de ces quatre stades de transformation. Le choix du stade de données à partager dépendra donc de la raison pour laquelle la personne souhaitera conserver ou partager ses données. Souhaitera-t-elle permettre à quelqu'un d'utiliser les mêmes données sources pour un nouveau projet ou pour recréer la même recherche du début ? Ou souhaitera-t-elle « témoigner de la manière dont s'est opéré le travail scientifique[79] » en conservant ses données traitées ou finalisées ? Christine Hadrossek m'a fait part d'une réflexion semblable en affirmant :

> « Voilà, donc on a plus décidé de diffuser des données moins importantes, mais avec par contre des documents techniques qui permettent à la personne qui va vouloir reproduire l'expérimentation, de pouvoir le faire

[79] Fayet, Sylvie. « ''Données'' de la recherche, les mal-nommées ». *op. cit.*

(cc) BY-NC-ND

> avec une démarche méthodologique. C'est-à-dire qu'on a expliqué comment on les avait extraites, comment elles ont… le contexte dans lequel elles ont été extraites… Il y avait beaucoup de documents techniques qui expliquaient la démarche en fait.[80] »

Somme toute, en *records management* tout comme lors de la rédaction d'un plan de gestion des données, il convient que toutes les informations n'auront pas vocation à être conservées / partagées. Une bonne évaluation des activités et du contexte permettra de faire un choix éclairé, qui sera de plus influencé par les motivations derrière la mise en place de ces démarches. Que veut-on prouver ?

Contexte de création du *record* / des données de la recherche

Un autre moment important dans la gestion des *records* et des données de la recherche est celui où les données/*records* sont créées. Ce moment doit être capturé adéquatement, sans quoi les données ou *records* perdent de leur signification. Est-ce qu'une donnée de recherche peut être réutilisée si l'on ne sait pas comment on l'a obtenue ? Est-ce qu'un *record* peut être probant si l'on ne sait pas quand ou comment il a été produit ?

Pour ce qui est des données de la recherche, une ou des données ne sont pas réutilisables ou exploitables si elles ne sont pas accompagnées du contexte de leur création : « Dans de nombreuses disciplines, les chercheurs s'accorderont probablement sur le fait qu'une donnée est indissociable de ces conditions de production (…) et que ces conditions de productions participent de la scientificité de la donnée[81] ». De plus, lors de notre entretien, Cynthia Pedroja a bien souligné l'importance de connaître la démarche du chercheur, sa problématique, la façon dont sont constitué ces données, car sinon, il devient difficile, voire impossible, d'interpréter les données correctement[82]. Cela fait d'ailleurs aussi écho aux propos de Madeleine Hubert : « un sociologue et un chimiste produisent tous les deux des données, mais cette donnée, même si elle respecte un cycle de vie plus ou moins commun, va évoluer différemment. Connaître le champ disciplinaire du chercheur est donc indispensable.[83] » On constate donc qu'il y a bien souvent un véritable

[80] Entretien avec C. Hadrossek, *op. cit.*

[81] Savonitto, Céline. « Introduction (quelles responsabilités ?) » **In** « Archiver la recherche : responsabilités partagées » *op. cit.* p. 21

[82] Entretien avec Cynthia Pedroja, chef de projet données de la recherche à SciencesPo Paris le mardi 27 février 2018

[83] Entretien avec Madeleine Hubert, *data librarian* à la Bibliothèque numérique de l'Université de Lorraine le mardi 20 mars 2018

(cc) BY-NC-ND

effort afin de conserver une partie du contexte de création des données, et que celui-ci contribue à l'interprétation des données et à leur valeur.

Dans le cas du *records management*, on se rend bien compte qu'une importante partie de la démarche consiste à s'assurer d'avoir les conditions idéales pour créer un *record* qui sera probant, et ensuite à capturer ce contexte de création. Concrètement, plusieurs outils et méthodes ont donc été développés afin de gérer le contexte de création des *records*. Prenons l'exemple de l'analyse de processus. Comme nous l'avons évoqué plus haut, cette méthode de travail permet au *records manager* de pouvoir identifier et comprendre toutes les étapes par lequel passe un *record* lors de son cycle de vie et ainsi être capable d'identifier quelles informations doivent être présentes et capturées lors de sa création. Le tableau de gestion, de façon similaire, permet de mieux comprendre les enjeux entourant les *records* (quels sont leurs valeurs, combien de temps doit-on les conservés, qui doit créer/valider/conserver le *record*) et prévoir une bonne gestion de tout son cycle de vie, et dans le cas qui nous intéresse, ses conditions de création. Le moment de la création est donc, sans aucun doute, au cœur de la démarche de *records management*. Ainsi nous le confirme Arnaud Jules qui écrivait en 2012, alors qu'il travaillait au *records management* chez Orange : « maîtriser les documents identifiés comme des documents d'activités suppose de maîtriser les conditions dans lesquelles ils sont créés pour optimiser les processus de travail.[84] »

La création des *records* et des données est donc un moment important, qu'il faut capturer si l'on souhaite produire des *records* de qualité. Cela fait partie intégrante du *records management* et il semblerait qu'une bonne gestion des données de la recherche implique d'y porter une grande attention. Toutefois, capturer le moment de création ne veut pas nécessairement dire figer ces *records*/données dans le temps. Ces derniers peuvent être réutilisés dans de nouveaux contextes, être changés d'environnement ou simplement devoir changer de format pour s'adapter aux évolutions technologiques. Cela représente un défi, car : « making data mobile requires separating them from the situation in which they became data. As a result, some amount of meaning is lost[85] ». Dans ce cas, comment faire pour conserver le

[84] Arnaud, Jules. « Gestion des documents d'activité : définition, principes et concepts » **In** *La Gazette des archives* **[en ligne]**, 2012/4 (n°228). *op. cit.,* p. 107.

[85] Borgman, Christine L. *Big Data, Little Data, No Data*. *op. cit.* p.219

(cc) BY-NC-ND

contexte de création du *record* ou des données de façon à pouvoir réutiliser les réutiliser ?

Les métadonnées

L'une des façons de bien capturer le contexte de création des *records* ou des données est d'établir des métadonnées nous renseignant sur les informations pertinentes. Cet aspect est fortement mis de l'avant pour ce qui est des données de la recherche, comme Borgman le souligne : « metadata can bridge distances from sources to resources by formalizing and standardizing how data are named and described[86] ». On le constate aussi par la place que prend l'information sur les métadonnées dans les plans de gestion des données[87], et avec raison.

Une étude ayant été réalisée aux États-Unis peu de temps après l'instauration de l'obligation de rédiger des PGD mentionnait : « The authors came to the conclusion that researchers often fail to develop clear, well-annotated datasets to accompany their research (i.e., metadata), and may lose access and understanding of the original dataset over time[88] ». Il existe/existait un manque au niveau des pratiques des chercheurs en ce sens, et le plan de gestion de données pourrait en partie y répondre. Cette conclusion trouve écho dans ce que m'a dit Cynthia Pedroja, à savoir que pour bien préserver des données de la recherche, il était important de bien les documenter[89] (en plus de bien les nommer et sélectionner de bons formats). Les PGD permettent de réfléchir et de prévoir les métadonnées qui accompagneront les données de la recherche, par exemple en précisant quel standard sera utilisé. Ainsi, on peut s'assurer que ces dernières conservent leur signification dans le temps.

En effet, le processus d'écriture d'un plan de gestion de données permet de réfléchir et d'organiser ces futures informations. Sans surprise, on constate une forte ressemblance entre les objectifs d'un PGD et les procédures pour le choix des métadonnées des *records* (toujours selon la norme ISO 23081-1) :

---

[86] *Ibid.,* p.68

[87] La section 2 du guide H2020 pour la rédaction d'un plan de gestion est principalement axée sur les métadonnées, le guide rédigé par le groupe de travail à l'université Paris-Diderot consacre une sous-section à cet aspect aussi.

[88] Tenopir, Carol., Allard, Suzie., Douglass, Kimberly., *et al.* « Data Sharing by Scientists : Practices and Perceptions ». *PLoS One*, 2011/6 (vol. 66), p.9. Disponible sur : https://www.researchgate.net/publication/51476349_Data_Sharing_by_Scientists_Practices_and_Perceptions (consulté le 09 décembre 2017)

[89] Entretien avec C. Pedroja, *op. cit.*

(cc) BY-NC-ND

> « Il convient que les acteurs, dont les professionnels de la gestion des documents d'activité, définissent et documentent les procédures et les règles de gestion des métadonnées et structurent ces métadonnées en fonction des impératifs de leur activité. Ces procédures et règles englobent des sujets tels que l'attribution des responsabilités, la sélection des métadonnées à créer et à capturer, le moment et les sources d'où les extraire, le choix des structures de métadonnées valides ainsi que les normes et les systèmes convenant à leur utilisation[90]. »

On l'aura constaté, les métadonnées sont aussi au cœur du *records management*. Comme l'indique la norme ISO 23081-1 : « Les métadonnées de gestion des documents d'activité peuvent servir à identifier, authentifier, contextualiser des documents d'activité et des personnes, des processus et des systèmes qui créent, gèrent, maintiennent et utilisent ces documents, ainsi que les règles qui les régissent.[91] » Ce sont les métadonnées de capture qui permettent de garantir l'authenticité (qui a créé le *record*, à quelle date, à quel endroit, etc.) et ce sont les métadonnées de gestion qui permettent de nous en garantir l'intégrité (le *record* a-t-il été modifié ? si oui, pourquoi ?).

Plus encore, un jeu de données, un *record*, bien que porteur d'informations, ne suffit pas en lui-même. Ainsi, si l'on considère que ces derniers sont les preuves d'une activité ou d'une expérimentation, leur présence en elle-même ne suffira probablement pas à la preuve. Il faudra connaître qui a produit ce *record*, ces données, comment il s'y est pris, à quel moment, etc. Toutes ces informations sont des métadonnées qui deviennent aussi importantes que les *records*/données en eux-mêmes, car c'est en grande partie ce contexte qui leur donne leur force de preuve.

En résumé, dans le cas des données comme dans celui des *records*, les métadonnées sont essentielles afin de comprendre ces derniers. C'est ce qui les rend exploitables et réutilisables.

Définir les responsabilités

En plus d'imposer une réflexion sur les métadonnées, le moment de rédiger un PGD ou une politique de *records management* demande aussi de bien établir les responsabilités concernant les données ou les *records* : « Ainsi, à chaque étape, une donnée peut faire l'objet d'un métier, d'une compétence particulière, d'une

[90] ISO, *Information et documentation - Processus de gestion des documents d'activité - Métadonnées pour les documents d'activité - Partie 1 : principes = Information and documentation - Records management processes - Metadata for records - Part 1 : principles*. ISO 23081. Paris : AFNOR, 2018, p. 9

[91] *Ibid.*, p.3

(cc) BY-NC-ND

responsabilité particulière…[92] » Dans les deux cas, on insiste sur cet aspect de la démarche afin d'assurer une bonne gestion et préservation des *records* et des données. En effet, sans responsabilité définie, il existe un risque que les *records*/données se perdent en cours de processus. Dans le cas des PGD, les niveaux de responsabilités ne sont pas toujours clairement définis : souvent, les différents modèles ne demandent qu'à connaître qui sera responsable des données, de façon générale. Pour le *records management*, cette définition des responsabilités est beaucoup plus précise et se fait lors de l'analyse (et l'amélioration) des processus de travail. En gérant tout le cycle de vie du *record*, on s'assure qu'à chaque étape soit associée une personne (ou plutôt une position, car les personnes peuvent changer de poste) qui aura la responsabilité du *record* à cette étape. Ainsi, chaque personne sait de qui elle recevra quel *record* et à qui elle devra l'envoyer une fois que les tâches pour lesquelles elle est responsable sont terminées.

Si toutefois la rédaction d'un PGD ne demande pas de connaître quelles sont exactement les responsabilités pour chaque étape de la vie d'une donnée, ce n'est pas pour autant que la notion de cycle de vie est ignorée, comme nous le verrons ci-dessous.

Le cycle de vie

Le cycle de vie est une notion qui est au cœur du *records management* : on ne peut envisager réellement d'appliquer cette démarche sans se positionner en amont du processus, afin de contrôler toutes « les étapes de l'existence d'un document de sa création à la mise en œuvre de son sort final[93] ». C'est ce contrôle qui permet de conserver les caractéristiques d'authenticité, de fiabilité, d'intégrité et d'exploitabilité du *record*.

De façon similaire, la rédaction d'un plan de gestion de données se fait généralement en amont du projet de recherche, et donc en amont de la création des données. On y prévoit la façon dont elles seront créées, comment elles seront exploitées et ce que l'on en fera une fois que leur usage primaire (le projet de recherche) sera terminé. Cet aspect de réutilisation est d'ailleurs mis de l'avant par

[92] Besnard, Myriam., Delobelle, Grégory. « Les archives du Bureau de recherches géologiques et minières en transition » *op. cit.*, p.69

[93] Kern, Gilliane., Holgado, Sandra. et Cottin, Michel. « Cinquantes nuances du cycle de vie : quelles évolutions possibles ? » **In** *Les cahiers du numériques* **[en ligne]**. 2015/2 (Vol. 11) p.40. Disponible sur : https://www-cairn-info.docelec.insa-lyon.fr/revue-les-cahiers-du-numerique-2015-2-page-37.htm (consulté le 24 mars 2017)

(cc) BY-NC-ND

la commission européenne dans son document *Lignes directrices pour le libre accès aux publications scientifiques et aux données de la recherche dans Horizon 2020* [94]. Il est aussi perçu comme un net avantage par ceux qui ont dû en rédiger : « mais aussi parce que c'est un vrai guide, un vrai outil qui permet aux chercheurs d'anticiper pas mal de choses dans le cycle de vie et dans la gestion de ses données… le cycle de vie des données.[95] » Plus encore, une bonne planification garantie la qualité des données de la recherche : « Ici, on entend par ''qualité'' de la donnée de la recherche son accès et son accessibilité, et donc aussi son exploitabilité, son usage approprié, mais également la pérennisation de ces qualités tout au long de son cycle de vie[96]. »

Mais quelles sont donc ces « étapes » ? En *records management*, elles ne font pas nécessairement objet d'une unanimité au sein de la communauté, au sens où cette notion a évolué au fil du temps. On parle au début d'un parcours des archives (sans nécessairement trop le définir), puis certains mettront de l'avant les valeurs primaires et secondaires des archives, reconnaissant que les archives jouent des rôles différents selon la valeur qui est mise de l'avant. Puis, Yves Pérotin articule la théorie des trois âges dans les années 60. Cette théorie dominera pendant longtemps le paysage archivistique français… Jusqu'à l'apparition du cycle de vie, notion très américaine qui accompagne souvent la théorie du *records management*[97]. Aujourd'hui, plusieurs experts français offrent leur propre définition des étapes du cycle de vie en *records management*. Par exemple, dans le guide de l'ADBS *Records management : mode d'emploi*, le cycle de vie est représenté comme étant toutes les étapes formant la « procédure » du *records management*[98] :

[94] Direction générale de la recherche et de l'innovation. *Lignes directrice pour le libre accès aux publications scientifiques et aux données de la recherche dans Horizon 2020. op. cit.*

[95] Entretien avec Christine Hadrossek, *op. cit.*

[96] Makhlouf-Shabou, Basma et Echernier, Lydie. « La gouvernance du cycle de vie des données de la recherche en Suisse : transfert et pérénnisation.» *op. cit.,* p. 90

[97] Kern, Gilliane., Holgado, Sandra. et Cottin, Michel. « Cinquantes nuances du cycle de vie : quelles évolutions possibles ? » *op. cit.*, p.43-54

[98] Drouhet, Geneviève., Keslassy, Georges., et Morineau, Elisabeth. *Records management : mode d'emploi*. *op. cit*.,. p.76

(cc) BY-NC-ND

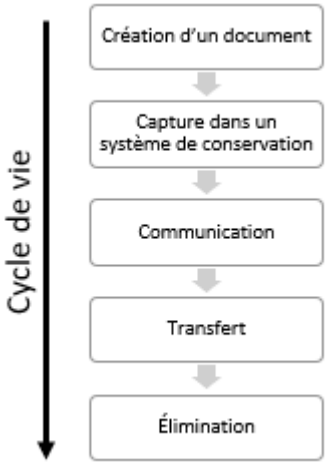

Figure 2 - Cycle de vie des documents en *records management* selon l'ADBS

Le rapport produit par l'APROGED en 2006 reprend des étapes semblables, tout en les explicitant dans leur document *La maîtrise du cycle de vie des documents numériques : présentation des concepts*[99], tel que démontré par la figure 4 ci-dessous.

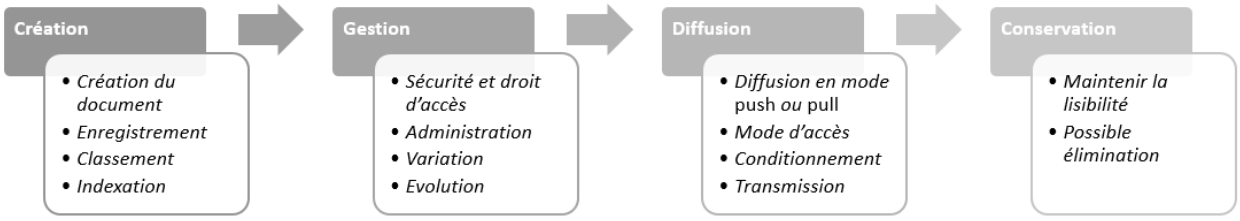

Figure 3 - Cycle de vie des documents numérique selon l'APROGED

Dans tous les cas, on constate qu'en France, le cycle de vie est envisagé de façon semblable : le *record* est créé, capturé et classé, diffusé puis finalement transféré pour mettre en œuvre son sort final (la conservation ou l'élimination). Il est intéressant de noter que ce cycle de vie est envisagé comme étant linéaire. Ce n'est pas un cycle qui se renouvelle.

Cela pourrait venir à changer, puisqu'on constate que certains pays anglo-saxons (en Australie d'abord et un peu au Canada aussi) se tournent de plus en plus vers la théorie du *records continum* qui permet « d'exprimer le fait qu'un document

[99] Groupe de travail DGME/SDAE – APROGED. *La maîtrise du cycle de vie du document numérique : présentation des concepts* (version 3 du 22 mai 2006) [en ligne]. Paris : APROGED, 2006, pp. 12-17. Disponible sur : https://documentslide.org/the-philosophy-of-money.html?utm_source=la-maitrise-du-cycle-de-vie-du-document-numerique (consulté le 10 janvier 2018)

(cc) BY-NC-ND

peut exister simultanément dans plusieurs dimensions[100] ». Ainsi, on reconnaît que le *record* ne cesse pas d'exister suite au sort final, mais qu'il peut, selon les circonstances, venir remplir d'autres rôles.

Du côté des données de la recherche, il n'y a pas cette notion de « sort final ». Au contraire, il est fortement encouragé de conserver les données afin de pouvoir par la suite les réutiliser. En ce sens, cette façon de voir le cycle de vie des données se rapproche beaucoup plus de la théorie du *records continum*. Par exemple, les cycles de vie sont presque toujours représentés sous forme de cycle continu, débutant par la création des données et se terminant par leur réutilisation, tel qu'on peut le constater dans les figures 5 et 6.

[100] Kern, Gilliane., Holgado, Sandra. et Cottin, Michel. « Cinquantes nuances du cycle de vie : quelles évolutions possibles ? » *op. cit.*, p.59

(cc) BY-NC-ND

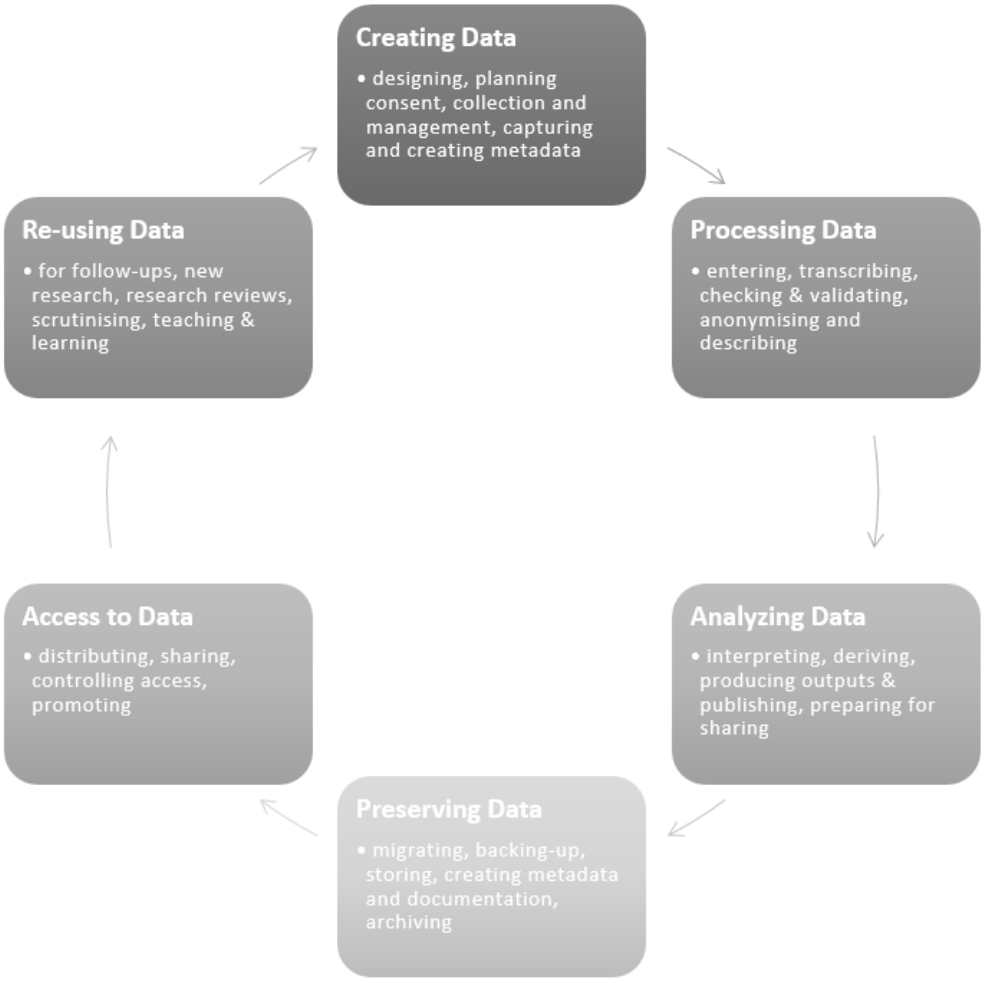

Figure 4 - Cycle de vie des données de la recherche selon le *UK Data Archive*[101]

[101] UK DATA ARCHIVE. *How we curate data*. **[en ligne]**. Disponible sur : http://www.data-archive.ac.uk/curate (consulté le 25 juillet 2018)

(cc) BY-NC-ND

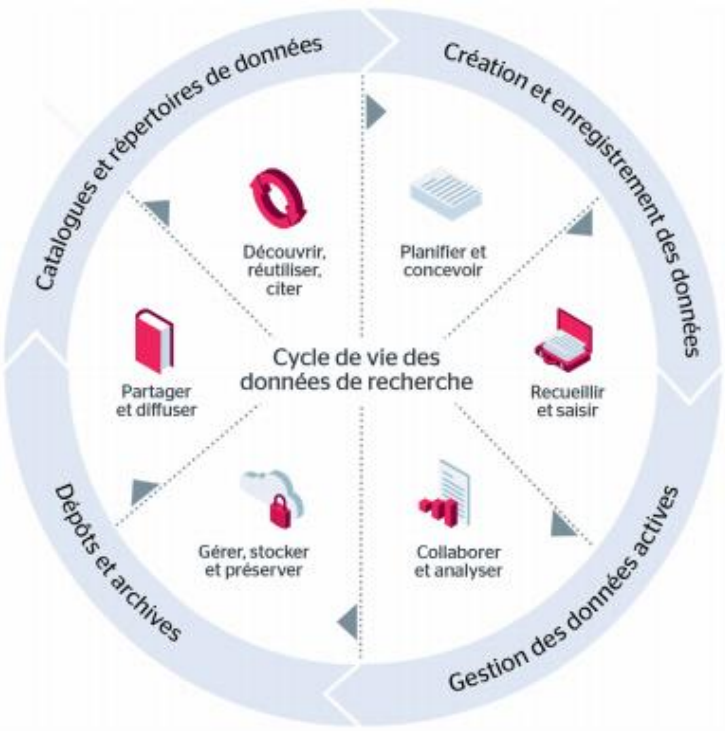

Figure 5 - Cycle de vie des données selon Données recherche Canada[102]

Cette différence entre les cycles de vie des *records* et ceux des données peut sembler importante, mais cela s'explique probablement par la façon dont on envisage ces démarches. Le *records management* s'intéresse surtout à la valeur primaire des *records*, celle qui fait que ces derniers sont probants. Ainsi, le cycle de vie a longtemps été construit autour de cette première valeur (cycle de vie linéaire). Toutefois, si le choix est fait de conserver les *records*, il n'est pas exclu que ceux-ci puissent être réutilisés dans un autre contexte (*records continuum*).

On peut sinon constater une forte similitude en ce qui a trait au cycle de vie. Que ce soit *record* ou données, ceux-ci suivent un cycle similaire : Création – Capture – Préservation – Communication – Réutilisation (pour *records*, s'il y a conservation pour valeur secondaire). Dans les deux cas, la démarche encourage de se placer en amont de la création des *records*/données et de prévoir les étapes jusqu'à leur sort final. Ce positionnement permet d'anticiper les étapes par lesquelles devront passer les *records*/données afin d'exercer un plus grand contrôle sur ces derniers et d'assurer leur qualité.

[102] Research Data Canada – Données de Recherche Canada. *Données de Recherche Canada.* **[en ligne]** Disponible sur : https://www.rdc-drc.ca/?wpdmdl=927 (consulté le 25 juillet 2018)

(cc) BY-NC-ND

Établir une durée de conservation

On constatera aussi que dans le cas des *records* comme celui des données de la recherche, le cycle de vie inclut inévitablement une période de préservation avant que son sort final soit appliqué… même si celui-ci est parfois une plus longue conservation.

Dans le cas du *records management*, on l'aura fait remarqué dans notre première partie, la période de conservation est souvent celle où les documents restent engageant. C'est donc la période pendant laquelle les *records* peuvent être utilisés pour contester une action judiciaire ou prouver la bonne mesure de leurs activités. Le Groupe-métiers AAF-ADBS notait toutefois que cette durée de conservation pouvait aussi varier selon l'analyse de risque réalisée pendant la mise en place de la démarche :

> « La prise en compte par la fonction *records management* du rapport risque/coût de la conservation peut l'amener à proposer des règles de conservation ne répondant pas totalement aux exigences de la réglementation. Par exemple, certaines banques ne conservent les chèques de faible montant que pendant quelques jours, soit une durée inférieure au délai de recours en justice.[103] »

Bref, il sera au *records manager* d'établir la durée de conservation des *records* en prenant en compte divers facteurs tels que la réglementation, l'environnement et le risque.

Dans le cas des DDR, puisque le but est souvent le dépôt dans un entrepôt de données pour en assurer le partage, la question des durées de conservation est moins présente. Il n'en reste pas moins qu'elle reste pertinente dans plusieurs cas. Pour n'en nommer que quelques-uns, on pourrait penser à un éditeur qui demande à un chercheur d'attendre une certaine période avant de rendre ses données publiques ou si la recherche doit rester confidentielle un certain temps pour des enjeux réglementaires ou économiques. Cette durée de conservation est donc bien souvent une période de rétention pendant lesquelles on doit conserver les données dans un endroit sécurisé avec des accès limités.

Bref, dans le *records management* tout comme lors de la rédaction d'un plan de gestion des données, il est important de reconnaître combien de temps nous allons

[103] Groupe métiers AAF – ADBS « Records management », « Comprendre et pratiquer le records management, Analyse de la norme ISO 15489 au regard des pratiques archivistiques françaises ». *op cit.*

(cc) BY-NC-ND

nous engager à conserver les données ou les *records.* Cette décision s'appuiera sur plusieurs facteurs tels que la gestion des risques, les enjeux de confidentialité ainsi que la réglementation de notre environnement.

La question du droit d'accès

Donner la bonne information, à la bonne personne, au bon moment.

Cette phrase simple est parfois employée pour définir, en résumé, ce qu'est le *records management*. Elle permet d'insister sur l'importance de bien gérer l'information afin d'atteindre l'objectif qu'elle décrit. Toutefois, en nous y attardant un peu plus, elle nous rappelle aussi que dans une démarche de *records management*, l'information n'a pas nécessairement vocation à être partagée à tous, et, qu'au contraire, cette démarche implique d'exercer un contrôle important sur l'information et son accès. Par ailleurs, l'ADBS en a effectivement fait un des objectifs d'un bon *records manager,* soit : « communiquer le document selon les droits d'accès associés.[104] »

Cette notion de droit d'accès se retrouve un peu partout dans la démarche : il est important de les noter dans les métadonnées des *records*, il est important de comprendre qui doit avoir accès au *record* (ou non) lorsqu'on analyse les processus d'une entreprise/d'un organisme, en mettant en place le système, etc.

Il en est ainsi, car en tant qu'actif informationnel ou document probant, il arrive bien souvent que les *records* doivent correspondre à un certain niveau de confidentialité et n'ont pas vocation à être diffusés largement.

Au contraire, on encourage fortement les chercheurs à diffuser leurs données de recherche, et les PGD sont vus comme une façon de les aider à les structurer afin d'assurer une meilleure diffusion. On pourrait donc supposer que la question du droit d'accès ne se pose pas, puisque les données ont vocation à être publiques. Raisonner ainsi serait par contre ignorer plusieurs enjeux autour des données de la recherche. Quelques-uns de ces enjeux sont bien résumés par Tenopir : « Depending on the nature of the data, issues of security, intellectual property, confidentiality, potential data misuse or misinterpretation, informed consent when future uses of the data are unknown, and inability to ensure human subject privacy via anonymization may

[104] *Ibid.*

(cc) BY-NC-ND

come into play[105] ». En somme, certaines données n'ont pas vocation à être partagées ou ne peuvent pas toujours être diffusées massivement, car plusieurs enjeux économiques ou de confidentialité peuvent entrer en conflit avec la diffusion des données. Spécialement aujourd'hui alors qu'on porte une plus grande attention aux données personnelles.

En rédigeant le plan de gestion de données, il est important de considérer ces enjeux, et celui-ci peut même nous y aider : « La démarche [de rédaction d'un PGD] elle est vraiment déjà dans un premier temps de se dire : bon, quelles sont les données que je vais mettre à disposition, pourquoi oui, pourquoi non ? Alors si on justifie que non, on le justifie aussi dans les plans de gestion de données. Et cette réflexion, elle a juste le mérite d'exister en amont…[106] »

De plus, en se posant la question du stockage et de la responsabilité des données, celle du droit d'accès se posera inévitablement : si on enregistre les données à un endroit X, est-ce que toutes les personnes devant y accéder seront en mesure de le faire ? Est-ce que stocker des données à cet endroit permet de protéger des données confidentielles ?

En résumé, on peut affirmer que la question du droit d'accès est une question transverse et essentielle aux démarches de *records management* et de rédaction de plan de gestion de données. Selon le contexte, divers enjeux de confidentialité peuvent être liés aux *records* et aux DDR, et il est indispensable d'en prendre compte lors que l'on met en place un SGDA ou que l'on rédige un PGD. En faire autrement pourrait compromettre l'intégrité de nos *records* ou de nos données.

LES POINTS DE DIVERGENCE

Des finalités différentes

Malgré tous leurs points communs, une des principales différences entre le *records management* et la rédaction d'un PDG est la finalité de la démarche. Dans le cas du premier, la démarche se base sur des objectifs très concrets : « la gestion des documents d'activité répond à une quintuple nécessité : permettre à un

[105]Tenopir, Carol., Dalton, Elizabeth D., Allard, Suzie., et al. « Changes in Data Sharing and Data Reuse Practices and Perceptions among Scientists Worldwide ». *PLoS One* **[en ligne]**, 2015/8 (vol. 10) p. 4 Disponible sur : https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4550246/#? (consulté le 09 décembre 2017).

[106] Entretien C. Hadrossek, *op. cit.*

(cc) BY-NC-ND

organisme de rendre compte de ses actions, de remplir ses obligations légales et réglementaires, de se défendre en cas de litige, de fonder ses prises de décisions sur des informations fiables et de participer à la bonne marche de ses activités.[107] » Ainsi, on parle plutôt d'objectifs de gestion, de contrôle et de maîtrise de l'information. Le *records management* n'est pas vu comme un moyen d'arriver à diffuser un document au plus de personnes possibles. Au contraire, cette démarche veut conserver les *records* (et leurs métadonnées) en faisant le moins de changements possible une fois les *records* validés.

Dans le cas des PGD, il est vrai que la démarche veut, dans un premier temps, s'assurer que les données créées soient bien conservées, que l'endroit de leur dépôt soit connu et sécurisé et identifier qui sera responsable de ces données et qui y aura accès, etc. Par contre, cela n'est pas vraiment dans le but de contrôler l'information et la façon elle sera réutilisée. Le programme Horizon2020 identifie la finalité des plans de gestion de données de façon simple : « L'objectif est de faciliter la gestion, la réutilisation et la protection, ainsi que l'archivage à long terme de données produites, grâce à des financements publics.[108] » Cet aspect est aussi porté par le mouvement de l'*open data* puisque les données de recherche sont vues comme des biens publics, lorsqu'elles sont produites à l'aide de financement d'origine public. Dans le cas des PGD donc, le but n'est pas de contrôler l'information, mais de s'assurer qu'elle pourra être disponible, dans le cas où quelqu'un souhaiterait la réutiliser, soit pour valider les résultats des recherches qui ont eu lieu avec les données ou alors afin de permettre de les utiliser dans le cadre d'une recherche différente, qui aurait tout de même besoin des mêmes données.

Le fait que les PGD et le RM ont des finalités différentes, nous l'aurons compris, influence la façon dont on perçoit le cycle de vie. Le *records management* s'assure que le document reste authentique, fiable, intègre et exploitable pendant tout son cycle de vie parce qu'il peut être utilisé comme preuve pendant ce cycle. Par contre, en rédigeant un plan de gestion de données, le rédacteur s'assure que ces données resteront retrouvables, identifiables, accessibles et réutilisables, non pas

[107] Dessolin-Baumann, Sylvie. Des chartiers aux bases de données : les enjeux de la gestion des documents d'activité. **In :** « Normalisation et gestion des documents d'activité (*records management*) : enjeux et nouvelles pratiques pour notre profession » **[en ligne]**. *La Gazette des archives*, 2012/4 (nº 228), p.95. Disponible sur : https://www.persee.fr/issue/gazar_0016-5522_2012_num_228_4 (consulté le 12 décembre 2017)

[108] Ministère de l'enseignement supérieur, de la recherche et de l'innovation. Publication : guide d'aide à la rédaction d'un plan de gestion de données. **In :** *Horizon2020 : Le portail français du programme européen pour la recherche et l'innovation*. Disponible sur : http://www.horizon2020.gouv.fr/cid87457/publication-guide-d-aide-a-la-redaction-d-un-plan-de-gestion-de-donnees.html (consulté le 27 juillet 2018)

(cc) BY-NC-ND

spécifiquement pour qu'elles servent de preuves pour sa recherche (quoique cela peut très bien être un objectif secondaire des plans de gestion des données), mais pour, qu'ultimement, quelqu'un puisse s'emparer de son travail, le faire vivre et évoluer.

Cette différence, quoiqu'importante, me semble peu influencer le déroulement de la démarche elle-même de rédaction d'un plan de gestion de données ou de mise en place d'un système de *records management*. Toutefois, cela influencera grandement les décisions qui seront prises au cours de cette démarche : la précision des métadonnées qui accompagneront les *records*/données, l'endroit où les *records*/données seront entreposés, la façon dont ils seront traités, etc.

Une mise en place permanente, un projet ponctuel

Au fur et à mesure des lectures, une importante différence me semble être la fréquence à laquelle il serait nécessaire de mettre en place une démarche de rédaction de plan de gestion de données et une de *records management*. Lorsque l'on met en place une démarche de RM, il faut commencer par faire une évaluation de l'entreprise (ou direction, etc.) où l'on souhaite mettre en place cette démarche. L'un des buts de cette démarche est éventuellement de connaître tout ce qui sera produit comme *records* dans le cadre des activités de l'entreprise et ainsi prévoir ce que l'on prévoit conserver, comment le préserver et quelle forme/format devront prendre ces *records*. Ultimement, le but est de mettre en place un système où l'on connaîtra à l'avance les *records* produits, et les instructions sur la façon de les gérer seront claires. Évidemment, le système aura à être évalué régulièrement et des changements devront probablement être apportés au fil du temps en fonction des évolutions des activités ou de la réglementation, mais, idéalement, un système de *records management* se veut un système répétitif, que l'on met en place une fois et que l'on ajuste au fil du temps.

Dans le cas des plans de gestion de données, ceux-ci sont liés à un projet, et non à un département, une direction ou une entreprise spécifique. Le plan de gestion mis en place n'a donc qu'une durée de vie temporaire, qui est celle du projet. Bien sûr, le fait de remplir ce document demandera d'évaluer la production à venir, de prendre des décisions concernant la forme ou le format des *records*, la façon dont ils seront préservés, puis évaluer ses décisions tout au long du projet. Toutefois, les décisions prises dans le cadre de ce plan de gestion de données n'auront pas vocation

(cc) BY-NC-ND

à être répétées *ad vitam æternam* (jusqu'à un changement de réglementation ou d'activités). Les plans de gestion de données sont plutôt créés dans la cadre d'un événement unique ; et si certains éléments peuvent être repris d'un plan à l'autre (on pense à la présentation de l'organisme, par exemple), cela ne peut se faire automatiquement, c'est-à-dire qu'il faudra presque assurément que le rédacteur ait tout de même une réflexion avant de reprendre des éléments d'un ancien PGD.

Mettre en place un système ou rédiger un document

Le fait que le *records management* soit quelque chose mis en place de façon plus permanente amène aussi une autre différence importante avec les plans de gestion des données. Dans le premier cas, on parle de mettre en place un système complet, où tous les éléments, du début à la fin, seront contrôlés. Dans l'autre, on parle de rédiger un document en lien avec un projet. Les *records manager* ont potentiellement beaucoup plus de pouvoir sur l'environnement et les conditions dans lesquels évoluent leurs *records* que les personnes responsables des données de la recherche l'ont sur leurs données. Effectivement, puisqu'on parle de mettre en place un *système* dans un cas, et dans l'autre de rédiger un *document*, le premier implique de prendre en compte beaucoup plus de moyens. Dans l'idéal, le *records manager* aura non seulement contrôle sur la façon dont les *records* seront produits et dans quels formats, mais aussi aidera à déterminer les spécificités informatiques du logiciel ou système qui conservera les *records*. Ce sera à lui d'en faire l'évaluation et proposer des améliorations, si nécessaire.

Quelqu'un qui est en charge de rédiger un plan de gestion de données (que ce soit un chercheur, un documentaliste, un ingénieur de recherche) n'aura pas nécessairement le même contrôle sur l'environnement et les conditions dans lesquels évoluent les données. On le place responsable des données d'un seul projet ; il n'a donc pas le même rôle de consultant auprès de la direction générale ou de la direction des systèmes informatiques qu'un *records manager* pourrait avoir. Bien souvent, le rédacteur aura à faire avec les moyens qui lui sont offerts, à supposer déjà que son centre de recherche / université / laboratoire offre bel et bien le support informatique ainsi que le matériel pour préserver des données.

Ceci dit, il n'est pas question ici d'affirmer qu'un rédacteur de PGD se préoccupe moins des questions de sécurité, de conservation à long terme ou de formats qu'un *records manager*. Simplement, qu'il n'aura peut-être pas

(cc) BY-NC-ND

nécessairement le même contrôle ou la même influence sur l'équipement et l'environnement du fait de la nature « ponctuelle » de son projet, puisqu'il est en charge de rédiger un document et non de mettre en place un système.

DES DONNEES AUTHENTIQUES ? DES *RECORDS* ACCESSIBLES ?

Il convient de rappeler que les données et les *records* (dans le cadre des PGD et du RM) peuvent être vus comme des entités semblables ; toutes deux étant la preuve de phénomènes/activités au sein d'activités d'une organisation. Puisque ces deux éléments partagent certaines similarités, il semble pertinent de se poser la question : est-ce que des *records* peuvent être « FAIR » (*findable, accessible, interoperable* et *reusable*) ? Est-ce que des données de la recherche peuvent être authentiques, fiables, intègres et exploitables ? Est-ce possible d'être tout cela en même temps ?

Comme on a pu le constater plus tôt, la norme 15489 concernant le *records management* définit de façon assez explicite ce que signifie les caractéristiques d'authenticité, de fiabilité, d'intégrité et d'exploitabilité. Dans le cas des PGD, il n'existe pas encore de norme internationale définissant les caractéristiques que doivent posséder les données. Toutefois, tel qu'il a été évoqué plus haut, les principes FAIR se veulent un guide à la création et à la gestion des données de la recherche.

En plus d'énoncer les grands principes auxquels devraient se conformer les données de la recherche, l'initiative GO FAIR a aussi précisé quelques critères pour chacune des caractéristiques[109]. Dans le but de comparer les caractéristiques « essentielles » des données de la recherche avec celles des *records*, chacun des critères FAIR sera présenté dans les tableaux 2 à 5 ci-dessous, regroupé par la caractéristique à laquelle il se rapporte, et étudié sous la loupe du *records management* ; est-ce qu'on peut y trouver des similarités ? est-ce qu'on peut appliquer ce critère aux *records* ?

Puis, la démarche inverse sera faite pour les données de la recherche dans les tableaux 6 à 9. Est-ce possible de leur appliquer les critères d'authenticité, de

[109] Pour ce mémoire, la version française utilisé par l'INRA a été utilisé : INRA. Principes FAIR. **In :** *Datapartage*. Disponible sur : https://www6.inra.fr/datapartage/Technologies/Principes-FAIR (consulté le 27 juillet 2018)

(cc) BY-NC-ND

fiabilité, d'intégrité et d'exploitabilité qui sont exigés par la démarche de *records management* ?

Des *records* FAIR ?

Retrouvable (*findable*)	
F1 Les données et les métadonnées sont identifiées par un identifiant global unique et pérenne.	**Similarités avec le *records management*** L'identifiant unique et pérenne permet d'identifier facilement et avec certitude les données, ainsi que de pouvoir les reconnaître sur une période temporaire étendue. Il faut donc que cet identifiant soit résistant face aux changements technologiques[110]. Ainsi, cela nous aide à reconnaître avec confiance un jeu de données ; cela fait donc échos aux principes de fiabilité et d'intégrité.
	Peut-on appliquer ce critère aux *records* ? Oui, attribuer un identifiant unique et pérenne (pas nécessairement global par contre, car les records n'ont pas à être partagés publiquement) aide à mieux gérer les *records*.
F2 Les métadonnées décrivant les données sont riches.	**Similarités avec le *records management*** Les métadonnées permettent de mieux connaître le contexte de création des données. Qui les a créées ? Comment ? Elles permettent aussi de connaître le traitement qu'a pu subir ces données. En ce sens, ce critère fait écho aux caractéristiques d'authenticité et d'intégrité.
	Peut-on appliquer ce critère aux *records* ? Absolument. Les normes ISO 15489 et particulièrement la norme ISO 23081 décrivent avec beaucoup de détails quelles métadonnées devraient accompagner les *records* et recommandent toutes deux d'inclure toutes les métadonnées pertinentes.
F3 Les données et les métadonnées sont enregistrées et indexées dans un dispositif permettant de les rechercher.	**Similarités avec le *records management*** Le *records management* n'est pas seulement une démarche, mais c'est aussi toute la mise en place d'un système qui permet d'exploiter le plus efficacement possible les *records*. En ce sens, on parle ici d'exploitabilité.
	Peut-on appliquer ce critère aux *records* ? Oui, même qu'on pourrait dire que ce critère est indispensable à une bonne démarche de RM.
F4 Les métadonnées spécifient l'identifiant de la donnée.	**Similarités avec le *records management*** Ce critère permet de s'assurer que l'on consulte / utilise les bonnes données. Si on ne peut retrouver l'identifiant des données facilement, on ne peut être certain de leur provenance. Autrement, cela permet d'assurer que les données consultées / utilisées sont authentiques et fiables.
	Peut-on appliquer ce critère aux *records* ?

[110] INSIT. « Définitions ». **In :** *Une introduction à la gestion et au partage des données de la recherche*. Disponible sur : http://www.inist.fr/donnees/co/module_Donnees_recherche_37.html (consulté le 27 juillet 2018)

(cc) BY-NC-ND

	Tout à fait. Ce critère est d'ailleurs mentionné par les normes ISO 15489 et ISO 23081[111].

Tableau 2 - Retrouvable

Cette caractéristique d'avoir des données retrouvables semble être bien liée au RM. Effectivement, les différents critères énoncés ci-dessous sont de bonnes pratiques que l'on doit mettre en place si l'on souhaite se conformer aux normes de *records management* dans un environnement électronique.

Accessible	
A1 Les données et les métadonnées sont accessibles par leur identifiant via un protocole de communication standardisé.	**Similarités avec le *records management*** Ce critère existe pour s'assurer que les données puissent être communiquées à toute personne en faisant la demande ou ayant l'intérêt d'y accéder. On parle ici plutôt de protocole informatique. Si le RM n'encourage pas une démarche aussi ouverte, le système de gestion des *records* demande tout de même que les personnes autorisées puissent accéder à un document facilement. On aborde ici un peu la question de l'exploitabilité.
	Peut-on appliquer ce critère aux *records* ? Oui et non. Bien que le RM implique que l'on devrait pouvoir retrouver un *record* par son identifiant, cela n'a pas à être par un protocole standardisé, mais plutôt par un système de gestion performant.
A1.1 Le protocole utilisé est ouvert, libre et peut être implémenté de manière universelle.	**Similarités avec le *records management*** Il est ici plus difficile de tisser un lien avec le *records management*. Ce critère existe pour que les données puissent être diffusées rapidement. Comme cela a été mentionné auparavant, le RM ne cherche absolument pas à partager une information au plus de gens possible, mais doit souvent, au contraire, en restreindre l'accès. En ce sens, il n'y a pas de similarités entre ce critère et la démarche du RM.
	Peut-on appliquer ce critère aux *records* ? Non.
A1.2 Le protocole utilisé permet l'authentification et l'autorisation si besoin.	**Similarités avec le *records management*** Puisque la question de la confidentialité des *records* est souvent importante lors d'une démarche de *records* management, oui, le système utilisé pour diffuser les données doit permettre l'authentification et l'autorisation. C'est une question de sécurité, mais aussi d'intégrité (afin que les *records* ne subissent pas de modifications indésirables).
	Peut-on appliquer ce critère aux *records* ? Oui, et pour être plus exact, il faudrait retirer le « si besoin » de ce critère.
A2 Les métadonnées sont accessibles même quand les données ne le sont plus.	**Similarités avec le *records management*** Bien qu'il n'y ait pas explicitement de consignes à ce sujet dans la norme du RM, il est facile d'argumenter que l'on doit garder une trace des *records*

[111] ISO 15489-1, *op. cit.* p.6 et ISO 23081-1, *op. cit* p.7

(cc) BY-NC-ND

	que l'on a supprimés. Effectivement, en gardant une trace de la destruction et de ce qu'était le *record* (les métadonnées donc), on peut prouver que l'on a bien respecté les différentes obligations ou réglementations qui peuvent exiger d'éliminer des *records*. Plus encore, si une copie illégitime d'un *record* venait à circuler, cela pourrait aider à prouver que le document est un faux.
	Peut-on appliquer ce critère aux *records* ? Bien qu'il ne soit pas exigé dans la norme, c'est probablement une bonne pratique de prendre en compte ce critère.

Tableau 3 - Accessible

La caractéristique d'accessibilité se porte un peu moins bien à la pratique du RM que celle précédente. Cela peut s'expliquer par les finalités des deux démarches, tel qu'il est abordé plus haut. Bien qu'une bonne démarche de *records management* demande que les personnes ayant l'autorisation à consulter/modifier un *record* y aient accès relativement aisément, l'accessibilité des caractéristiques FAIR comporte un aspect plus universel. On souhaite que les données soient accessibles à tous, et non à un groupe restreint d'individu. Ainsi, bien que certains critères plus précis correspondent bien aux *records*, on ne peut affirmer que ceux-ci doivent être accessibles.

Interopérable	
I1 Les données et les métadonnées utilisent un langage formel, accessible, partagé et largement applicable pour la représentation des connaissances.	**Similarités avec le *records management*** Ce critère est important pour les données dans la mesure où celles-ci ont vocation à être partagées entre plusieurs organisations. Le fait d'utiliser un langage tel qu'il est décrit ici permet au plus de gens possible de comprendre et de pouvoir utiliser les données si nécessaires. Ici, on se rapproche un peu de l'exploitabilité.
	Peut-on appliquer ce critère aux *records* ? Oui… si on « limite » ce critère à l'organisation mettant en place une démarche de *records management*. C'est-à-dire que les *records* doivent être présentés de façon à ce que les individus y ayant accès puissent les comprendre.
I2 Les données et les métadonnées utilisent des vocabulaires qui respectent les principes FAIR.	**Similarités avec le *records management*** Ce critère se rapproche du *records management* au sens où les métadonnées et autres éléments qui accompagne les *records* doivent eux aussi être authentiques, fiables, intègres et exploitables. C'est donc un concept qui est semblable au RM dans son esprit.
	Peut-on appliquer ce critère aux *records* ? Comme on le constate, les critères FAIR ne s'appliquent pas totalement à une démarche RM, donc il est difficile d'appliquer ce critère en entier aux *records*.

(cc) BY-NC-ND

I3 Les données et les métadonnées incluent des liens vers d'autres (méta)données.	**Similarités avec le *records management***
	Ce critère fait référence au web de données « visant à favoriser la publication de données structurées sur le web, non pas sous la forme de silos de données isolés les uns des autres, mais en les reliant entre elles pour constituer un réseau global d'informations.[112] » En ce sens, ce critère existe pour que les données d'un projet de recherche puissent facilement être liées avec d'autres projets et comprises par d'autres organisations. Le RM n'a pas vocation à être aussi ouvert. Par contre, si l'on prend ce critère au premier degré, oui on y constate une ressemblance avec le *records management* ou plusieurs éléments doivent être réunis afin de démontrer qu'une organisation est conforme dans ses activités. Les liens entre ces éléments doivent donc être bien clairs, pour avoir une information complète et exacte, une information fiable.
	Peut-on appliquer ce critère aux *records* ?
	Encore une fois oui, si on limite la portée de ce critère à la seule organisation mettant en place une démarche de *records management*.

Tableau 4 - Interopérable

En somme, l'interopérabilité est présentée ici pour assurer que les données de la recherche soient utilisables par le plus de systèmes différents possible, sans qu'il y ait de trop gros freins au partage. Cette caractéristique ne peut donc s'appliquer totalement au *records management*, qui n'a pas vocation à diffuser largement les *records*. Toutefois, il semble que l'on puisse affirmer que l'interopérabilité soit importante dans la limite d'une organisation mettant en place une démarche de *records management*. En effet, il faut que les *records* puissent être regroupés et utilisés au même moment et dans le même système afin que l'on puisse démontrer notre conformité.

Réutilisable	
R1 Les données et les métadonnées ont des attributs multiples et pertinents.	**Similarités avec le *records management***
	La gestion des données et des métadonnées occupe une place importante au sein du *records management* (la norme ISO 23081 y est entièrement consacrée) et à cet effet on insiste sur l'importance d'accompagner les *records* de toutes les informations pertinentes.
	Peut-on appliquer ce critère aux *records* ?
	Oui, car les *records* doivent être accompagnés de toutes les informations nécessaires à la démonstration de la preuve.
	Similarités avec le *records management*

[112] BnF. Web sémantique, web de données : définitions. **In :** *Bibliothèque nationale de France*. Disponible sur : http://www.bnf.fr/fr/professionnels/anx_web_donnees/a.web_donnees_definitions.html (consulté le 27 juillet 2018)

(cc) BY-NC-ND

R1.1 Les données et les métadonnées sont mises à disposition selon une licence explicite et accessible.	Pour faire une réelle comparaison avec le RM, il faudrait ici demander pour qui les données sont mises à disposition. Ici, il est sous-entendu que les données doivent être mises à disposition au plus de gens possible. Par contre, on peut y trouver une ressemblance au sens où il est vrai que les droits d'accès doivent être clairs et compris par tous, afin de conserver l'intégrité de nos *records*.
	Peut-on appliquer ce critère aux *records* ? On ne peut l'appliquer tel quel, mais ce critère rappelle l'importance d'avoir des droits d'accès explicites.
R1.2 Les données et les métadonnées sont associées à leur provenance.	**Similarités avec le *records management*** Ici, on peut tracer un lien très clair avec le RM. Effectivement, afin de démontrer qu'un *record* est authentique, on doit pouvoir savoir d'où il provient, c'est-à-dire qui l'a créé, comment, et à quel moment. Les données et métadonnées de provenance sont donc essentielles dans les deux cas.
	Peut-on appliquer ce critère aux *records* ? Totalement.
R1.3 Les données et les métadonnées correspondent aux standards des communautés indiquées.	**Similarités avec le *records management*** Ce critère est un peu flou, et ne précise pas les standards (car bien sûr ceux-ci peuvent changer selon les communautés). En RM, ce critère peut facilement se traduire pour affirmer que les *records* doivent correspondre aux normes et réglementations en vigueur dans le cadre des activités de l'entreprise, afin qu'ils puissent être exploitables.
	Peut-on appliquer ce critère aux *records* ? Oui.

Tableau 5 - Réutilisable

Dans le cadre de l'initiative FAIR, la réutilisation est envisagée pour que des personnes, autres que le ou les créateurs des données, puissent se réapproprier lesdites données sans trop de difficultés. Si le RM n'envisage pas cette même finalité, les critères énoncés ci-dessus sont tout de même applicables, et sont mêmes implicites à une démarche de *records management*. Ces critères permettent de conserver le contexte de création d'un *record* et de s'assurer que ces derniers restent exploitables tout au long de la durée de leur exploitation primaire.

En conclusion, les caractéristiques FAIR ne s'appliquent pas automatiquement aux *records* et à leur gestion. Bien que la plupart soient pertinentes et que certains critères offrent de bonnes pratiques à mettre en place dans une démarche de RM, il convient qu'une bonne partie de ces derniers doivent être adaptés pour être conforme au *records management*.

Bien sûr, les finalités de ces deux démarches influencent les critères par lesquelles sont évaluées les différentes caractéristiques. Puisque l'un des objectifs

(cc) BY-NC-ND

des PGD est de diffuser les données de la recherche le plus largement possible, il est certain que les caractéristiques de « bonnes » données de la recherche prendront cet objectif en ligne de compte ; cela explique donc les ajustements à faire.

Des données de la recherche authentiques, fiables, intègres et exploitables ?

Inversement, on peut se demander si les caractéristiques d'authenticité, de fiabilité, d'intégrité et d'exploitabilité peuvent s'appliquer aux données de la recherche. Pour ce faire, les tableaux ci-dessous reprennent les précisions que la norme ISO 15489 apporte à chaque caractéristique pour s'en servir comme critère pour évaluer ladite caractéristique. Similairement à ce qui a été fait ci-dessus, on explicite par la suite si l'on peut retrouver des similarités avec la rédaction d'un plan de gestion de données et si ces critères peuvent s'appliquer aux données de la recherche[113].

<table>
<tr><th colspan="2">Authenticité</th></tr>
<tr><td rowspan="2">« Il convient que les règles, les processus, les politiques et les procédures qui contrôlent la création, la capture et la gestion des documents d'activité soient mis en œuvre et documentés pour garantir l'authenticité des documents d'activité. »</td><td>Similarités avec le la rédaction d'un PGD
La rédaction d'un plan de gestion de données est l'acte de documenter « la création, la capture et la gestion des [données] ». Par contre, bien que le plan de gestion de données inclut normalement le contexte du projet de recherche, celui-ci ne va pas jusqu'à inclure les politiques et la réglementation entourant les projets de recherche.</td></tr>
<tr><td>Peut-on appliquer ce critère aux DDR ?
Oui, en partie ou totalement. Les politiques et règles n'ont pas à être documentées lors de la rédaction d'un plan de gestion de données. Toutefois, il n'est pas non plus « interdit » de les inclure… Celles-ci peuvent être pertinentes dans le cas où, par exemple, le projet utilise des données personnelles, car la réglementation peut influencer la façon dont on gérera les données. Un autre exemple où documenter ces règles serait pertinent est dans un cas où l'organisation où œuvre un chercheur lui demande de verser ses données dans un dépôt en particulier, ou d'attendre un délai avant le versement.</td></tr>
<tr><td rowspan="2">« Il convient que les créateurs des documents d'activité soient autorisés et identifiés. »</td><td>Similarités avec la rédaction d'un PGD
Ce critère fait écho au critère R1.2 de l'initiative FAIR. Dans les deux cas, il est important de connaître la provenance des données / records.</td></tr>
<tr><td>Peut-on appliquer ce critère aux DDR ?
Absolument.</td></tr>
</table>

Tableau 6 - Authenticité

[113] ISO 15489-1, *op. cit.*, p.5

(cc) BY-NC-ND

Il semble évident que les données de la recherche se doivent d'être authentiques ; en effet, le contraire remettrait en question tout un projet de recherche. Toutefois, ces données n'ont pas nécessairement à être authentiques au sens où l'entend le *records management*. Bien que certaines politiques et certains règlements peuvent influencer la façon dont on créera ou gérera les données de la recherche, il n'y a pas une obligation aussi forte de documenter pourquoi cela est le cas.

Fiabilité	
« Il convient que les documents d'activité soient créés au moment de l'événement auquel ils se rapportent ou juste après, par des personnes qui ont une connaissance directe des faits ou par des systèmes utilisés systématiquement pour réaliser l'opération. »	**Similarités avec le la rédaction d'un PGD** Ce critère s'applique totalement aux données de la recherche (les résultats de la recherche pourraient être invalidés si cela n'était pas le cas). Puisque, pour être réutilisables, les données (et leurs métadonnées) doivent indiquer leur provenance, on peut supposer qu'une personne dirigeant un projet de recherche s'assurera d'indiquer le moment de création des données, leur auteur, etc. dans ses métadonnées.
	Peut-on appliquer ce critère aux DDR ? Oui, car sans être explicitement nommé dans les caractéristiques FAIR, il n'en reste pas moins que cela est un aspect important des PGD.

Tableau 7 - Fiabilité

Puisque les données de la recherche peuvent être vues comme une preuve du déroulement d'un projet de recherche, d'une expérimentation, il est important que les métadonnées de celles-ci puissent nous assurer que les DDR aient été obtenus de façon légitime.

Intégrité	
« Il convient de protéger un document d'activité contre toute altération non autorisée. »	**Similarités avec le la rédaction d'un PGD** Bien que ce critère ne soit pas explicitement mentionné dans les principes FAIR, il n'en reste pas moins applicable aux données de la recherche. En effet, la plupart des modèles de plan de gestion de données incluent une section sur la sécurité et l'accès des données.
	Peut-on appliquer ce critère aux DDR ? Oui.
« Il convient que les politiques et les procédures de gestion des documents d'activité précisent quels ajouts ou annotations peuvent être apportés à un	**Similarités avec la rédaction d'un PGD** Ce critère fait écho au critère R.1.1, qui demande à ce que les données et leurs métadonnées soient « mises à disposition selon une licence explicite et accessible.[114] »
	Peut-on appliquer ce critère aux DDR ?

[114] INRA. Principes FAIR. ***op. cit.***

(cc) BY-NC-ND

document d'activité après sa création, dans quelles circonstances ces ajouts ou ces annotations peuvent être autorisés et qui est habilité à les appliquer. »	Sans aucun doute.
« Il convient que toute annotation, tout ajout ou suppression autorisé apporté à un document d'activité soit explicitement indiqué et traçable. »	**Similarités avec la rédaction d'un PGD** On ne peut retrouver d'écho de ces critères dans les principes FAIR. Les données de la recherche peuvent être retravaillées plusieurs fois avant d'être utilisables par le chercheur, et ce ne sont pas tous les PGD qui demandent à ce qu'un historique des versions soit conservé[115]. Si l'on considère que l'on ignore tout ce travail sur les versions et que les données que l'on choisit d'inclure dans le PGD sont les données « validées » (au même titre que des *records* sont bien souvent des documents qui auront eu plusieurs versions avant d'être validés), ce critère devient important au sens où, même si on réutilise les données dans le cadre d'un nouveau projet, il serait important de conserver les données originales quelque part.
	Peut-on appliquer ce critère aux DDR ? Ce n'est pas obligatoire, mais il pourrait être considéré comme une bonne pratique de documenter les différentes versions.

Tableau 8 - Intégrité

La caractéristique d'intégrité amène la question suivante : à partir de quand considère-t-on que les données doivent être conservées sans plus de modifications ? Dans une démarche de *records management* nous avons tendance à considérer les documents validés comme ceux ayant une véritable valeur de *records*. Toutefois, il est plus difficile de reconnaître le moment où des données de la recherche peuvent être « validées », et c'est cela qui, en partie, les rend si difficiles à définir[116]. Le PGD peut aider cette démarche, au sens où, une fois que les données ont atteint la forme décrite dans ce document, on les considère comme « validées » et on conserve leur caractéristique d'intégrité à partir de ce moment.

Exploitabilité	
« Il convient qu'un document d'activité exploitable soit relié au processus ou à l'opération à l'origine de sa création. »	**Similarités avec le la rédaction d'un PGD** Encore une fois, cela fait écho au critère FAIR selon lequel on doit connaître la provenance des données afin de pouvoir les réutiliser. Effectivement, le simple fait de verser des données sans savoir comment on les aurait obtenues n'est pas suffisant en lui-même ; cela ne fait pas partie d'une bonne méthode scientifique.

[115] Cela n'est pas explicitement mentionné dans le modèle proposé par le programme Horizon2020. Toutefois, le modèle suivant inclut une partie avec une description des versions : DCC. Checklist for a Data Management Plan. v.4.0. **In :** *Data management plans.* Disponible sur : http://www.dcc.ac.uk/resources/data-management-plans (consulté le 27 juillet 2018)

[116] Fayet, Sylvie. « ''Données'' de la recherche, les mal-nommées ». *op. cit.*

(cc) BY-NC-ND

	Peut-on appliquer ce critère aux DDR ? Il est même essentiel.
« Il convient que les liens entre les documents d'activité et les opérations associées qu'ils documentent soient maintenus. »	**Similarités avec la rédaction d'un PGD** Comme il est mentionné plus haut, on doit connaître la provenance des données/métadonnées pour que celles-ci soient réutilisables. Ainsi, si un chercheur réutilise un jeu de données qui a servi dans le cadre d'un autre projet, il devra en indiquer la provenance dans son plan de gestion de données. De plus, les principes FAIR demandent à ce qu'une licence explicite accompagne chaque jeu de données, cette licence servira à conserver un lien entre un jeu de données et le projet original desquelles elles émanent. Finalement, si les principes FAIR cherchent à incorporer les données de la recherche au web de données, ce genre de lien devrait exister et être conservé grâce à cette structure de l'information.
	Peut-on appliquer ce critère aux DDR ? Oui, et plus particulièrement, ce critère rejoint plusieurs principes FAIR (F3, I3, R1.2).
« Il convient que les métadonnées d'un document d'activité contribuent à son exploitabilité en fournissant toute l'information qui peut être nécessaire à sa récupération et sa communication, par exemple des identifiants, des informations de format ou de stockage. »	**Similarités avec la rédaction d'un PGD** Toutes ces informations contribuent à ce qu'un *record* ou une donnée soit facilement retrouvable et surtout réutilisable.
	Peut-on appliquer ce critère aux DDR ? Oui.

Tableau 9 - Exploitabilité

L'exploitabilité semble aussi être une caractéristique essentielle des données de la recherche. En résumé, l'exploitabilité sous-entend que nos données/*records* sont bien gérés, c'est-à-dire que l'on peut les retrouver aisément (*findable*), que l'on sache d'où ils proviennent et qu'on puisse avoir confiance en leur utilisation (*reusable*).

En somme, on peut affirmer que les caractéristiques d'authenticité, de fiabilité, d'intégrité et d'exploitabilité soient tout à fait applicables aux données de la recherche. Ces caractéristiques permettent d'assurer le caractère probant des *records* : s'assurer de leur maintien tout au long du cycle de vie permet d'avoir confiance dans les documents et de pouvoir bien les utiliser pour démontrer que nos activités sont bien conformes. Au final, c'est aussi l'un des objectifs que l'on cherche à atteindre en rédigeant un plan de gestion de données : être confiant que nos données puissent démontrer notre conformité à la conclusion de notre projet de

(cc) BY-NC-ND

recherche ; les rédacteurs de plan de gestion de données ont donc tout intérêt à s'inspirer des caractéristiques du RM pour assurer la qualité de leurs données...

Ainsi, ce petit exercice de comparaison aura démontré qu'il n'est pas possible de faire des parallèles directs entre chaque principe FAIR et les caractéristiques des *records* exigées par la norme ISO 15489. Toutefois, en observant chacun des éléments de façon plus attentive, on peut trouver un peu de toutes les caractéristiques des *records* dans chacun des principes FAIR, et vice-versa. Si les deux ne sont pas totalement identiques, certains parallèles existent, sans aucun doute. Les différences que l'on constate entre les deux groupes de caractéristiques tirent son origine dans une partie des finalités différentes des deux démarches : le fait que l'une cherche à diffuser le plus largement possible les données, tandis que l'autre cherche à protéger les éléments probants. Ainsi, comme cela est un élément central de la rédaction d'un plan de gestion de données, il est certain que cela sera aussi un élément important dans la façon dont on évalue la qualité des données qui font partie de cette démarche ; c'est donc pourquoi les *records* ne peuvent pas se conformer totalement aux caractéristiques FAIR.

UNE DEMARCHE POUVANT EN AIDER UNE AUTRE ?

Malgré quelques différences entre le *records management* et les plans de gestion de données, on a pu constater que ces derniers ont beaucoup en commun. Les deux sont une façon de gérer différents types d'information pour qu'ils conservent certaines qualités que l'on juge indispensables. Plus encore, ces mêmes qualités ne sont pas si différentes l'une de l'autre ! Logiquement, on peut donc en conclure que la façon dont on déploie un de ces deux éléments pourrait apporter des notions intéressantes à l'autre. Est-ce que s'inspirer du *records management* peut nous aider à améliorer les plans de gestion de données ? Est-ce que le contraire est possible ?

Faciliter la rédaction d'un plan de gestion des données

Plusieurs textes, ainsi que les entretiens que j'ai menés, ont fait remarquer que les compétences en gestion de l'information étaient presque indispensables pour bien gérer les données et facilitaient le travail avec les plans de gestion de données. Cela est d'autant plus vrai que, selon les études de Tenopir de 2011 et 2015, les problèmes limitant la bonne gestion des données de la recherche sont justement des

(cc) BY-NC-ND

problèmes pouvant être résolus par une démarche de *records management* ; les études en question mentionnaient que l'absence de politique générale sur la gestion des données, les lieux de stockage inappropriés, la confusion autour de la propriété intellectuelle ainsi que le manque de métadonnées accompagnant les jeux de données freinaient le partage des données des chercheurs[117]. Or, le *records management* peut justement clarifier ces points.

Ainsi, en utilisant la démarche du RM pour écrire notre PGD, peut-être pourrions-nous faciliter le processus et clarifier certaines des étapes nécessaires à la production d'un tel document. À quoi ressemblerait donc une démarche de *records management* lorsqu'appliquée à la gestion des données, et plus précisément, à la rédaction d'un plan de gestion des données ? Pour l'imaginer, nous avons repris la structure proposée par le guide pratique de l'ISO 15489 et intégré les éléments du mode d'emploi de l'ADBS. Cette démarche serait donc la suivante.

Étape A – Enquête préliminaire

Cette étape est déjà, d'une certaine façon, mise en œuvre par les universités françaises. Effectivement, les enquêtes effectuées par les universités de Lorraine, de Lille 3, de Rennes 2 de ou Lyon 1 dont nous avons parlé plus haut font partie de cette catégorie que sont les enquêtes préliminaires. Elles ont permis de mieux connaître les pratiques actuelles ainsi que d'évaluer les connaissances des collaborateurs qui seraient touchés par les plans de gestion des données. Si l'on cherche à appliquer cette démarche à un projet beaucoup plus précis, cette enquête préliminaire serait plus semblable à celle que l'on effectue avant de rencontrer des chercheurs dans le cadre d'un *data curation profile*. Dans tous les cas, elle permet de mieux connaître le sujet du plan de gestion de données ainsi que l'environnement dans lequel il devra être produit.

Étape B – Analyse des activités

De façon générale, l'analyse des activités permet de connaître le déroulement « typique » d'un projet de recherche, de la même façon qu'on analyserait les activités de n'importe quel organisme. Dans le cadre de la rédaction d'un plan de gestion de données, cette analyse des activités permettrait de connaître exactement la façon

[117] Tenopir, Carol., Allard, Suzie., Douglass, Kimberly., *et al.* « Data Sharing by Scientists : Practices and Perceptions ». *op. cit.* et Tenopir, Carol., Dalton, Elizabeth D., Allard, Suzie., et al. « Changes in Data Sharing and Data Reuse Practices and Perceptions among Scientists Worldwide ». *op. cit.*

(cc) BY-NC-ND

dont se déroulera un projet de recherche précis. Quelles étapes sont essentielles et par qui seront-elles exécutées ? Quelle sera la méthodologie employée ? À quel moment exactement seront produites les données ?

Étape C – Identification des exigences archivistiques

Tout comme les *records,* de nombreuses exigences peuvent être liées aux données de la recherche, nous n'avons qu'à penser aux organismes de financement qui exigent la rédaction d'un PGD, ou même demandent carrément la dissémination des données. On risque aussi de devoir prendre compte des données à caractère personnel, des projets qui touchent à des enjeux économiques ou de propriété intellectuelle. Les exigences concernant la conservation et la publication des données de la recherche vont certainement guider les décisions que quelqu'un pourrait prendre lors de la rédaction d'un plan de gestion de données, il est donc important de s'y attarder.

Étape D – Évaluation des systèmes existants

Afin de bien gérer et partager les données de la recherche, spécialement lorsque ces dernières sont sous formes numériques, il est important d'avoir une bonne infrastructure. L'évaluation des systèmes existants permettra de savoir ce qu'il en est pour l'instant. Est-ce que le chercheur utilise Dropbox ou Google Drive ? Fait-il des sauvegardes supplémentaires par sécurité ? Inclut-il des métadonnées suffisantes ? Est-ce que son laboratoire possède une méthode particulière pour gérer les données de la recherche ? Cette évaluation permettra de savoir si le plan de gestion de données consistera à décrire ce qui se fait présentement ou si l'on doit changer les méthodes et les systèmes.

Étape E – Identification de la stratégie pour la satisfaction des exigences archivistiques

Cette étape pourrait être le moment où la personne rédigeant le PGD et le chercheur discutent afin de trouver le meilleur moyen d'organiser, conserver et même partager les données. C'est à ce moment où on peut se poser la question de quels changements il faudra apporter, concrètement, dans les façons de faire du chercheur. Devrait-on sélectionner un standard de métadonnées en particulier ? Est-ce que l'on peut se tourner vers un dépôt de données sécurisé ? Ou alors a-t-on accès à un disque dur externe ? Doit-on adapter le format dans lequel les données sont

(cc) BY-NC-ND

créées ? Doit-on organiser des séances de formations pour les personnes qui seront impliquées dans le projet ? Ainsi de suite…

Étape F – Conception d'un système d'archivage

Concrètement, cette étape correspond à la rédaction du plan de gestion de données. C'est ici que l'on met sur papier qui sera responsable des données, comment celles-ci devront être gérées, à quel moment et si elles devront être partagées, etc. Bref, tout ce qui a été énoncé plus haut.

Étape G – Mise en œuvre d'un système d'archivage

Cette étape est tout simplement la mise en œuvre du plan de gestion des données, soit l'exécution du projet de recherche en suivant les principes et les décisions qui ont été prises lors de la rédaction du plan.

Étape H – Contrôle a posteriori

Cette étape doit être exécutée une fois le projet de recherche terminé. Est-ce que les principes et les consignes que nous y avons énoncés ont été respectés ? Sinon, pourquoi ? Doit-on modifier quelque chose dans notre plan de gestion de données pour que celui-ci corresponde mieux à la réalité ? Ce qui est intéressant, c'est que cette étape est déjà intégrée, d'une certaine manière, aux consignes données par le programme Horizon 2020, puisque celui-ci demande d'avoir plusieurs versions du PGD, soit une en début de projet, et de le mettre à jour régulièrement, lors de changement important ou d'évaluations du projet[118].

En somme, sans être complètement révolutionnaire, ce guide semble offrir une marche à suivre qui s'applique relativement bien au processus de rédaction d'un plan de gestion de données. Celle-ci permet d'envisager en amont les problématiques qui pourraient être soulevées par la préservation ou le partage des données de la recherche. Au lieu de se retrouver face à un document, seul, qui peut sembler intimidant aux personnes qui sont moins familières avec la gestion documentaire, on peut retrouver dans un guide comme celui-ci une marche à suivre qui peut aider à la rédaction du PGD.

De plus, ce guide pourrait même potentiellement être adapté à plus grande échelle, c'est-à-dire servir de référence à des personnes cherchant à développer des

[118] Directorate-General for Research & Innovation, *H2020 Programme: Guidelines on FAIR Data management in Horizon 2020. op. cit.*, p.5

(cc) BY-NC-ND

offres de service en matière de données de la recherche. Effectivement, les PGD ne sont souvent, au final, qu'un morceau d'un plus grand système permettant de produire des données de bonne qualité et de les conserver adéquatement.

Un plan de gestion de données pour identifier et gérer nos *records*

Il semblerait donc que d'appliquer une démarche de *records management* au plan de gestion de données soit une option intéressante, permettant de mieux s'y retrouver dans la rédaction d'un tel document. Cependant, le contraire est difficilement envisageable. Comme nous l'avons dit plus haut en discutant des points de divergences, une des grandes différences entre le *records management* et les plans de gestion de données est que l'un consiste en la mise en place d'un système et l'autre, en la rédaction d'un document. Si on ne peut appliquer globalement les principes du plan de gestion de données au RM, peut-on néanmoins envisager le plan de gestion de données comme une aide au *records management ?*

Les plans de gestion de données servent principalement à prévoir la façon dont seront gérées les données dès leur création jusqu'à leur réutilisation, c'est-à-dire tout au long de leur cycle de vie. En ce sens, leurs objectifs se rapprochent de l'outil tableau de gestion en *records management*, qui, au minimum, explicite les différentes typologies de document ainsi que leur sort final. On y retrouve donc un peu, aussi, la notion de cycle de vie. Toutefois, il est très possible que ces outils soient beaucoup plus détaillés, sans qu'il n'existe de modèles précis quant à savoir quoi inclure dans ce genre de document. S'inspirer des plans de gestion de données pourrait donc aider un *records manager* à créer des tableaux de gestion, et éventuellement des référentiels de métadonnées aussi (puisque les métadonnées sont une part importante des PGD) plus détaillés.

Pour en donner un exemple, vous trouverez ci-dessous un PGD adapté aux *records* inspiré à la fois de ceux qui ont été développés par l'Université Paris-Diderot et le programme H2020, qui semblent pour l'instant les plus connus en France, selon les entretiens que j'ai réalisés et la littérature que j'ai consultée. Comme le but est d'en faire un outil ou une « check-list » facilitant la création de tableau de gestion et de référentiels de métadonnées, j'ai éliminé les parties concernant l'administration des projets pour ne garder que les éléments permettant de décrire les jeux de données ou, dans ce cas-ci, les typologies de *records*.

(cc) BY-NC-ND

Pour chaque typologie de *records*
Objets visés par ce plan • D'où proviennent les *records* ? (activités) • À qui appartiennent les *records* ? (quel service est responsable de le verser dans le SGDA) • De quel type de *records* parle-t-on (lettre, formulaire, courrier, etc.) • Quel est le format des *records* ? • Quel est le volume des *records* ?
Organisation des *records* • Quels standards seront utilisés pour les métadonnées des *records* ? • Quelle convention sera utilisée pour nommer les fichiers ? (incluant versionning) • Comment seront organisés les fichiers (plan de classement) • Est-ce qu'un identifiant sera utilisé ? Comment ? • Est-ce que vos *records* seront accompagnés de mots clés ? Comment ceux-ci seront-ils sélectionnés ?
Support et stockage des *records* • Quel sera le support de conservation des *records* • Où seront stockés les *records* pendant leur durée de validité ?
Sécurité et accès des *records* • Y a-t-il des enjeux de confidentialité à prendre en compte pour vos *records* ? • À qui les *records* seront-ils accessibles ? • Comment allez-vous gérer les droits d'accès aux *records* ? • Quelles seront les mesures de protection utilisées pour assurer la sécurité de vos *records* ?
Sort final • Quel sera le sort final des *records* (conservation, destruction, tri) • Si les *records* sont conservés à la fin de durée de validité ? Où seront-ils conservés ? • Lors de la conservation, est-ce que les *records* devront rester confidentiels, pourquoi ?

Tableau 10 - PGD adapté aux *records*

Ces questions, qui peuvent être répondues sous forme de textes ou organisées en tant que colonnes d'un tableau, permettent de fournir une sorte de liste de rappel quant à savoir si l'on a pensé aux éléments qui influenceront le cycle de vie des *records* ainsi qu'aux informations (métadonnées) qui devront les accompagner. Sans être un guide infaillible, ces questions que l'on retrouve normalement dans les PGD semblent pouvoir aider un *records manager* à s'assurer que les éléments dont il sera responsable conserveront les caractéristiques d'authenticité, d'intégrité, de fiabilité et d'exploitabilité.

Finalement, malgré leur échelle d'application différentes (un document versus un système) il reste que chacune de ces deux démarches, de ces outils ont quelque

(cc) BY-NC-ND

chose à apporter à l'autre. Leurs nombreux points communs et leur cycle de vie semblable font que l'on retrouve un peu de chaque démarche dans l'une, et que l'on peut même en améliorer certains aspects de ces démarches en s'inspirant de la deuxième… Sans être identique, comparer ces deux outils aura permis de faire ressortir les points forts de l'une et de l'autre ainsi que de s'en inspirer pour les améliorer l'une et l'autre. Somme toute, peut-être avons-nous effectivement affaire à des démarches complémentaires.

(cc) BY-NC-ND

CONCLUSION

Le dernier siècle en fut l'un où la quantité d'information et la quantité des échanges se sont grandement accrus. Différents milieux ont été touchés par cette multiplication exponentielle à différents moments.

Au milieu du 20e siècle, l'administration a vu le papier s'accumuler dans divers bureaux jusqu'à en devenir problématique. Face à cette quantité de documents, plusieurs pays anglo-saxons se sont tournés vers le *records management* afin de mieux gérer les *records*, soit les documents engageants émanent des activités d'une organisation. Ce système de gestion a aussi été introduit en France au moment où les administrations « s'électronisaient ». Par conséquent, elles devaient aussi faire face à une multiplication de l'information en plus de devoir gérer des supports de stockage de plus en plus instables. En se plaçant en amont du cycle de vie du document, les *records manager* pouvaient anticiper les informations qu'ils convenaient de conserver, la période durant laquelle il fallait les préserver ainsi la forme et le contenu de cette information pour en assurer la valeur probante, soit s'assurer de leurs caractéristiques d'authenticité, de fiabilité, d'intégrité et d'exploitabilité. Cette méthode de travail exige la mise en place de tout un système et d'outils de travail tels qu'un tableau de gestion des *records*, un plan de classement des activités et un référentiel des métadonnées.

Durant les dernières décennies, le monde scientifique a dû faire face à un problème semblable. L'évolution de la recherche a facilité la capacité à produire et échanger des données et les mouvements d'*open access*, *open science* et *open data* ont poussé le domaine scientifique à diffuser de plus en plus ces données de la recherche. Face à ces circonstances, plusieurs organismes de financement américains se sont mis à exiger des plans de gestion de données lorsqu'ils accordent des subventions. Ces documents permettent de structurer les données qui seront créées dans le cadre d'un projet de recherche pour éventuellement en favoriser le partage. Idéalement, les données d'un projet de recherche devraient correspondre aux principes FAIR, soit être *findable, accessible, interoperable* et *reusable*.

Le *records management* et les plans de gestion de données sont donc des réponses à des problèmes qui ne sont, sommes toutes, pas tellement différents. Plus

(cc) BY-NC-ND

encore, ces deux outils de gestion partagent plusieurs aspects : les deux ont permis de revaloriser le travail des sciences de l'information, les deux misent sur une démarche d'évaluation avant d'implanter les outils, sélectionnent soigneusement quelles informations deviendront *records* ou données, se préoccupent grandement du contexte de création de l'information, mettent de l'avant l'utilisation de métadonnées, mentionnent l'importance de définir les responsabilités dans les processus qui leur sont propres, incorporent les notions de cycle de vie, de durée de conservation et portent une attention particulière aux droits liés à la diffusion de l'information. Néanmoins, on ne peut affirmer que ces outils sont identiques : leurs finalités sont différentes (protection de l'information versus diffusion de l'information) et l'un consiste à mettre en place tout un système de façon permanente tandis que l'autre n'est qu'un document associé à un projet ponctuel.

De plus, nous aurons démontré que sans être totalement équivalentes, les caractéristiques exigées de « bonnes » données de la recherche et de « bons » *records* sont assez semblables et peuvent, presque dans tous les cas, être utilisé dans les deux outils pour nous aider à créer des données et des *records* de meilleure qualité.

Plus encore, les nombreux guides publiés pour mettre en place une démarche de *records management* peuvent relativement facilement être adapté à la rédaction d'un plan de gestion de données, ce qui pourrait éventuellement aider les personnes qui sont moins familières avec cette démarche. De même, l'idée d'instaurer tout un système n'est pas non plus à négliger ; on sait déjà que les chercheurs éprouvent certaines appréhensions face à l'exigence de plus en plus commune de rédiger des plans de données[119]. Mettre en place des systèmes informatiques, des processus de gestion des données et offrir les connaissances nécessaires à la bonne gestion des données pourrait être une façon de les rassurer et d'encourager leur participation.

De même, les plans de gestion des données offrent l'occasion de se poser concrètement la question : quelles informations choisit-on de conserver, et pourquoi ? Ceci pourrait permettre à un *records manager* de mieux connaître ses

[119] Notamment dans les articles suivants :

TENOPIR, Carol., ALLARD, Suzie., DOUGLASS, Kimberly., *et al.* « Data Sharing by Scientists : Practices and Perceptions »., *op.cit*
TENOPIR, Carol., DALTON, Elizabeth D., ALLARD, Suzie., *et al.* « Changes in Data Sharing and Data Reuse Practices and Perceptions among Scientists Worldwide ». *op. cit.*
Ce point a aussi été abordé lors de mes entretiens.

(cc) BY-NC-ND

records… et pourquoi pas, l'aider à déterminer lesquels sont plus essentiels que d'autres.

Finalement, il semblerait donc que ces deux outils puissent apporter des notions intéressantes l'un à l'autre, que la connaissance des PGD puisse être avantageuses pour un *records manager* de même que la connaissance du *records management* peut aider un rédacteur de PGD dans sa démarche. Il existe donc ici une certaine complémentarité.

Tout cela reste pour l'instant théorique. Malheureusement, nous ne pouvions, dans le cadre de ce mémoire, tester l'adaptation du *records management* au travail des chercheurs et à la rédaction des plans de gestion des données. Puisque la pratique risque de se développer au courant des prochaines années, ils pourraient être intéressants de revenir sur le sujet, afin de voir si, effectivement, la démarche de rédaction finirait par intégrer quelques notions de gestion des *records*... De même, l'inverse pourra aussi être vrai, puisse Christine Nougaret proposait dans son rapport *Une stratégie nationale pour la collecte et l'accès aux archives publiques à l'ère numérique*[120] que les plans de gestion de données pourraient être utilisé dans la gestion des archives en France. On peut donc constater que des liens commencent à se tisser ; il sera intéressant de suivre l'évolution du sujet…

120 Nougaret, Christine. *Une stratégie nationale pour la collecte et l'accès aux archives publiques à l'ère numérique* **[en ligne].** Rapport à Madame Audrey Azoulay, Ministre de la Culture et de la Communication. Paris : Ministère de la culture et de la Communication, 2017, p.8 Disponible sur : https://francearchives.fr/file/b0d6555950508ab637adb10ece33d381644d6d37/2017_03_24_RAPPORT_DEFINITIF_NOUGARET.compressed.pdf (consulté le 30 janvier 2018)

(cc) BY-NC-ND

SOURCES

Entretiens

Entretien avec Alina Dancieu, chargée d'études et Anne Cornilleau, chargée d'études, du CDSP le 15 janvier 2018

Entretien avec Christine Hadrossek, chargée de l'information scientifique et technique à l'UMS2414 Riate - CNRS le 17 janvier 2018

Entretien avec Cynthia Pedroja, chef de projet données de la recherche à SciencesPo Paris le mardi 27 février 2018

Entretien avec Madeleine Hubert, *data librarian* à la Bibliothèque numérique de l'Université de Lorraine le mardi 20 mars 2018

Textes de loi

Article 30 de la *LOI n° 2016-1321 du 7 octobre 2016 pour une République numérique (1)*. Legifrance **[en ligne].** 7 octobre 2016. Disponible sur : https://www.legifrance.gouv.fr/affichTexte.do;jsessionid=298A6C25EFA7003656E88AAFB3D19609.tplgfr25s_2?cidTexte=JORFTEXT000033202746&categorieLien=id (consulté le 16 avril 2018)

Article L211-1. **[en ligne]** Code du patrimoine. Disponible sur : https://www.legifrance.gouv.fr/affichCode.do;jsessionid=986188EC1F33C71FEB56F633683887FE.tplgfr30s_3?idSectionTA=LEGISCTA000006159940&cidTexte=LEGITEXT000006074236&dateTexte=20180813 (consulté le 13 août 2018)

Loi no 2000-230 du 13 mars 2000 portant adaptation du droit de la preuve aux technologies de l'information et relative à la signature électronique **[en ligne]**. JORF n°62 du 14 mars 2000 page 3968. Disponible sur : https://www.legifrance.gouv.fr/affichTexte.do?cidTexte=JORFTEXT000000399095&categorieLien=id (consulté le 13 août 2018)

Publications officielles

Guidelines for a Research Data Management Plan **[en ligne]**. Paris : Intergovernmental Oceanographic Commission of UNESCO, 16 p, 2016. (IOC Manuals and Guides, 73) Disponible sur : http://unesdoc.unesco.org/images/0025/002565/256544e.pdf (consulté le 14 novembre 2017)

(cc) BY-NC-ND

Déclaration de l'OCDE sur l'accès aux données de la recherche financée par des fonds publics **[en ligne]**. Paris : OCDE, 2004. Disponible sur : http://openaccess.inist.fr/?Declaration-de-l-OCDE-sur-l-acces (consulté le 14 novembre 2017)

COMMISSION DES COMMUNAUTÉS EUROPÉENNES, *Communication de la commission au parlement européen, au conseil et au comité économique et social européen sur l'information scientifique à l'ère numérique : accès, diffusion et préservation* **[en ligne]**. Bruxelles : Commission européenne, 2007, 12 p. Disponible sur : http://eur-lex.europa.eu/legal-content/FR/TXT/PDF/?uri=CELEX:52007DC0056&from=FR (consulté le 11 novembre 2017)

[COMMISSION EUROPEENNE] HIGH LEVEL-EXPERT GROUP ON SCIENTIFIC DATA, *Riding the wave : how Europe can gain from the rising tide of scientific data* **[en ligne]**. Bruxelles : Union Européenne, 2010, 36 p. Disponible sur : https://ec.europa.eu/digital-single-market/en/news/digital-agenda-unlock-full-value-scientific-data-high-level-group-presents-report (consulté le 20 octobre 2017)

DIRECTION GÉNÉRALE DE LA RECHERCHE ET DE L'INNOVATION, *Lignes directrice pour le libre accès aux publications scientifiques et aux données de la recherche dans Horizon 2020* **[en ligne]**. Bruxelles : Commission européenne, 2016 (v 3.1), 12 p. Disponible sur : http://www.donneesdelarecherche.fr/IMG/pdf/lignes-directrices_libre-acces_horizon2020_version3.1_tr-fr.pdf (consulté le 20 octobre 2017)

DIRECTORATE-GENERAL FOR RESEARCH & INNOVATION, *H2020 Programme: Guidelines on FAIR Data management in Horizon 2020* **[en ligne]**. Bruxelles : Commission européenne, 2016 (v 3.0), 12 p. Disponible sur : http://ec.europa.eu/research/participants/data/ref/h2020/grants_manual/hi/oa_pilot/h2020-hi-oa-data-mgt_en.pdf (consulté le 20 octobre 2017)

ETALAB. *Pour une action publique transparente et collaborative : plan d'action national pour la France 2018-2020* **[en ligne].** Paris : Direction interministérielle du numérique et du système d'information et de communication de l'Etat, 2018, 76 p. Disponible sur : https://www.etalab.gouv.fr/wp-content/uploads/2018/04/PlanOGP-FR-2018-2020-VF-FR.pdf (consulté le 13 août 2018)

NATIONAL SCIENCE FOUNDATION, *Scientists Seeking NSF Funding Will Soon Be Required to Submit Data Management Plans* **[en ligne]** (communiqué diffusé le 10 mai 2010). Disponible sur : https://www.nsf.gov/news/news_summ.jsp?cntn_id=116928 (consulté le 09 décembre 2017)

MINISTÈRE DE L'ENSEIGNEMENT SUPÉRIEUR, DE LA RECHERCHE ET DE L'INNOVATION. *Le Plan national pour la science ouverte : les résultats de la recherche scientifique ouverts à tous, sans entrave, sans délai, sans paiement* **[en ligne]** (modifié le 12 juillet 2017)**.** Disponible sur : http://www.enseignementsup-recherche.gouv.fr/cid132529/le-plan-national-pour-la-science-ouverte-les-resultats-de-la-recherche-scientifique-ouverts-a-tous-sans-entrave-sans-delai-sans-paiement.html (consulté le 13 août 2018)

(cc) BY-NC-ND

Normes

ISO, *Information et documentation - Gestion des documents d'activité Partie 1 : concepts et principes = Information and documentation - Records management Part 1 : Concepts and principles*. ISO 15489-1. Paris : AFNOR, 2016, 20 p.

ISO, *Information et documentation - "Records management" Partie 2 : guide pratique = Information and documentation - "Records management" Part 2 : guidelines*. ISO 15489-2. Paris : AFNOR, 2002, 39 p.

ISO, *Information et documentation – Système de gestion des documents d'activité Principes essentiels et vocabulaire = Information and documentation – Management systems for records – Fundamentals and vocabulary*. ISO 30300. Paris : AFNOR, 2011, 16 p.

ISO, *Information et documentation - Analyse des processus pour la gestion des informations et documents d'activités = Information and documentation - Work process analysis for records*. ISO 26122. Paris : AFNOR, 2014, 15 p.

ISO, *Information et documentation - Processus de gestion des documents d'activité - Métadonnées pour les documents d'activité - Partie 1 : principes = Information and documentation - Records management processes - Metadata for records - Part 1 : principles*. ISO 23081. Paris : AFNOR, 2018, 23 p.

(cc) BY-NC-ND

BIBLIOGRAPHIE

Méthodologie d'entretiens

BARBILLON, Enrick., LE ROY, Jeanne. *Petit manuel méthodologique de l'entretien de recherche : de la problématique à l'analyse*. Paris : Enrick B. Editions, 2012, 64 p.

Références générales

Abrégé d'archivistique: principes et pratiques du métier d'archiviste. 3e édition revue et augmentée. Paris : Association des archivistes français, 2012, 346 p.

BANAT-BERGER, Françoise., « Les archives et la révolution numérique », *Le Débat* **[en ligne]**, n° 158, 2010/1 (n° 158), Gallimard, p. 70-82. Disponible sur : http://www.cairn.info.docelec.insa-lyon.fr/revue-le-debat-2010-1-page-70.htm (consulté le 13 décembre 2017).

BANAT-BERGER, Françoise., DUPLOUY, Laurent., HUC, Claude., *L'archivage numérique à long terme, Les débuts de la maturité ?* Paris : La Documentation française, Direction des Archives de France, 2009, 284 p.

SUBER, Peter. *Qu'est-ce que l'accès ouvert ?* Nouvelle édition **[en ligne]**. Marseille : OpenEdition Press, 2016. Disponible : http://books.openedition.org/oep/1600 (consulté le 05 avril 2018).

WILLIAMS, Caroline. « Records and Archives : Concepts, Roles and Definitions ». **In** : BROWN, Caroline (edited by). *Archives and Recordkeeping : Theory into practice.* Londres : Facet Publishing, 2014 p. 1-29

NOUGARET, Christine. *Une stratégie nationale pour la collecte et l'accès aux archives publiques à l'ère numérique* **[en ligne].** Rapport à Madame Audrey Azoulay, Ministre de la Culture et de la Communication. Paris : Ministère de la culture et de la Communication, 2017, 52 p. Disponible sur : https://francearchives.fr/file/b0d6555950508ab637adb10ece33d381644d6d37/2017_03_24_RAPPORT_DEFINITIF_NOUGARET.compressed.pdf (consulté le 30 janvier 2018)

Données de la recherche

« Archiver la recherche : responsabilités partagées » *Actes des journées d'études de la section Aurore - archivistes des universités, rectorats, organismes de recherche et mouvement étudiants - de l'Association des archivistes français du 23 juin 2016* **In** *La Gazette des archives*, 2017/2 (n° 246), 107 p.

ANAGNOSTOU, Paolo., CAPOCASA, Marco., MILIA, Nicola. *et al.* «When Data Sharing Gets Close to 100% : What Human Paleogenetics Can Teach the Open Science

(cc) BY-NC-ND

Movement » *PLoS One* **[en ligne]**, 2015/3 (vol. 10). Disponible sur : http://journals.plos.org/plosone/article?id=10.1371/journal.pone.0121409 (consulté le 09 avril 2017)

BORGMAN, Christine L. *Big Data, Little Data, No Data : Scholarship in the Networked World* **[en ligne]**. Cambridge, London : MIT Press, 2015, 411 p. Disponible sur : http://unr-ra.scholarvox.com/book/88841741 (consulté le 27 octobre 2017).

CURTY, Renata Gonçalves., CROWSTON, Kevin., SPECHT, Alison., *et al.* « Attitudes and norms affecting scientist's data reuse ». *PLoS One* **[en ligne]**, 2017/12 (vol. 12). Disponible sur : http://journals.plos.org/plosone/article?id=10.1371/journal.pone.0189288 (consulté le 09 avril 2018)

DIETRICH, Diane., ADAMUS, Trisha., MINER, Alison., *et al*. « De-Mystifying the Data Management Requirements of Research Funders ». *Issues in Science & Technology Librarianship* **[en ligne]**, 2012 (n° 70). Disponible sur : http://istl.org/12-summer/refereed1.html (consulté le 14 janvier 2017)

FAYET, Sylvie. « ''Données'' de la recherche, les mal-nommées ». **In** : *Urfistinfo* **[en ligne]** (publié le 15 novembre 2013). Disponible sur : https://urfistinfo.hypotheses.org/2581 (consulté le 05 avril 2018).

GROUPE D'EXPERTS SUR LA PRÉSERVATION DE PORTAGE. *La préservation des données de recherche au Canada : Livre blanc* **[en ligne]**. Ottawa : Association des bibliothèques de recherche au Canada, 2018, 17 p. Disponible sur : https://portagenetwork.ca/wp-content/uploads/2018/05/Portage-PEG-WhitePaper-FR.pdf (consulté le 21 mai 2018)

HICKSON, Susan., POULTON, Kylie Ann., CONNOR, Marie. *et al.* « Modifying researchers' data management practices: A behavioural framework for library practitioners ». *IFLA Journal*, 2016/4 (vol. 42), p. 266 à 277

PROST, Hélène., SCHÖPFEL, Joachim. *Les données de la recherche en SHS. Une enquête à l'Université de Lille 3 : Rapport final.* **[en ligne]** Rapport de recherche. Lille : Université de Lille 3, 2015, 36 p. Disponible sur : http://hal.univ-lille3.fr/hal-01198379 (page consulté le 27 avril 2018)

SCHÖPFEL, Joachim. KERGOSIEN, Eric. PROST, Hélène. *" Pour commencer, pourriez-vous définir 'données de la recherche' ? " Une tentative de réponse*. **[en ligne]** Atelier VADOR : Valorisation et Analyse des Données de la Recherche ; INFORSID 2017, May 2017, Toulouse, 2017. Disponible sur : http://hal.univ-lille3.fr/hal-01530937/document (consulté le 27 avril 2018)

SERRES, Alexandre (dir.) *Données de la recherche en SHS. Pratiques, représentations et attentes des chercheurs : une enquête à l'Université Rennes 2.* **[en ligne]** Rapport de recherche. Rennes : Université de Rennes 2, 2017, 160 p. Disponible sur : https://hal.archives-ouvertes.fr/hal-01635186 (consulté le 13 décembre 2017).

(cc) BY-NC-ND

TENOPIR, Carol., ALLARD, Suzie., DOUGLASS, Kimberly., *et al.* « Data Sharing by Scientists : Practices and Perceptions ». *PLoS One* **[en ligne]**, 2011/6 (vol. 6) 21 p. Disponible sur : https://www.researchgate.net/publication/51476349_Data_Sharing_by_Scientists_Practices_and_Perceptions (consulté le 09 décembre 2017).

TENOPIR, Carol., DALTON, Elizabeth D., ALLARD, Suzie., *et al.* « Changes in Data Sharing and Data Reuse Practices and Perceptions among Scientists Worldwide ». *PLoS One* **[en ligne]**, 2015/8 (vol. 10) 24 p. Disponible sur : https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4550246/#? (consulté le 09 décembre 2017).

TIBOR, Koltay. « Data governance, data literacy and the management of data quality ». *IFLA Journal*, 2016/4 (vol. 42), p. 303-312

Plan de gestions des données

Une introduction à la gestion et au partage des données de la recherche **[en ligne]**. Vandoeuvre-lès-Nancy : Institut de l'information scientifique et technique – CNRS. Module de formation, 2014. Disponible sur : http://www.inist.fr/donnees/co/module_Donnees_recherche.html (consulté le 12 novembre 2017)

BISCHOFF, Carolyn., JOHNSTON, Lisa. « Approaches to Data Sharing: An Analysis of NSF Data Management Plans from a Large Research University ». *Journal of Librarianship and Scholarly Communication* **[en ligne]**, 2015/2 (vol. 3) 27 p. Disponible sur : https://jlsc-pub.org/articles/10.7710/2162-3309.1231/galley/90/download/ (consulté le 09 décembre 2017)

CARTIER, Aurore., MOYSAN, Magalie., et REYMONET, Nathalie. *Réaliser un plan de gestion de données*. Paris : Université Paris Diderot, Sorbonne Paris Cité, Université Paris Descartes, 2015, 30 p.

CARTIER, Aurore., MOYSAN, Magalie., et REYMONET, Nathalie. Gestion des données, partage et conservation pérenne avec le Data Management Plan. **In :** CAVALIÉ, Étienne., CLAVERT, Frédéric., LEGENDRE, Olivier., *et al. Expérimenter les humanités numériques. Des outils individuels aux projets collectifs* **[en ligne].** Montréal : Les presses de l'université de Montréal, 2017 (Parcours numériques). Disponible sur : http://parcoursnumeriques-pum.ca/gestion-des-donnees-partage-et-conservation-perenne-avec-le (consulté le 14 août 2018)

DIEKEMA, Anne R., WESOLEK, Andrew., WALTERS, Cheryl D. « The NSF/NIH Effect: Surveying the Effect of Data Management Requirements on Faculty, Sponsored Programs, and Institutional Repositories ». *Journal of Academics Librarianship* **[en ligne]**, 2014/3-4 (vol. 40) p. 322-331 Disponible sur : http://www.sciencedirect.com/science/article/pii/S0099133314000640 (consulté le 09 décembre 2017)

(cc) BY-NC-ND

LATHAM, Bethany. « Research Data Management: Defining Roles, Prioritizing Services, and
Enumerating Challenges ». *Journal of Academics Librarianship*, 2017/3 (vol. 43) p. 263-265 Disponible sur : http://www.sciencedirect.com/science/article/pii/S0099133317301453 (consulté le 09 décembre 2017)

MORLOCK, Emmanuelle., ROMIER, Geneviève. *Les enjeux du « data management plan » pour les chercheurs* [présentation]. Paris : CNRS, 2016, 62 diapositives.

CARTIER, Aurore., MOYSAN, Magalie., REYMONET, Nathalie. Construire des outils pour la gestion des données de la recherche dans une communauté d'universités. **In** : LABORATOIRE INVISU (CNRS-INHA) en partenariat avec l'ADBS SECTEUR RECHERCHE. *La gestion des données de la recherche : retours d'expérience, Paris, 16 janvier 2015* **[en ligne]**. 10 p. Disponible sur : https://hal.archives-ouvertes.fr/hal-01138663 (consulté le 27 octobre 2017)

PUREN, Marie., RIONDET, Charles. *Research data management, a chance for Open Science. Methods and tutorials to create a Data Management Plan (DMP)*. **[en ligne]** Prague : DARIAH's Humanities at Scale Winter School, 2016, 81 diapositives. Disponible sur : https://hal.inria.fr/hal-01416978 (consulté le 24 octobre 2017)

VAN LOON, James E., AKERS, Katherin G., HUDSON, Cole. *et al.* « Quality evaluation of data management plans at a research university ». *IFLA Journal*, 2017/1 (vol. 43), p. 98-104

Records Management

« Normalisation et gestion des documents d'activité (*records management*) : enjeux et nouvelles pratiques pour notre profession » **[en ligne]**. *La Gazette des archives*, 2012/4 (n° 228), 299 p. Disponible sur : https://www.persee.fr/issue/gazar_0016-5522_2012_num_228_4 (consulté le 12 décembre 2017)

« Records Management : Gérer les documents et l'information ». *Documentaliste-Sciences de l'Information*, 2009/2 (vol 46). Disponible sur : https://www-cairn-info.docelec.insa-lyon.fr/revue-documentaliste-sciences-de-l-information-2009-2.htm (consulté le 03 janvier 2018)

COTTIN Michel. « La mise en œuvre de la nouvelle norme ISO 15489-2016 », *I2D – Information, données & documents*, 2016/4 (Volume 53), p. 60-60. Disponible sur : https://www-cairn-info.docelec.insa-lyon.fr/revue-i2d-information-donnees-et-documents-2016-4-page-60.htm (consulté le 01 mai 2018)

COUTURE, Cynthia., ROY, Julie. « La norme ISO 15489 : principes et application » **[en ligne]**. *Archives*, 2006-2007/2 (vol. 38) p. 143-177. Disponible sur : https://www.archivistes.qc.ca/revuearchives/vol38_2/38_2_Couture_Roy.htm (consulté le 10 janvier 2018)

(cc) BY-NC-ND

DHÉRENT, Catherine. « Le pilotage et l'évaluation du système de *records management* de la Bibliothèque nationale de France » **[en ligne]**. *La Gazette des archives*, n°216, 2009-4. Chiffrer et évaluer les archives. p. 131-144. Disponible sur : http://www.persee.fr/doc/gazar_0016-5522_2009_num_216_4_4647 (consulté le 27 avril 2018)

DROUHET, Geneviève., KESLASSY, Geogres., et MORINEAU, Elisabeth. *Records Management : mode d'emploi*. Paris, ADBS Éditions, 2000, 125 p.

FUZEAU, Pierre (dir.). *Archivage, records management et conformité*. Paris : Serda édition-IDP, 2014, 56 p. (Archimag. Guide pratique, n° 49)

GROUPE DE TRAVAIL DGME/SDAE – APROGED. *La maîtrise du cycle de vie du document numérique : présentation des concepts* (version 3 du 22 mai 2006) **[en ligne]**. Paris : APROGED, 2006, 32 p. Disponible sur : https://documentslide.org/the-philosophy-of-money.html?utm_source=la-maitrise-du-cycle-de-vie-du-document-numerique (consulté le 10 janvier 2018)

GROUPE MÉTIERS AAF – ADBS « Records management », « Comprendre et pratiquer le records management, Analyse de la norme ISO 15489 au regard des pratiques archivistiques françaises ». *Documentaliste-Sciences de l'Information*, 2/2005 (Vol.42). Disponible sur : https://www-cairn-info.docelec.insa-lyon.fr/revue-documentaliste-sciences-de-l-information-2005-2-page-106.htm (consulté le 10 janvier 2018)

KERN, Gilliane., HOLGADO, Sandra., et COTTIN, Michel. « Cinquantes nuances du cycle de vie : quelles évolutions possibles ? » *Les cahiers du numériques* **[en ligne]**. 2015/2 (Vol. 11) pp. 37-76. Disponible sur : https://www-cairn-info.docelec.insa-lyon.fr/revue-les-cahiers-du-numerique-2015-2-page-37.htm (consulté le 24 mars 2017)

NSW STATE ARCHIVES. *DIRKS Methodology and Manual.* **[en ligne]**. Disponible sur : https://www.records.nsw.gov.au/recordkeeping/advice/dirks/methodology (consulté le 31 juillet 2018)

OTT, Florence (dir.). « Apport de la gestion documentaire à la gouvernance de l'information ». *Les Cahier du numérique*, 2015/2 (vol. 11) 205 p.

PÉROTIN, Yves. « Le *records management* et l'administration anglaise des archives », *La Gazette des archives* **[en ligne]**, 1964 (n°44), p. 5-17. Disponible sur : http://www.persee.fr/doc/gazar_0016-5522_1964_num_44_1_1753?q=records+management+et+l%27administration+anglaise (consulté le 24 mars 2018)

(cc) BY-NC-ND

Mémoire

GAILLARD, Rémi. *De l'Open data à l'Open research data : quelle(s) politique(s) pour les données de recherche ?*. **[en ligne]** Mémoire d'étude, diplôme de conservateur de bibliothèque. Villeurbanne : Enssib, 2014. Disponible sur : http://www.enssib.fr/bibliotheque-numerique/documents/64131-de-l-open-data-a-l-open-research-data-quelles-politiques-pour-les-donnees-de-recherche.pdf consulté le 26 avril 2018.

Sitographie

*Open Data 5** **[en ligne]**. Disponible sur : https://5stardata.info/fr/ (consulté le 14 août 2018)

AUSTRALIAN NATIONAL DATA SERVICE. *What is research data* **[en ligne].** Disponible sur : http://www.ands.org.au/guides/what-is-research-data (consulté le 14 août 2017)

DATA CURATION PROFILES. *Home* **[en ligne]**. Disponible sur : http://datacurationprofiles.org/ (consulté le 14 août 2018)

DORANum. *Apprentissage numérique sur les données de la recherche.* **[en ligne].** Disponible sur : http://doranum.fr/ (consulté le 03 janvier 2017)

ÉCOLE POLYTECHNIQUE FÉDÉRALE DE LAUSANNE. *Research Data at EPFL.* **[en ligne].** (modifié le 18 décembre 2017). Disponible sur : https://researchdata.epfl.ch/ (consulté le 03 janvier 2018)

FOSTER. *Open Science Definition* **[en ligne]**. Disponible sur : https://www.fosteropenscience.eu/foster-taxonomy/open-science-definition (consulté le 14 août 2018)

GO FAIR. *GO FAIR: a bottom-up international approach.* **[en ligne]** Disponible sur : https://www.go-fair.org/ (consulté le 14 mars 2018)

INIST, *DMP OPIDoR* **[en ligne]**. Disponible sur : https://dmp.opidor.fr/ (consulté le 03 janvier 2018)

INRA. Principes FAIR. In : Datapartage. Disponible sur : https://www6.inra.fr/datapartage/Technologies/Principes-FAIR (consulté le 27 juillet 2018)

MINISTÈRE DE L'ENSEIGNEMENT SUPÉRIEUR, DE LA RECHERCHE ET DE L'INNOVATION. *Le Plan national pour la science ouverte : les résultats de la recherche scientifique ouverts à tous, sans entrave, sans délai, sans paiement* **[en ligne]** (modifié le 12 juillet 2017)**.** Disponible sur : http://www.enseignementsup-recherche.gouv.fr/cid132529/le-plan-national-pour-la-science-ouverte-les-resultats-de-

(cc) BY-NC-ND

la-recherche-scientifique-ouverts-a-tous-sans-entrave-sans-delai-sans-paiement.html (consulté le 13 août 2018)

OPIDoR. *Cat OPIDoR, wiki des services dédiés aux données de la recherche*. **[en ligne].** (modifié le 21 décembre 2017). Disponible sur : https://cat.opidor.fr/index.php/Cat_OPIDoR,_wiki_des_services_d%C3%A9di%C3%A9s_aux_donn%C3%A9es_de_la_recherche (consulté le 03 janvier 2018)

UK DATA ARCHIVE. *Home* **[en ligne]**. Disponible sur : www.data-archive.ac.uk (consulté le 14 août 2018)

WITHMIRE, Amanda., CARLSON, Jake., WESTRA, Brian., *et al.* The DART Project: using data management plans as a research tool. **In :** *Open Science Framework* **[en ligne]** (modifié le 24 mars 2017). Disponible sur : https://osf.io/kh2y6/ (consulté le 14 août 2018)

(cc) BY-NC-ND

ANNEXES

Table des annexes

GRILLES DES ENTRETIENS.....102
RESUME DES ENTRETIENS 105

(cc) BY-NC-ND

GRILLES DES ENTRETIENS

Pour les entretiens avec des personnes ayant déjà travaillé sur un PGD :

Thèmes et sous-thèmes	**Réponses**
Nom, Prénom Lieu de travail	
Contexte des données	
Dans quel contexte/projet avez-vous rédigé un plan de gestion des données (PGD) ? Était-ce obligatoire dans le cadre d'un financement ?	
Comment décrieriez-vous les données (*qualitative, quantitative, volume, etc*) que vous aviez à inclure dans votre PGD?	
Sur la rédaction de plan de gestion des données	
Quelle démarche avez-vous suivie pour la rédaction du PGD ?	
Vous êtes-vous servi d'un modèle en particulier ?	
Quelles étapes avez-vous suivies ? Y a-t-il certaines étapes que vous avez trouvées particulièrement difficiles ?	
Avez-vous travaillé en collaboration avec d'autres collègues/métiers ?	
Quelles sont, pour vous, les différentes sections d'un PGD ? Laquelle ou lesquelles sont les plus importantes ?	
Impression par rapport aux plans de gestion des données	
Que pensez-vous des nouvelles obligations (commission européenne, Horizon 2020) par rapport à la rédaction de PGD ?	

(cc) BY-NC-ND

Est-ce que cette nouvelle exigence changera vos pratiques ?
Pensez-vous que les PGD offrent des avantages ? Comportent des désavantages ?

Pour les personnes n'ayant jamais travaillé sur un PGD :

Thèmes et sous-thèmes
Nom, Prénom Lieu de travail
Contexte du travail sur les données
Qu'est-ce qui vous motive à vous attarder à la préservation des données ?
Avec quels types de données travaillez-vous (*qualitative, quantitative, volume, etc*) ?
Sur la préservation des données
Quel type de démarche suivez-vous afin de préserver les données ?
Quelles étapes avez-vous suivies ? Y a-t-il certaines étapes que vous avez trouvées particulièrement difficiles ?
Travaillez-vous en collaboration avec d'autres collègues/métiers ?
Quelles sont, pour vous, les informations importantes à conserver pour une bonne préservation des données ?
Impression par rapport aux plans de gestion des données
Que pensez-vous des nouvelles obligations (commission européenne, Horizon 2020) par rapport à la rédaction de PGD ?

(cc) BY-NC-ND

Est-ce que cette nouvelle exigence changera vos pratiques ?
Pensez-vous que les PGD offrent des avantages ? Comportent des désavantages ?

(cc) BY-NC-ND

RESUME DES ENTRETIENS

Les paragraphes en italique sont des citations directes.

ENTRETIEN ALINA DANCIEU ET ANNE CORNILLEAU – LE 15 JANVIER 2018

Thèmes et sous-thèmes	**Réponses**
Nom, Prénom Lieu de travail	Alina Dancieu et Anne Cornilleau, chargées d'études au CDSP
Contexte des données	
Dans quel contexte/projet avez-vous rédigé un plan de gestion des données (PGD) ? Était-ce obligatoire dans le cadre d'un financement ?	*Test d'un plan de gestion des données. Vous trouverez le plan de gestion des données, ainsi que notre test dans le chapitre 11 de cette publication. Nous l'avons testé pour l'une des enquêtes du panel ELIPSS : http://quanti.dime-shs.sciences-po.fr/fr/le-projet/* *On fait de la documentation et de la mise à disposition de données pour la recherche et du coup c'est très proche en fait de notre activité, de notre façon de travailler… Et c'est aussi une façon de voir quelles seraient les difficultés ou les facilités, au contraire, que l'on aurait pour remplir ce type de plan de gestion de données, si on est amené à l'avenir à accompagner des équipes qui auraient ce besoin-là.* *L'enjeu était de tester nos pratiques versus ces plans de gestion de données.*
Comment décrieriez-vous les données (*qualitative, quantitative, volume, etc*)	Voir réponse ci-dessus pour plus de précision.

(cc) BY-NC-ND

que vous aviez à inclure dans votre PGD?	
Sur la rédaction de plan de gestion des données	
Quelle démarche avez-vous suivi pour la rédaction du PGD ?	Démarche résumée dans l'article cité ci-dessus.
Vous êtes-vous servi d'un modèle en particulier ?	Oui, celui cité dans l'article, soit le PGD élaboré par l'Université Paris-Diderot.
Quelles étapes avez-vous suivies ? Y a-t-il certaines étapes que vous avez trouvées particulièrement difficiles ?	*Lecture du plan, discussion avec mes collègues, ensuite rédaction des réponses. Également, discussion téléphonique et par e-mail avec l'équipe qui a rédigé le plan de gestion des données.* ***Questions posées à l'équipe auteure du plan de gestion de données*** *Dans le cas d'un panel (comme le panel ELIPSS) devrait-on remplir le plan de gestion des données pour le panel ou pour chaque enquête ?Nous avons rempli ce plan pour l'Enquête annuelle, qui est conçu par l'Equipe ELIPSS et le CST. Un exemple concernant une enquête sélectionnée à la suite de l'appel à projets ELIPSS serait plus utile ?* ***3 - Ressources nécessaires à la mise en œuvre du DMP / 3 - Resources needed to implement the DMP*** *Matériel, Personnel : Devrait-faire une estimation pour un jeu de données ? Pour tout le projet ?* ***4.1 - Description des données – Jeu de données n°…/ 4.1 - Data description – Dataset #…*** ***Méthode de production des données**** *Pourquoi mentionner la norme de documentation utilisée ici, alors que la section 4.3 - Métadonnées : documentation et organisation des données lui est dédiée ?*

(cc) BY-NC-ND

	4.2 - Au cours du projet - Stockage, accès et sécurité des données /4.2 - During the project - Storage, access and security - De quelle manière l'ensemble des chercheurs partenaires du projet auront-ils accès aux données pendant la recherche ? *Qu'est-ce qu'on entend par recherche ? La période de terrain ? Le terrain et la période d'apurement et d'enrichissement des données (ce qu'on appelle la post-production pour ELIPSS) ?* ***Echanges et partage**** *: pourquoi poser cette question dans cette section ? Redondance avec la section 4.4 - A l'issue du projet – Dissémination / 4.4 - At the end of the project – Dissemination ? Dans le cadre du projet ELIPSS, pour les projets autre que l'Enquête annuelle, nous diffusons les données aux équipes de recherche. Devrait-on parler de la diffusion à l'équipe de recherche dans la section 4.2 et de la diffusion à la communauté scientifique dans la section 4.4 ?* *Autres remarques : le fait que les exemples sont parfois en anglais, peut, dans le cas d'une lecture rapide, pourrait prêter à confusion au sujet de la langue qui doit être utilisée pour remplir le plan de gestion des données. Une traduction serait utile.*
Avez-vous travaillé en collaboration avec d'autres collègues/métiers ?	Oui, tous les gens qui sont impliqués dans le projet.
Quelles sont, pour vous, les différentes sections d'un PGD ? Laquelle ou lesquelles sont les plus importantes ?	*Il me semble que les sections obligatoires ont été établies dans le cadre du programme européen pour la recherche et l'innovation H2020. Le modèle que nous avons testé suit ces recommandations.* *Toutes les sections ont leur importance.*
Impression par rapport aux plans de gestion des données	
Que pensez-vous des nouvelles obligations	Elles ne sont pas touchées pour l'instant.

(cc) BY-NC-ND

(commission européenne, Horizon 2020) par rapport à la rédaction de PGD ?	
Est-ce que cette nouvelle exigence changera vos pratiques ?	*S'il s'agit des obligations concernant la rédaction d'un plan de gestion de données pour tout projet ayant obtenu un financement européen, je ne suis pas forcément concernée par cette mesure.* *C'était plutôt facile en fait… Car l'on avait la majorité des réponses et ça nous a permis de nous poser d'autres questions… Comme elle disait, on n'est pas forcément un cas classique parce qu'on a l'habitude de la gestion des données.* *Alors on n'est pas allé aussi loin dans la réflexion, puisqu'on a beaucoup d'enjeux justement sur ce qu'on va faire dans les prochains… prochaines années au CDSP, ça fait partie des questions qu'on s'est posées, mais pour l'instant on n'a pas de réponses très précises là-dessus. Mais on sait que c'est quelque chose qu'on pourrait faire.*
Pensez-vous que les PGD offrent des avantages ? Comportent des désavantages ?	*Comme avantages, le fait qu'ils permettent de se poser les bonnes questions au bon moment et peuvent s'avérer très utiles au moment de passer une certification, par exemple dans notre cas (centre de données qui documente et diffuse des données qualitatives et quantitatives).* *Je ne vois aucun désavantage pour le moment :)*

(cc) BY-NC-ND

ENTRETIEN AVEC CHRISTINE HADROSSEK – 17 JANVIER 2018

Thèmes et sous-thèmes	Réponses
Nom, Prénom Lieu de travail	**Christine Hadrossek** Chargée de l'Information Scientifique et Technique au RIATE
Contexte du travail sur les données	
Dans quel contexte/projet avez-vous rédigé un plan de gestion des données (PGD) ? Était-ce obligatoire dans le cadre d'un financement ?	*Oui oui absolument. Bon… c'est en fait… Une démarche qui était dans un premier temps expérimentale. Il n'y avait pas particulièrement obligation de faire… il ne s'agit pas d'un projet H2020, ou… d'un projet européen qui nécessite absolument la production d'un DMP. On se retrouvait, dans mon équipe, dans une démarche de vouloir expérimenter… le plan de gestion de données, pour pouvoir le moment venu, être capable de le réaliser dans les meilleures conditions. Donc… en fait il s'agit d'un projet qui a été réalisé… une étude, en fait, qui a été réalisée pour le compte du CGET, qui est le Commissariat général à l'égalité des territoires, c'est les services de gouvernement. Et donc il s'agissait d'une étude, une typologie socio-économique des régions frontalières de l'Union européenne. Voilà. Et donc c'est dans ce cadre-là qu'on a… dans le cadre de cette étude, il y a un certain nombre de données qui ont été produites. À partir desquels évidemment l'étude a été réalisée… et c'est dans ce cadre-là qu'a été rédigé ce plan de gestion de données.*
Comment décrieriez-vous les données (*qualitative, quantitative, volume, etc*) que vous aviez à inclure dans votre PGD?	Données quantitatives : socio-économiques / statistiques Gros volume de données, mais pas tout n'est diffusé. Une partie des données étaient extraites d'Open Street Map. *Et en fait elles sont… bon… on s'est posé la question de savoir effectivement s'il était intéressant ou pas de les extraire et de les mettre à disposition… Chacun peut aussi les extraire au moment voulu, quoi enfin. Voilà, donc on a plus décidé de diffuser des données moins importantes, mais avec par contre des documents techniques qui permettent à la personne qui va vouloir reproduire l'expérimentation, de pouvoir le faire avec une démarche méthodologique. C'est-à-dire qu'on a expliqué comment on les avait extraites, comment elles ont… le contexte dans lequel elles ont été extraites… Il y avait beaucoup de documents techniques qui expliquaient la démarche en fait.*

(cc) BY-NC-ND

Sur la rédaction de plan de gestion des données	
Quelle démarche avez-vous suivie pour la rédaction du PGD ?	C. Hadrossek a choisi l'option de travailler en binôme avec ingénieur 1. Faire connaître le DMP à l'équipe qui travaille, réunir l'équipe et présenter le plan, à quoi il sert et pourquoi il est important 2. Mettre en pratique, travail en binôme avec les producteurs de données.
Vous êtes-vous servi d'un modèle en particulier ?	C'est un plan de données qui a été préparé par l'Université Paris Descartes en collaboration avec Paris Diderot.
Quelles étapes avez-vous suivies ? Y a-t-il certaines étapes que vous avez trouvées particulièrement difficiles ?	*Au niveau de la rédaction non… De manière générale, oui ça s'est bien passé. Ça a été un travail qui a été vraiment productif. Parce que moi j'arrivais, en fait, dans une équipe que je connaissais mal, et ça m'a permis déjà d'une part de connaître la nature de leurs travaux et puis d'autre part vraiment de partager, de mettre en commun des pratiques différentes et… et on se rend compte de... du fait qu'on a des cultures différentes… et c'était vraiment la convergence des cultures différentes avec des ingénieurs qui travaillent d'une certaine façon, moi qui travaille d'une façon différente… donc on a appris, en fait, les uns des autres, ça c'était un point positif. Ensuite les difficultés c'étaient quand même aussi de parler le même langage. Donc parfois, quand on met derrière métadonnées, qu'est-ce qu'on met comme représentation derrière ? Est-ce qu'on parle de la même chose ? Pour moi quand je parlais de métadonnées en tant que documentaliste, je parle plutôt de métadonnées Dublin Core… voilà, enfin, c'est celles que je connais un peu. Or, eux ont leurs propres métadonnées que je ne connais pas nécessairement. Voilà donc il faut… c'est difficile aussi pour nous, documentaliste, de connaître toutes les métadonnées, quoi. Surtout quand elles sont propres à un domaine d'application vraiment très spécifique. Donc voilà… mais globalement non je n'ai pas rencontré de grosses difficultés. Ce qui pouvait parfois être difficile, c'était de se dire dans la rédaction du plan de gestion de données, à quel niveau de description on se place ? Est-ce que...jusqu'où on va dans la description de nos données ? Et avec quel niveau de détails ? Ce qui fait poser la question, globalement aussi de savoir « à qui on s'adresse ? » finalement. Qui est la personne qui va lire ce plan de gestion de données, qui va avoir envie de réutiliser ces données. Est-ce que c'est une personne lambda, qui a aucune connaissance du domaine ou est-ce que c'est quelqu'un qui a priori, a une bonne connaissance et va rechercher une information précise… Vous voyez ce que je veux dire ? Alors on a fait des choix.* *C'était plutôt à priori on se disait que c'était à peu près le même profil que ceux qui ont produit les données quand même… à priori. En se disant que si c'est…on peut imaginer que ces données peuvent peut-être resservir aussi... je ne sais pas… dans la*

(cc) BY-NC-ND

	manière dans laquelle elles ont été créées à un historien des sciences, j'en sais rien je dis peut-être n'importe quoi. Je me dis que bon, voilà, si jamais il veut pouvoir recréer l'expérience et voir ce que… analyser comment a été fait le travail, peut-être qu'il pourrait s'accompagner d'un spécialiste à côté. Voilà on a pris cette… pour adapter un peu adapter le niveau de langage, parce qu'on n'allait pas non plus peut-être… voilà… expliquer absolument tout tout tout. Voilà, on se pose cette question quant à savoir à quel niveau de langage, de degrés description aussi on va. Mais on donnait aussi un maximum d'indications. C'est-à-dire pour créer ces données-là, il nous faut tel outil, tel outil, tel outil… Voilà en fait la méthodologie associée à chacun des outils qui nous permettent d'obtenir ces résultats-là.
Avez-vous travaillé en collaboration avec d'autres collègues/métiers ?	Avec les ingénieurs de recherche, en binôme
Quelles sont, pour vous, les différentes sections d'un PGD ? Laquelle ou lesquelles sont les plus importantes ?	*Je ne sais pas s'il y avait des sections plus ou moins importantes… Non je… Je pense qu'elles ont vraiment toutes leur utilité… Si, alors peut-être... il y en a une sur laquelle on a peut-être le moins travaillé..., ou enfin du moins sur lequel on a peut-être… peut-être moins accordé de temps on va dire… Non pas qu'elle soit moins importante… c'est l'archivage. L'archivage final des données. Parce qu'on avait en fait beaucoup de données nativement numériques, on n'a pas de données papiers, ce qui fait qu'il n'y avait pas forcément de nécessité de les archiver dans un centre… voyez… et puis peut-être aussi parce qu'on était dans une phase d'expérimentation et que, voilà, peut-être une prochaine fois on lui accordera plus d'importance, mais en l'occurrence si vous voulez c'est… et puis on avait du mal à évaluer quand même, voilà… finalement… est-ce que ces données on les archive à long terme, à moyen terme, à long terme… est-ce que ça a du sens de le faire… bon… et puis en plus on n'avait pas non plus de lieu de stockage, d'archivage en fait… autre que les nôtres, que notre serveur, voilà… Donc...L'université Paris-Diderot ne fournit pas nécessairement un espace d'archivage à long terme, ou à moyen terme, pour les outils numériques, vous voyez ?*
Impression par rapport aux plans de gestion des données	
Que pensez-vous des nouvelles obligations (commission européenne,	Impression assez positive, très important de le faire maintenant. *Moi je pense que ça l'est ! Ça l'est, mais… parce que je pense qu'il faut vraiment aller dans ce sens-là quoi. Enfin on est convaincu par l'intérêt, si vous voulez, de mettre à disposition des données qui puissent être reproductibles, qui puissent être*

(cc) BY-NC-ND

Horizon 2020) par rapport à la rédaction de PGD ?	*réutilisées par d'autres. Donc ça c'est important je pense. Si chaque producteur de données se situent dans une démarche comme celle-ci, ce sera une bonne chose, quoi. Parce que sinon de toute façon sinon c'est d'autres qui le feront, et qui le feront en le monnayant. Si nous on le fait pas, ce seront les grands éditeurs qui vont s'approprier les données et qui vont demande, comme c'est le cas déjà pour les articles scientifiques… Je veux dire voilà maintenant pour les gens qui publient dans des revues au facteur d'impact extrême n'ont pas forcément toujours… je veux dire la loi pour le numérique ça change un petit peu… mais n'ont pas forcément la possibilité tout de suite de les mettre à libre disposition des autres… Ça veut dire qu'il faut payer pour voir, et tout le monde ne peut pas payer pour voir. Et si… ça va être pareil pour les données. Je veux dire si nous on pend pas... enfin, nous les producteurs, les chercheurs, etc. on ne décide pas d'entrer dans une démarche open access ça va être les éditeurs qui vont prendre la main sur les données et qui vont devoir… enfin qui vont faire payer pour voir, vous voyez ce que je veux dire ?* Vigilance sur données personnelles, sur les divers enjeux économiques, les brevets... Le temps de mettre ses données à disposition ?
Est-ce que cette nouvelle exigence changera vos pratiques ?	*En tout cas nous en plus au RIATE, si vous voulez, on se situe vraiment dans une démarche très Open Science, donc on a la volonté, très claire, de mettre à disposition ces données à chaque fois que c'est possible.*
Pensez-vous que les PGD offrent des avantages ? Comportent des désavantages ?	*Je serais quand même un peu embêtée pour vous répondre à cette question parce que je pense que je n'ai pas suffisamment d'expérience là-dessus encore… c'est encore trop tôt. Pour le moment, moi je vois vraiment beaucoup d'avantages et quasiment pas d'inconvénients. Je… ça nous… ça ne mange pas de pain, voyez. C'est sûr que c'est du boulot, mais c'est du bon boulot. Je veux dire ça facilite, ça facilite beaucoup les choses de les prévoir en amont. Donc... non je serais vraiment partielle là-dessus en disant que c'est vraiment une bonne chose.*

(cc) BY-NC-ND

ENTRETIEN AVEC CYNTHIA PEDROJA – 27 FEVRIER 2018

Thèmes et sous-thèmes	Réponses
Nom, Prénom Lieu de travail	Cynthia Pedroja Chef de projet Gestion des données de la recherche - Direction des Ressources et de l'information Scientifique – Sciences Po Paris
Contexte du travail sur les données	
Qu'est-ce qui vous motive à vous attarder à la préservation des données ?	Projet qui développe une infrastructure de services et d'outils pour la communauté de recherche, aide à la gestion des données *Spire est représentatif de la préoccupation de Sciences Po de mettre à disposition les travaux des chercheurs au plus grand nombre, il y a déjà une culture de l'open access bien ancrée dans l'institution et particulièrement à la bibliothèque. De plus, il y a une complémentarité entre les publics et les données*. Question : Qu'est-ce qui pousse la DRIS à s'engager dans un tel projet ? *Il y a plusieurs mouvements, si on se place au niveau de la bibliothèque il y a une espèce de métamorphose des bibliothèques avec une volonté d'offrir plus de services au niveau de la recherche et notamment au niveau des données de la recherche. Il y a une « envie » de redéfinir ce que sont les bibliothèques… Dans le contexte international, on pense aux bibliothèques anglo-saxonnes qui ont présence très forte sur des thématiques comme celle-ci. Par contre, leur positionnement par rapport à leurs universités est très différent de celui des bibliothèques universitaires françaises donc on ne peut pas faire exactement la même chose.* *On est dans le contexte à Sciences Po avec une conscience assez aiguisée des problèmes de gestion de données rencontrés par les chercheurs*

(cc) BY-NC-ND

	CDSP et Medialab = *fer de lance, car un est une unité de service qui travaille beaucoup avec les données et l'autre en est une composée de beaucoup d'ingénieurs, développeurs, etc. qui sont confrontés à une problématique de gestion des données et ils sont vraiment dans l'optique de diffuser des données* Contexte national européen, conférence, programme, H2020, formation et sensibilisation, on sent qu'il y a quelque chose qui est en train de se passer de ce côté-là *Problématiques de données de la recherche, la France s'en est saisi un peu de la même façon que les pays anglo-saxons, c'est-à-dire que tout cela est essentiellement de l'information à gérer… et c'est déjà dans le mandat des bibliothèques* *Sachant que dans notre contexte, il y a aussi des archives. Sciences Po a deux centres de recherche qui ont des archives, le CEVIPOF et le centre d'histoire, en plus des archives de la structure Sciences Po elle-même qui sont gérées par la bibliothèque. Dans d'autres lieux, ce sont les archives qui se sont emparées de ces problématiques (des données de la recherche), mais ici comme la bibliothèque a un peu le double rôle, elle s'est saisie de la question.*
Avec quels types de données travaillez-vous (*qualitative, quantitative, volume, etc*) ?	FDSP travaille essentiellement sur des données qualitatives, et sinon des données quantitatives à travers des enquêtes *Université de sciences sociales (Économie, politique, histoire, anthropologie, etc…) donc l'enquête est un mode de production de données privilégié, que ce soit avec des données primaires ou des données qu'ils réexploitent*. *Parfois issues de base de données économique, des données d'archives (constitution de base de données avec par exemple des registres de commerces historiques), documentation politique au travers des professions de foi, image (réexploitée ou de terrain), medialab qui récolte des données de réseaux socios ou de sites web.* Données textuelles, des données sons, des données images, des données de logs *En ce qui concerne le volume, ce n'est pas « big » si on compare avec d'autres disciplines, mais si on les compare entre elles, si. Par exemple les économistes ont besoin de force de calcul, comparativement à certains chercheurs en sociologie qui vont travailler sur des données de moindres envergures.*
Sur la préservation des données	
Quel type de démarche	*On n'en est pas là*

(cc) BY-NC-ND

suivez-vous afin de préserver les données ?	D'abord phase de diagnostic : essayer de comprendre ce dont aurait besoin le chercheur (il y a eu une phase d'étude pendant deux ans, pour étudier les besoins, développer des scénarios et une offre de service et ils sont maintenant dans une phase de test de ces services qui consistait à travailler avec projets pilotes pour tester ce qu'ils étaient capables de faire, les outils dont ils avaient besoin, etc. Ne peut donc pas généraliser la réponse à cette question) On ne parle pas encore de préservation ni de diffusion… C'est-à-dire qu'elle en parle sans en parler… Elle ne pose pas cette problématique face à ce chercheur… Car il n'est pas documentaliste, la notion de préservation est assez vague et la perspective de diffusion n'est pas un objectif en soi. Considère que c'est un biais d'envisager les DDR comme ça, car ça nous coupe de la communauté des chercheurs. *Donc il faut comprendre les projets de recherche des chercheurs. Il faut comprendre le contexte pour comprendre les données qui sont produites. Phase de discussion qui va partir de la problématique de la recherche, parler du moment ou de la phase de constitution des données et demander où se trouvent ces objets, quelles formes ils ont, quelle est leur quantité, qu'est-ce qu'ils en font et quelles sont leurs étapes.* Phase de diagnostic qui permet de discuter avec le chercheur et permet d'identifier des problématiques (ex. ils stockent uniquement sur leur ordinateur) et éventuellement de connaître ce qu'ils veulent en faire… les laisser aborder eux-mêmes la diffusion des données. On est à chaque fois sur du cas spécifique. On lui fait comprendre que si on fait des projets ensemble c'est pour que lui s'y retrouve. *Sciences Po a pour objectif de faciliter également le stockage des données, un travail d'identification de solution est en cours.* *Parle essentiellement de stockage et de s'assurer de la pérennité de leur donnée donc s'assurer qu'ils sont dans un bon format, bien nommer, qu'elles sont bien documentées et que les différents traitements qu'ils apportent à leurs données sont aussi documentés.* Réutilisation pour eux de leurs propres objectifs.
Quelles étapes avez-	Le temps et une partie des compétences.

(cc) BY-NC-ND

vous suivies ? Y a-t-il certaines étapes que vous avez trouvées particulièrement difficiles ?	Temps parce que : 1 – C'est du travail long surtout, car on n'est pas encore outillé 2 – On est à la fois contraint par nos propres agendas, mais aussi celui des chercheurs Cela pourrait être moins important une fois qu'il y aura Compétences, car elle est pour l'instant la seule à travailler dans ce programme et elle ne maîtrise pas elle-même toutes les compétences (par exemple dans un langage informatique qui permettrait d'automatiser certaines tâches). Il faut comprendre les objets informatiques.
Travaillez-vous en collaboration avec d'autres collègues/métiers ?	Essaie aussi de travailler en collaboration avec les ingénieurs dans les laboratoires (pas le même profil, pas le même type de poste) *Une équipe de travail s'est constituée depuis février 2018 à la bibliothèque et nous commençons à travailler régulièrement avec des ingénieurs des centres.*
Quelles sont, pour vous, les informations importantes à conserver pour une bonne préservation des données ?	Bons formats de fichier, bons encodages, bon nommage et une bonne documentation. Stockage sécurisé quand c'est nécessaire (au niveau de la sensibilité des données et de la perte des fichiers).
Impression par rapport aux plans de gestion des données	
Que pensez-vous des nouvelles obligations (commission européenne, Horizon 2020) par rapport à la	Sentiment double : *c'est à la fois très bien, car c'est un moteur, permet une prise de conscience au niveau institutionnel et surtout des moyens qui seront alloués à ses problématiques.* *De l'autre côté, se concentrer sur le plan de gestion de données seulement fait passer à côté d'autres choses. Crois qu'il*

(cc) BY-NC-ND

rédaction de PGD ?	*y a une distorsion terrible entre les acteurs institutionnels et les chercheurs, on oublie que pour eux c'est un livrable parmi d'autres.* Trouve que c'est un outil intéressant, mais qu'il faut oublier l'aspect livrable demandé par un financier, mais qu'il faut le ramener à ce qu'il est : un outil de travail collectif. Se pose la question pourquoi ce serait seulement des projets européens qui le ferait ? Sinon elle a commencé à travailler sur les projets avec le data curation profil de l'université de Purdue qui selon pose des questions semblables… Se dit qu'elle pourrait faire des plans de gestion des données à partir de ces profils. Elle pense qu'il ne faut pas confronter le chercheur à ces documents-là à moins qu'il ne le demande ou qu'il y soit obligé pour ne pas les obliger.
Est-ce que cette nouvelle exigence changera vos pratiques ?	Sciences Po débutera son premier projet Horizon 2020 en juin depuis l'exigence des PGD et donc elle y travaillera. Par contre, Sciences Po a déjà fait partie d'autres projets où il n'était pas *lead partner* mais parties prenantes. C. Pedroja a donc relu des plans de gestions de données afin de voir s'il n'y avait pas eu des engagements qui avaient été pris de la part de leur chercheur auquel ils auront un rôle à jouer où voir s'il y avait des questions de propriétés intellectuelles. Oui cela change les pratiques donc. *Pour information, Sciences Po a décidé de proposer une structure de DMP adapté au contexte institutionnel. C'est un outil que nous allons proposer à l'automne.*
Pensez-vous que les PGD offrent des avantages ? Comportent des désavantages ?	

Données variées amènent une complexité au programme.

« Quand je me compare avec mes collègues dans d'autres universités, j'ai l'impression que ben… on est tous un peu dans une phase expérimentale »

(cc) BY-NC-ND

« Et pourtant quand on parle de données de la recherche, on a beau être confronté à deux enquêtes qualitatives, les besoins peuvent être assez variables. Selon les besoins du chercheur et la façon dont lui-même gère ses données, les compétences qu'il apporte lui et son équipe certains sont très ‘’documentalistes’’ tandis que d'autres maîtrise les contenus de ses données, mais pas ses formes »

(cc) BY-NC-ND

ENTRETIEN AVEC MADELEINE HUBERT – 20 MARS 2018

Malheureusement, un problème technique est survenu pendant l'enregistrement de notre entretien, ce qui fait que je n'ai pu le retranscrire par la suite. Ainsi, j'ai produit un résumé de notre discussion qui a été revu par Madeleine Hubert, que vous retrouverez ci-dessous. Celle-ci s'étant aussi présenté par mail, j'ai aussi inclus cet échange.

Extrait du mail de Madeline Hubert du 18 mars 2018

Je travaille en tant que data librarian, chargée de la gestion, du signalement, de l'archivage et de la diffusion des données de la recherche. J'ai intégré la section "Bibliothèque numérique" pour laquelle Monsieur Jouneau, Thomas est responsable du dossier "Données de la recherche". J'ai été donc engagée sous sa supervision et j'ai également intégré le réseau RAR : réseau d'appui et de recherche qui se compose de différents bibliothécaires référents de laboratoires sur le territoire lorrain. Cette fonction de bibliothécaire référent est de devenir un interlocuteur privilégié pour toutes les questions documentaires des chercheurs du laboratoire en question.

C'est dans ce contexte que j'effectue mes missions principales depuis janvier 2017 à savoir :

- développer une expertise sur les outils, solutions et services sur la thématique de la gestion et la valorisation des DR (Données de la Recherche). = se créer un réseau de partenaire, se familiariser avec des outils, se former via des plateformes, être curieuse sur les disciplines scientifiques, ...

- participer au recueil des besoins en lien avec le réseau d'appui (le RAR) : nous avons fait une enquête pour amorcer la réflexion : une brève enquête auprès des laboratoires pour connaitre l'état de leur connaissance et pratique sur les DR.

- mettre son expertise à disposition des chercheurs : suite à l'enquête, nous avons pu repérer des laboratoires qui portaient un vif intérêt pour les DR pour initier un dialogue plus construit. Nous avons donc quelques laboratoires pour tester la réception de nos idées. Nous essayons également de sensibiliser un public à différents niveaux : dans la communauté scientifique : les doctorants (dans la formation Doctorant Domptez La Doc!) et les chercheurs / directeurs de labo avec des sessions de sensibilisation sur la question. Et également dans la communauté des pro de l'IST : demi-journée, dialogue avec les collègues pour les faire monter en compétence sur le sujet.

- participer au projet d'archivage et de diffusion des publications sous l'angle du lien publications et données : bien entendu, il est important de considérer également les publications, productions scientifiques au cœur de leur métier, de connaitre leur modèle d'édition et de pouvoir proposer d'ajouter les DR comme matériel pouvant valoriser les publications.

Concernant mes attentes sur la préservation des DR :

- je fais un focus sur la gestion avant tout (la valorisation qui inclut l'ouverture pour moi est secondaire dans mon discours). A partir d"une bonne gestion, les efforts à fournir pour les publier / les ouvrir sera moindre. Après quand il y a des exigences politiques de la part des éditeurs ou financeurs, ben on est là pour les aider aussi !

(cc) BY-NC-ND

- c'est une nouvelle entité documentaire aux frontières un peu mouvantes: il faut délimiter au cas par cas à partir de quand la donnée est "partageable". La donnée évolue au fil d'un projet et à partir de quand faut-il envisager sa pérennisation et/ou sa publication ? Quel est le coût humain et technique pour y parvenir ?

- Créer un réseau de partenaires : une très bonne connaissance du paysage universitaire est indispensable : connaitre ses différents acteurs au sein de l'université. J'ai la chance d'être à proximité géographique avec l'INIST-CNRS donc je peux facilement m'y rendre pour collaborer (sur le PGD).

Je trouve ça passionnant car la construction est bilatérale : chaque laboratoire a un niveau de connaissance et un intérêt différent sur les DR avec des besoins différents, ce qui m'oblige à chercher et contacter d'autres acteurs de l'Unif. Et de mon côté j'essaye de leur apporter une expertise documentaire : je ne peux pas leur donner un appui opérationnel (donc coder ou développer des solutions) mais je peux les diriger vers ces compétences/services/outils adéquats.

Maintenant, après un an de travail : c'est très expérimental ! Tout ce projet est pilote : nous essayons de développer une roue de service au plus proche des besoins des chercheurs qui est façonnée par nos propres limites techniques et humaines.

Les attentes de l'université est de pouvoir développer une expertise sur les DR pour remplir ce rôle d'appui aux institutions de recherche / aux chercheurs. Il a fallu d'abord une année d'exploration pour vraiment commencer à façonner une gamme de service. Nous sommes actuellement en discussion avec les vice-présidents de l'université pour qu'ils nous soutiennent et orientent notre démarche. (pour ce volet, je peux vous renvoyer vers Thomas Jouneau, mon superviseur). Il y a déjà une très belle initiative de Pierre-Yves ARnould qui a développé un répertoire institutionnel pour le pôle de labo OTELO (thématique : sciences de la terre). Mais donner un accès à des jeux de donnée structuré n'est pas encore une pratique commune.

Actuellement, je travaille :

- à encourager les sollicitations des chercheurs (être visible et se faire connaitre !, création d'outil comme un PGD, formation de sensibilisation, rencontre sur le terrain...)

- répondre aux sollicitations qui sont pas très nombreuses pour le moment.

Désolée pour cette longue introduction et ce grand tableau dont certains coups de pinceaux sont expérimentaux et le résultat de l'ordre de l'abstrait !

Extrait de notre échange de mail du 22 mars 2018

***** Les paragraphes en italique ont été rédigées par Madeleine Hubert, les autres sont les miens.

- Discussion sur le records management

- Commencer par une petite enquête (Numilab), un peu dans le style de ce qu'on fait Lille 3 et Rennes 2. Rôle d'accompagnateur auprès des chercheurs. La démarche consiste à discuter avec les chercheurs, détecter quels sont leurs besoins, les aider à trouver des répertoires où entreposer leurs données, éventuellement les aider à rédiger des DMPs … Il faut faire de la promotion auprès des laboratoires, car les services offerts ne sont pas encore bien connus, et pour l'instant les sollicitations restent rares.

(cc) BY-NC-ND

J'ajouterai que la thématique n'est pas connue. Notre offre de service n'est pas encore élaborée dans sa version définitive, nous ne pouvons que proposer des accompagnements. Par contre, nous pouvons sans problème sensibiliser les chercheurs avec cette thématique par des formations. Nous espérons qu'avec ces sensibilisations, ils viendront vers nous pour qu'on puisse les accompagner dans leurs besoins documentaires.

- Les DMP sont un moyen d'aller chercher les chercheurs, d'amorcer un dialogue avec ces derniers.

Oui, c'est un bon document qui demande au chercheur de formaliser leur activité de recherche d'une manière générale / générique et il nous permet de prendre connaissance de l'état de la situation plus facilement (ou même un nouveau doctorant arrivant au laboratoire !). On peut les aider à créér la structure (donc la table des matières avec les points à aborder) ou simplement le relire (et utiliser des checklists pour savoir si le DMP abordent tous les aspects. Exemples : https://qdr.syr.edu/guidance/managing/dmp-checklist ou http://www.dcc.ac.uk/news/new-checklist-data-management-plan)

- Pour l'instant, les défis sont des défis d'adaptation, les chercheurs travaillent avec des données qui diffèrent les uns des autres, les programmes utilisés pour traiter ces données diffèrent aussi, leurs méthodes de travail… Il faut bien connaître leur environnement et leur problématique afin d'être capable de les aider.

Exact ! la nature et le type de données sont différentes : un sociologue et un chimiste produisent tous les deux des données mais cette donnée, même si elle respecte un cycle de vie plus ou moins commun, va évoluer différemment. Connaitre le champ disciplinaire du chercheur est donc indispensable. Par exemple : un linguiste utilisera des corpus de texte qu'il faudra peut-être océriser.

- Pour la préservation des données, ce qui peut être très bien c'est d'avoir un fichier de type ReadMe qui accompagne le jeu de données, afin que l'on puisse savoir dans quels formats sont les données, comment elles ont été collectées, avec quel programme, etc.

Tout à fait ! Dans la préservation des données, j'identifier 3 niveaux de documentation qui permette de décrire et signaler les jeux de données.

En 1er : les métadonnées (de type Dublin Core) sur un jeu de données, en 2ème un fichier de type Readme pour la description du contenu technique (il peut contenir un glossaire, des recommandations sur les logiciels à utiliser pour l'ouvrir,..) et le 3ème (le plus élaboré intellectuellement) : le data paper qui est une publication scientifique qui décrit le jeu de données, sa provenance, son potentiel de réutilisation, les techniques utilisées pour produire les résultats...

Plus d'info sur : http://doranum.fr/wp-content/uploads/gabarit_readme.txt (super site au passage)

(cc) BY-NC-ND

- Pour l'instant en France il ne semble pas y avoir d'obligation de créer des plans de gestion de données dans le cadre d'une recherche (hors H2020 et quelques éditeurs/publications). ANR n'a pas d'obligation ni les universités ou autres programmes de subventions.

Effectivement, il n'y a pas d'obligation à proprement parler.

Au niveau politique : un plan d'action poind à l'horizon, il spécifie un encouragement :

Engagement 14 du Plan d'action gouvernement ouvert[121]

« Dans le cadre du soutien public aux revues, recommander l'adoption d'une politique de données ouvertes associées aux articles et le développement des data papers. »

« Généraliser progressivement via un accompagnement la mise en place de plans de gestion des données dans les appels à projets de recherche, et inciter à une ouverture des données produites par les programmes financés (2019 et en continu). »

D'un point de vue des organismes financiers en France, l'ANR a une politique d'accès mais pas de publication des productions scientifiques : http://v2.sherpa.ac.uk/id/funder/30

- Discussion sur le travail avec l'INIST, la volonté de créer plusieurs modèles de plan de gestion des données, et les rendre disponibles sur DMP Opidor.
  Observation du mouvement de Data Activism.

Je pense que certains modèles vont devoir être plus spécifiques à certains laboratoires (à vérifier mais je n'ai pas le recul nécessaire). Si c'est le cas, la bonne collaboration avec l'INIST permettra de faciliter ce travail. DMP Opidor est un bon outil pour l'aide à la rédaction, il faut juste parfois le modifier pour le faire correspondre aux besoins pratiques.

Pour Dataactivist : il s'agit d'une organisation qui aide les collectivités territoriales à rendre les données administratives "exploitables" et les mettre en ligne.

[121] Etalab. *Pour une action publique transparente et collaborative : plan d'action national pour la France 2018-2020, op. cit.*

(cc) BY-NC-ND

TABLE DES ILLUSTRATIONS

Figures

Figure 1 – Structure d'un SGDA selon la norme ISO 30300 22
Figure 2 - Cycle de vie des documents en *records management* selon l'ADBS.... 58
Figure 3 - Cycle de vie des documents numérique selon l'APROGED 58
Figure 4 - Cycle de vie des données de la recherche selon le *UK Data Archive* ... 60
Figure 5 - Cycle de vie des données selon Données recherche Canada 61

Tableaux

Tableau 1 - Les niveaux d'ouverture de l'*open data* 32
Tableau 2 - Retrouvable..... 70
Tableau 3 - Accessible..... 71
Tableau 4 - Interopérable..... 72
Tableau 5 - Réutilisable..... 73
Tableau 6 - Authenticité 74
Tableau 7 - Fiabilité 75
Tableau 8 - Intégrité 76
Tableau 9 - Exploitabilité 77
Tableau 10 - PGD adapté aux *records*..... 83

(cc) BY-NC-ND

TABLE DES MATIERES

SIGLES ET ABREVIATIONS 7

INTRODUCTION 9

COMPRENDRE LE *RECORDS MANAGEMENT* ET LES PLANS DE GESTION DE DONNEES 11

Le *records management* 12

Plus que de la gestion de documents d'activité 12

Historique du records management 14

Quelques clés pour mieux comprendre le records management 16

Qu'est-ce qu'un record ? 16

Authenticité, fiabilité, intégrité, exploitabilité 18

Autour du record : les métadonnées 18

Gérer les documents probants et leurs métadonnées 20

Le système de gestion des documents d'activités (SGDA) 20

L'analyse de processus 22

Les outils du records management 23

Mettre en place un système de *records management* 24

La norme ISO 15489-2 25

Le mode d'emploi de l'ADBS 26

Les plans de gestion de données 29

L'objet d'une longue réflexion 29

D'abord, l'accès libre 29

Du partage des publications vers le partage des données 30

Les mouvements « open » 31

Donner sens aux « données de la recherche » 33

Le contenu d'un plan de gestion des données 35

La pratique concernant les PGD 37

L'exemple d'un pionnier : les États-Unis 37

Et en France ? 39

Mettre en place les plans de gestion de données 42

Bâtir une offre de service concernant les données de la recherche 42

Data curation profiles 44

LE RECORDS MANAGEMENT ET LES PLANS DE GESTION DE DONNEES, COMPLEMENTAIRES ? 47

Les points communs 47

Revalorisation des métiers des sciences de l'information 47

(cc) BY-NC-ND

Démarche d'évaluation *49*
Tout n'est pas records, tout n'est pas données de la recherche *50*
Contexte de création du record / des données de la recherche *52*
Les métadonnées *54*
Définir les responsabilités *55*
Le cycle de vie *56*
Établir une durée de conservation *62*
La question du droit d'accès *63*
Les points de divergence **64**
Des finalités différentes *64*
Une mise en place permanente, un projet ponctuel *66*
Mettre en place un système ou rédiger un document *67*
Des données authentiques ? Des *records* accessibles ? **68**
Des records FAIR ? *69*
Des données de la recherche authentiques, fiables, intègres et exploitables ? *74*
Une démarche pouvant en aider une autre ? **78**
Faciliter la rédaction d'un plan de gestion des données *78*
Un plan de gestion de données pour identifier et gérer nos records *82*
CONCLUSION **85**
SOURCES **89**
BIBLIOGRAPHIE **93**
ANNEXES **101**
TABLE DES ILLUSTRATIONS **123**
TABLE DES MATIERES **125**

(cc) BY-NC-ND