



UFR de sciences humaines et arts

—

Master esDOC – information, documentation, bibliothèque

Année universitaire 2020-2021

Les Plans de Gestion de Données entre les mains des chercheurs, dans le contexte de l'ouverture des données de la recherche

Mémoire pour l'obtention du grade de master
mention « Information, communication »

Présenté par
Alicia MAUBAILLY

Le 6 juillet 2021

Sous la direction de
Mme Marina DINET-DUMAS

Université de Poitiers





UFR de sciences humaines et arts

—

Master esDOC – information, documentation, bibliothèque

Année universitaire 2020-2021

Les Plans de Gestion de Données entre les mains des chercheurs, dans le contexte de l'ouverture des données de la recherche

Mémoire pour l'obtention du grade de master
mention « Information, communication »

Présenté par
Alicia MAUBAILLY

Le 6 juillet 2021

Sous la direction de
Mme Marina DINET-DUMAS

Université de Poitiers



Remerciements

Ce mémoire de recherche est issu d'une réflexion débutée au cours de l'année de M1 et concrétisée durant l'année de M2. Au cours de cette période, de nombreuses personnes m'ont accompagnée dans ma réflexion et je souhaite maintenant les remercier.

Dans un premier temps, je tiens à remercier ma directrice de mémoire, Mme Marina DINET DUMAS, pour sa disponibilité et son accompagnement sans failles. Je remercie également M. David GUILLEMIN pour l'encadrement de cette année de M2. Et je tiens à les remercier tous les deux pour leurs remarques constructives tout au long de la rédaction de ce mémoire.

De même, je voudrais exprimer ma reconnaissance à Mme Emmanuelle VAREILLE, M. Yanis DELMAS et Mme Armelle GIGLIO-JACQUEMOT pour leurs interventions dans le cadre de l'UE4 « Pratiques de recherche, pratiques de terrain » et les conseils qu'ils ont su m'apporter à cette occasion.

Par ailleurs, je souhaite montrer ma gratitude à ma tutrice de stage, Mme Pascale EBNER, ainsi que l'ensemble du service MiCOEST pour leurs encouragements et leur implication dans la préparation de mon oral.

Je voudrais également remercier toutes les personnes, chercheurs, ingénieurs, doctorants et post-doctorants, qui ont pris le temps de répondre à ma sollicitation et qui m'ont accordé de leur temps pour un entretien nécessaire à la bonne marche de ce mémoire.

Et enfin, je remercie ma famille pour sa compréhension durant les derniers mois de rédaction du mémoire. Et en particulier, j'offre mon entière gratitude à mon époux Valentin RIBEIRO, qui m'a soutenu et encouragé tout au long de ce master.

Glossaire

ANR – Agence Nationale de Recherche

CERN – Conseil Européen pour la Recherche Nucléaire

Cirad – Centre de Coopération Internationale en Recherche Agronomique et pour le développement

CNRS – Centre National de Recherche Scientifique

CoSO – Comité pour la Science Ouverte

Couperin – Consortium Unifié des Établissements Universitaires et de Recherche pour l'Accès aux Publications Numériques

DMP ou **PGD** – Data Management Plan / Plan de Gestion de Données

FOSTER – *Facilitate Open Science Training for European Research* : projet européen proposant des formations en ligne pour les chercheurs, les étudiants et les professionnels de l'information-documentation sur la Science Ouverte et le libre accès.

GTSO-Couperin – Groupe de travail Science ouverte-Couperin

H2020 : Programme européen Horizon 2020

IFSTTAR – Institut français des sciences et technologies des transports, de l'aménagement et des réseaux

OCDE – Organisation de Coopération et de Développement Économique

OpenAIRE – Open Access Infrastructure for Research in Europe : projet européen promouvant la Science ouverte et diffusant les publications, données ou autres produits de la recherche en libre accès.

Principes FAIR – *Findable, Accessible, Interoperable, Reusable* : il s'agit de préconisations européennes pour permettre que la donnée soit facile à trouver, accessible, interoperable et réutilisable.

PROGEDO – Production et Gestion des Données

RGPD – Règlement Général pour la Protection des Données

TGIR – Très Grande Infrastructure de Recherche

Sommaire

Remerciements.....	5
Glossaire.....	7
Introduction.....	11
Chapitre 1 – État de l’art.....	15
1. Le contexte : la science ouverte et les données de la recherche.....	15
1.1. La science ouverte, un mouvement récent.....	15
1.2. Des données aux données de la recherche.....	22
1.3. Présentation des structures d’appui qui accompagnent le mouvement et participent à l’acculturation de la communauté universitaire.....	25
2. Les données de la recherche au travers du Plan de Gestion de données.....	28
2.1. Les enjeux des données de la recherche.....	28
2.2. Le Plan de Gestion de Données : la formalisation d’un processus.....	31
2.3. L’appropriation des plans de gestion de données par les chercheurs.....	35
Chapitre 2 – Méthodologie.....	40
1. Présentation du terrain.....	40
2. Le corpus.....	41
2.1. La sélection des PGD.....	41
2.2. La grille d’analyse.....	42
3. Les entretiens.....	44
3.1. La construction de l’échantillon.....	44
3.2. Le déroulement.....	45
3.3. La grille d’entretien.....	45
Chapitre 3 – Les plans de gestion de données et leurs usages : une analyse des observations.....	48
1. La mise en œuvre d’un PGD.....	48
1.1. Le modèle : un choix ou une obligation.....	48

1.2. Le rôle potentiel du parcours des intervenants.....	49
1.3. Les raisons de la rédaction d'un PGD.....	50
2. La rédaction du PGD.....	51
2.1. La manière de remplir le PGD comme indicateur de la compréhension des enjeux.....	51
2.2. Difficultés rencontrées et solutions appliquées.....	53
2.3. Un soutien nécessaire à la bonne compréhension.....	55
3. Les apports de l'expérience de rédaction d'un PGD.....	55
3.1. Compréhension des enjeux véhiculés par les PGD.....	55
3.2. La sensibilité aux enjeux d'ouverture.....	56
3.3. La perception du PGD par les chercheurs.....	57
Conclusion.....	59
Bibliographie.....	62
Table des annexes.....	69
Annexes.....	70
Table des matières.....	112

Introduction

« Dans un monde interdépendant et interconnecté, la politique de recherche ne peut plus se faire à la seule échelle nationale, c'est tout l'enjeu d'un programme-cadre européen de recherche ambitieux, lui-même inscrit, comme notre politique nationale, dans un cadre mondial », a souligné Geneviève Fioraso, ministre de l'Enseignement supérieur et de la Recherche, à l'occasion de la cérémonie de lancement du programme européen Horizon 2020, lundi 16 décembre 2013, à Paris, en Sorbonne.¹

En France, le lancement de ce programme Horizon 2020 (H2020) a notamment donné lieu à l'organisation d'une nouvelle dynamique d'ouverture, dans le domaine de la recherche, des travaux et publications scientifiques, ainsi que des données produites dans le cadre de missions publiques ou lors de recherches. Dans la continuité, le projet de loi Valter (2015), relatif à la gratuité et aux modalités de la réutilisation des informations du secteur public, fait le premier pas vers l'ouverture des données. En 2016, la loi CADA de 1978, qui portait notamment sur la liberté d'accès aux documents administratifs et la réutilisation d'informations publiques², est révisée. Et dans le même temps est mise en place la *Loi pour une République Numérique*, dite Loi Lemaire. Puis, par la mise en place du *Plan National pour la Science Ouverte* en 2018, le gouvernement français entend « rendre obligatoire la diffusion ouverte des données de recherche issues de programmes financés par appels à projets sur fonds publics »³. Ce Plan National prévoit notamment de soutenir les axes suivants : la généralisation de l'accès ouvert aux publications, une structuration et l'ouverture des données de la recherche, ainsi que l'inscription du mouvement de la science ouverte dans une dynamique durable, européenne et internationale.

¹ La France donne le coup d'envoi du programme européen Horizon 2020. *Horizon 2020* [en ligne]. 16 décembre 2013. [Consulté le 12 janvier 2021]. Disponible à l'adresse : <https://www.horizon2020.gouv.fr/cid75832/la-france-donne-coup-envoi-programme-europeen-horizon-2020.html>.

² *Loi n° 78-753 du 17 juillet 1978 portant diverses mesures d'amélioration des relations entre l'administration et le public et diverses dispositions d'ordre administratif, social et fiscal*. 17 juillet 1978. [Consulté le 02 avril 2020]. Disponible à l'adresse : <https://www.legifrance.gouv.fr/affichTexte.do?cidTexte=JORFTEXT000000339241#LEGISCTA000006095904>.

³ MINISTÈRE DE L'ENSEIGNEMENT SUPÉRIEUR DE LA RECHERCHE ET DE L'INNOVATION. *Plan National pour la Science Ouverte* [en ligne]. 4 juillet 2018. [Consulté le 1^{er} avril 2020]. Disponible à l'adresse : <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>.

Le programme européen Horizon 2020⁴ est l'un des premiers à mettre en place de manière structurée l'usage des Plans de Gestion de Données (PGD) ou *Data management Plan* (DMP) dans le but d'ouvrir les données de la recherche. Il impose notamment le recours au PGD à tout projet souhaitant solliciter un financement du programme. Au niveau national, l'Agence Nationale de Recherche (ANR)⁵ s'inscrit dans la continuité du programme européen : en 2019, elle rend ces plans de gestion de données obligatoires pour demander un financement de projet, se plaçant ainsi dans une optique de science ouverte. En effet, le plan de gestion de données est un outil qui peut servir aux financeurs. Mais il fait également sens pour les organismes de recherche et pour les chercheurs⁶, qui peuvent ainsi réutiliser plus facilement des données issues de la recherche pour leurs propres études. Il consiste en un document formel explicitant la façon dont s'obtiennent, se documentent, s'analysent et s'utilisent des données, au cours de la recherche mais aussi une fois le projet terminé⁷. Les plans de gestion des données prennent donc de plus en plus d'importance dans le domaine de la recherche, et sont en lien direct avec l'ouverture des données de la recherche.

Les données de la recherche ne font pas l'objet d'un consensus au niveau de leur définition. Malgré cela, de manière générale et en suivant une approche juridique, on peut désigner les données de la recherche comme « *toutes les données issues d'une recherche financée en tout ou partie sur fonds publics* »⁸. Le récent mouvement de la science ouverte a permis que les données de la recherche prennent une place majeure dans la communauté scientifique, jusqu'à devenir un élément incontournable. En effet, l'ouverture de ces données s'inscrit dans la démarche

⁴ Il s'agit du programme de financement de la recherche et de l'innovation de l'Union Européenne pour la période 2014-2020.

⁵ L'ANR est l'agence française de financement de la recherche publique, placée sous la tutelle du Ministère chargé de la Recherche.

⁶ SERGI, Gianna. *Le Data Management Plan, DMP ou le Plan de Gestion de données, PGD et OPIDoR* [en ligne]. 2018. Collège de France – Direction des réseaux et partenariats documentaires – Département de bibliométrie et des ressources en ligne. [Consulté le 26 mars 2020]. Disponible à l'adresse : https://www.college-de-france.fr/media/bibliotheques-archives/UPL3412676772100941154_TUTORIEL_DMPOPIDOR_gs_v07092018.pdf.

⁷ DEBOIN, Marie-Claude. *Se familiariser avec les plans de gestion de données de la recherche, en 6 points* [en ligne]. Montpellier : CIRAD, 2018. p. 2. [Consulté le 13 janvier 2021]. Disponible à l'adresse : <https://coop-ist.cirad.fr/donnees-recherchehttps://coop-ist.cirad.fr/donnees-recherchehttps://coop-ist.cirad.fr/donnees-recherche>.

⁸ STÉRIN, Anne-Laure et NOÛS, Camille. Ouverture des données de la recherche : les mutations juridiques récentes. *Tracés. Revue de Sciences humaines*. 31 décembre 2019. N° 19, pp.37-50. Disponible à l'adresse : <http://journals.openedition.org/traces/10603>.

d'ouverture de la science et consiste à vouloir rendre accessibles à tous les travaux et recherches des chercheurs, mais aussi le fruit de leur réflexion et les données utilisées.

C'est en analysant ce contexte actuel et l'articulation entre les concepts de science ouverte, d'ouverture et de données de la recherche que nous en sommes arrivés à nous questionner sur le sujet. Dans la mesure où les PGD sont devenus largement obligatoires, en quoi peut-on dire qu'ils jouent un rôle dans la diffusion des enjeux de l'ouverture des données de la recherche auprès des chercheurs ?

Nous avons émis trois hypothèses pour réfléchir à la problématique posée et tenter d'y répondre. La première consiste à suggérer que le plan de gestion de données aurait une fonction qui va au-delà de son rôle initial de cadrage opérationnel. En effet, le document cible les enjeux des données et permettrait au chercheur de mieux les comprendre, mais cette sensibilisation aux enjeux pourrait varier selon le cadrage demandé par exemple par l'organisme de recherche. Nous présumons également que le cadre dans lequel l'élaboration du plan de gestion de données se réalise peut modifier cette perception, selon que l'élaboration est le résultat d'une démarche volontaire ou imposée. Nous supposons, en deuxième hypothèse, qu'une meilleure compréhension des enjeux de l'ouverture des données de la recherche, obtenue par l'expérience d'usage du plan de gestion de données, pourrait engendrer un phénomène d'adhésion à l'ouverture des données chez les chercheurs. Et enfin la dernière hypothèse soulevée est que, dans le cadre de l'élaboration des Plans de Gestion de Données, les professionnels de l'information-documentation auraient un rôle de médiateur à jouer auprès des chercheurs qui sont amenés à utiliser ces plans de gestion de données. Nous faisons plus précisément l'hypothèse que certaines rubriques des plans de gestion de données mobilisant une expertise info-documentaire amèneraient les chercheurs et les professionnels de l'information-documentation à collaborer. Nous pouvons également imaginer que cette collaboration permettrait, au-delà de la rédaction méthodique du document, d'échanger les opinions quant aux enjeux de l'ouverture des données de la recherche.

Afin de soumettre ces hypothèses, nous envisageons d'étudier un corpus de modèles de plans de gestion de données, tirés de la plateforme *DMP Opidor*. Cela permettra d'étudier plus précisément la composition d'un PGD et ce qui est requis de manière prioritaire, mais également

de déterminer s'il existe des différences entre les multiples modèles et de chercher à en comprendre les raisons. Dans un deuxième temps, nous envisageons également de procéder à des entretiens individuels de chercheurs, afin de saisir leurs usages des PGD, et à travers eux, de mesurer l'influence de cette expérience sur leurs positions quant à l'ouverture des données.

Ainsi, pour tenter de répondre à notre questionnement et tester les hypothèses avancées, nous présenterons dans une première partie le contexte des données de la recherche, les enjeux puis les questions autour de l'appropriation des plans de gestion de données. La seconde partie présentera la mise en place de la méthodologie de recherche appliquée pour tenter de répondre aux hypothèses émises. Enfin, une troisième partie portera sur le résultat de ces recherches.

Chapitre 1 – État de l’art

1. Le contexte : la science ouverte et les données de la recherche

Dans cette première partie, nous tenterons de faire le point sur le contexte entourant la mise en place de la science ouverte et le développement des données de la recherche.

1.1. La science ouverte, un mouvement récent

La science ouverte ou *open science* est un mouvement international qui existe depuis les années 1990. On attribue souvent à Jean-Claude Bradley, chimiste américain, le rôle d’initiateur du mouvement, avec la mise en place des « carnets de laboratoire publics »⁹ au sein de sa discipline, la chimie. La Commission Européenne définit la science ouverte comme « *une nouvelle approche du développement scientifique, basée sur le travail coopératif et la distribution de l’information via des réseaux utilisant des technologies avancées et des outils collaboratifs. Elle vise à faciliter l’acquisition de connaissances collectives et à encourager l’émergence de solutions basées sur l’ouverture et sur le partage* »¹⁰. En effet, tous les secteurs de la recherche n’ont pas la même implication dans la diffusion des travaux scientifiques ni les mêmes moyens, c’est pourquoi aux prémises de la science ouverte, on pouvait observer des disparités entre les disciplines et les moyens mis en œuvre dans le cadre des recherches scientifiques. Le mouvement de la science ouverte est assez complexe. En effet, il prend forme pour des raisons scientifiques, développées dans la définition ci-dessus, mais également pour des raisons citoyennes tel que le rétablissement d’une confiance entre la science et les citoyens, économiques, sociales et politiques telle que la recherche d’un certain contrôle dans la production scientifique¹¹. Pour citer la Commission Européenne, la science ouverte doit être « *aussi ouverte que possible, aussi fermée que nécessaire* »¹². Il faut comprendre par là qu’il y a un besoin d’ouvrir les données issues de la

⁹ CHARTRON, Ghislaine. L’Open science au prisme de la Commission européenne. *Éducation et sociétés*. 25 septembre 2018. Vol. n° 41, n° 1, p. 177. Disponible à l’adresse : <https://www.cairn.info/revue-education-et-societes-2018-1-page-177.htm>.

¹⁰ RENTIER, Bernard. *Science Ouverte, le défi de la transparence* [en ligne]. Bruxelles : Académie Royale de Belgique, 2018. [Consulté le 9 mars 2020]. Disponible à l’adresse : <https://orbi.uliege.be/handle/2268/230014>.

¹¹ CNRS. Le mouvement pour la science ouverte. *Science ouverte* [en ligne]. 2020. [Consulté le 1 novembre 2020]. Disponible à l’adresse : <https://www.science-ouverte.cnrs.fr/le-mouvement-pour-la-science-ouverte/>

¹² *Ibid.*

recherche, mais que les institutions doivent aussi pouvoir garder un minimum la main sur ce qui est fait de ces données.

La science ouverte a pris de l'importance avec le développement du web, mais surtout avec les questions autour du libre accès. Il est donc intéressant de retracer le parcours entre l'organisation du libre accès dans la communauté scientifique et la mise en place de la science ouverte. Ensuite, nous présenterons la législation qui entoure peu à peu ce mouvement initial de la science ouverte. En France notamment, la science ouverte s'est développée avec la *Loi pour la République numérique* (2016), et davantage encore après l'émergence du *Plan National pour la Science Ouverte*. Enfin, nous détaillerons les principes FAIR, qui sont des bonnes pratiques mises en avant pour faciliter la diffusion des données, et que la science ouverte tend à pérenniser.

1.1.1. Du libre accès à la science ouverte

Comme cela a été présenté précédemment, les questions de libre accès et de science ouverte sont intimement liées. Celles-ci sont notamment issues d'une volonté internationale d'ouvrir la science, et particulièrement les données résultantes de la recherche.

Le site consacré à la science ouverte proposé par le consortium Couperin définit le libre accès, ou *open access*, comme étant la mise à disposition immédiate, gratuite et permanente sur Internet des publications scientifiques issues de la recherche et de l'enseignement¹³. Selon Stevan Harnad, chercheur et professeur de *psychologie* à l'université de Québec, les débuts du libre accès sont à situer dès les années 1980 avec l'archivage et le partage *File Transfer Protocol* (FTP) anonyme¹⁴. Il s'agit d'un mouvement initié par des chercheurs et développé avec l'arrivée du web, puis repris dans les années 2000 par les États. Son but premier est de remettre en question le monopole que possèdent certains éditeurs sur la diffusion des résultats produits par la recherche¹⁵. Par la suite, d'autres questionnements sont apparus, comme celui de la pérennité des articles scientifiques disponibles au travers des abonnements numériques : en effet, la conservation de ces articles dépend alors du renouvellement ou non des abonnements par

¹³ <https://scienceouverte.couperin.org/comment-definir-lopen-access/>

¹⁴ HARNAD, Stevan. *Mouvement de l'Open Access*. CC.Webcast [en ligne]. 24 janvier 2013. [Consulté le 18 janvier 2021]. Disponible à l'adresse : https://webcast.in2p3.fr/video/mouvement_de_l_open_access.

¹⁵ BENEVOLO, Sylvie. *Dynamiques des services d'information scientifique et technique face à la politique nationale d'évaluation de la recherche: accompagner la gestion des publications scientifiques au sein des UMR de l'IRD* [en ligne]. Mémoire. CNAM-INTD, 2019. Disponible à l'adresse : https://memsic.ccsd.cnrs.fr/mem_02903694.

l'établissement, ce qui pose des problèmes d'accès aux publications scientifiques¹⁶. Or, le libre accès a justement pour objectif de favoriser la disparition des barrières d'accès, aussi bien économiques que juridiques, aux publications et aux données de la recherche¹⁷. Face à cela des mesures ont été adoptées pour promouvoir le libre accès, avec trois déclarations présentées entre 2002 et 2003 : l'initiative de Budapest, la déclaration de Béthesda et la déclaration de Berlin¹⁸. Ces textes fondateurs donnent au libre accès la définition suivante :

Par « accès libre » à cette littérature, nous entendons sa mise à disposition gratuite sur l'Internet public, permettant à tout un chacun de lire, télécharger, copier, transmettre, imprimer, chercher ou faire un lien vers le texte intégral de ces articles, les disséquer pour les indexer, s'en servir de données pour un logiciel, ou s'en servir à toute autre fin légale, sans barrière financière, légale ou technique autre que celles indissociables de l'accès et l'utilisation d'Internet. La seule contrainte sur la reproduction et la distribution, et le seul rôle du copyright dans ce domaine devrait être de garantir aux auteurs un contrôle sur l'intégrité de leurs travaux et le droit à être correctement reconnus et cités.¹⁹

C'est par l'intérêt grandissant porté au libre accès qu'est né ce mouvement de la science ouverte²⁰, afin d'aller au-delà de la question de la barrière d'accès aux publications et tendre vers une science plus efficace et innovante, répondant à des enjeux scientifiques, sociaux et économiques. Il a été adopté par les gouvernements, qui ont mis en œuvre assez rapidement une politique de la science ouverte. Celle-ci s'est organisée en suivant deux arguments majeurs, que relèvent Ghislaine Chartron et Joachim Schöpfel dans leur article « *Open access et Open science en débat* »²¹. Le premier est celui de faciliter le transfert des résultats de la recherche à la société, et

¹⁶ Ce point a été soulevé par un texte de la *Public Library of Science* en 2001 (<https://openaccess.inist.fr/la-lettre-ouverte-de-public-library-of-science/>).

¹⁷ FÉRET, Romain, et al. Améliorer son projet ANR grâce à la Science Ouverte [en ligne]. Zenodo. 9 mai 2020. p. 8. [Consulté le 10 décembre 2020]. Disponible à l'adresse : <https://lilloa.univ-lille.fr/handle/20.500.12210/20359.2>.

¹⁸ Ces déclarations ainsi que d'autres textes de références sur le libre accès sont disponibles à l'adresse <https://openaccess.inist.fr/category/textes-de-references/>.

¹⁹ Initiative de Budapest pour l'accès ouvert. 2002. Disponible à l'adresse : <https://openaccess.inist.fr/initiative-de-budapest-pour-laccès-ouvert/>.

²⁰ FÉRET, Romain, et al. Améliorer son projet ANR grâce à la Science Ouverte. op.cit. p. 8.

²¹ CHARTRON, Ghislaine et SCHÖPFEL, Joachim. Open access et Open science en débat. *Revue française des sciences de l'information et de la communication* [en ligne]. Août 2017. N° 11. [Consulté le 2 janvier 2021]. Disponible à l'adresse : <http://journals.openedition.org.ressources.univ-poitiers.fr/rfsic/3331>.

le second entend faciliter les liens entre ces résultats et l'économie dans sa globalité. Mais, pour impulser cette politique de la science ouverte, un rééquilibrage de l'accès ouvert s'est avéré nécessaire. Et parmi les points nécessaires à traiter, il a fallu fixer les lignes directrices du libre accès, conditionner les financements, et travailler la législation et notamment les questions de droit liées à la propriété intellectuelle²². De plus, de nouveaux moyens de communication se sont développés, introduisant par la même occasion de nouveaux formats d'écriture, ce qui a entraîné la mise en œuvre de plateformes dédiées, pour publier et mettre à disposition de la communauté, comme la plateforme *Openedition* citée par les auteurs²³. De la même manière, le mouvement voit également arriver de nouvelles méthodes de travail scientifiques, avec de nouveaux outils numériques. Aujourd'hui, il est important de retenir que l'innovation se trouve au cœur de la science ouverte.

Aussi, une nouvelle volonté apparaît en France, avec la mise en œuvre du *Plan national pour la science ouverte*, illustré par le discours de Frédérique Vidal :

*La Science Ouverte n'est pas une mode, ce n'est pas une discipline, c'est un nouveau paradigme. Elle comporte donc ses nouvelles pratiques et ses nouvelles compétences. J'attacherai une attention particulière à ce que ces compétences soient considérées comme faisant partie du bagage initial des jeunes chercheurs. La Science Ouverte ne sera pas l'affaire d'un petit groupe de spécialistes, elle devra imprégner l'ensemble du monde de la recherche.*²⁴

En résumé, pour reprendre les mots de Ghislaine Chartron, on peut dire que « *l'accès libre s'est focalisé sur la communication et la circulation des publications scientifiques, [et que] la science ouverte a élargi la problématique à la transformation de la pratique de la science et à son évaluation* »²⁵.

²² BENEVOLO, Sylvie. *Dynamiques des services d'information scientifique et technique face à la politique nationale d'évaluation de la recherche*. *op. cit.* p. 32.

²³ CHARTRON, Ghislaine et SCHÖPFEL, Joachim. *Open access et Open science en débat*. *op.cit.*

²⁴ FÉRET, Romain, et al. *Améliorer son projet ANR grâce à la Science Ouverte*. *op.cit.* p. 9.

²⁵ CHARTRON, Ghislaine. *L'Open science au prisme de la Commission européenne*. *op. cit.* p. 178.

1.1.2. L'environnement juridique de la science ouverte

Tout d'abord, il est nécessaire de faire le point sur les textes importants qui entourent le libre accès. Dans l'ordre chronologique, on trouve d'abord la *Déclaration de Budapest* (2002), qui propose le libre accès aux publications de recherche. En 2003, la *Déclaration de Berlin* propose d'ouvrir l'accès à la littérature scientifique mondiale, aux données de la recherche et aux outils qui permettent de collecter ces données. Et cette même année, le Parlement Européen vote pour l'ouverture des données institutionnelles dans la perspective de leur réutilisation (*Directive Public Sector Information*). Cette directive européenne a d'ailleurs donné lieu en France à la création d'un département de la direction interministérielle du numérique, *Étalab*²⁶, en 2011, pour l'ouverture et le partage de ces données.²⁷

Ensuite, d'autres textes ont été publiés concernant plus particulièrement la science ouverte. Après la Commission Européenne du 17 juillet 2012²⁸, qui place les données au même niveau que les publications²⁹, on peut présenter deux lois relativement importantes. Promulguée le 28 décembre 2015, la *Loi Valter* porte sur la gratuité et les modalités de réutilisation des informations issues du secteur public. Cette loi entérine, d'après le guide réalisé par le groupe *Données* du GTSO-Couperin³⁰, le principe de l'ouverture des données publiques (*Open data*). Et il en est de même pour la *Loi pour une République numérique* du 7 octobre 2016, aussi appelée *Loi Lemaire* car proposée par la secrétaire d'État au numérique, Axelle Lemaire. En effet, cette loi vise à faciliter la libre diffusion de résultats de recherche.

Enfin, à la suite d'une recommandation faite par la Commission européenne en 2018, encourageant l'ouverture des données de la recherche, la Ministre de l'Enseignement supérieur, de

²⁶ Pour plus d'informations sur les missions d'*Étalab*, consulter <https://www.etalab.gouv.fr>.

²⁷ JACQUEMIN, Bernard, SCHÖPFEL, Joachim et FABRE, Renaud. Libre accès et données de recherche. De l'utopie à l'idéal réaliste [en ligne]. *Études de communication*. 19 décembre 2019. N° 52, p. 13. [Consulté le 9 février 2020] Disponible à l'adresse : <https://www.cairn.info/revue-etudes-de-communication-2019-1-page-11.htm>.

²⁸ COMMISSION EUROPÉENNE. *Communication de la Commission au Parlement européen, au Conseil, au Comité économique et social européen et au comité des régions. Pour un meilleur accès aux informations scientifiques : dynamiser les avantages des investissements publics dans le domaine de la recherche*. 17 juillet 2012. [Consulté le 22 février 2021]. Disponible à l'adresse : <https://eur-lex.europa.eu/legal-content/FR/TXT/?uri=CELEX%3A52012DC0401>.

²⁹ REBOUILLAT, Violaine. *Ouverture des données de la recherche : de la vision politique aux pratiques des chercheurs* [en ligne]. Thèse de doctorat en Sciences de l'information et de la communication. Conservatoire national des arts et métiers – CNAM, 2019. p. 71. [Consulté le 23 janvier 2020]. Disponible à l'adresse : <https://tel.archives-ouvertes.fr/tel-02447653>.

³⁰ FÉRET, Romain, et al. *Améliorer son projet ANR grâce à la Science Ouverte*. op.cit.

la Recherche et de l'Innovation, Frédérique Vidal, lance le *Plan national pour la science ouverte* le 4 juillet 2018. Ce plan s'organise autour de trois axes, dont le premier est la publication, le second les données et enfin les actions transverses. Et à travers ce plan national, « la France s'engage pour que les résultats de la recherche scientifique soient ouverts à tous, chercheurs, entreprises et citoyens, sans entrave, sans délai, sans paiement.³¹». Dans le même temps, il est prévu la constitution d'un Comité pour la Science Ouverte (CoSo)³².

La Commission européenne a également mis en place des principes, lignes directrices auxquelles les projets sollicitant des aides financières doivent se conformer, dans le but de rendre plus facile la réutilisation des données.

1.1.3. Les principes FAIR

Comme le présente Violaine Rebouillat dans sa thèse, les principes FAIR sont donc un ensemble de préconisations reprises par la Commission Européenne pour le contexte des projets, afin de rendre les données trouvables (*findable*), accessibles (*accessible*), interopérables (*interoperable*) et réutilisables (*reusable*), autant pour les humains que pour les machines³³. Ce qui signifie que le principe *Findable* vise à faciliter la découverte des données, le principe *Accessible* a pour objectif d'accéder et de télécharger facilement les données, le principe *Interoperable* doit permettre d'exploiter les données quel que soit l'environnement informatique utilisé, et enfin le principe *Reusable* sert à réutiliser des données pour de futures recherches³⁴. Il s'agit donc de définir les bonnes pratiques applicables pour la gestion des données. Ces principes avaient été présentés dès 2016 par Mark D. Wilkinson et d'autres chercheurs, dans un article paru dans la revue *Nature*³⁵.

Ces quatre principes visent à faciliter l'accès aux publications scientifiques et l'ouverture des données qui en sont issues. Ils sont recommandés dans le cadre des plans de gestion de données, que ce soit au niveau du programme européen *Horizon 2020* ou du programme national de l'ANR.

³¹ CNRS. Le mouvement pour la science ouverte. *op.cit.*

³² *Ibid.*

³³ REBOUILLAT, Violaine. *Ouverture des données de la recherche: de la vision politique aux pratiques des chercheurs.* *op.cit.* p. 76.

³⁴ <https://doranum.fr/enjeux-benefices/principes-fair/>.

³⁵ Cet article du collectif FORCE11 est disponible à l'adresse suivante : <https://www.nature.com/articles/sdata201618>.

Et, si l'on en croit Joachim Schöpfel, ces principes sont sur le point de devenir des standards du fait qu'ils s'imposent comme cadre de référence unique dans l'environnement des données de la recherche³⁶.

Parmi les finalités visées par ces principes FAIR, il peut être intéressant de présenter les entrepôts de données, qui s'inscrivent dans la notion de « données trouvables » et jouent un rôle important dans la conservation et le partage des données. On peut notamment présenter une étude faite par Hélène Prost et Joachim Schöpfel en 2019 à propos de ces entrepôts de données³⁷, qui sont « *des services en ligne permettant la collecte, la description, la conservation, la recherche et la diffusion de jeux de données* ». Les entrepôts de données font donc partie intégrante de la science ouverte, et sont particulièrement concernés par l'application des principes FAIR. Or, ces chercheurs ont démontré qu'il existe un manque de normalisation, notamment à propos des métadonnées, pour une grande majorité des sites qu'ils ont évalués. En effet, ils signalent que seuls les « grands entrepôts génériques » semblent suivre ces recommandations internationales pour la gestion des données de la recherche. Cela montre donc que les principes, n'ayant pas force de loi, ne sont pas encore strictement adoptés ou qu'ils sont en cours d'assimilation par les entrepôts de données. Comme grand entrepôt on peut nommer Zenodo, qui est un entrepôt international créé lors du projet OpenAIRE par le CERN.

La mise en place du mouvement de la science ouverte a donné lieu à l'organisation progressive d'une législation visant à tenter d'encadrer les nouvelles pratiques. De celle-ci sont issus les principes FAIR, bonnes pratiques proposées et soutenues par la science ouverte et qui ont pour objectif de faciliter la diffusion des données.

³⁶ SCHÖPFEL, Joachim. Hors norme? Une approche normative des données de la recherche. *Revue COSSI* [en ligne]. 2018. N° 5. [Consulté le 10 décembre 2020]. Disponible à l'adresse : <https://revue-cossi.info/numeros/n-5-2018-processus-normalisation-durabilite-information/730-5-2018-schopfel>

³⁷ PROST, Hélène et SCHÖPFEL, Joachim. Les entrepôts de données en sciences de l'information et de la communication (SIC). Une étude empirique [en ligne]. *Études de communication*. 2019. N° 52, p. 89. [Consulté le 17 janvier 2021]. Disponible à l'adresse : <https://www.cairn.info/revue-etudes-de-communication-2019-1-page-71.htm>.

1.2. Des données aux données de la recherche

Les données de la recherche occupent une place centrale dans la politique de la science ouverte. Au sein même de la science ouverte et par le mouvement du libre accès, les données issues des recherches ont pris toujours plus d'importance dans le milieu de la recherche scientifique, se distinguant progressivement des données de manière générale. Il nous faut donc maintenant expliquer ce que sont les données de la recherche. Mais avant cela, nous allons nous attacher à déterminer ce que désigne le terme général de « donnée », et l'appliquer au contexte du numérique et de la recherche.

1.2.1. Qu'est-ce que la « donnée » ?

Pour définir la donnée, on peut citer dans un premier temps la définition de la CNIL : « *Une donnée correspond à la représentation d'une information sous une forme conventionnelle destinée à faciliter son traitement (en anglais : data)* »³⁸. Or, la donnée est souvent placée au sein du triptyque « donnée, information, connaissance ». Et, l'articulation entre ces trois termes a souvent fait l'objet de réflexions chez les chercheurs, comme ce fut notamment le cas d'Evelyne Broudoux en 2018³⁹.

L'étymologie du terme « donnée » est présentée dans la thèse de Violaine Rebouillat⁴⁰, qui s'est appuyée en partie sur les travaux de Daniel Rosenberg sur le terme « *data* ». V. Rebouillat formule l'idée que le mot « donnée » a une « dimension rhétorique », c'est-à-dire qu'il est lié au discours et conserve sa valeur, que la donnée soit vraie ou fausse. Et elle explique que « *selon Rosenberg, c'est cette dimension rhétorique qui a rendu le terme si prégnant dans notre monde d'aujourd'hui, où la communication occupe une place centrale* ». Toujours d'après V. Rebouillat, Evelyne Broudoux considère que, de nos jours, la définition informatique de la donnée prend le pas sur celle de la donnée classique. Ce qui l'amène à suggérer que les données doivent être traitées de la même manière que les documents.

³⁸ Open Data : comprendre les enjeux de l'ouverture des données publiques culturelles. [en ligne]. p. 7. [Consulté le 2 janvier 2021]. Disponible à l'adresse : <https://dataactivist.coop/ministere-culture/jour1.html#1>.

³⁹ BROUDOUX, Evelyne. *Dispositif info-communicationnels numériques : éditorialisation et autorité*. Habilitation à diriger des recherches. Université Paris 8. 2018

⁴⁰ REBOUILLAT, Violaine. *Ouverture des données de la recherche*. op.cit. p. 35 et suivantes.

*Données, informations et connaissances sont des concepts que l'on relie habituellement ensemble selon un principe d'intégration et de construction mais les données acquièrent leur territoire propre et sont susceptibles de traitement au même titre que les documents.*⁴¹

Anne-Laure Stérin et Camille Noûs rappellent que certains font également la distinction entre les données brutes, primaires, enrichies, élaborées ou secondaires, ou encore les métadonnées⁴². Cette absence de consensus autour de la définition des données rend d'ailleurs fragile, aux yeux de Sylvie Fayet, la distinction entre données et publications⁴³. C'est d'ailleurs le constat que fait Francisca Cabrera dans son mémoire, en montrant qu'il existe parfois une confusion voire une ambiguïté du terme « données » chez les chercheurs⁴⁴, qui en arrivent parfois à les confondre avec les publications.

1.2.2. Quelle définition pour les données de la recherche ?

Définir ce que sont les données de la recherche n'est donc pas simple, étant donné que la définition de la donnée reste elle-même difficile à établir. On l'a vu, pour Violaine Rebouillat la donnée est une entité informationnelle qui possède un lien étymologique avec le discours. Pour cette raison, elle présente la notion de « donnée de la recherche » comme étant liée au contexte. Deux définitions, fournies par des organismes officiels, permettent de se faire une idée assez précise de ce que le terme doit désigner pour la communauté scientifique. En 2016, la Commission Européenne annonce que

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 d'expériences, des mesures, des observations sur le terrain, des

⁴¹ Ibid. p. 37.

⁴² STÉRIN, Anne-Laure et NOÛS, Camille. Ouverture des données de la recherche : les mutations juridiques récentes. *op.cit.*

⁴³ FAYET, Sylvie. « Données » de la recherche, les mal-nommées. *UrfistInfo* [en ligne]. [Consulté le 23 septembre 2020]. Disponible à l'adresse : <https://urfistinfo.hypotheses.org/2581>.

⁴⁴ CABRERA, Francisca. *Les données de la recherche en Sciences humaines et sociales : enjeux et pratiques. Enquête exploratoire*. [en ligne]. Mémoire. Institut National des techniques et sciences de la documentation - Conservatoire National des Arts et des Métiers, 2015. p. 52. [Consulté le 3 février 2020]. Disponible à l'adresse : https://memsic.ccsd.cnrs.fr/mem_01117375.

*résultats d'enquêtes, des enregistrements d'entretiens ou des images. Il s'agit plus spécifiquement de données disponibles sous forme numérique.*⁴⁵

Le terme « information » qu'emploie la Commission européenne est très vague et ne permet pas, en réalité, de clarifier ce qui peut être désigné comme « données de la recherche ». D'ailleurs, cela se retrouve dans la suite de la définition, puisqu'elle ne fait que lister des exemples de données de la recherche, laissant supposer un large spectre de données. L'OCDE, dès 2007, avait formulé sa propre définition, désignant les données de la recherche 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* »⁴⁶. De son côté, cette définition présente les données de la recherche comme les sources utilisées par les chercheurs. Or, là encore les sources utilisées par les chercheurs ne sont pas les mêmes selon la discipline concernée. En réalité, on ne peut pas dire qu'il existe un consensus de la part des acteurs des politiques publiques autour de ce que sont les « données de la recherche », et l'on voit fleurir une grande variété de définitions et de versions, dans la littérature scientifique autant que dans le milieu professionnel. Par exemple, pour Lionel Maurel les données de la recherche désignent les informations qui sont générées dans le contexte d'un processus de recherche⁴⁷, tandis que V. Rebouillat les définit comme « des données qui sont collectées par les chercheurs dans la perspective d'être utilisées comme preuve d'une théorie » et qui pour cela sont rendues « portables »⁴⁸. Enfin, le Cirad propose de définir les données de la recherche de la manière suivante :

⁴⁵ COMMISSION EUROPÉENNE-DIRECTION GÉNÉRALE DE LA RECHERCHE ET DE L'INNOVATION. Lignes directrices pour le libre accès aux publications scientifiques et aux données de la recherche dans Horizon 2020 [en ligne]. 15 février 2016, version 2.1. [Consulté le 03/03/2020]. Disponible à l'adresse : <https://www.enssib.fr/bibliotheque-numerique/documents/66433-lignes-directrices-pour-le-libre-acces-aux-publications-scientifiques-et-aux-donnees-de-la-recherche-dans-horizon-2020.pdf>.

⁴⁶ OCDE. *Principes et lignes directrices pour l'accès aux données de la recherche financée sur fonds publics* [en ligne]. Paris : Éditions OCDE, 2007. [Consulté le 08/02/2020]. Disponible à l'adresse : <https://doi.org/10.1787/9789264034020-en-fr>.

⁴⁷ MAUREL, Lionel. La réutilisation des données de la recherche après la loi pour une République numérique. *La diffusion numérique des données en SHS – Guide de bonnes pratiques éthiques et juridiques* [en ligne]. Presses Universitaires de Provence, 2018. [Consulté le 20 janvier 2020]. Disponible à l'adresse : <https://hal.archivesouvertes.fr/hal-01908766>.

⁴⁸ REBOUILLAT, Violaine. *Ouverture des données de la recherche*. op.cit. p. 46.

*Les données de la recherche, ou données de recherche, ou données scientifiques, peuvent être définies comme l'ensemble des informations collectées, observées ou créées sous une forme numérique dans le cadre d'un projet de recherche*⁴⁹.

Ainsi, on peut dire que les données deviennent ce que l'on nomme « données de la recherche » à partir du moment où elles sont utilisées dans des travaux de recherche, qu'elles soient sources ou produits de la recherche du scientifique. Et Thierry Fournier, conservateur des bibliothèques et actuellement en charge de la documentation en ligne et de l'*open access* au SCD de l'Université de Rennes 1, va même plus loin, puisqu'il avance qu'elles peuvent également être des documents qui servent à comprendre « comment la science se fait » mais qui ne sont pas « nécessaires pour valider les résultats de recherche »⁵⁰. Ainsi, pour lui, de nombreuses données peuvent entrer dans la désignation « données de la recherche ». D'autant qu'il rappelle que les données ne prennent pas la même forme selon la discipline, et n'ont donc probablement pas le même traitement, ce qui joue certainement sur l'usage qui en est fait. En effet, les données en sociologie, plutôt issues des enquêtes, ne sont pas comparables aux rapports de fouilles auxquels se réfèrent les archéologues dans le cadre de leurs recherches, pour prendre seulement deux exemples disciplinaires.

Aujourd'hui, les données sont dans leur grande majorité de nature numérique, et se trouvent en grande quantité, ce qui oblige la communauté de la science ouverte à s'intéresser de plus près aux questions de traitement, de stockage et de diffusion. Et pour cela, des structures s'organisent pour accompagner ce mouvement et aider à la prise de conscience de ces questions par la communauté scientifique.

1.3. Présentation des structures d'appui qui accompagnent le mouvement et participent à l'acculturation de la communauté universitaire

Pour accompagner ce mouvement de la Science Ouverte et principalement le Plan National pour la Science Ouverte, plusieurs organismes sont nés. Des sites web ont notamment été mis en place afin d'informer les chercheurs et les scientifiques, et les guider dans les démarches de

⁴⁹ DEBOIN, Marie-Claude. *S'initier en ligne aux données de la recherche et à leur gestion, en 4 points*. op.cit. p. 1.

⁵⁰ FOURNIER, Thierry. Les données de la recherche : définition et enjeux. *Arabesques* [en ligne]. 22 août 2019. [Consulté le 18 février 2020]. Disponible à l'adresse : <https://publications-prairial.fr/arabesques/index.php?id=985>.

Science ouverte. Selon les disciplines et les besoins, les chercheurs et les organismes de recherche peuvent avoir recours à l'accompagnement offert par ces institutions, qui proposent essentiellement des contenus et ressources pédagogiques autour de la science ouverte, du libre accès et des données de la recherche.

1.3.1. *CoopIST-Cirad⁵¹, spécialisé en Information Scientifique et Technique*

CoopIST (*Coopérer en Information Scientifique et Technique*) est un site internet géré par le Cirad, organisme français de recherche agronomique et de coopération internationale pour le développement durable des régions tropicales et méditerranéennes. Il propose des contenus en libre accès, pour aider à rédiger, publier et diffuser, évaluer et gérer les publications et les données de la recherche. On peut par exemple y trouver des fiches, des bases de données, des actualités, etc. Et ce site est principalement destiné à des professionnels de l'IST et à des chercheurs.

1.3.2. *Le CoSO⁵², concernant la science ouverte en France*

Le Comité pour la Science Ouverte (CoSO) s'inscrit dans le cadre du plan national pour la science ouverte et est organisé par le Ministère de l'enseignement supérieur, de la recherche et de l'innovation. Il soutient donc l'ouverture des publications et des données de la recherche. Le CoSO a créé le site web « Ouvrir la Science », pour répondre aux objectifs suivant : faire connaître la Science Ouverte aux chercheurs mais aussi à tout public intéressé, expliquer les défis à relever pour sa réalisation, et sensibiliser aux enjeux de la Science Ouverte. Ce site propose des contenus variés : de l'actualité, des ressources (textes fondateurs, rapports, etc.), l'agenda des manifestations en lien avec les axes développés par le Plan National, et un blog regroupant l'actualité et une veille de la Science Ouverte.

1.3.3. *Science Ouverte France, Couperin⁵³ : pour promouvoir l'Open Access et les données de la recherche*

Le Consortium Couperin (Consortium unifié des établissements universitaires et de recherche pour l'accès aux publications numériques) est une association à but non lucratif qui s'organise autour de l'information scientifique et technique. Il regroupe essentiellement des

⁵¹ Adresse du site : <https://coop-ist.cirad.fr/>.

⁵² Adresse du site : <https://www.ouvrirlascience.fr/>.

⁵³ Adresse du site : <https://scienceouverte.couperin.org/>.

universités, des organismes de recherche et des grandes écoles, et a pour volonté d'améliorer la communication scientifique. Pour cela, le consortium a mis en place le GTSO, qui est un Groupe de Travail pour la science ouverte animé par Christine Ollendorff (Directrice de la Documentation et de la Prospective chez Arts et Métiers ParisTech), et qui se divise lui-même en cinq groupes thématiques. Parmi ceux-ci, il y a notamment un groupe « données », animé par Romain Féret, ancien chef du service Science ouverte à l'université de Lille. Le consortium est par ailleurs partenaire du projet FOSTER, qui cherche à promouvoir la science ouverte en valorisant les contenus didactiques français et européen. Dans ce cadre, le GTSO a d'ailleurs créé un support en ligne⁵⁴, afin notamment de sensibiliser les jeunes chercheurs aux questions de science ouverte et de données de la recherche.

1.3.4. *DORANum⁵⁵, outil proposant des formations aux données de la recherche*

Doranum est une plateforme mise en place par Inist-CNRS et le réseau des Urfist (Unités Régionales de Formation à l'Information Scientifique et Technique), et qui s'inscrit dans les projets du CoSO. Outil de formation à distance, elle met à disposition de la communauté universitaire des ressources de formation ou d'auto-formation sur la gestion des données de la recherche et sur leur partage.

Ces diverses ressources sont à destination des chercheurs, des enseignants-chercheurs, des doctorants ainsi que des professionnels de l'information évoluant dans le secteur de l'enseignement supérieur.

Après avoir posé le contexte encadrant l'intérêt croissant pour les données de la recherche et après avoir travaillé à définir les données de la recherche, la présentation des différentes plateformes mises en place par des structures d'appui à la Science ouverte permet ensuite de s'intéresser à la manière dont les utilisateurs sont guidés dans la gestion des données produites par la recherche.

⁵⁴ *Ibid.*

⁵⁵ Adresse du site : <https://doranum.fr/>.

2. Les données de la recherche au travers du Plan de Gestion de données

2.1. Les enjeux des données de la recherche

Avec l'intérêt pour les données de la recherche et leur mise en avant dans les mouvements d'*open access* puis d'*open science*, de multiples enjeux sont apparus. Si l'on reprend l'inventaire qu'en fait Alexandre Serres de l'Université de Rennes 2⁵⁶, aux enjeux épistémologiques et scientifiques originels, viennent s'ajouter des enjeux sociétaux, techniques, juridiques, économiques, professionnels et de formation, de la connaissance des pratiques. Plus importants, et servant ces enjeux, sont ceux de l'ouverture et de la diffusion. En effet, ces enjeux sont primordiaux pour la science ouverte, puisqu'il s'agit de donner les moyens de partager des données. Toutefois, il ne faut pas non plus oublier que le partage des données et la science ouverte doivent servir l'intégrité scientifique. Et ces enjeux impliquent l'ensemble des acteurs de la recherche⁵⁷, c'est-à-dire autant les financeurs que les établissements, les laboratoires et leurs infrastructures et plateformes, et les réseaux disciplinaires.

2.1.1. L'ouverture des données de la recherche

L'ouverture des données de la recherche découle du mouvement de science ouverte vu plus tôt, et est aujourd'hui devenue un élément obligatoire dans le milieu de la recherche. Toutefois, Viviane Clavier et Céline Paganelli expliquent bien que « *dans les faits, le choix des données, des modalités de mise à disposition n'est pas simple et nécessite une connaissance du cadre juridique et réglementaire des projets d'open data* »⁵⁸. Il faut savoir que l'ouverture dépend d'abord de ce que contiennent les données : le droit peut amener à ouvrir certaines données et à en « entrouvrir » d'autres, tout en en gardant fermées certaines⁵⁹. De fait, l'expression « ouvrir des données » recouvre plusieurs réalités, comme le présentent les deux auteurs⁶⁰ : il peut s'agir de

⁵⁶ SERRES, Alexandre (dir.). *Données de la recherche en SHS. Pratiques, représentations et attentes des chercheurs : une enquête à l'Université Rennes 2* [en ligne]. Research Report. Université Rennes 2, 2017. p. 6-7. [Consulté le 29 janvier 2021]. Disponible à l'adresse : <https://hal.archives-ouvertes.fr/hal-01635186>

⁵⁷ FOURNIER, Thierry. Les données de la recherche : définition et enjeux. *op.cit.*

⁵⁸ CLAVIER, Viviane et PAGANELLI, Céline. L'ouverture des données de la recherche dans le cadre d'un projet pluridisciplinaire entre SIC et Informatique : le cas des médias sociaux de santé. *Études de communication* [en ligne]. 2019. N° 52, p.130. Disponible à l'adresse : <https://www.cairn.info/revue-etudes-de-communication-2019-1-page-117.htm>.

⁵⁹ STÉRIN, Anne-Laure et NOÛS, Camille. Ouverture des données de la recherche : les mutations juridiques récentes. *Tracés. Revue de Sciences humaines*. 31 décembre 2019. N° 19, pp. 37-50.

⁶⁰ *Ibid.*

donner accès à certaines personnes, sans qu'elles puissent utiliser les données, ou encore de donner accès à tout le monde mais sans possibilités de réutilisation voire avec une réutilisation soumise à condition, et enfin, cela peut aussi signifier l'autorisation pour toutes les personnes à accéder et à utiliser les données. Or, les modalités d'ouverture établies par le droit dépendent de la catégorie de données, par exemple si elles relèvent du droit d'auteur ou encore des données personnelles. Ce droit diffère donc selon les informations contenues par ces données.

En France, la politique d'ouverture et de partage des données publiques est coordonnée par Etalab, département de la Direction Interministérielle du Numérique (DINUM)⁶¹. Aujourd'hui, les données de la recherche qui sont soumises à ouverture par le droit sont les données publiques, qui ne se trouvent pas couvertes par le droit d'auteur. Ces données doivent donc être accessibles et réutilisables par chacun, sous couvert du respect des règles légales telles que la citation de la référence. L'exception à la diffusion de données publiques, et ainsi à leur ouverture, sont les données personnelles, encadrées et protégées par la mise en place du RGPD en 2018⁶².

Ainsi, l'ouverture des données, au travers des déclarations, directives et lois citées plus haut, suppose une réflexion autour de la gestion des données avant et durant le projet de recherche.

2.1.2. *La gestion des données de la recherche*

La politique de gestion des données est notamment portée par PROGEDO, qui est une Très Grande Infrastructure de Recherche chargée d'impulser et structurer une politique publique des données pour la recherche en sciences sociales⁶³. Pour cette TGIR la gestion des données prend la forme d'une chaîne. Et cette chaîne de gestion des données s'organise de la manière suivante : il s'agit d'abord de définir les « *conditions éthiques et juridiques de production des données* », ensuite de stocker ces données produites puis de les archiver et enfin, de les diffuser en les mettant à disposition de tous les citoyens⁶⁴.

⁶¹ *La gestion des données de recherche à l'IFSTTAR : Guide méthodologique pour répondre aux questions pratiques et juridiques sur la gestion des données* [en ligne]. Avril 2018. p. 6. [Consulté le 11 septembre 2020]. Disponible à l'adresse : https://www.ifsttar.fr/fileadmin/redaction/5_ressources-en-ligne/politiques/politique_data_web.pdf.

⁶² Ibid.

⁶³ Adresse du site : <http://www.progedo.fr/>.

⁶⁴ OLIVEAU, Sébastien. Chaîne de gestion des données. *Progedo* [en ligne]. 24 septembre 2020. [Consulté le 23 janvier 2021]. Disponible à l'adresse : <http://www.progedo.fr/promouvoir/chaine-de-gestion-des-donnees>.

Une enquête a été menée en 2018 par le groupe de recherche CommonData⁶⁵ auprès de chercheurs, enseignants-chercheurs et ingénieurs de recherche actifs issus de laboratoires académiques, toutes disciplines confondues. Elle a été conduite par un questionnaire auto-administré en ligne qui a abouti à 159 réponses complètes permettant l'analyse. Cette enquête fait ressortir que les pratiques de gestion des données du côté des chercheurs apparaissent peu standardisées. Le groupe constate en effet que pour une même recherche, il existe bien souvent plusieurs moyens de collecte utilisés (par exemple : utilisation d'un appareillage, consultation de bases de données, observation sur site, lecture de documentation, consultation d'archives, etc.). Par la même occasion, l'enquête met en exergue une forte méconnaissance du régime juridique des données utilisées au cours de la recherche. Elle montre également des pratiques peu organisées en termes de stockage, puisque bien souvent les chercheurs signalent utiliser des supports individuels pour stocker les données de leur recherche. Enfin, pour les deux tiers des répondants aucun standard, norme ou procédure d'organisation des données n'est utilisé. Toutefois, il semble que les pratiques de gestion varient en fonction des disciplines concernées.

L'équipe de l'Institut français des sciences et technologies des transports, de l'aménagement et des réseaux (IFSTTAR) présente la nécessité, de la part de l'institut, de mettre en place une politique d'établissement pour la gestion des données⁶⁶. En effet, le volume des données à traiter augmente continuellement, en raison des avancées technologiques, ce qui nécessite de structurer la gestion. Mais l'organisation de cette politique semble en réalité dépendre de la volonté de chaque organisme. Pour l'IFSTTAR, « *la politique de gestion des données de la recherche a pour objectifs de permettre aux chercheurs de répondre aux questions pratiques et juridiques que pose la gestion des données de recherche aux différentes phases d'un projet* »⁶⁷. Pour une gestion raisonnée des données de la recherche, le programme de recherche CommonData préconise donc une « *acculturation des chercheurs, et surtout des ingénieurs, aux grandes notions juridiques* ». Et pour ce faire les professionnels de l'information-documentation du Groupe Données du GTSO-

⁶⁵ AMIEL, Philippe, *et al.* Pratiques de gestion des données de la recherche : une nécessaire acculturation des chercheurs aux enjeux de la science ouverte ? Résultats d'une enquête exploratoire dans le bassin montpelliérain (juin 2018)[en ligne]. *Cahiers Droit, Sciences & Technologies*. 11 mai 2020. N° 10, pp. 147-168. [Consulté le 3 septembre 2020]. Disponible à l'adresse : <http://journals.openedition.org/cdst/2061>.

⁶⁶ *La gestion des données de recherche à l'IFSTTAR : Guide méthodologique pour répondre aux questions pratiques et juridiques sur la gestion des données. op.cit.*

⁶⁷ *Ibid.* p. 6.

Couperin montrent l'intérêt croissant d'ouvrir des postes de data manager et d'instaurer des « correspondants données » aux seins des universités⁶⁸.

L'ouverture des données, mais principalement leur gestion, se fait à travers la mise en place d'outils, tels que le Plan de gestion de données, document formel au centre de cette étude.

2.2. Le Plan de Gestion de Données : la formalisation d'un processus

Parmi les outils et les documents qui sont apparus avec le mouvement de la science ouverte et l'intérêt croissant pour l'ouverture des données, notamment les données de la recherche, les plans de gestion de données occupent une place importante. « *Document qui synthétise la description des données de la recherche d'un projet et la manière dont elles seront gérées tout au long du projet, afin, notamment, de préparer leur partage, leur réutilisation et leur pérennisation* »⁶⁹, le plan de gestion de données est devenu un document indispensable pour tout projet sollicitant un financement auprès de l'ANR ou du programme Horizon 2020.

2.2.1. Le plan de gestion de données en détail

Pour aller plus loin que la définition citée précédemment, les plans de gestion des données de recherche (PGD en français, et *Data Management Plan* ou DMP en anglais) sont présentés par Heinz Pampel et Maxi Kindling comme :

*des instruments où sont consignées par écrit les données traitées lors d'un projet de recherche depuis le stade de sa programmation jusqu'à son achèvement. Les DMP sont des documents électroniques qui sont mis à jour tout au long du déroulement des projets de recherche. Ils offrent aux porteurs et partenaires d'un projet un cadre opérationnel pour le traitement des données produites et exploitées pendant toute sa durée. Ces plans de gestion précisent, par exemple, le type et la forme de ces données, leur degré de standardisation (par exemple, les métadonnées utilisées, les formats), des indications sur leur accessibilité et les possibilités de réutilisation, ainsi que sur les procédés employés pour leur conservation à long terme.*⁷⁰

⁶⁸ FÉRET, Romain, et al. Améliorer son projet ANR grâce à la Science Ouverte. *op.cit.*

⁶⁹ *Ibid.* p. 5.

⁷⁰ PAMPEL, Heinz et KINDLING, Maxi. Le potentiel des infrastructures d'information pour les données de la recherche. *Études et communication*, 2019, n° 52, p. 34. Également disponible à l'adresse : <https://www.cairn.info/revue-etudes->

Tout l'intérêt de ces plans est de mieux organiser la gestion de données, qui sont de plus en plus nombreuses, particulièrement avec le développement de la sphère numérique et du mouvement de l'*open data*. Ainsi, le PGD doit suivre les principes FAIR recommandés par la Commission Européenne, et traite aussi bien des données que des métadonnées, ces dernières facilitant notamment l'interopérabilité. Il s'agit d'un document qui doit servir d'aide pour organiser les données issues d'un projet, et les préparer pour le cycle de vie (de leur création à leur archivage et communication). Le plan a donc également pour but d'être utilisé dans le cas de l'ouverture et de la réutilisation des données.

Pour élaborer un PGD, il est nécessaire de suivre plusieurs étapes : créer et collecter des (méta)données, documenter les données, stocker et utiliser les données, partager ces données, et enfin les conserver⁷¹. Et pour ce faire, il peut être utile de bénéficier d'un accompagnement méthodologique de la part des professionnels de l'information. Celui-ci prend notamment la forme de guides ou de modèles de PGD. Cet accompagnement répond à un besoin né avec le développement de l'usage de ces plans de gestion de données et qui fait suite à l'obligation d'y recourir pour demander un financement à certaines structures.

Le plan de gestion de données pourrait aussi avoir une finalité autre que la gestion des données. En effet, Nathalie Reymonet suggère de l'utiliser pour produire un *data paper*, c'est-à-dire « *un article dans une revue à comité de lecture, décrivant les données d'un projet de recherche* »⁷², et qui serait un moyen pour exposer les données.

Il est un document utile pour les financeurs de projets, mais qui sert également aux chercheurs ainsi qu'aux organismes de recherche. Il est notamment devenu obligatoire pour le programme Horizon 2020, et depuis 2019 l'ANR le requiert pour toute demande de financement de projet.

Dans un premier temps, le plan de gestion de donnée était un document de cinq questions prescrit par le programme Horizon 2020 mais à caractère volontaire. Aujourd'hui, le PGD est

[de-communication-2019-1-page-27.htm](#).

⁷¹ SERGI, Gianna. *Le Data Management Plan, DMP ou le Plan de Gestion de données, PGD et OPIDoR* [en ligne]. op.cit. p. 6 et suivantes.

⁷² REYMONET, Nathalie. Améliorer l'exposition des données de la recherche : la publication de *data papers*. [en ligne]. 6 janvier 2017. [Consulté le 30 janvier 2020]. Disponible à l'adresse : https://archivesic.ccsd.cnrs.fr/sic_01427978.

devenu un document nécessaire pour tout dépôt de demande de financement auprès de la commission européenne, et il doit également répondre aux critères FAIR que celle-ci a mis en place. Toutefois, Joachim Schöpfel souligne qu'aucune sanction n'existe pour le moment contre ceux qui ne suivent pas ces directives⁷³. Au niveau national, l'ANR a également fait le choix de rendre ces PGD obligatoires pour chaque demande de financement, dans l'optique de promouvoir la science ouverte et de l'encourager auprès de la communauté scientifique. Ce PGD doit donc être fourni dans les six mois suivant le début du projet. Puis, il doit être mis à jour à la moitié du parcours dans le cas de projets supérieurs à trente mois, et une version finale est nécessaire pour clôturer le projet⁷⁴.

La qualité des intervenants potentiels d'un PGD est très diverse : ils peuvent être chercheurs, ingénieur-projets, informaticiens ou encore professionnels de l'information scientifique et technique. Mais d'autres personnes peuvent également être concernées, puisque « *l'accession de ce document au statut de livrable récurrent des projets de recherche peut donner l'opportunité d'y familiariser un nombre croissant d'acteurs de la recherche, y compris [...] les doctorants ou les directeurs d'unités [...]* »⁷⁵. Il nous faut donc présenter brièvement le rôle de chacun des principaux acteurs, à l'instar d'Aurore Cartier, Magalie Moysan et Nathalie Reymonet lors de la journée d'étude sur les données de la recherche de l'ADBS⁷⁶. Pour commencer, le chercheur est la personne en charge de la coordination du projet ainsi que le responsable de la gestion des données. C'est donc à lui qu'incombe la création du PGD et sa mise en œuvre, mais également la description des données. L'ingénieur-projet intervient quant à lui dès le début du projet, afin d'aider le chercheur sur la partie des informations concernant le projet. En ce qui concerne les informaticiens, ceux-ci sont présents pour aider les chercheurs à propos des questions de stockage et de sécurité des données. Enfin, les référents IST, qui peuvent être documentalistes, archivistes ou bibliothécaires, sont chargés d'accompagner les chercheurs dans la gestion de leurs données (principalement sur les parties de description et de partage de celles-ci), en fournissant notamment des outils pour les

⁷³ SCHÖPFEL, Joachim. Hors norme? Une approche normative des données de la recherche. *op.cit.*

⁷⁴ FÉRET, Romain, *et al.* Améliorer son projet ANR grâce à la Science Ouverte. *op.cit.*

⁷⁵ ANCELIN-FABRE, Justine. Travailler à distance sur la médiatisation des plans de gestion de données : retours d'expérience. *UrfistInfo* [en ligne]. 24 août 2020. [Consulté le 24 septembre 2020]. Disponible à l'adresse : <https://urfistinfo.hypotheses.org/3555>.

⁷⁶ CARTIER, Aurore, MOYSAN, Magalie et REYMONET, Nathalie. Construire des outils pour la gestion des données de la recherche dans une communauté d'universités. In : *Journée sur les données de la recherche* [en ligne]. Paris, France : ADBS, janvier 2015. [Consulté le 16 décembre 2020]. Disponible à l'adresse : <https://hal.archives-ouvertes.fr/hal-01138663>.

guider. Les professionnels de l'information-documentation cités ci-dessus présentent d'ailleurs leur méthodologie d'accompagnement de la manière suivante : « *notre démarche a donc été de fournir aux chercheurs et ingénieurs-projets de nos établissements des outils pratiques et harmonisés, qui soient susceptibles de répondre aux exigences des organismes financeurs, au-delà de la seule CE, et de constituer une base de coopération avec les organismes de recherche partenaires.* ».

Pour aider à la rédaction des plans de gestion de données, les professionnels de l'information et de la documentation mettent au point des modèles qu'ils rendent accessibles à la communauté scientifique au travers de plateformes dédiées⁷⁷. Cela peut passer entre autres par des modèles de plan de dissémination des résultats, des plans de formation pour les ingénieurs, mais également des modèles de plan de gestion de données, recensés notamment sur la plateforme *DMP OPIDoR*.

2.2.2. Les modèles de plans de gestion de données sur DMP OPIDoR

Le site *DMP OPIDoR* est un outil d'aide à la rédaction de plans de gestion de données, mis en ligne par le CNRS et géré par Inist-CNRS. Ce site web, mis en ligne en 2016, a été créé spécifiquement pour la communauté scientifique française. Depuis sa création, DMP OPIDoR est en constante évolution⁷⁸, afin de s'adapter au mieux aux besoins des chercheurs. Justine Ancelin-Fabre rapporte notamment une volonté d'éviter la double saisie des informations, qui devrait prochainement se traduire par l'intégration de référentiels structurés⁷⁹. Sans avoir besoin de se connecter, le site permet l'accès à des PGD rendus publics par leurs propriétaires, des modèles de PGD, des exemples, ainsi que des ressources.

À la date de cette étude, on trouve sur *DMP OPIDoR* 33 modèles de plan de gestion de données, avec bien souvent une version française et une version anglaise du même modèle. Si l'on s'appuie sur le rapport du Collège de France fait en 2018, le site présentait à ce moment seize modèles. Le doublement du nombre de modèles disponibles démontre une volonté certaine de la part des organismes, qu'ils soient financeurs ou producteurs, d'orienter les chercheurs dans la pratique de gestion des données de la recherche. De manière générale, les modèles sont

⁷⁷ cf. *supra* 1. 3. Présentation des structures d'appui qui accompagnent le mouvement et participent à l'acculturation de la communauté universitaire. p. 25-27.

⁷⁸ Sur les évolutions de la plateforme voir la présentation du service Partage des données de Inist-CNRS : https://f.hypotheses.org/wp-content/blogs.dir/78/files/2020/07/20200625_Inist_Nouveautes_OPIDoR.pdf.

⁷⁹ ANCELIN-FABRE, Justine. Travailler à distance sur la médiatisation des plans de gestion de données . *op.cit.*

composés de six points qui varient selon l'organisme faisant la recommandation. Par exemple, le modèle français proposé par l'ANR s'organise de la manière suivante : un premier point pour la description des données et collecte ou réutilisation de données existantes, le second à propos de la documentation et qualité des données, un troisième pour faire état du stockage et des sauvegardes pendant le processus de recherche, le quatrième se rapportant aux exigences légales et éthiques, codes de conduites, le cinquième s'intéressant au partage des données et la conservation à long terme, et enfin le dernier pour regrouper les responsabilités et ressources en matière de gestion des données. Tandis que le modèle « Horizon 2020 FAIR DMP (français) » demande de renseigner un résumé descriptif des données, les données FAIR, l'allocation de ressources, la sécurité des données, les aspects éthiques et laisse un dernier champ « autres ».

La formalisation du processus d'ouverture des données de la recherche se met en place par l'usage progressif du PGD, devenu un document officiel auprès des financeurs. Et le développement de l'outil *DMP OPIDoR*, qui permet à la fois d'avoir accès à un répertoire actualisé de modèles de PGD mais aussi de créer ses propres PGD à partir de l'un des modèles proposés, vise à aider à la prise en main du document. DMP OPIDoR a notamment pour objectif de faciliter l'usage et l'appropriation de ces plans par les chercheurs ayant l'obligation d'y recourir.

2.3. L'appropriation des plans de gestion de données par les chercheurs

Le terme d'appropriation s'inscrit dans le champ de la sociologie, et particulièrement de la sociologie des usages. Afin d'explorer la manière dont les chercheurs s'approprient ce processus d'élaboration du PGD, il nous faut prendre la mesure de ce contexte. Pour cela, nous puiserons dans les concepts de la sociologie des usages pour interroger ensuite la pratique des professionnels ou chercheurs.

En France, la littérature place Michel de Certeau comme un pionnier de l'approche des usages, lui qui s'est intéressé à la « *manière de faire* » des individus et qui leur reconnaît une capacité d'autonomie. Dans l'article « *Retour critique sur la sociologie des usages* »⁸⁰, Josiane Jouët, professeure émérite en SIC (Science de l'Information et de la Communication) de l'université Paris-II, balaye rapidement les évolutions qu'a connu la sociologie des usages depuis ses débuts,

⁸⁰ JOUËT, Josiane. Retour critique sur la sociologie des usages. *Réseaux. Communication - Technologie - Société*. 2000. Vol. 18, n° 100, pp. 487-521. DOI [10.3406/reso.2000.2235](https://doi.org/10.3406/reso.2000.2235).

principalement concentrée sur les TIC (Technologies de l'Information et de la Communication). Elle y présente notamment la notion d'appropriation, qui nous intéresse particulièrement dans le cadre de cette étude. La sociologie des usages désigne une préoccupation pour un problème relevant de trois disciplines de sociologie différentes : la sociologie de la technique, de la communication et des modes de vie⁸¹. Et plusieurs approches en découlent donc : la sociologie de la diffusion, la sociologie de l'innovation et la sociologie de l'appropriation.

Le terme « usage » est souvent assimilé à celui d'emploi, d'utilisation, de pratique ou encore d'assimilation selon Florence Millerand, ce qui le rend difficile à cerner. Pour cela, elle reprend la distinction faite par J. Jouët entre les notions d'usage et de pratique : « *l'usage est [...] plus restrictif et renvoie à la simple utilisation tandis que la pratique est une notion plus élaborée qui recouvre non seulement l'emploi des techniques (l'usage) mais les comportements, les attitudes et les représentations des individus qui se rapportent directement ou indirectement à l'outil* »⁸². Et elle en conclut que l'usage renvoie à une utilisation faite d'un média ou d'une technologie et que l'on remarque au travers de pratiques. Dans le cas qui nous intéresse, l'usage renverrait donc au simple fait qu'un chercheur recourt à un PGD, qu'il utilise le document. À savoir que « *l'intensité de l'usage est d'ailleurs souvent un indicateur d'appropriation forte qui cristallise des enjeux d'identité* »⁸³ selon J. Jouët. Tandis que la pratique, dans le cadre des PGD, serait ici ce que les chercheurs font avec, c'est-à-dire surtout de quelle manière ils les utilisent et dans quel contexte. Pour ce qui est de l'appropriation, Pierre Chambat considère que cela consiste principalement à « *analyser comment se constituent les usages différenciés selon les groupes sociaux et le sens qu'ils revêtent pour ceux-ci* »⁸⁴. En revanche, pour Philippe Breton et Serge Proulx, le terme « appropriation » renvoie plutôt à une « orientation » prise par des travaux de recherche⁸⁵. Serge Proulx considère d'ailleurs que quatre conditions sont nécessaires à la réalisation de

⁸¹ CHAMBAT, Pierre. Usages des technologies de l'information et de la communication (TIC) : évolution des problématiques. *TIS*. 1994. Vol. 6, n° 3, p. 254.

⁸² MILLERAND, Florence. Usages des NTIC : les approches de la diffusion, de l'innovation et de l'appropriation (1ère partie). *Composite*. 1998. Vol. 2, n° 1, p. 4. Disponible à l'adresse : <http://www.composite.org/index.php/revue/article/view/21>.

⁸³ JOUËT, Josiane. Retour critique sur la sociologie des usages. *op.cit.* p. 504.

⁸⁴ CHAMBAT, Pierre. Usages des technologies de l'information et de la communication (TIC) : évolution des problématiques. *op.cit.* p. 258.

⁸⁵ BRETON, Philippe et PROULX, Serge. 11/ Usages des technologies de l'information et de la communication. In : *L'explosion de la communication. Introduction aux théories et aux pratiques de la communication* [en ligne]. La Découverte. Paris : La Découverte, 2012. p. 282. Repères. [Consulté le 31 janvier 2021]. Disponible à l'adresse : <https://www.cairn.info/l-explosion-de-la-communication—9782707173829-page-263.htm>.

l'appropriation : une maîtrise technique et cognitive, une intégration de l'objet dans la pratique quotidienne, un usage répété de la technologie qui ouvre vers de la création, et enfin une représentation des usagers dans l'établissement de politiques publiques ainsi qu'une prise en compte dans les processus d'innovation⁸⁶. Cette description très complète s'appliquerait assez bien aux questionnements autour du PGD et des pratiques des chercheurs que nous nous posons ici. En effet, si notre objet d'étude n'est pas un outil technique en tant que tel, la réflexion autour de la notion d'appropriation s'avère assez riche dans la mesure où elle permet de s'interroger sur l'intensité avec laquelle les PGD sont utilisés par les chercheurs dans le cadre de leurs projets. On comprend ici que tant qu'il y a un usage régulier et une pratique régulière des PGD par les chercheurs, il est possible de parler d'appropriation. En réalité, il faut donc plutôt s'intéresser au degré d'appropriation de chaque chercheur pour en tirer des informations plus précises. Finalement, ces travaux autour du triptyque « usages, pratiques et appropriation » montrent que les trois notions sont liées entret-elles. Pour étudier l'appropriation des PGD par les chercheurs il faut donc également penser aux usages et aux pratiques.

La question de l'appropriation des plans de gestion de données par les chercheurs est une chose, mais il ne faut pas non plus oublier que les professionnels de l'IST jouent également un rôle important auprès de ces derniers. En effet, il est intéressant d'appréhender la gestion des données de la recherche des deux points de vue : celui des chercheurs, et celui des professionnels de l'information.

Comme mentionné plus tôt, le programme H2020 et l'ANR requièrent maintenant la mise en place de plans de gestion de données pour faire une demande de financement. On peut donc imaginer que les chercheurs font usage, certes parfois de manière contrainte, de ces plans de gestion de données au moment d'établir leur demande de projet. Toutefois, l'idée d'appropriation reste à déterminer car, si l'on se réfère aux conditions définies par S. Proulx, cela sous-entend notamment d'avoir une pratique quotidienne de l'objet. Dans le cas de notre étude, il serait plus pertinent de s'intéresser aux expériences antérieures, car cela permettrait de soulever la question de la familiarité plutôt que la fréquentation. Néanmoins, on l'a vu précédemment, aucune

⁸⁶ PROULX, Serge. Penser les usages des TIC aujourd'hui : enjeux, modèles, tendances. In : *Enjeux et usages des TIC : aspects sociaux et culturels* [en ligne]. Presses universitaires de Bordeaux. Bordeaux, 2005. p. 11. [Consulté le 27 janvier 2021]. Disponible à l'adresse : <https://sergeproulx.uqam.ca/wp-content/uploads/2010/12/2005-proulx-penser-les-usa-43.pdf>

législation n'existe pour encadrer ce dispositif et les chercheurs sont donc en réalité libres d'utiliser ou non ces PGD. Éventuellement, la crainte de ne pas recevoir de financement peut obliger les chercheurs à recourir à un PGD. D'ailleurs, les normes de gestion qui encadrent les données de la recherche peuvent être perçues par les chercheurs comme contraignantes : c'est notamment ce qui ressort de l'étude menée par V. Rebouillat, laissant apparaître également l'idée que chacun fait plus ou moins comme cela l'arrange⁸⁷. La chercheuse va jusqu'à parler de « non-appropriation des normes légales » par les chercheurs⁸⁸. Pour sa part, Ben Kaden présente la publication des données de la recherche comme une « *charge de travail supplémentaire [...] sans aucune récompense pour la carrière scientifique* »⁸⁹. Et ce ressenti peut également se retrouver dans le cas de l'obligation à recourir à des plans de gestion de données, souvent perçus par les chercheurs comme une « *simple contrainte administrative* » s'ajoutant à leur travail⁹⁰.

Cependant, les professionnels de l'information et de la documentation peuvent avoir un rôle à jouer dans l'utilisation et l'appropriation des plans de gestion de données, notamment en mettant en place un accompagnement des chercheurs dans la rédaction de ces plans de gestion de données⁹¹. Ces professionnels de l'information se placent en effet au premier plan pour sensibiliser les chercheurs à l'usage des PGD, puisqu'ils sont les plus à même de comprendre les enjeux autour du partage de données et de les communiquer à la communauté scientifique. Les questions autour de l'accompagnement et de la sensibilisation des chercheurs ont notamment donné lieu à la création du groupe de travail « GT-DMP » présenté par la co-responsable de l'URFIST de Paris⁹². Enfin, V. Rebouillat et C. Chartron ont mené une enquête auprès des chercheurs de l'université de Strasbourg afin de savoir si les services développés et mis en avant par les professionnels de l'IST étaient utilisés par la communauté scientifique⁹³. Il en ressort que peu de services sont réellement

⁸⁷ REBOUILLAT, Violaine. *Ouverture des données de la recherche : de la vision politique aux pratiques des chercheurs*. *op.cit.* p. 195-196.

⁸⁸ *Ibid.* p. 196.

⁸⁹ KADEN, Ben. Pourquoi les données de recherche ne sont-elles pas publiées ? *op.cit.* p. 144.

⁹⁰ AMIEL, Philippe, *et al.* Pratiques de gestion des données de la recherche : une nécessaire acculturation des chercheurs aux enjeux de la science ouverte ? *op.cit.*

⁹¹ DUPRAT, Julie. *Les données de la recherche à l'Université Bordeaux Montaigne* [en ligne]. Research Report. Université Bordeaux Montaigne, 2019. p. 29. [Consulté le 1 février 2021]. Disponible à l'adresse : <https://hal.archives-ouvertes.fr/hal-02020141>.

⁹² ANCELIN-FABRE, Justine. Travailler à distance sur la médiatisation des plans de gestion de données . *op.cit.*

⁹³ REBOUILLAT, Violaine et CHARTRON, Ghislaine. Services de gestion et de partage des données de recherche : ce qu'en pensent les chercheurs. In : *12ème Colloque international d'ISKO-France : Données et mégadonnées ouvertes en SHS : de nouveaux enjeux pour l'état et l'organisation des connaissances ?* [en ligne]. Montpellier, France, octobre 2019. p. 2.

utilisés, ce qui peut signifier que la communauté n'est pas encore totalement familiarisée avec le concept des données de la recherche et les enjeux de gestion. Mais, cela peut également montrer que les professionnels de l'information ne parviennent pas encore à avoir autant d'influence sur les chercheurs que l'édition scientifique⁹⁴.

Pour conclure, la science ouverte est donc un mouvement relativement récent, mais qui s'entoure progressivement d'un cadre juridique ainsi que de bonnes pratiques promouvant l'ouverture des données et en particulier des données de la recherche. On a vu notamment la proposition des principes FAIR par la Commission Européenne, principes qui prennent aujourd'hui valeur de normes aux yeux de la communauté scientifique. Ce sont également divers établissements qui se sont organisés autour du mouvement de l'ouverture des données, en proposant via leurs interfaces du contenu pédagogique, des formations ou de l'accompagnement. Ces ressources numériques pourraient constituer pour les professionnels de l'information et de la communication un outil de médiation visant à amener une appropriation du sujet, et en l'occurrence de l'usage du plan de gestion de données, chez les membres de la communauté scientifique. En effet, la mise en place des PGD, d'abord facultatifs puis obligatoires, a entraîné l'apparition de nouvelles pratiques au sein de la communauté scientifique, que les professionnels de l'information pourraient tenter d'accompagner afin de favoriser l'appropriation des chercheurs. Tout cela nous amène à nous interroger sur le rôle de ces PGD dans la diffusion des enjeux de ces données de la recherche auprès des chercheurs.

[Consulté le 26 mars 2020]. Disponible à l'adresse : <https://hal.archives-ouvertes.fr/hal-02307085>.

⁹⁴ *Ibid.* p. 10.

Chapitre 2 – Méthodologie

Nous allons maintenant nous intéresser à la mise en place de la méthodologie visant à répondre aux hypothèses émises. Pour rappel, nous nous sommes interrogée sur le rôle que les PGD peuvent avoir dans la diffusion des enjeux d'ouverture des données de la recherche auprès des chercheurs, maintenant qu'ils sont largement obligatoires. Dans un premier temps, nous avons avancé l'idée que le PGD aurait une fonction qui va au-delà de son rôle initial de cadrage opérationnel. De plus, nous présumons que la perception du PGD pourrait également varier selon le caractère volontaire ou non de sa rédaction. Dans un second temps, nous supposons que l'obtention d'une meilleure compréhension des enjeux de l'ouverture des données de la recherche par la rédaction d'un PGD permettrait aux chercheurs d'adhérer davantage à l'ouverture des données. Enfin, le troisième point soulevé est le rôle de médiation que joueraient les professionnels de l'information-communication auprès des chercheurs. Ce serait notamment l'expertise info-documentaire que demanderaient certaines rubriques qui pourrait entraîner une collaboration des deux corps, et permettrait en même temps d'élargir les horizons de chacun autour des données et des questions de l'ouverture.

1. Présentation du terrain

Pour tenter de répondre aux questions que pose le sujet étudié, nous avons fait le choix d'analyser un corpus de PGD disponibles et accessibles sur la plateforme DMP Opidor. En parallèle, nous avons également procédé à des entretiens avec des chercheurs ayant rédigés eux-mêmes des PGD. Un champ ensuite élargi aux doctorants et ingénieurs de recherche. L'étude de ces deux corpus nous a semblé nécessaire pour obtenir une analyse complète. En effet, examiner un corpus de PGD permet de se rendre compte de la manière dont les documents sont remplis, tandis que les entretiens vont davantage servir à recenser les opinions des personnes autour du PGD, de sa rédaction et plus largement des données de la recherche.

Lors des recherches pour rédiger l'état de l'art, nous avons eu l'occasion de consulter divers articles mentionnant ou renvoyant à DMP Opidor, ce qui nous a amené à visiter régulièrement cette plateforme. Ce site étant relativement bien fait et semblant plébiscité par la

communauté scientifique française, il nous a semblé intéressant de le sélectionner pour fonder notre étude. DMP Opidor est intéressant, car il recense de nombreux modèles de PGD et propose également des PGD rédigés et mis à la disposition de tous par leur auteur. Notre questionnement autour des données de la recherche nous a amené à nous tourner vers la recherche publique, et donc des chercheurs affiliés aux organismes publics tels que les universités ou encore le CNRS. Pour cela, il nous a semblé pertinent de débiter notre étude en prenant comme point de départ la plateforme DMP Opidor. En effet, cette plateforme est gérée et hébergée par l'Inist, unité du CNRS, et est destinée principalement à des chercheurs de l'enseignement supérieur et de la recherche. Pour une question de cohésion de corpus, nous avons donc fait le choix de partir de ce site pour solliciter différentes personnes pour les entretiens nécessaires à l'étude.

Pour étudier ces deux corpus, nous avons donc mis en place une grille d'entretien ainsi qu'une grille d'analyse pour le corpus. Au vu des ajustements qui ont dû être fait vis-à-vis des entretiens, nous avons décidé de décliner la grille en trois. Cela a permis d'avoir une grille adaptée pour chaque catégorie de personnes interrogées. Par la suite nous allons donc présenter chacune des grilles construites, ainsi que le contexte dans lequel elles ont été utilisées.

2. Le corpus

Puisque l'étude porte sur les PGD, il nous a semblé utile d'étudier un corpus composé d'un ensemble de ces Plans de gestion des données. Ce corpus a été composé sur la base de choix qui ont été faits et que nous allons expliquer par la suite. Ensuite, la grille pensée pour aider à l'analyse de ce corpus sera présentée.

2.1. La sélection des PGD

Pour cela nous avons pu consulter de nombreux modèles de PGD ainsi que des PGD remplis et déposés sur la plateforme DMP Opidor. Nous en avons donc tiré le corpus nécessaire à notre étude, c'est-à-dire différents modèles et PGD rédigés. Et afin de nous assurer d'avoir le temps de procéder à une étude sérieuse des PGD, nous avons choisi de sélectionner parmi les PGD déposés sur DMP Opidor uniquement les modèles français. Mais, avant d'étudier ce corpus, une analyse rapide des modèles proposés sur DMP Opidor a été faite. L'idée était d'étudier la structure et les différents points pouvant être demandés par les organismes ou les financeurs. Pour ce corpus

également nous avons préféré nous centrer uniquement sur les modèles en français. Cette première analyse, avec ce double corpus, avait pour but de mieux comprendre les PGD du corpus mais également de pouvoir ensuite en discuter avec les chercheurs interrogés. Enfin, cette première approche des PGD nous permettait de nous rendre compte des ressemblances ou dissemblances pouvant exister entre les différents modèles proposés pour les projets, et d'avoir déjà une idée de l'articulation entre ce qui est demandé et ce qui est fait.

2.2. La grille d'analyse

La grille d'analyse⁹⁵ a été composée afin de cibler les informations qui pourraient permettre de répondre aux trois hypothèses posées. Divisée en deux grandes parties, il s'agit dans un premier temps d'analyser le corpus dans son ensemble et dans un second temps de mener une analyse davantage en profondeur, c'est-à-dire portant sur le contenu du corpus.

2.2.1. *Un état des lieux*

La première partie de l'analyse du corpus se propose de faire un état des lieux chiffré. L'objectif est de montrer concrètement la part du corpus étudié par rapport à l'ensemble des modèles et PGD déposés. Cela permet de situer l'échantillon, de le délimiter et d'avoir une idée de sa représentativité. Cette première partie donne à comparer le corpus entier des modèles et le corpus restreint de PGD déposés sélectionné pour notre étude.

Ensuite il s'agit d'analyser la répartition des PGD du corpus. Le premier tableau permet de rendre compte du nombre de modèles utilisés, du type de modèle semblant être préféré par les chercheurs. Éventuellement, il peut aussi permettre de relever une préférence de certaines structures pour un modèle.

2.2.2. *La structure des PGD*

Le second tableau a été construit afin de voir si l'on peut retrouver des points, des portions similaires entre les différents modèles du corpus. L'idée est de faire un constat rapide sur la construction des modèles et d'en comparer plusieurs. Aussi, ce tableau permet de déterminer le niveau de détail demandé, et fonctionne avec les deux tableaux suivants. L'un de ceux-ci a été

⁹⁵ cf. Annexe D. Grilles d'analyse.

pensé afin de regrouper les points qui semblent revenir le plus souvent dans les modèles et ceux qui sont plutôt rares, dans le but de faire ressortir plus facilement les informations recueillies grâce au tableau précédent. L'autre tableau complète le premier, puisqu'il précise la structure du PGD, reprend les thèmes traités et précise les contenus attendus pour chaque modèle du corpus. Cela montre le niveau de détail attendu ainsi que le niveau de description demandé par le créateur du PGD (financeur ou établissement).

2.2.3. *Le contenu du corpus*

Initialement, un tableau de comparaison avait été pensé pour montrer de possibles liens entre le modèle choisi, l'organisme finançant le projet, la discipline et la fonction de l'auteur. Au final, il a été décidé de laisser ce tableau de côté car non seulement sa pertinence ne nous semblait pas évidente, mais aussi car certains points étaient plutôt difficiles à compléter, notamment la discipline et la fonction de l'auteur, qui ne sont pas mentionnés dans le PGD rédigé.

Le dernier tableau construit pour l'analyse du corpus est constitué de différents points qui devraient permettre d'indiquer comment sont remplis les PGD. Découpé par modèle, il s'agit de traiter chaque PGD sélectionné pour l'étude et de relever les nombres de pages, si les informations semblent être mises là où elles sont attendues, d'indiquer un niveau de complétion et éventuellement des spécificités relevées lors de la lecture du PGD. L'organisation du tableau permet de mieux comparer les PGD faits sur un même modèle. Ce tableau a été pensé pour essayer de déterminer si les personnes rédigeant le PGD comprennent ce qui est attendu. Il a également été construit pour analyser le taux de complétion des PGD sur l'ensemble de corpus. On cherche par là à montrer l'intérêt des chercheurs, ou éventuellement des rédacteurs, et peut-être leur implication pour les données.

Afin de compléter les données que peut nous apporter l'analyse d'un corpus de PGD, nous avons choisi d'interroger directement les personnes concernées, c'est-à-dire les auteurs de ces PGD. Après avoir présenté le corpus et la grille d'analyse qui permettra de l'étudier, nous allons maintenant expliquer les choix relatifs à la mise en œuvre des entretiens et l'organisation de la grille pour guider ces entretiens.

3. Les entretiens

Pour la deuxième partie de l'analyse, nous avons donc voulu mener des entretiens avec les rédacteurs des PGD. Notre idée était dans un premier temps de chercher à interroger des chercheurs ayant déjà rédigé eux-mêmes un PGD. Dans cette dernière partie nous présenterons donc la préparation de l'échantillon, la mise en œuvre des entretiens et la grille qui a été établie pour mener ces interviews.

3.1. La construction de l'échantillon

Après avoir envisagé différents terrains possibles pour l'expérimentation, nous avons initié nos recherches à partir des PGD de DMP Opidor, afin de tenter de contacter des chercheurs. La première étape a consisté à contacter l'équipe d'Opidor afin de savoir s'ils avaient une liste ou des contacts spécifiques à nous recommander. La réponse qui a suivi nous a finalement enjoint à consulter les PGD publiés sur la plateforme, et nous a fourni le contact d'un ingénieur de la MSHS de Poitiers qui avait justement organisé récemment un séminaire sur les PGD. La seconde étape a donc consisté à contacter cet ingénieur de la MSHS. Lors de l'échange, lui aussi a recommandé de contacter directement les personnes ayant déposé leur PGD sur DMP Opidor, afin d'avoir davantage de chance de trouver des chercheurs. De plus, après notre entretien, M. Vera Cruz a également fourni le contact de plusieurs personnes qui étaient susceptibles de répondre à notre demande selon lui. Suite à cet échange, il est apparu assez vite qu'interroger uniquement des chercheurs ayant rédigé eux-mêmes leur PGD risquait d'être ambitieux, au vu de la pratique encore plutôt balbutiante, et qu'il risquait d'être difficile de trouver suffisamment de monde pour obtenir un corpus raisonnable.

Dans un second temps, nous avons donc décidé d'élargir le champ des personnes à interroger, en ajoutant aux chercheurs les doctorants, post-doctorants et ingénieurs de recherche. Ce nouveau panel présentait l'avantage d'avoir une plus grande variété de points de vue. Il nous a donc semblé intéressant de les inclure dans les personnes à interroger, afin d'avoir une approche plus complète de la tâche de rédaction du PGD notamment. Bien qu'il aurait été également intéressant d'inclure dans notre expérimentation des chercheurs n'ayant pas rédigé de PGD, afin peut-être de nuancer les propos recueillis lors des entretiens avec ceux ayant rédigé des PGD, nous avons décidé de ne pas les solliciter. Cela a principalement été décidé pour éviter d'avoir trop de

bruits parmi les résultats obtenus et ne pas se retrouver ensuite pris par le temps au moment d'analyser. D'autant que nos questionnements portaient en priorité sur l'utilisation du PGD par les chercheurs. Après une première prise de contact avec les personnes sélectionnées sur DMP Opidor et celles fournies par l'ingénieur contacté, le panel s'est progressivement étoffé. Il s'est finalement constitué au fil de l'eau, en suivant les recommandations des uns et des autres, et en sélectionnant de nouveau quelques personnes pas encore contactées sur DMP Opidor.

Le résultat de la prospection a abouti à un total de 8 entretiens, répartis de la manière suivante : trois ingénieurs, deux chercheurs et trois doctorants ou post-doctorants. En plus d'une diversité de postes, ce panel nous donne également une disparité de parcours et un mélange de disciplines qui ne peuvent qu'enrichir notre échantillon.

3.2. Le déroulement

L'ensemble des entretiens se sont déroulés du 4 mai au 4 juin. En dehors du dernier, qui a eu lieu en présentiel, tous les autres ont été passés en visioconférence. Cela s'est fait sous cette forme à cause du contexte sanitaire particulier de cette année, mais également en raison du stage qui permettait difficilement de se déplacer. Et pour l'organisation de ces réunions, nous avons utilisée la plateforme Zoom. Les entretiens ont duré entre 22 minutes pour le plus court et 1h13 pour le plus long. Mais la majorité des entretiens ont duré entre 40 et 45 minutes. Les échanges ont été enregistrés, avec l'accord des personnes interrogées. Toutefois pour notre usage, il a été décidé de procéder à une retranscription partielle. Pour cela nous avons simplement complété la grille des entretiens avec les réponses obtenues, soit en les reformulant soit sous forme de *verbatim*. Par ailleurs, nous avons fait le choix d'anonymiser les personnes, en indiquant dans les retranscriptions le statut de la personne suivi d'un numéro. Pour cela, lorsque nous nous référons, lors de l'analyse, aux propos tenus lors des entretiens nous indiquerons « Le chercheur 1, l'ingénieur 3, etc. ».

3.3. La grille d'entretien

Les entretiens devaient être réalisés auprès de chercheurs, cible première de notre étude. Mais comme nous l'avons déjà dit, par manque de chercheurs ayant réellement été impliqués dans la production de PGD, nous avons dû nous résoudre à assouplir notre recherche. Ce qui nous a

amené à inclure dans les personnes sollicitées les doctorants et post-doctorants, ainsi que les ingénieurs de recherche. Ce choix a finalement fait sens, puisqu'il permettait de varier les visions, les opinions, en incluant des personnes du domaine de la recherche ayant également une expérience de rédaction de PGD.

Le guide pour mener à bien ces entretiens a pris la forme d'une grille d'entretien, disponible en annexe de ce travail⁹⁶. Cette grille a été construite en gardant à l'esprit les 3 hypothèses que nous nous sommes posées au début de ce travail. Elle a plutôt été pensée pour un entretien directif, avec des questions précises déjà définies pour chacun des points qui nous intéressent. Ainsi, les questions sont regroupées en 4 thématiques, qui correspondent plus ou moins aux hypothèses soulevées. Et pour chacune des questions, une précision de ce que l'on souhaite savoir est donnée. Cela permet éventuellement d'ajuster la question au moment de l'entretien, si jamais l'on se rend compte que la réponse ne correspond pas à ce que l'on cherche à savoir. De plus, cette colonne est aussi un moyen de garder sous les yeux le lien avec les hypothèses.

La grille commence par le parcours de la personne : cette approche permet d'ouvrir la discussion et éventuellement de rendre compte de l'expérience antérieure ou du domaine d'étude. Ce qui devait permettre de voir si un lien pourrait éventuellement exister entre le fait d'avoir rédigé un PGD et le parcours ou le domaine d'étude de la personne. Mais elle permet également de rendre compte des différents profils interrogés.

Ensuite, la grille d'entretien comprend trois grandes parties et a été construite dans l'objectif de répondre à nos différentes hypothèses. La première s'intéresse au Plan de gestion de données de manière générale. Elle vise à déterminer la raison qui a entraîné la personne à rédiger un PGD, ainsi que ce que cette expérience a pu laisser comme impression de la finalité du PGD.

La deuxième partie se concentre quant à elle sur la réalisation du PGD. Il s'agit donc ici d'identifier d'abord les différentes démarches de la construction du plan, puis de déterminer s'il y a eu des difficultés, et le cas échéant comment la personne y a réagi. Enfin, il s'agit également de vérifier si certains points demandent une connaissance particulière. La grille destinée aux ingénieurs inclut pour sa part une autre question, afin de savoir s'ils ont été sollicités pour un PGD, et si cela est le cas, dans quel but.

⁹⁶ cf. Annexes A. Grille d'entretiens : chercheurs, B. Grille d'entretiens : doctorants et C. Grille d'entretiens : ingénieurs.

Dans la troisième partie de la grille, les objectifs restent les mêmes pour tous, mais la formulation des questions a été modifiée selon que l'on s'adressait à un chercheur, à un doctorant ou à un ingénieur. Cette partie cherche à comprendre l'opinion et la perception des enjeux autour des données de la recherche et des PGD. Dans un premier temps, il s'agit donc d'identifier les enjeux liés à l'ouverture des données de la recherche pour la personne interrogée. Ensuite, nous voulons vérifier si le PGD peut permettre une meilleure compréhension de ces enjeux. Finalement, nous nous intéressons à la position du chercheur vis-à-vis du PGD, mais nous cherchons également à connaître celle des futurs-chercheurs et celle des ingénieurs. Ce dernier point est particulièrement intéressant du fait des différents parcours et spécialités des interlocuteurs.

Les différentes grilles ont été conçues comme des outils dont le but est de faciliter le recueil des données nécessaires à l'analyse, qu'il s'agisse des entretiens ou du corpus de PGD. Après avoir présenté la méthodologie définie dans le cadre de cette étude, nous allons nous intéresser aux données recueillies et à l'analyse de celles-ci.

Chapitre 3 – Les plans de gestion de données et leurs usages : une analyse des observations

Après avoir présenté le corpus choisi pour l'étude ainsi que les outils construits pour bâtir l'analyse, nous allons consacrer ce chapitre à l'analyse des résultats obtenus lors des entretiens et de l'examen du corpus de PGD. Ainsi, nous tenterons de déterminer si le PGD joue effectivement un rôle auprès de la communauté de chercheurs dans la diffusion des enjeux des données de la recherche. Pour cela, nous allons d'abord observer les points nécessaires à la mise en œuvre d'un PGD. Ensuite, nous nous intéresserons à la rédaction en elle-même du PGD par les chercheurs. Et finalement, nous tâcherons d'analyser ce que l'expérience de rédaction peut apporter aux chercheurs du point de vue de l'ouverture des données.

1. La mise en œuvre d'un PGD

Dans cette première partie, nous nous intéresserons à ce qui entoure l'établissement d'un PGD, et particulièrement ce qui précède. En effet, nous allons étudier la raison qui justifie qu'un modèle soit préféré à un autre, en tentant de déterminer de qui relève la décision. Ensuite, nous nous pencherons sur le parcours des personnes interrogées et le potentiel lien avec les PGD. Finalement, nous verrons pour quels motifs les personnes sollicitées en sont venues à produire un PGD.

1.1. Le modèle : un choix ou une obligation

Pour reprendre les quelques chiffres présentés dans la grille d'analyse, il existe 36 modèles disponibles sur DMP Opidor, dont dix-huit français et dix-huit anglais. Notre corpus se compose lui-même de dix-huit PGD et seulement cinq modèles y sont utilisés. On peut alors se demander la raison d'une si faible diversité parmi les plans rédigés. On remarque une préférence pour le modèle ANR, car on dénombre dix PGD faits sur le modèle de l'ANR, pour un corpus total de dix-huit. De plus, le modèle ANR est également fortement plébiscité par les personnes interrogées lors des entretiens. En effet, la doctorante 2 dit avoir sélectionné ce modèle, car il correspondait le mieux à son besoin, et une autre doctorante annonce qu'elle devait rendre un PGD fait sur le

modèle ERC mais qu'elle a préféré travailler sur celui de l'ANR car « il donnait des précisions de ce qui était attendu pour chaque partie ». On en déduit donc que le modèle ANR est plus facilement utilisé par les personnes rédigeant les PGD, car plus explicite que d'autres modèles proposés. En observant sa structure, on remarque d'ailleurs des similitudes avec les quatre autres modèles étudiés⁹⁷, ce qui laisse penser qu'il a pu servir de référence pour créer certains modèles. Et cela pourrait conforter l'idée que le modèle proposé par l'ANR est plus explicite que d'autres, raison pour laquelle les autres modèles s'appuieraient dessus. On peut également remarquer que 80 % des PGD basés sur un modèle INRAE ont été rédigés par des personnes relevant de cet organisme. Nous pouvons donc imaginer que la structure demande (ou recommande) à ce que les projets qu'elle finance présentent un PGD INRAE. Mais il est également possible que ces modèles soient particulièrement adaptés aux besoins des agents d'INRAE. Cependant, le taux de complétion n'est pas exceptionnel puisqu'il est seulement de 40 %. Au regard de ce chiffre, nous pouvons nous interroger sur la facilité de compréhension du plan, surtout lorsque l'on sait qu'en comparaison, 80 % des modèles ANR sont complets.

1.2. Le rôle potentiel du parcours des intervenants

Parmi toutes les personnes sollicitées, huit ont répondu présents et ont participé à un entretien avec nous. L'échantillon est constitué de trois hommes et cinq femmes. Et il s'agit d'un panel très varié, puisque trois ingénieurs, trois doctorants et deux chercheurs nous ont répondu.

Grâce à la première question que nous avons posée lors de l'échange avec ces personnes, nous avons pu nous rendre compte que la majorité des personnes interrogées étaient issues de disciplines de Sciences humaines. En effet, si l'on regarde les différentes formations suivies, on trouve des parcours en psychologie, en géographie, en histoire, en lettres, en langues et en histoire de l'art. Reste que deux personnes sont formées aux « sciences dures » : on remarquera qu'il s'agit des deux seuls chercheurs qui ont répondu présents à notre appel. En effet, l'un est formé en physique-chimie, à l'interface avec la biologie, et travaille actuellement dans un laboratoire de chimie des polymères organique, et l'autre est chercheuse en biologie végétale⁹⁸. Ce constat peut également être le signe que les chercheurs issus des sciences dites « dures » sont plus en avance dans le domaine de la science ouverte que les autres disciplines. Toutefois, cela est à nuancer, car

⁹⁷ cf. Annexe H. Grilles d'analyse complétées

⁹⁸ cf. Annexe E. Retranscriptions des entretiens avec les chercheurs

notre observation ne s'appuie que sur deux réponses, ce qui ne peut pas être considéré comme représentatif de la communauté scientifique. Il faut relever par ailleurs que les sciences humaines semblent commencer à davantage s'impliquer dans la science ouverte et les données de la recherche. On peut voir cela par le fait que la majorité des doctorants interrogés sont formés en sciences humaines. L'exception est une doctorante avec un profil parfaitement atypique, même à l'échelle du panel. Son parcours est totalement hétéroclite : formée initialement en droit, ce qui lui a permis de travailler comme avocate spécialisée en droit de l'environnement, elle a depuis peu repris ses études et a suivi une formation à la croisée de plusieurs disciplines (sciences, sciences humaine et droit) et toujours en lien avec l'environnement⁹⁹.

Il est intéressant de noter qu'il n'y a pas un seul parcours qui se ressemble, et que toutes les personnes interrogées proviennent de disciplines différentes. Si cela est le fruit du hasard et non pas de notre volonté, c'est également ce qui fait la richesse de notre échantillon. En revanche cette diversité laisse à penser que le parcours initial n'est pas réellement en lien avec la compréhension et l'appropriation des PGD. Ou du moins, qu'il n'y a pas un parcours type que semble pousser les personnes à davantage utiliser les PGD.

1.3. Les raisons de la rédaction d'un PGD

On peut imaginer plusieurs raisons à la rédaction d'un PGD : soit qu'une personne estime que cela lui apportera quelque chose et fait le choix d'en produire un, soit qu'elle est simplement curieuse, ou bien encore qu'il s'agit d'une demande faite par un financeur du projet. C'est en tout cas les trois idées qui nous ont semblé ressortir de l'ensemble de nos entretiens.

Pour la moitié de notre panel interrogé, il semble que la rédaction du PGD n'était pas un choix. En effet, quatre personnes ont répondu qu'elles avaient produit ou aidé à produire un PGD à la suite d'une « demande » d'un financeur du projet. Le plus souvent, il est signalé que le financeur est l'ANR, mais pour un cas la demande était liée à un projet ERC. Le doctorant 1, dont le projet était également financé par l'ERC, précise qu'elle a fait un PGD car cela était inscrit dans sa fiche de poste. Toutefois, nous avons compris par l'entretien que la fiche de poste a justement été créée pour engager une personne en charge du PGD et de sa rédaction. Cette même personne a fait un second PGD, lors d'un autre contrat doctoral : il s'agissait cette fois-ci d'une « volonté d'appliquer

⁹⁹ cf. Annexe F. Entretien doctorant 2.

ce qui avait été appris » lors de son premier post-doctorat¹⁰⁰. Les deux autres doctorants interrogés ont également montré une curiosité à l'égard du PGD ou des questions de protection et gestion des données, ce qui les a amenés à choisir d'en rédiger un. Selon la composition de notre corpus et les réponses obtenues, on peut donc avancer l'idée que les PGD intéressent particulièrement les doctorants ou post-doctorants, et que cette population est pour le moment plus sensible aux enjeux liés à la science ouverte et aux données de la recherche. C'est d'ailleurs le constat fait par l'ingénieur 1, qui estime qu'au sein de son laboratoire ce sont en effet les doctorants et post-doctorants qui sont « les plus actifs au niveau des PGD ». Ce qui laisse à penser qu'il existe actuellement un manque d'intérêt des chercheurs pour ces questions¹⁰¹. Et le fait que nous ayons eu du mal à identifier des chercheurs ayant rédigé eux-mêmes leur PGD amène également à cette conclusion.

2. La rédaction du PGD

Dans cette seconde partie, nous allons étudier la rédaction des PGD. Et pour cela, nous nous pencherons dans un premier temps sur la manière dont ceux-ci sont remplis, en partant de notre corpus de PGD. Dans un second temps, nous analyserons les entretiens afin de voir les éventuels problèmes qui se posent au moment de la rédaction, avant d'étudier les solutions mises en œuvre pour résoudre ces problématiques. Enfin, nous examinerons davantage les moyens mis à disposition des chercheurs pour aider à la rédaction de ce PGD.

2.1. La manière de remplir le PGD comme indicateur de la compréhension des enjeux

Pour cette partie, nous allons essentiellement nous appuyer sur le tableau d'analyse consacré au « remplissage » du PGD, que l'on complétera si besoin avec des entretiens. Il nous semble intéressant de nous pencher sur la façon dont les PGD sont remplis et ce que cela peut montrer.

Tout d'abord, on signalera qu'il existe des personnes qui font un PGD mais qui ne le remplissent pas. C'est le cas, au sein de notre corpus, pour les projets « Data_DiaPhen » et

¹⁰⁰ cf. *Ibid.* Entretien doctorant 1.

¹⁰¹ cf. partie 3.2. La perception du PGD par les chercheurs.

« BIOSWOT-Med cruise », dans lesquels on ne trouve aucune information en dehors de celles concernant le projet, sur la première page. Cette absence d'informations se remarque d'autant plus que l'on trouve la mention « question sans réponse » dans toutes les rubriques du plan. Cela nous fait d'ailleurs penser à un élément qui serait inséré automatiquement lorsqu'aucune information n'est précisée. Il s'agit donc probablement d'une fonction de DMP Opidor, qui s'active lorsque le PGD est créé à l'aide de la plateforme. On distinguera cette mention « question sans réponse » de celle « non applicable ». On imagine facilement que cette dernière signifie réellement que la question ne s'applique pas pour le projet concerné. Tandis que la première mention laisse entendre que la personne n'a pas souhaité remplir la rubrique. Il peut s'agir dans ce cas-là d'un problème de compréhension des informations à renseigner, ou encore d'un manque d'implication de l'auteur du plan. En revanche, on imagine plus difficilement un oubli de la part du rédacteur.

Cependant, lorsque les PGD sont remplis, ce qui est le cas de la majorité des plans de notre corpus, nous avons relevé que les informations se trouvent au bon endroit. Ce qui indique que les personnes rédigeant sérieusement leur PGD semblent être à même de les compléter correctement. Toutefois, on observe que le niveau de complétion varie d'un plan à l'autre. Si l'on reprend les chiffres fournis en annexe¹⁰², sur la totalité du corpus, 67 % des PGD sont complets. Cela montre que les personnes vont au bout du PGD lorsqu'elles commencent à en rédiger un, mais que ce n'est pas systématique. Ces PGD qui ne sont pas rédigés complètement peuvent être le fait de chercheurs tels que le Chercheur 2 que nous avons interrogé. En effet, celle-ci nous a indiqué avoir rempli son plan du mieux qu'elle pouvait, en s'abstenant lorsqu'elle ne trouvait pas de réponse. Cette manière de faire montre une volonté de la part des chercheurs de produire un effort quant aux nouvelles directives, mais également une certaine difficulté au moment de renseigner les données¹⁰³. Le modèle de l'ANR, le plus utilisé dans notre corpus, est également celui pour lequel le taux de complétion est le plus élevé. En effet, huit PGD au modèle ANR sur dix sont entièrement remplis. Même en additionnant les deux modèles INRAE, on constate uniquement 40 % de complétion des plans. Ces chiffres illustrent encore une fois la supériorité du modèle ANR par rapport aux autres, et vont de pair avec les allégations des personnes interrogées. Parmi les personnes ayant choisi leur modèle, nous en avons effectivement relevé plusieurs ayant choisi de travailler avec le modèle ANR, car elles l'estimaient « plus précis » ou encore parce qu'il

¹⁰² cf. Annexe I. Tableaux d'aide à l'analyse.

¹⁰³ cf. infra partie 2.2. Difficultés rencontrées et solutions appliquées.

« convenait le mieux » aux besoins¹⁰⁴. Toutefois, il faut également rappeler que les deux PGD non remplis sont issus d'un modèle ANR et d'un modèle INRAE. On peut donc nuancer le fait que les PGD mieux construits entraînent systématiquement une meilleure adhésion de la part des chercheurs, car cela dépend également de l'état d'esprit du chercheur ou de l'auteur du PGD.

2.2. Difficultés rencontrées et solutions appliquées

Nous allons maintenant porter notre attention sur les difficultés soulevées par les individus interviewés dans le cadre de ce mémoire, afin de tenter de déterminer ce qui a pu empêcher de compléter correctement un PGD ou simplement ce qui a pu poser problème au moment de la rédaction.

Pour les chercheurs interrogés, les modèles de PGD proposés sont trop génériques, et ne sont pas adaptés aux différentes disciplines et aux différents projets. En effet, certaines disciplines requièrent des jeux de données de plusieurs natures, qui par exemple ne sont pas stockées au même endroit et n'ont pas les mêmes règles d'archivage. Il peut donc s'avérer compliqué de renseigner ces informations pour un chercheur qui ne possède pas les données avec lesquelles il travaille. Le manque de clarté est également un point qui revient souvent, tant du côté des chercheurs que des doctorants, et même de la part des ingénieurs. On a en effet pu relever un problème de vocabulaire, souligné par l'ingénieur 1, et confirmé par le chercheur 2, qui a indiqué avoir eu des difficultés « à bien cerner le vocabulaire ». Ces problèmes montrent que les PGD ne sont pas forcément adaptés pour des personnes n'ayant pas de notions sur la gestion des données. Il est alors possible d'y voir le besoin de créer un lien entre les chercheurs, amenés à produire des PGD, et les ingénieurs, qui disposent des connaissances techniques nécessaires. De plus, au-delà de l'obstacle du vocabulaire, des difficultés plus techniques ont été relevées. Difficultés qui impactent directement la manière de compléter le PGD. En effet, le chercheur 2 a eu du mal « à savoir comment remplir » son PGD, tandis que le doctorant 1 nous a signalé un manque d'information quant au degré de précision requis. Pour résumer, tel qu'il est actuellement construit, le PGD présente encore des zones d'ombres pour les chercheurs qui les complètent. Pour remédier à cet obstacle, nous avons noté que le chercheur 1 a « bricolé en essayant de répondre aux questions », tandis que l'autre chercheur interrogé a mis ce qui lui semblait répondre à la

¹⁰⁴ cf. Annexe F. Retranscriptions des entretiens avec les doctorants/post-doctorants.

question, en laissant le champ vide lorsqu'elle ne savait pas quoi répondre¹⁰⁵. Cette manière de réagir de la part des chercheurs sollicités laisse penser à un manque de formation ou de relation avec les professionnels de l'information-communication. Mais il est également possible d'y voir un problème concernant la construction du modèle. En effet, les principaux destinataires des modèles sont les chercheurs, il semblerait donc évident que les modèles soient adaptés, et donc décrits en conséquence, selon le niveau de connaissances de ces derniers. Or, le fait que les chercheurs soient obligés de s'adapter eux-mêmes signifie une perte de temps pour eux et probablement une mauvaise assimilation des informations que devrait transmettre le plan. En effet, le risque est que les chercheurs bâclent leur rédaction, simplement pour rendre quelque chose. Une autre façon de compenser les modèles peu clairs a été utilisée par la doctorante 1 : il consistait à choisir un autre modèle plus détaillé pour rédiger, puis de rebasculer ensuite sur le modèle obligatoire¹⁰⁶. En revanche, dans notre corpus de personnes interrogées, les ingénieurs n'ont pas eu affaire à ces problèmes. On peut attribuer cela au poste même d'ingénieur, qui implique une meilleure connaissance du vocabulaire et des enjeux autour des données. L'ingénieur 2 précise qu'elle n'a pas rencontré de difficultés avec le plan en lui-même. Cependant, elle admet en avoir eu pour accéder aux données nécessaires à la rédaction du PGD, en indiquant un manque d'intérêt de la part des chercheurs du projet.

Si les chercheurs qui nous ont répondu signalent un problème de compréhension générale du vocabulaire et de ce qu'il faut réellement renseigner, de leur côté les ingénieurs relèvent également un manque de connaissances de la part des chercheurs, nécessaire à la bonne mise en œuvre d'un PGD. Ce que les ingénieurs estiment pouvoir solutionner en apportant leur aide aux chercheurs et doctorants les sollicitant. Une autre solution possible, appliquée par la doctorante 2 notamment, consiste à regarder les modèles déposés et à s'en inspirer afin de rédiger le sien. Ce moyen pose quelques limites, car si les plans à disposition ne sont pas bien remplis, cela peut induire la personne en erreur. Toutefois, l'analyse de notre corpus de PGD nous permet de montrer que douze des dix-huit PGD sont complets, ce qui correspond à plus de 65 % du corpus. Et on rappellera que tous les PGD sélectionnés et remplis, même partiellement, disposent des bonnes informations au bon endroit. Ainsi, utiliser les PGD déposés comme source d'inspiration peut être

¹⁰⁵ cf. Annexe E. Retranscriptions des entretiens avec les chercheurs.

¹⁰⁶ cf. Annexe F. Entretien doctorant 1.

considéré comme un bon départ dans la rédaction. Mais par la suite, un soutien de la part de professionnels permettrait de s'assurer de la bonne mise en œuvre.

2.3. Un soutien nécessaire à la bonne compréhension

Pour revenir à l'aide disponible ou demandée par les chercheurs, on retrouve assez régulièrement dans les propos des personnes interrogées l'idée de consulter des ingénieurs ou d'autres professionnels de l'information à leur disposition. Et si l'on s'intéresse au discours des ingénieurs interrogés, deux sur les trois ont été sollicités pour accompagner et aider à la rédaction d'un ou de plusieurs PGD. Enfin, si l'aide de la troisième ingénieure n'a pas été requise par des chercheurs, elle est elle-même amenée à faire des formations pour les chercheurs ou encore à participer à des tables rondes. Les trois interventions¹⁰⁷ illustrent donc le rôle certain qu'a l'ingénieur dans la compréhension de ces enjeux et dans leur transmission à la communauté scientifique. Et plus largement, on peut dire que les professionnels de l'information-communication sont un peu les gardiens de la bonne mise en place des pratiques, notamment par la communication qu'ils peuvent faire sur le sujet, mais également par le soutien qu'ils sont à même d'apporter. Cela peut prendre la forme de formations, mais aussi de documents d'aide ou encore de séminaires sur les données de la recherche par exemple.

3. Les apports de l'expérience de rédaction d'un PGD

Nous allons dans cette troisième partie tenter de voir ce que l'expérience de rédaction d'un PGD peut apporter à un chercheur, en termes de compréhension des enjeux qui y sont liés notamment. Pour cela, nous ferons le point dans une première partie sur les enjeux véhiculés par les PDG. Dans un second temps, nous chercherons à savoir si le PGD est un bon moyen pour sensibiliser les chercheurs aux enjeux des données de la recherche. Et dans un troisième temps, il faudra étudier la façon dont les chercheurs perçoivent ces PGD.

3.1. Compréhension des enjeux véhiculés par les PGD

De manière générale, on semble trouver dans tous les modèles étudiés les mêmes informations requises, avec une variation principalement autour des renseignements sur le PGD et

¹⁰⁷ cf. Annexe G. Retranscriptions des entretiens avec les ingénieurs.

le projet, ou encore la présentation des données. Ces points sont particulièrement développés dans les modèles proposés par INRAE et le semblent moins pour les autres. C'est d'ailleurs l'une des raisons qui expliquent que les PGD produits à partir d'un modèle INRAE se composent globalement d'un plus grand nombre de pages que les autres. Toutefois, le nombre de pages reste plutôt variable entre les modèles et au sein d'un même modèle. Les notions les plus régulièrement relevées dans les modèles étudiés sont celles de partage des données, de conservation, et de description de celles-ci. En revanche, la notion de données FAIR est moins visible directement, lorsque l'on étudie uniquement la structure du modèle. Pourtant on l'a vu, les données FAIR tendent à devenir une norme dans la question d'ouverture des données. Cependant, si l'on étudie plus en détail le contenu attendu¹⁰⁸, on remarque que ce thème est également présent dans le modèle ANR.

3.2. La sensibilité aux enjeux d'ouverture

Sur l'ensemble des personnes interrogées, les avis autour de la sensibilité qu'aurait amenée la production d'un PGD sont assez disparates. En effet, « c'est un peu l'œuf et la poule, moi c'est plutôt parce que je suis sensible que j'en ai fait un sinon j'aurais fait le minimum »¹⁰⁹ a répondu le Chercheur 1 à la question « Vous sentez-vous plus sensible au principe d'ouverture des données de la recherche en ayant fait un PGD ? ». Et cette idée resurgit en diverses occasions : plusieurs personnes se sont prêtées au jeu de rédiger un PGD, car elles étaient déjà intéressées par les questions de données, de gestion ou de partage de celles-ci. Mais l'autre partie du panel s'est vu clairement sensibiliser par la réflexion menée autour de ce PGD. Chez les doctorants, on remarque que la production d'un PGD leur a notamment permis d'acquérir une nouvelle vision de l'ouverture des données¹¹⁰. Enfin, de manière générale tous les intervenants semblent estimer que le PGD sensibilise les chercheurs : soit parce qu'ils ont été eux-mêmes sensibilisés en produisant un PGD, soit parce que la rédaction de leur PGD les a confortés dans leur perception de l'intérêt du partage des données, ou encore parce qu'ils ont constaté un intérêt nouveau chez des personnes qui ont été amenées à faire un PGD dans le cadre d'un projet.

¹⁰⁸ cf. Annexe H. Tableau « structure et thèmes des parties ».

¹⁰⁹ cf. Annexe E. Entretien chercheur 1.

¹¹⁰ cf. Annexe F. Retranscriptions des entretiens avec les doctorants/post-doctorants.

Il semble toutefois que tout le monde ne se soit pas senti sensibilisé pour les mêmes choses. La chercheuse 2 ne s'est pas sentie particulièrement concernée par le partage en faisant son PGD, mais davantage aux questions de stockage et de sécurité. Un doctorant a quant à lui ressenti une meilleure compréhension de ce qu'est l'ouverture¹¹¹, tandis que la doctorante 1 y a vu un changement dans sa manière de produire des données. Ainsi, on voit ici que les PGD peuvent sensibiliser les personnes selon les forces et faiblesses de chacun.

3.3. La perception du PGD par les chercheurs

Lorsque l'on s'intéresse aux finalités perçues du PGD, on remarque que les réponses sont assez diversifiées. L'un des chercheurs interrogés parle de la « finalité affichée », qui est la science ouverte, l'intégrité scientifique, etc., mais qui selon lui ne peut se concrétiser, du fait d'un manque de moyens accordés à sa mise en œuvre¹¹². On retrouve également dans plusieurs entretiens l'idée que le PGD est un document qui a un usage sur le long terme. Par exemple, le doctorant 3 nous a présenté son PGD comme « une boussole », un moyen pour lui de s'orienter tout au long de sa recherche. Et cette idée rejoint celle d'un outil utile à la gestion de projet¹¹³ ou qui permet de mieux se projeter vers la fin du projet en cours¹¹⁴. D'autres encore voient le PGD comme un moyen d'obliger les chercheurs à réfléchir sur les données qu'ils produisent, comment les conserver et ce que cela implique. Pour ces personnes, la mise en place de l'obligation de produire un PGD viendrait donc essentiellement de ce besoin de sensibiliser les chercheurs, de les obliger à se poser des questions qu'ils ne se poseraient pas d'ordinaire, « qu'ils éviteraient », pour reprendre les mots de l'ingénieur 2.

Selon les dires des personnes interrogées, il semble donc que le PGD soit perçu par beaucoup comme un outil utile, qui permet de prendre conscience, entre autres, des questions relatives à la gestion des données, au stockage et au partage. Mais cet avis est essentiellement celui de personnes qui ont produit, d'une manière ou d'une autre, un PGD. En effet, notre échantillon ne comprend pas de personnes qui n'ont pas rédigé de PGD, ce qui ne nous permet pas de savoir quelle est l'opinion de ces autres personnes sur les PGD. Toutefois, notre grille d'entretien

¹¹¹ cf. *Ibid.* Entretien doctorant 3.

¹¹² cf. Annexe E. Entretien Chercheur 1.

¹¹³ cf. Annexe F. Entretien Doctorant 2.

¹¹⁴ cf. Annexe G. Entretien Ingénieur 2.

nous a permis de recueillir l'avis des personnes interrogées sur leur perception de l'opinion générale des chercheurs. Il est alors apparu avec les différents échanges que nous avons pu avoir, que beaucoup des chercheurs sont encore réticents à la question du partage de leurs données, voire qu'ils ne comprennent pas qu'ils produisent des données¹¹⁵. Aussi, si l'on en croit les informations recueillies au cours des entretiens, de nombreux chercheurs donnent l'impression de vivre cette obligation de rédiger un PGD comme une contrainte et du travail en plus. Ce qui a d'ailleurs été corroboré par les chercheurs interrogés¹¹⁶ mais également par le reste de l'échantillon. L'unique exception à cela est le doctorant 3, qui ne connaît que des chercheurs vraiment intéressés par les PGD.

Au moins à deux occasions, les personnes interrogées ont fait remarquer qu'ils étaient pour le principe d'ouverture des données, en précisant toutefois qu'ils estimaient nécessaire de laisser des droits sur les données aux chercheurs qui les ont produites¹¹⁷. En effet, ils ont soulevé l'idée que le chercheur a besoin de recueillir les fruits de son travail, ce qui pour eux n'empêche pas d'ouvrir les données, mais dans un second temps.

Pour conclure, l'étude du corpus sélectionné et des entretiens passés a permis de mieux nous rendre compte de l'usage qui est fait des PGD. En premier lieu, il est apparu que le modèle utilisé et la raison qui amène à produire un PGD ont davantage d'impact sur la rédaction du plan que le parcours de la personne. Ensuite, il a été déterminé que la manière dont un PGD est rempli donne un aperçu de la compréhension des attendus et des enjeux liés à l'ouverture. Toutefois, par les entretiens, il est apparu que l'intervention d'un ingénieur reste utile et parfois nécessaire afin d'aider les chercheurs à comprendre les notions et à s'approprier les enjeux. Enfin, il semble que la rédaction d'un PGD soit une expérience profitable pour mieux se rendre compte des enjeux de gestion, stockage et partage des données. Cependant, le manque d'intérêt de la part de la part des chercheurs, encore important, complique la sensibilisation qui pourrait s'opérer au travers du PGD.

¹¹⁵ cf. Annexe F : entretien doctorant 1.

¹¹⁶ cf. Annexe E : entretiens chercheur 1 et 2.

¹¹⁷ cf. Annexe F et G : entretiens doctorant 3 et ingénieur 3.

Conclusion

Nous rappelons que l'objet de notre travail de mémoire a porté sur le rôle joué par les PGD dans la diffusion des enjeux de l'ouverture des données de la recherche auprès des chercheurs. Pour ce faire, nous avons sélectionné un corpus de PGD mis en ligne sur DMP Opidor, et en parallèle nous avons également mené des entretiens avec des chercheurs, des doctorants et des ingénieurs.

L'analyse du corpus de PGD ainsi que celle des entretiens passés nous ont fourni des éléments permettant d'avancer une réponse pour chacune des trois hypothèses que nous avons établies au début de ce travail afin de tenter de répondre à notre problématique. La première hypothèse consistait à suggérer que le plan de gestion de données aurait une fonction qui va au-delà de son rôle initial de cadrage opérationnel. En effet, il nous semblait que le document cible les enjeux des données et permet au chercheur de mieux les comprendre, mais cette sensibilisation aux enjeux pourrait varier selon le cadrage demandé par exemple par l'organisme de recherche. Ainsi nous présumons également que le cadre dans lequel l'élaboration du plan de gestion de données se réalise peut modifier cette perception, selon que l'élaboration est le résultat d'une démarche volontaire ou imposée. Au cours de notre analyse, il est apparu que le PGD est effectivement un outil transmettant les enjeux liés à la gestion à l'ouverture des données. Toutefois, le degré de sensibilisation semble dépendre en partie de la volonté de la personne qui le rédige. On a vu en effet que les personnes qui nous ont répondu ont été sensibilisées en partie, mais que beaucoup l'étaient déjà auparavant. Au vu des informations recueillies aux cours de nos entretiens, il semble toutefois possible d'avancer que la démarche volontaire a un meilleur impact au niveau de l'enseignement des enjeux, sans que l'on ait pu remarquer par ailleurs un cadrage particulier de la part de l'organisme de recherche.

Nous supposons, en deuxième hypothèse, qu'une meilleure compréhension des enjeux de l'ouverture des données de la recherche, obtenue par l'expérience d'usage du plan de gestion de données, pourrait engendrer un phénomène d'adhésion à l'ouverture des données chez les chercheurs. Parmi les personnes interrogées, l'expérience du PGD a effectivement engendré une meilleure compréhension des enjeux et les chercheurs ont semblé adhérer à la question de

l'ouverture des données. Mais, pour parvenir à obtenir une adhésion, il est nécessaire que les chercheurs s'intéressent déjà aux questions, ce qui d'après les propos recueillis n'est pas le cas. Pour le moment, au sein de notre corpus, la majorité des personnes qui se sont impliquées dans la rédaction des PGD l'ont fait parce qu'elles étaient déjà en partie sensibles et qu'elles estimaient que ce travail leur apporterait quelque chose de plus. D'après nos données, il semble cependant qu'une fois que les chercheurs ont produit un PGD, ils comprennent mieux ce qu'implique l'ouverture des données et sont donc davantage à même d'adhérer au principe.

Et enfin la dernière hypothèse soulevée était que, dans le cadre de l'élaboration des Plans de Gestion de Données, les professionnels de l'information-documentation auraient un rôle de médiateur à jouer auprès des chercheurs qui sont amenés à utiliser ces plans de gestion de données. Nous faisons plus précisément l'hypothèse que certaines rubriques des plans de gestion de données mobilisant une expertise info-documentaire amèneraient les chercheurs et les professionnels de l'information-documentation à collaborer. Nous imaginions également que cette collaboration permettrait, au-delà de la rédaction méthodique du document, d'échanger les opinions quant aux enjeux de l'ouverture des données de la recherche. Et en effet, les professionnels de l'information-communication que nous avons pu questionner sont apparus assez largement sollicités. On a pu remarquer une sorte d'entraide entre les chercheurs ou doctorants et les ingénieurs, afin de mieux comprendre ce qui est attendu dans certaines parties du plan. Toutefois, selon les dires des ingénieurs interrogés, il est encore courant que les chercheurs contactent des ingénieurs pour davantage, c'est-à-dire afin de leur demander de rédiger leur PGD. L'idée de la collaboration ne semble donc pas encore totalement en place, mais s'organise tout de même, notamment par des contacts pris lors de formations ou de colloques. La collaboration entre les chercheurs et les professionnels de l'information-communication n'apparaît pas toujours simple, mais semble plus facilement développée chez les doctorants de notre corpus.

L'impression générale qui ressort des entretiens est un intérêt pour l'ouverture des données de la recherche. Toutes les personnes interrogées ont compris l'utilité d'avoir fait un PGD. Ce qui montre que l'on peut effectivement tirer de ces PGD un certain nombre de réflexions autour des enjeux liés aux données de la recherche. On remarque néanmoins quelques nuances, et notamment l'idée que les conditions actuelles ne sont pas optimales pour que les chercheurs puissent ouvrir leurs données. Il a également été souligné que les chercheurs sont loin d'être

encore tous prêts pour l'ouverture des données de la recherche. Le rôle des professionnels de l'information et de la communication est donc de savoir informer correctement les chercheurs, en leur faisant comprendre que le PGD est de toute manière un document obligatoire mais qu'il peut aussi leur servir. En effet, le moyen employé par les ingénieurs pour sensibiliser les chercheurs semble être de leur expliquer en quoi ce travail pourra ensuite leur être bénéfique, notamment plus tard, s'ils ont besoin de s'y référer. Encore une fois, on constate par notre corpus une certaine volonté de la part des jeunes doctorants, qui semblent se lancer plus facilement dans la construction d'un PGD.

Pour conclure, s'il est vrai que le PGD joue un certain rôle dans la transmission des enjeux des données de la recherche, il n'est en réalité qu'un maillon de la chaîne. En effet, pour que les enjeux soient transmis, il faut déjà que les personnes rédigent des PGD. Or, malgré l'obligation d'en produire un pour la majorité des projets, beaucoup de chercheurs ne comprennent pas ou ne souhaitent pas partager leurs données. Cette réticence et l'incompréhension des chercheurs restent donc le principal obstacle à la transmission des enjeux de la science ouverte de manière générale. Il semble donc qu'un travail bien en amont du PGD soit nécessaire afin de pouvoir donner au PGD ce rôle de diffuseur des enjeux auprès de la communauté des chercheurs.

Bibliographie

- AMIEL, Philippe, *et al.* Pratiques de gestion des données de la recherche : une nécessaire acculturation des chercheurs aux enjeux de la science ouverte ? Résultats d'une enquête exploratoire dans le bassin montpelliérain (juin 2018)[en ligne]. *Cahiers Droit, Sciences & Technologies*. 11 mai 2020. N° 10, pp. 147-168. [Consulté le 3 septembre 2020]. Disponible à l'adresse : <http://journals.openedition.org/cdst/2061>
- ANCELIN-FABRE, Justine. Travailler à distance sur la médiatisation des plans de gestion de données : retours d'expérience. *UrfistInfo* [en ligne]. 24 août 2020.[Consulté le 24 septembre 2020]. Disponible à l'adresse : <https://urfistinfo.hypotheses.org/3555>
- BENEVOLO, Sylvie. *Dynamiques des services d'information scientifique et technique face à la politique nationale d'évaluation de la recherche: accompagner la gestion des publications scientifiques au sein des UMR de l'IRD* [en ligne]. Mémoire. CNAM-INTD, 2019. Disponible à l'adresse : https://memic.ccsd.cnrs.fr/mem_02903694
- BRETON, Philippe et PROULX, Serge. 11/ Usages des technologies de l'information et de la communication. In : *L'explosion de la communication. Introduction aux théories et aux pratiques de la communication* [en ligne]. La Découverte. Paris : La Découverte, 2012. pp. 263-287. Repères. [Consulté le 31 janvier 2021]. Disponible à l'adresse : <https://www.cairn.info/l-explosion-de-la-communication--9782707173829-page-263.htm>
- CABRERA, Francisca. *Les données de la recherche en Sciences humaines et sociales : enjeux et pratiques. Enquête exploratoire* [en ligne]. Mémoire. Institut National des techniques et sciences de la documentation - Conservatoire National des Arts et des Métiers, 2015. [Consulté le 3 février 2020]. Disponible à l'adresse : https://memic.ccsd.cnrs.fr/mem_01117375
- CARTIER, Aurore, MOYSAN, Magalie et REYMONET, Nathalie. Construire des outils pour la gestion des données de la recherche dans une communauté d'universités. In : *Journée sur les données de la recherche* [en ligne]. Paris, France : ADBS, janvier 2015. [Consulté le 16 décembre 2020]. Disponible à l'adresse : <https://hal.archives-ouvertes.fr/hal-01138663>
- CHAMBAT, Pierre. Usages des technologies de l'information et de la communication (TIC) : évolution des problématiques. *TIS*. 1994. Vol. 6, n° 3, pp. 249-270

CHARTRON, Ghislaine et SCHÖPFEL, Joachim. Open access et Open science en débat. *Revue française des sciences de l'information et de la communication* [en ligne]. 1 août 2017. N° 11. [Consulté le 2 janvier 2021]. Disponible à l'adresse :

<http://journals.openedition.org.ressources.univ-poitiers.fr/rfsic/3331>

CHARTRON, Ghislaine. L'Open science au prisme de la Commission européenne. *Education et sociétés*. 25 septembre 2018. Vol. n° 41, n° 1, pp.177-193. Disponible à l'adresse :

<https://www.cairn.info/revue-education-et-societes-2018-1-page-177.htm>

CLAVIER, Viviane et PAGANELLI, Céline. L'ouverture des données de la recherche dans le cadre d'un projet pluridisciplinaire entre SIC et Informatique : le cas des médias sociaux de santé. *Études de communication* [en ligne]. 2019. N° 52. Disponible à l'adresse :

<https://www.cairn.info/revue-etudes-de-communication-2019-1-page-117.htm>

CNRS. Le mouvement pour la science ouverte. *Science ouverte* [en ligne]. 2020.

[Consulté le 1 novembre 2020]. Disponible à l'adresse : <https://www.science-ouverte.cnrs.fr/le-mouvement-pour-la-science-ouverte/>

COLCANAP, Grégory. *Science ouverte France – Le site couperin de la science ouverte en France*.

[en ligne]. Mis à jour le 1^{er} février 2021. [Consulté le 1 février 2021]. Disponible à l'adresse : <https://scienceouverte.couperin.org/>

COMMISSION EUROPÉENNE-DIRECTION GÉNÉRALE DE LA RECHERCHE ET DE L'INNOVATION. Lignes directrices pour le libre accès aux publications scientifiques et aux données de la recherche dans Horizon 2020 [en ligne]. 15 février 2016, version 2.1. [Consulté le 03/03/2020]. Disponible à l'adresse :

<https://www.enssib.fr/bibliotheque-numerique/documents/66433-lignes-directrices-pour-le-libre-acces-aux-publications-scientifiques-et-aux-donnees-de-la-recherche-dans-horizon-2020.pdf>

DEBOIN, Marie-Claude. *Se familiariser avec les plans de gestion de données de la recherche, en 6 points* [en ligne]. Montpellier : CIRAD, 2018. [Consulté le 13 janvier 2021]. Disponible à l'adresse :

<https://coop-ist.cirad.fr/donnees-recherche><https://coop-ist.cirad.fr/donnees-recherche>

DUPRAT, Julie. *Les données de la recherche à l'Université Bordeaux Montaigne* [en ligne]. Research Report. Université Bordeaux Montaigne, 2019. [Consulté le 1 février 2021]. Disponible à l'adresse :

<https://hal.archives-ouvertes.fr/hal-02020141>

EDDI, Michel. *Coopérer en information scientifique et technique – Cirad*. [en ligne]. Mis à jour le 29 janvier 2021. [Consulté le 1 février 2021]. Disponible à l'adresse : <https://coop-ist.cirad.fr/>

FAYET, Sylvie. « Données » de la recherche, les mal-nommées. *UrfistInfo* [en ligne]. [Consulté le 23 septembre 2020]. Disponible à l'adresse : <https://urfistinfo.hypotheses.org/2581>

FÉRET, Romain, et al. *Améliorer son projet ANR grâce à la Science Ouverte* [en ligne]. Zenodo. 9 mai 2020. [Consulté le 10 décembre 2020]. Disponible à l'adresse : <https://lilloa.univ-lille.fr/handle/20.500.12210/20359.2>.

FOURNIER, Thierry. Les données de la recherche : définition et enjeux. *Arabesques* [en ligne]. 22 août 2019. [Consulté le 18 février 2020]. Disponible à l'adresse : <https://publications-prairial.fr/arabesques/index.php?id=985>

HARNAD, Stevan. Mouvement de l'Open Access. *CC.Webcast* [en ligne]. 24 janvier 2013. [Consulté le 18 janvier 2021]. Disponible à l'adresse : https://webcast.in2p3.fr/video/mouvement_de_l_open_access

Inist-CNRS. *Plan de gestion de données : Fiche synthétique* [en ligne]. 10 février 2017. Mise à jour le 18/09/2018. [Consulté le 4 avril 2020]. Disponible à l'adresse : <https://dorandum.fr/plan-gestion-donnees-dmp/fiche-synthetique/>

JACQUEMIN, Bernard, SCHÖPFEL, Joachim et FABRE, Renaud. Libre accès et données de recherche. De l'utopie à l'idéal réaliste. *Études de communication*. 19 décembre 2019. Vol. n° 52, n° 1, pp. 11-26. Disponible à l'adresse : <https://www.cairn.info/revue-etudes-de-communication-2019-1-page-11.htm>.

JOUËT, Josiane. Retour critique sur la sociologie des usages. *Réseaux. Communication - Technologie - Société*. 2000. Vol. 18, n° 100, pp. 487-521. Disponible à l'adresse : https://www.persee.fr/doc/reso_0751-7971_2000_num_18_100_2235

La France donne le coup d'envoi du programme européen Horizon 2020. *Horizon 2020* [en ligne]. 16 décembre 2013. [Consulté le 12 janvier 2021]. Disponible à l'adresse :

<https://www.horizon2020.gouv.fr/cid75832/la-france-donne-coup-envoi-programme-europeen-horizon-2020.html>

La gestion des données de recherche à l'IFSTTAR : Guide méthodologique pour répondre aux questions pratiques et juridiques sur la gestion des données [en ligne]. Avril 2018. [Consulté le 11 septembre 2020]. Disponible à l'adresse : https://www.ifsttar.fr/fileadmin/redaction/5_ressources-en-ligne/politiques/politique_data_web.pdf.

LARROUTUROU, Bernard. *Ouvrir la Science*. [en ligne]. Mis à jour le 1^{er} février 2021. [Consulté le 1 février 2021]. Disponible à l'adresse : <https://www.ouvrirlascience.fr/>

Loi n° 78-753 du 17 juillet 1978 portant diverses mesures d'amélioration des relations entre l'administration et le public et diverses dispositions d'ordre administratif, social et fiscal. 17 juillet 1978. [Consulté le 02 avril 2020]. Disponible à l'adresse : <https://www.legifrance.gouv.fr/affichTexte.do?cidTexte=JORFTEXT000000339241#LEGISCTA000006095904>

MAUREL, Lionel. La réutilisation des données de la recherche après la loi pour une République numérique. In : GINOUVÈS Véronique et GRAS Isabelle (dir.). *La diffusion numérique des données en SHS - Guide de bonnes pratiques éthiques et juridiques*. Aix-en-Provence : Presses Universitaires de Provence, 2018. p. 49-60. Disponible à l'adresse : <https://hal.archives-ouvertes.fr/hal-01908766>

MILLERAND, Florence. Usages des NTIC : les approches de la diffusion, de l'innovation et de l'appropriation (1ère partie). *Composite*. 1998. Vol. 2, n° 1, pp. 1-19. Disponible à l'adresse : <http://www.composite.org/index.php/revue/article/view/21>

MILLERAND, Florence. Usages des NTIC : les approches de la diffusion, de l'innovation et de l'appropriation (2e partie). *Composite*. 1999. Vol. 3, n° 1, pp. 54-73. Disponible à l'adresse : <http://www.composite.org/index.php/revue/article/view/17>

MINISTÈRE DE L'ENSEIGNEMENT SUPÉRIEUR DE LA RECHERCHE ET DE L'INNOVATION. *Plan National pour la Science Ouverte [en ligne]*. 4 juillet 2018. [Consulté le 1^{er} avril 2020]. Disponible à l'adresse : <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>

OCDE. *Principes et lignes directrices pour l'accès aux données de la recherche financée sur fonds publics [en ligne]*. Paris : Éditions OCDE, 2007. [Consulté le 08/02/2020]. Disponible à l'adresse : <https://doi.org/10.1787/9789264034020-en-fr>.

OLIVEAU, Sébastien. Chaîne de gestion des données. *Progedo [en ligne]*. 24 septembre 2020. [Consulté le 23 janvier 2021]. Disponible à l'adresse : <http://www.progedo.fr/promouvoir/chaine-de-gestion-des-donnees>

Open Data : comprendre les enjeux de l'ouverture des données publiques culturelles. [en ligne]. [Consulté le 2 janvier 2021]. Disponible à l'adresse : <https://dataactivist.coop/ministere-culture/jour1.html#1>

PAMPEL, Heinz et KINDLING, Maxi. Le potentiel des infrastructures d'information pour les données de la recherche. *Études de communication*. 19 décembre 2019. Vol. n° 52, n° 1, pp. 27-50. Disponible à l'adresse : <https://www.cairn.info/revue-etudes-de-communication-2019-1-page-27.htm>

PETIT, Antoine. DMP OPIDoR [en ligne]. Mis à jour le 1^{er} février 2021. [Consulté le 1 février 2021]. Disponible à l'adresse : <https://dmp.opidor.fr/>

PETIT, Antoine. DoRANum – *Données de la recherche : Apprentissage Numérique*. [en ligne]. Mis à jour le 1^{er} février 2021. [Consulté le 1 février 2021]. Disponible à l'adresse : <https://doranum.fr/>

PROST, Hélène et SCHÖPFEL, Joachim. Les entrepôts de données en sciences de l'information et de la communication (SIC). Une étude empirique. *Études de communication*. 2019. N° 52. [Consulté le 17 janvier 2021]. Disponible à l'adresse : <https://www.cairn.info/revue-etudes-de-communication-2019-1-page-71.htm>

PROULX, Serge. Penser les usages des TIC aujourd'hui : enjeux, modèles, tendances. In : *Enjeux et usages des TIC : aspects sociaux et culturels [en ligne]*. Presses universitaires de Bordeaux. Bordeaux, 2005. pp. 7-20. [Consulté le 27 janvier 2021]. Disponible à l'adresse : <https://sergeproulx.uqam.ca/wp-content/uploads/2010/12/2005-proulx-penser-les-usa-43.pdf>

REBOUILLAT, Violaine et CHARTRON, Ghislaine. Services de gestion et de partage des données de recherche : ce qu'en pensent les chercheurs. *12ème Colloque international d'ISKO-France : Données et mégadonnées ouvertes en SHS : de nouveaux enjeux pour l'état et l'organisation des connaissances ?* [en ligne]. Montpellier, octobre 2019. [Consulté le 26 mars 2020]. Disponible à l'adresse : <https://hal.archives-ouvertes.fr/hal-02307085>

REBOUILLAT, Violaine. *Ouverture des données de la recherche : de la vision politique aux pratiques des chercheurs* [en ligne]. Thèse. Conservatoire national des arts et métiers – CNAM, 2019. [Consulté le 23 janvier 2020]. Disponible à l'adresse : <https://tel.archives-ouvertes.fr/tel-02447653>

RENTIER, Bernard. *Science Ouverte, le défi de la transparence* [en ligne]. Académie Royale de Belgique, 2018. [Consulté le 9 mars 2020]. Disponible à l'adresse : <https://orbi.uliege.be/handle/2268/230014>

REYMONET, Nathalie, MOYSAN, Magalie, CARTIER, Aurore et DÉLÉMONTEZ, Renaud. *Réaliser un plan de gestion de données « FAIR » : modèle.* 2018. Disponible à l'adresse : https://archivesic.ccsd.cnrs.fr/sic_01690547v2

REYMONET, Nathalie. *Améliorer l'exposition des données de la recherche : la publication de data papers.* [en ligne]. 6 janvier 2017. [Consulté le 30 janvier 2020]. Disponible à l'adresse : https://archivesic.ccsd.cnrs.fr/sic_01427978

SCHÖPFEL, Joachim. Hors norme? Une approche normative des données de la recherche [en ligne]. *Revue COSSI.* 2018. N° 5. [Consulté le 10 décembre 2020]. Disponible à l'adresse : <https://revue-cossi.info/numeros/n-5-2018-processus-normalisation-durabilite-information/730-5-2018-schopfel>

SERGI, Gianna. *Le Data Management Plan, DMP ou le Plan de Gestion de données, PGD et OPIDoR* [en ligne]. 2018. Collège de France – Direction des réseaux et partenariats documentaires – Département de bibliométrie et des ressources en ligne. [Consulté le 26 mars 2020]. Disponible à l'adresse : https://www.college-de-france.fr/media/bibliotheques-archives/UPL3412676772100941154_TUTORIEL_DMPOPIDOR_gs_v07092018.pdf

SERRES, Alexandre (dir.). *Données de la recherche en SHS. Pratiques, représentations et attentes des chercheurs : une enquête à l'Université Rennes 2* [en ligne]. Research Report. Université Rennes 2, 2017. [Consulté le 29 janvier 2021]. Disponible à l'adresse : <https://hal.archives-ouvertes.fr/hal-01635186>

STÉRIN, Anne-Laure et NOÛS, Camille. Ouverture des données de la recherche : les mutations juridiques récentes. *Tracés. Revue de Sciences humaines*. 31 décembre 2019. N° 19, pp. 37-50. Disponible à l'adresse : <http://journals.openedition.org/traces/10603>

Table des annexes

A. Grille d'entretiens : chercheurs.....	70
B. Grille d'entretiens : doctorants.....	72
C. Grille d'entretiens : ingénieurs.....	74
D. Grilles d'analyse.....	76
E. Retranscriptions des entretiens avec les chercheurs.....	78
F. Retranscriptions des entretiens avec les doctorants/post-doctorants.....	82
G. Retranscriptions des entretiens avec les ingénieurs.....	87
H. Grilles d'analyse complétées.....	93
I. Tableaux d'aide à l'analyse.....	111

Annexes

A. Grille d'entretiens : chercheurs

Thématique	Questions	Ce que je veux savoir
Parcours de la personne	Pouvez-vous me parler de votre parcours ?	En quoi le parcours peut être lié à la compréhension et à l'appropriation des PGD
Le Plan de gestion de Données	Quelles ont été les principales raisons qui vous ont amenées à produire un PGD ?	Identifier les raisons qui nécessitent de faire un PGD
	Selon vous, et d'après votre expérience, quelle est la fonction, la finalité d'un Plan de Gestion de Données ?	Perception de l'utilité du PGD
La réalisation du PGD	Lorsque vous avez fait votre (vos) PGD, comment avez-vous procédé ?	Identifier les démarches
	Quelles ont été les difficultés que vous avez rencontrées, si vous en avez rencontrées ? Et comment les avez-vous surmontées ?	Déterminer les difficultés Appréhender le degré d'autonomie de la personne
	Avez-vous bénéficié d'une aide dans la réalisation ? Pour quoi (thème, rubrique, etc.) ?	Vérifier si certains points demandent des connaissances particulières
Opinion et perception des enjeux	Avez-vous un avis concernant l'ouverture des données de la recherche ? Vous sentez vous plus sensible au principe d'ouverture des données de la recherche en ayant fait un PGD ? Selon vous, dans un PGD, qu'est-ce qui peut permettre l'adhésion ou non à l'ouverture des données ?	Identifier les enjeux Vérifier si les enjeux sont mieux compris grâce à ces PGD

	Quel est votre avis sur les PGD ?	Connaître leur rapport aux PGD
--	-----------------------------------	--------------------------------

B. Grille d'entretiens : doctorants

Thématique	Questions	Ce que je veux savoir
Parcours de la personne	Pouvez-vous me parler de votre parcours ?	En quoi le parcours peut être lié à la compréhension et à l'appropriation des PGD
Le Plan de gestion de Données	Quelles ont été les principales raisons qui vous ont amenées à produire un PGD ?	Identifier les raisons qui nécessitent de faire un PGD
	Selon vous, et d'après votre expérience, quelle est la fonction, la finalité d'un Plan de Gestion de Données ?	Perception de l'utilité du PGD
La réalisation du PGD	Lorsque vous avez fait votre (vos) PGD, comment avez-vous procédé ?	Identifier les démarches
	Quelles ont été les difficultés que vous avez rencontrées, si cela a été le cas ? Et comment les avez-vous surmontées ?	Déterminer les difficultés Appréhender le degré d'autonomie de la personne
	Avez-vous bénéficié d'une aide dans la réalisation ? Pour quoi (thème, rubrique, etc) ?	Vérifier si certains points demandent des connaissances particulières
Opinion et perception des enjeux	Avez-vous un avis concernant l'ouverture des données de la recherche ? Vous sentez vous plus sensible au principe d'ouverture des données de la recherche en ayant fait un PGD ? Selon vous, dans un PGD, qu'est-ce qui peut permettre l'adhésion ou non à l'ouverture des données ? Selon vous, quelle position les chercheurs ont-ils vis-à-vis des PGD ?	Identifier les enjeux Vérifier si les enjeux sont mieux compris grâce à ces PGD Percevoir la position des chercheurs d'après un regard externe

	En tant que futur chercheur, que pensez-vous des PGD et des données de la recherche ?	Connaître la position de futurs chercheurs
--	---	--

C. Grille d'entretiens : ingénieurs

Thématique	Questions	Ce que je veux savoir
Parcours de la personne	Pouvez-vous me parler de votre parcours ?	En quoi le parcours peut être lié à la compréhension et à l'appropriation des PGD
Le Plan de gestion de Données	Quelles ont été les principales raisons qui vous ont amenées à produire un PGD ?	Identifier les raisons qui nécessitent de faire un PGD
	Selon vous, et d'après votre expérience, quelle est la fonction, la finalité d'un Plan de Gestion de Données ?	Perception de l'utilité du PGD
La réalisation du PGD	Lorsque vous avez fait votre (vos) PGD, comment avez-vous procédé ?	Identifier les démarches
	Quelles ont été les difficultés que vous avez rencontrées, si vous en avez rencontrées ? Et comment les avez-vous surmontées ?	Déterminer les difficultés Appréhender le degré d'autonomie de la personne
	Avez-vous bénéficié d'une aide dans la réalisation ? Si oui, Pour quoi (thème, rubrique, etc) ? Ou au contraire, avez-vous été sollicité pour aider à la rédaction d'un PGD ? Si oui, pour quoi ?	Vérifier si certains points demandent des connaissances particulières
Opinion et perception des enjeux	Avez-vous un avis concernant l'ouverture des données de la recherche ? Pensez-vous que faire un PGD permette de sensibiliser les chercheurs à l'ouverture des données de la recherche ? Selon vous, dans un PGD, qu'est-ce qui peut permettre l'adhésion ou non à l'ouverture des données ? Selon vous, quelle position les chercheurs ont-ils vis-à-vis des PGD ?	Identifier les enjeux Vérifier si les enjeux sont mieux compris grâce à ces PGD Percevoir la position des chercheurs d'après un regard externe

	En tant qu'ingénieur, que pensez-vous des PGD et des données de la recherche ?	Avoir un point de vue différent de celui des chercheurs
--	--	---

D. Grilles d'analyse

- **Analyse du corpus**

Chiffres globaux :

Modèles sur DMPOpidor	PGD déposés
Nombre total	Nombre total
Nombre de modèles différents	→ période : le plus ancien
Nombre de modèles anglais	le plus récent
Nombre de modèles français	Nombre total de PGD avec modèle anglais
→ période : le plus ancien	Nombre total de PGD avec modèle français
le plus récent	→ Période sélectionnée : le plus ancien
	le plus récent
	Nombre de modèles français différents

Répartition des PGD déposés :

N°	Modèles utilisés	Organismes	Nombre de PGD	Total
	Ex :Modèle Horizon 2020 Modèle cirad	Ex : CIRAD	Ex : 3	

Structure des PGD (étude : ressemblances/différences)

Modèle	Parties		
Ex : ANR	1. Description des données	2. Archivage	3. Partage des données

Plans : partie / sujet que l'on retrouve le plus souvent – partie ou sujet rares

	Partie/sujet que l'on retrouve le plus souvent	Partie ou sujet rares
	Ex : « protection des données »	« le partage des données »
Parmi les PGD déposés		

Modèles : structure (nb parties) – thèmes des parties

Modèle	Structure (nombre de parties)	Thèmes traités (titres de parties et contenus attendus)
Ex : modèle ANR	5 parties	Données, métadonnées, protection, ...

- **Analyse des contenus**

Comparaisons

Modèle	Organisme	Discipline	Fonction de l'auteur principal
ANR	ANR	x	Chercheur / ingénieur

Remplissage

(par modèle)

DMP	Pages	Niveau de complétion	Informations au bon endroit	Spécificité
Nom	3 p.	Ex : rien / incomplet / complet	oui/non	« sans informations »

E. Retranscriptions des entretiens avec les chercheurs

Thématique	Questions	Chercheur 1	Chercheur 2
Parcours de la personne	Pouvez-vous me parler de votre parcours ?	Formation de physico-chimiste, à l'interface avec la biologie. Chercheur en science des matériaux au CNRS depuis 20 ans. A rejoint le laboratoire de chimie des polymères organiques à l'école d'ingénieurs de physique-chimie Bordeaux il y a 10 ans.	Chercheuse en biologie végétale. Dans son laboratoire depuis 2017.
Le Plan de gestion de Données	Quelles ont été les principales raisons qui vous ont amenées à produire un PGD ?	Demande de l'ANR, dans le cadre d'un projet financé.	Demande du financeur (ANR JCJC). Projet commencé en 2020.
	Selon vous, et d'après votre expérience, quelle est la fonction, la finalité d'un Plan de Gestion de Données ?	Il y a la finalité affichée par l'ANR, qui est la science ouverte, l'intégrité scientifique et la reproductibilité (ce à quoi le chercheur souscrit). Mais en réalité, ces finalités lui paraissent difficiles à mettre en place, car il y a un manque de moyens (ex : réseau mais pas serveur)	Bonne intention, pour s'assurer que les données soient bien archivées, référencées. Le PGD cherche à assurer que l'on puisse remonter à la source des informations
La réalisation du PGD	Lorsque vous avez fait votre (vos) PGD, comment avez-vous procédé ?	A suivi quelques séminaires (université de Bordeaux) et une formation du CNRS sur les données ouvertes A pris le <i>template</i> ANR en anglais sur Opidor	Définition des données du projet. Puis pour chaque jeu de données cités, réflexion autour des questions de stockage, de partage

		<p>A répondu aux questions en réfléchissant à quel type de données allaient être produites (en a défini 4)</p> <p>A également inséré des images, conservée sur une plateforme de l'université</p>	
	<p>Quelles ont été les difficultés que vous avez rencontrées, si vous en avez rencontrées ?</p>	<p>Pas de modèle pour le PGD en chimie.</p> <p>Pas de serveur qui centralise les données des expériences : rend compliqué la question de l'archivage</p>	<p>PGD pas toujours très clair : difficultés à savoir comment remplir le PDG, ou encore de bien cerner le vocabulaire. Plein de modèles. On y trouve des choses assez génériques, pas forcément adaptées au projet</p>
	<p>Et comment les avez-vous surmontées ?</p>	<p>« Pour l'instant j'ai pas trouvé un template qui est exactement ce dont j'avais besoin, donc du coup j'ai bricolé en essayant de répondre aux questions »</p>	<p>En mettant un peu ce qui lui semblait répondre à la question. Mais n'a pas répondu lorsqu'elle ne savait pas quoi mettre.</p>
	<p>Avez-vous bénéficié d'une aide dans la réalisation ? Pour quoi (thème, rubrique, etc.) ?</p>	<p>Oui : a reçu des conseils d'ingénieurs de l'Inist (demande sur Opidor) après avoir fait une première proposition de PGD. Et a demandé de l'aide aux ingénieurs de la plateforme hébergeant les images utilisées (Bordeaux Imagine Center) : ceux-ci ont rédigé un paragraphe à mettre dans le PGD à propos de leur serveur.</p>	<p>Non. Elle a simplement regardé des PGD déposés sur DMP Opidor, pour s'inspirer. Elle aurait pu faire une formation, proposée par le CNRS, mais n'a pas eu le temps. A reçu de l'aide de la part d'un collègue qui en remplissait un en même temps.</p>
<p>Opinion et perception</p>	<p>Avez-vous un avis concernant l'ouverture des données de la recherche ?</p>	<p>« Je pense que pour l'objectif de l'éthique et de la reproductibilité c'est tout à fait intéressant de pouvoir partager les données brutes »</p>	<p>Elle est pour l'ouverture des données. D'autant que les labos s'orientent de plus en plus vers la science ouverte et le libre accès</p>

des enjeux		Actif en ce qui concerne la science ouverte : il est référent HAL pour son laboratoire. Trouve ça normal d'ouvrir, car en tant que chercheurs publics les citoyens doivent pouvoir contrôler ce qui est produit. Pense que le libre accès profite à tout le monde (auteurs et lecteurs) Seulement, le fait que ça ne soit pas obligatoire pour le privé lui semble compliquer les choses.	
	Vous sentez vous plus sensible au principe d'ouverture des données de la recherche en ayant fait un PGD ?	Non, pas plus sensible puisque déjà sensibilisé : « c'est un peu l'oeuf et la poule, moi c'est plutôt parce que je suis sensible que j'en ai fait un sinon j'aurai fait le minimum ».	Pas vraiment. S'est surtout sentie plus sensibilisée à la question du stockage et de la sécurité des données. Ne s'est pas sentie spécialement sensibilisé au partage avec ce PGD, mais peut-être parce qu'elle y était déjà sensibilisée avec le labo.
	Selon vous, dans un PGD, qu'est-ce qui peut permettre l'adhésion ou non à l'ouverture des données ?	Le manque d'adhésion est la conséquence du manque de temps. Le PGD est perçu comme du travail en plus, de l'administratif, par les chercheurs. Ce qui pourrait permettre une meilleure adhésion, ce serait d'avoir les moyens informatiques en local, que les tutelles investissent de manière intelligente dans l'open data, et pourquoi pas quelques formations efficaces Aussi une déclinaison des modèles par discipline pourrait faciliter l'adhésion des chercheurs	Pour des personnes qui n'ont pas pensé au partage des données le PGD peut « donner un petit déclic ». Mais il manque de pistes pratiques pour orienter les personnes vers les bonnes démarches.
	Quel est votre avis sur les	Sur le principe, il n'est pas contre puisqu'il en a fait	Le PGD c'est une bonne intention. Mais donne

	PGD ?	un. Et il trouve utile de pouvoir partager facilement les données.	l'impression aux chercheurs de rajout de strates administratives, de temps. Elle aimerait bien que ça soit un peu plus simple à remplir, et souhaiterait qu'il y ait plus de soutien.
--	-------	--	---

F. Retranscriptions des entretiens avec les doctorants/post-doctorants

Pour simplifier la rédaction, on parlera de « doctorant » pour les trois personnes interrogées.

Thématique	Questions	Doctorant 1	Doctorant 2	Doctorant 3
Parcours de la personne	<i>Pouvez-vous me parler de votre parcours ?</i>	Elle a fait deux postdoc : un premier à l'ENS de Lyon avec pour mission de faire un PGD, et le second à Lyon 3.	Formation d'avocate. A travaillé 6 ans en cabinet d'avocats, en droit de l'environnement. A fait des études à l'étranger. Elle a également travaillé dans une entreprise faisant que génie écologique, puis à <i>Deloitte développement durable</i> . À ce poste, elle travaillait pour la commission européenne et certains projets H2020 (menait des études à partir de réponses à appel d'offre). Ensuite, elle a repris ses études pour faire un M2 Biodiversité Territoire Environnement, puis actuellement un doctorat dans le labo LADYSS.	Doctorant de géographie en 2e année, spécialité rurale.
Le Plan de gestion de Données	<i>Quelles ont été les principales raisons qui vous ont</i>	Pour le premier postdoc : dans la fiche de poste. Pour le second : volonté d'appliquer ce qui avait été appris lors du premier	En s'intéressant aux questions de RGPD dans le cadre de sa thèse, elle a découvert le PGD et a décidée d'en faire un.	A fait un PGD dans le cadre de sa thèse, parce qu'il voulait le faire, qu'il voulait apprendre à en faire et voir ce que cela pourrait lui apporter dans la

	<i>amenées à produire un PGD ?</i>	postdoc, sans que cela soit une demande du poste.		suite de son travail.
	<i>Selon vous, et d'après votre expérience, quelle est la fonction, la finalité d'un Plan de Gestion de Données ?</i>	Attirer l'attention des chercheurs sur les questions de gestion, de conservation, d'archivage des données.	« C'est quelque chose de positif ... qui fait gagner énormément de temps à la fin ». Pour elle, c'est aussi une manière de se démarquer un peu. C'est « un outil de gestion de projet en réalité ».	La protection et la réutilisation de ces données.
La réalisation du PGD	<i>Lorsque vous avez fait votre (vos) PGD, comment avez-vous procédé ?</i>	D'abord, sélection du modèle et rédaction du PGD. Ensuite, rendez-vous avec personnes de la MSH pour relecture et correction éventuelles. Puis, il y a eu une étape de nettoyage des infos qui avaient été mises pour aider les membres du projet à la compréhension, mais qui ne devaient pas figurer dans le PGD. Enfin, il y a eu un travail de déplacement des informations du modèle ANR au modèle ERC, avec la MSH. Et finalement la traduction du plan en	À d'abord suivi d'un séminaire. Ensuite, a consulté des exemples de modèles pour regarder, tester et sélectionner celui qui semblait convenir le mieux (ANR). Après elle a lu quelques PGD. Puis s'est mise à la rédaction du PGD à partir d'Opidor. Enfin, elle a fait relire la première version par l'ingénieure de recherche du labo.	Travail au sein d'une équipe choisie pour participer au à la zone test de l'entrepôt de données de l'IRD. Il a décidé de faire un PGD pour alimenter cet entrepôt. Puis a reçu l'apport et le soutien (dans tous les domaines de compétences) des agents de l'IRD de cette équipe.

		anglais et révision de celle-ci. La dernière étape a été la validation du PGD.		
	<i>Quelles ont été les difficultés que vous avez rencontrées, si vous en avez rencontrées ?</i>	Le modèle ERC (demandé pour un projet ERC) n'était pas clair. Difficulté à savoir le degré de précision requis.	Pour comprendre certaines notions, notamment la différence entre stockage et archivage. Et de savoir quelles plateformes existent, pour quoi.	Oui. Au début il a tâtonné. Difficulté à savoir comment articuler les compétences dans le PGD (gestion droit d'auteur, sauvegarde et stockage). Et aussi à comprendre le vocabulaire.
	<i>Et comment les avez-vous surmontées ?</i>	En travaillant sur un autre modèle : La chercheuse a donc travaillé avec le modèle pour l'ANR, car le modèle donnait des précisions de ce qui est attendu pour chaque partie.	En regardant ce qui a été fait par les autres sur DMP Opidor, en suivant une formation et en interrogeant une ingénieure de son labo.	A surmonté ces difficultés en sollicitant l'équipe de l'IRD avec laquelle il travaillait, et les responsables informatiques.
	<i>Avez-vous bénéficié d'une aide dans la réalisation ? Pour quoi (thème, rubrique, etc.) ?</i>	A reçu de l'aide de la part de personnes relevant de la MSH pour la relecture de son PGD. A aussi eu besoin d'aide, car il lui a fallu produire des données à présenter dans le PGD (comme il s'agissait d'une VO tout était à faire). A également participé à une formation.	S'est aidée des modèles et PGD déposés sur DMP Opidor. A également participé à un colloque, une formation, et était suivie par une ingénieure de recherche.	A reçu l'aide du service informatique. A également fait une formation pour apprendre à faire PGD. De plus, il a reçu de l'aide de la part d'autres personnes, telles que les documentalistes ou informaticiens.

Opinion et perception des enjeux	<i>Avez-vous un avis concernant l'ouverture des données de la recherche ?</i>	Lui semble important que les chercheurs soient sensibilisés aux données, à la gestion des données.	C'est une bonne chose. Indispensable pour faire avancer la recherche.	C'est une très bonne chose. Mais en même temps c'est compliqué, car il n'est pas toujours facile d'ouvrir les données lorsque l'on a un projet financé. Il lui semble qu'il y a un besoin de redéfinir les droits d'auteur pour permettre que l' <i>open data</i> soit une réalité.
	<i>Vous sentez vous plus sensible au principe d'ouverture des données de la recherche en ayant fait un PGD ?</i>	Oui, énormément. Ça lui a changé sa vision et la manière de produire des données.	Pas spécialement car déjà sensibilisée aux questions de l'ouverture de la donnée avant même de faire son PGD.	Oui. Avant de faire le PGD il ne comprenait pas concrètement ce que voulait dire « l'ouverture des données » .
	<i>Selon vous, dans un PGD, qu'est-ce qui peut permettre l'adhésion ou non à l'ouverture des</i>	<i>Pas de réponse</i>	Le PGD entier permet de sensibiliser aux obligations notamment liées au RGPD. Il y a aussi les questions de partage des données, pour des équipes.	<i>Pas de réponse</i>

	<i>données ?</i>			
	<i>Selon vous, quelle position les chercheurs ont-ils vis-à-vis des PGD ?</i>	Pour elle, « certains chercheurs ont du mal avec le principe de la science ouverte ». Par rapport aux PGD : ils sont dans « l'inconscience », car ils ne se rendent pas compte des données qu'ils produisent.	Elle sent une sorte de « résistance au changement ». Beaucoup de chercheurs ont encore l'impression qu'il s'agit d'une sorte d'entrave à leur liberté de recherche. Et c'est chronophage. Elle sent de la réticence sur la mise en œuvre, surtout « de la part de chercheurs bien installés ».	Ne connaît que 2 chercheurs : ils sont vraiment pour les PGD, ils pensent que c'est une très bonne chose car cela facilite la recherche de financement, et permet de montrer le sérieux du chercheur.
	<i>En tant que futur chercheur, que pensez-vous des PGD et des données de la recherche ?</i>	« Le PGD c'est merveilleux et indispensable ». Les données de la recherche : il y a un besoin de former les gens qui en produisent, pour qu'ils aient bien conscience de produire des données en faisant leurs recherches.	C'est nécessaire et ça va devenir quelque chose de commun, qui a un moment donné devra rentrer dans les mœurs.	Il s'agit d'une bonne chose si ça peut lui permettre plus tard de décrocher un financement. L'ouverture de la donnée « c'est une bonne chose si et seulement si les droits des auteurs des chercheurs sont sauvegardés, garantis ». S'il n'y a pas d'amélioration dans ce sens, lui ne serait pas trop pour l'ouverture, car il estime que le travail du chercheur a besoin d'être reconnu.

G. Retranscriptions des entretiens avec les ingénieurs

Thématique	Questions	Ingénieur 1	Ingénieur 2	Ingénieur 3
Parcours de la personne	<i>Pouvez-vous me parler de votre parcours ?</i>	A fait des études d'anglais, et a été professeur de français en Écosse puis en Espagne, et enfin professeur de français dans une association auprès de réfugiés malades. Suite à un bilan de compétence, elle s'est inscrite à l'École bibliothèques et documentation Elle est documentaliste depuis 10 ans : d'abord au Service ingénierie documentaire à la MSH de Nanterre, puis pour le Labo LADYSS, depuis 2 ans (Engagée pour des questions de RGPD).	Ingénieure contractuelle pour un projet ERC (base de donnée épigraphique sur l'Asie du Sud et l'Asie du Sud-Est, du VIe au XVe siècle). En parallèle, elle est doctorante d'histoire de l'art en sculpture médiévale.	A fait un master et un doctorat de Psychologie. A été correspondant AFP, puis maître de conférence, ATER, Chercheur. A également fait une formation en data science. Occupe actuellement un poste d'ingénieur.
Le Plan de gestion de Données	<i>Quelles ont été les principales raisons qui vous ont amenées à</i>	Les questions de loi, ainsi que le fait de travailler sur le RGPD. Cela l'a amené à se poser la question de la protection des données.	Parce que qu'il était obligatoire d'en faire un pour un projet ERC.	L'accompagnement des données dans le cadre de projets ANR.

	<i>produire un PGD ?</i>			
	<i>Selon vous, et d'après votre expérience, quelle est la fonction, la finalité d'un Plan de Gestion de Données ?</i>	La protection et la réutilisation de ces données.	Ça permet de mettre sur la table des sujets qui ont tendance à être évités avant la fin du projet voire après (en particulier les questions juridiques) Permet de découvrir les corpus et les sujets Il force aussi à se projeter vers la fin du projet et ce qu'on est censé rendre, pose des jalons de management de projet	Sert pour enregistrer les données. Le PGD permet de décrire des données et de les réutiliser plus facilement. Il permet d'expliquer comment seront sécurisés ces données, ainsi que d'éviter de financer un projet pour lequel il existe déjà des données.
La réalisation du PGD	<i>Lorsque vous avez fait votre (vos) PGD, comment avez-vous procédé ?</i>	N'a pas fait à proprement parlé de PGD, s'est simplement chargée de la partie administrative pour un suite à la demande d'un chercheur du laboratoire.	D'abord, a cherché à accéder aux données, avec une tentative de petit questionnaire sur les données et leurs usages (finalement pas autorisé à être envoyé) Ensuite, est passée à la rédaction : finaliser le workflow, conceptualiser les étapes et les chaînes de travail.	S'est renseigné auprès du CNRS. A consulté DMP Opidor, pour s'entraîner et regarder les autres PGD. Puis, a rédigé un PGD pour un projet qui n'a pas abouti, ce qui lui a servi d'entraînement. Et enfin a rédigé un PGD ANR.
	<i>Quelles ont été les difficultés que vous avez rencontrées, si vous en avez</i>	Il existe un problème de vocabulaire des chercheurs (par exemple, ils ne savent pas ce que sont les métadonnées) Aussi, il y a un problème d'acceptation : les	Pas spécialement de difficultés, à part pour l'accès aux données et pour avoir des réponses aux questions visant à cerner un peu plus les données du projet. A eu affaire à un	Les chercheurs n'ont pas une idée précise de ce que sont les PGD. Et le PGD demande des connaissances techniques. La rédaction du PGD a également soulevé des questions de

	<i>rencontrées ?</i>	chercheurs ne veulent pas qu'on les oblige à faire quelque chose, ni rendre leurs données publiques.	problème de désintérêt, et même de la part des chefs de projet.	stockage pendant et après le projet, et de la création de dépendance entre les chercheurs et la structure de stockage.
	<i>Et comment les avez-vous surmontées ?</i>	Accompagne les chercheurs quand cela est nécessaire. Et essaye de faire comprendre aux chercheurs que cela pourra leur servir plus tard.	<i>Pas de réponse</i>	La solution a été d'informer les chercheurs sur les objectifs et les aspects techniques, et ensuite de les amener à se mettre d'accord A suggéré de laisser le stockage des données sensibles à la charge de l'INSERM. Et a également suggéré les différentes plateformes de stockage (Nakala, Sharedoc) .
	<i>Avez-vous bénéficié d'une aide dans la réalisation ? Pour quoi (thème, rubrique, etc.) ?</i>	–	PGD rédigé avec une autre ingénieure (écriture à deux mains). Elle a suivi une formation organisée par EHESS. A également lu d'autres PGD pour voir comment ils étaient remplis. Enfin, le PGD a été relu par une autre ingénieure du projet	–
	<i>Avez_ vous été sollicité pour</i>	Sollicitée par une doctorante pour relire son PGD. Et cette doctorante lui	Pas strictement dans le projet, mais fait aussi des formations pour les	A été sollicité par des chercheurs pour faire le PGD. Car les chercheurs

	<i>aider à la rédaction d'un PGD ? Si oui, pour quoi ?</i>	a aussi posé plusieurs questions. Aussi sollicitée par un chercheur pour remplir un PGD : elle a complété la partie administrative mais pas plus, rédaction d'une partie en anglais.	chercheurs : participation à une table ronde organisée par la MSH de Lyon Saint-Étienne, participation à la journée organisée par la MSH de Poitiers, et doit participer aux journées MATE-SHS en juin.	ne savaient pas ce qu'était le PGD, ce qu'il fallait faire.
Opinion et perception des enjeux	<i>Avez-vous un avis concernant l'ouverture des données de la recherche ?</i>	Elle semble trouver que l'ouverture est nécessaire puisqu'elle dit vouloir une plateforme nationale pour les données	« C'est bien, c'est beau, c'est une belle utopie ». Des progrès ont été faits. Mais en Sciences Humaines, c'est peut-être moins évident que pour les sciences dures.	Plutôt favorable à l'ouverture : car « l'objectif de la recherche c'est faire avancer les connaissances, [...] améliorer la compréhension et la vie des êtres humains ». Il lui semble que partager les données que l'on produit ne peut être que bénéfique pour toute la communauté scientifique. Et si la recherche est financée par de l'argent public, il semble normal que les résultats qui en sont issus puissent servir à tout le monde. Mais pour lui, il faut « laisser à la décision des chercheurs de mettre un embargo [...] pour qu'ils puissent d'abord eux-mêmes utiliser les données ».
	<i>Pensez-vous que faire un</i>	« Le problème, c'est que les données ne sont pas bien comprises et	Oui, car l'obligation va obliger les chercheurs à s'y intéresser	Oui. Car de son expérience les chercheurs, qui au début ne

	<p><i>PGD permette de sensibiliser les chercheurs à l'ouverture des données de la recherche ?</i></p>	<p>perçues par les chercheurs ».</p>		<p>comprenaient pas pourquoi il leur était demandé de faire un PGD, en ont compris la nécessité après en avoir fait un. Permet d'anticiper les choses, ce qui donne une certaine consistance au projet. Et « permet de chiffrer en termes financiers ce que tu vas faire ». Ils y ont aussi vu l'intérêt pour de futures recherches.</p>
	<p><i>Selon vous, dans un PGD, qu'est-ce qui peut permettre l'adhésion ou non à l'ouverture des données ?</i></p>	<p>Il est nécessaire de structurer les données avant de pouvoir se lancer dans les questions de PGD.</p>	<p>Ce qui y est demandé reste des informations relativement techniques (format, métadonnées). Elle se demande donc jusqu'à quel point un chercheur peut en rédiger un seul. Pour elle, l'adhésion peut apparaître en jouant sur ce que cela peut apporter au chercheur qui le fait.</p>	<p>Dans son cas, l'adhésion des chercheurs à l'ouverture des données semble être apparu après avoir fait un PGD.</p>
	<p><i>Selon vous, quelle position les chercheurs ont-ils vis-à-vis des PGD ?</i></p>	<p>Pour la plupart, ils n'ont aucun intérêt pour les PGD. Pour le moment les plus actifs au niveau des PGD, dans le labo, ce sont les doctorants et post-doctorants.</p>	<p>Ne sont pas intéressés par les PGD. Et cela même en ayant organisé un meeting autour des DMP, de l'importance de la gestion de données, etc.</p>	<p>« Pour le moment je pense que la majorité voit ça comme une contrainte. A l'exception peut-être de ceux qui ont eu une expérience et qui commencent à comprendre la nécessité de faire cela [les PGD] ». Mais « ils pensent toujours que pour</p>

				eux c'est une charge et que cette partie là doit être sous-traitée par des ingénieurs » par exemple.
	<i>En tant qu'ingénieur, que pensez-vous des PGD et des données de la recherche ?</i>	Pour l'instant s'occupe surtout des questions de RGPD, qu'elle estime très liées aux PGD. Elle sait qu'il est important de s'occuper des données.	Le PGD est plus fait pour les ingénieurs, pour les personnes qui ont l'habitude de manipuler de la donnée. Il s'agit d'un super outil, qui valorise le métier des « petites mains », des personnes auxiliaires dans la recherche.	Cela facilite la vie des ingénieurs, pour savoir où se trouvent les données et pour leur travail d'assistance aux chercheurs.

H. Grilles d'analyse complétées

- Analyse du corpus

Chiffres globaux

◆ Modèles sur DMPOpidor	◆ PGD déposés
Nombre total : 36 Nombre de modèles différents : 27 Nombre de modèles anglais : 18 Nombre de modèles français : 18 → période : le plus ancien → 26/09/2019 le plus récent → 11/06/2021	Nombre total : 39 → période : le plus ancien → 05/08/2019 le plus récent → 16/06/2021 Nombre total de PGD avec modèle anglais : 21 Nombre total de PGD avec modèle français : 18 Période des PGD sélectionnés : le plus ancien → 05/08/2019 le plus récent → 19/03/2021 Nombre de modèles français différents : 5

Répartition des PGD déposés :

	Modèles utilisés	Organismes	Nombre de PGD	Total
	Ex :Modèle Horizon 2020 Modèle cirad	Ex : CIRAD	Ex : 3	
1	ANR	INRAE	1	10
	ANR	CNRS	5	
	ANR	IRD	1	
	ANR	ENS (dont Lyon)	2	
	ANR	Université d'Aix-Marseille	1	
2	DCC Template	IRD	1	2
	DCC Template	Dataactivist	1	
3	Horizon 2020 FAIR DMP	Bibracte EPCC	1	1
4	INRAE trame générique projet	INRAE/ INRA	2	3
	INRAE trame générique projet	ENS Lyon	1	
5	INRAE trame structure	INRAE	2	2
				18

Structure des PGD (étude : ressemblances/différences)

Modèle	Parties								
ANR	1. Description des données et collecte ou réutilisation de données existantes	2. Documentation et qualité des données	3. Stockage et sauvegarde pendant le processus de recherche	4. Exigences légales et éthiques, codes de conduite	5. Partage des données et conservation à long terme	6. Responsabilités et ressources en matière de gestion des données			
HORIZON 2020 FAIR DMP	1. Résumé descriptif des données	2. Données FAIR	3. Allocation de ressources	4. Sécurité des données	5. Aspects éthiques	6. Autres			
DCC TEMPLATE	Collecte des données	Documentation et métadonnées	Aspects éthiques et juridiques	Stockage et sauvegarde	Sélection et préservation	Partage des données	Responsabilités et ressources		
INRAE – TRAME GÉNÉRIQUE PROJET	Informations sur le plan de gestion	Informations sur le projet	Présentation générale des données du projet	Droits de propriété intellectuelle	Confidentialité	Partage des données à l'issue du projet	Description et organisation des données	Stockage et sécurité des données	Archivage et conservation des données après la fin du projet
INRAE – TRAME STRUCTURE	Information sur la structure	Informations sur le plan de gestion	Présentation générale des données	Droits de propriété intellectuelle	Confidentialité	Partage des données	Organisation et documentation des données	Stockage et sécurité des données	Archivage et conservation des données

Plans : partie / sujet que l'on retrouve le plus souvent – partie ou sujet rares

	Partie/sujet que l'on retrouve le plus souvent	Partie ou sujet rares
	Ex : « protection des données »	« le partage des données »
Parmi PGD déposés	Partage des données Conservation des données Description/collecte/documentation des données	Données FAIR

Modèles : structure (nb parties) – thèmes des parties

Modèle	Structure (nombre de parties)	Thèmes traités (titres de parties et contenus attendus)
Ex : modèle ANR	5 parties	Données, métadonnées, protection, ...
ANR	6	<p>1. Description des données et collecte ou réutilisation de données existantes</p> <ul style="list-style-type: none"> • Quelles seront les ressources (budget et temps alloués) dédiées à la gestion des données permettant de s'assurer que les données seront FAIR (Facile à trouver, Accessible, Interopérable, Réutilisable) ? • Quelles données (types, formats et volumes par ex.) seront collectées ou produites ? <p>2. Documentation et qualité des données</p> <ul style="list-style-type: none"> • Quelles métadonnées et quelle documentation (par exemple méthodologie de collecte et mode

		<p>d'organisation des données) accompagneront les données ?</p> <ul style="list-style-type: none"> • Quelles mesures de contrôle de la qualité des données seront mises en œuvre ?
		<p>3. Stockage et sauvegarde pendant le processus de recherche</p> <ul style="list-style-type: none"> • Comment les données et les métadonnées seront-elles stockées et sauvegardées tout au long du processus de recherche ? • Comment la sécurité des données et la protection des données sensibles seront-elles assurées tout au long du processus de recherche ?
		<p>4. Exigences légales et éthiques, codes de conduite</p> <ul style="list-style-type: none"> • Si des données à caractère personnel sont traitées, comment le respect des dispositions de la législation sur les données à caractère personnel et sur la sécurité des données sera-t-il assuré ? • Comment les autres questions juridiques, comme la titularité ou les droits de propriété intellectuelle sur les données, seront-elles abordées ? Quelle est la législation applicable en la matière ? • Comment les éventuelles questions éthiques seront-elles prises en compte, les codes déontologiques respectés ?
		<p>5. Partage des données et conservation à long terme</p> <ul style="list-style-type: none"> • Comment et quand les données seront-elles partagées ? Y a-t-il des restrictions au partage des données ou des raisons de définir un embargo ? • Comment les données à conserver seront-elles sélectionnées et où seront-elles préservées sur le long terme (par ex. un entrepôt de données ou une archive) ? • Quelles méthodes ou quels outils logiciels seront nécessaires pour accéder et utiliser les données ? • Comment l'attribution d'un identifiant unique et pérenne (comme le DOI) sera-t-elle assurée pour

		chaque jeu de données ?
		<p>6. Responsabilités et ressources en matière de gestion des données</p> <ul style="list-style-type: none"> • Qui (par exemple rôle, position et institution de rattachement) sera responsable de la gestion des données (c'est-à-dire le gestionnaire des données) ? • Quelles seront les ressources (budget et temps alloués) dédiées à la gestion des données permettant de s'assurer que les données seront FAIR (Facile à trouver, Accessible, Interopérable, Réutilisable) ?
Horizon 2020 FAIR DMP	6	<p>1. Résumé descriptif des données</p> <ul style="list-style-type: none"> • Fournir un résumé descriptif des données en abordant les points suivants : objectif de la collecte/génération de données, lien avec les objectifs, type et formats de données, origine, estimation du volume, potentiel de réutilisation <p>2. Données FAIR</p> <ul style="list-style-type: none"> • Rendre les données faciles à trouver, en incluant la mise à disposition des métadonnées • Rendre les données librement accessibles • Rendre les données interopérables • Accroître la réutilisation des données (au moyen de licences) <p>3. Allocation de ressources</p> <ul style="list-style-type: none"> • Expliquer l'allocation des ressources en abordant les points suivants : estimer les coûts, identifier les responsabilités, définir la valeur des données et les coûts pour une conservation à longs termes <p>4. Sécurité des données</p> <ul style="list-style-type: none"> • Examiner la récupération des données ainsi que le stockage sécurisé et le transfert de données

		<p>sensibles</p> <p>5. Aspects éthiques</p> <ul style="list-style-type: none"> • Doivent être abordés dans le cadre de l'évaluation éthique, dans la section éthique de la Description de l'action et dans les livrables éthiques. Inclure des références et les aspects techniques liés s'ils ne sont pas couverts par l'évaluation énoncée précédemment <p>6. Autres</p> <ul style="list-style-type: none"> • Mentionner les autres procédures nationales, sectorielles, départementales ou issues de financeurs utilisées pour la gestion des données (le cas échéant)
DCC Template	7	<p>Collecte des données</p> <ul style="list-style-type: none"> • Quelles données allez vous collecter ou créer ? • Comment les données seront-elles collectées ou créées ? <p>Documentation et métadonnées</p> <ul style="list-style-type: none"> • Quelle documentation et quelles métadonnées seront associées aux données ? <p>Aspects éthiques et juridiques</p> <ul style="list-style-type: none"> • Comment allez-vous gérer les aspects éthiques ? • Comment allez-vous prendre en compte les aspects de droits d'auteur (copyright) et de propriété intellectuelle <p>Stockage et sauvegarde</p> <ul style="list-style-type: none"> • Comment les données seront-elles stockées et sauvegardées pendant le projet de recherche ? • Comment allez-vous gérer l'accès et la sécurité ?

		<p>Sélection et préservation</p> <ul style="list-style-type: none"> • Quelles données doivent être conservées, partagées et/ou préservées ? • Quel est le plan de préservation à long terme des jeux de données ?
		<p>Partage des données</p> <ul style="list-style-type: none"> • Comment allez-vous partager les données ? • Existe-t-il des restrictions au partage des données ?
		<p>Responsabilités et ressources</p> <ul style="list-style-type: none"> • Qui sera responsable de la gestion des données ? • De quelles ressources aurez-vous besoin pour mettre en œuvre votre plan ?
INRAE – trame générique projet	9	<p>Informations sur le plan de gestion</p> <ul style="list-style-type: none"> • Responsable • affiliation de l’auteur • version en cours • date de la dernière version
		<p>Informations sur le projet</p> <ul style="list-style-type: none"> • ID de l’AAP • Financeur(s) • Nom du programme de recherche • référence de la convention de financement

		<ul style="list-style-type: none"> • Acronyme du projet • Nom du projet • Institution leader • ...
		<p>Présentation générale des données du projet</p> <ul style="list-style-type: none"> • Brève présentation des données générées, collectées ou réutilisées
		<p>Droits de propriété intellectuelle</p> <ul style="list-style-type: none"> • Qui détiendra les droits sur les données et les autres informations créées lors du projet ? • Du matériel protégé par des droits spécifiques sera-t-il utilisé au cours du projet ? Dans ce cas, qui s'occupe des formalités à accomplir, obtient les autorisations d'utilisation et de diffusion éventuelle ... ?
		<p>Confidentialité</p> <ul style="list-style-type: none"> • Identification des jeux de données confidentielles • Quelles sont les mesures prises et les normes auxquelles il est nécessaire de se conformer pour garantir cette confidentialité ? • Le cas échéant, comme la confidentialité des données fournies par des personnes sera garantie lorsque les données seront partagées ou rendues disponibles pour une analyse de second niveau ?
		<p>Partage des données à l'issue du projet</p> <ul style="list-style-type: none"> • Y a-t-il une obligation de partage (ou à l'inverse une interdiction ou une restriction) ? • Quelles données seront partagées à l'issue du projet ? Si toutes les données ne sont pas disponibles

		<p>de la même façon, ou en même temps, le préciser</p> <ul style="list-style-type: none"> • Quelles sont les réutilisations potentielles de ces données ? • La lecture des données nécessite-t-elle le recours à un logiciel ou un outil spécifique ? Si oui, lequel et comment y accéder ? • Comment les données seront-elles partagées ? • Avec qui ? Sous quelle licence ? • À partir de quand ? Pendant combien de temps ? ...
		<p>Description et organisation des données</p> <ul style="list-style-type: none"> • Quels méthodes et outils sont utilisés pour acquérir et traiter les données ? Précisez les différents formats dans lesquels les données seront disponibles aux différentes phases de la recherche • Documentation associée aux données • Quels types de métadonnées seront produites pour accompagner les données ? Quels sont les standards et les vocabulaires ou taxonomies qui seront utilisés pour décrire les données ? • Comment les métadonnées seront-elles produites ? • Comment les fichiers de données sont-ils gérés et organisés au cours du projet : contrôle des versions, conventions de nommage des fichiers, organisation des fichiers.. • Quelle est la procédure de contrôle qualité des données ? joindre éventuellement le plan d'assurance qualité
		<p>Stockage et sécurité des données</p> <ul style="list-style-type: none"> • Stockage : quels seront les supports utilisés pour les données au cours du projet ? Quels seront les types de flux empruntés par les données au cours du projet ? Quelle est la volumétrie prévisionnelle ? Où sont hébergées physiquement les données, sur quel type d'hébergement ? Où

		<p>sont localisées géographiquement les données ?</p> <ul style="list-style-type: none"> • Sécurité : L'identité hébergeant physiquement les données a-t-elle une politique de sécurité pour son système d'information ? • Sécurité – confidentialité : les données feront-elles l'objet d'échange ou de partage avec de tiers acteurs ? Comment sont déterminés les droits d'accès aux données pendant les recherches ? De quelle manière l'ensemble des chercheurs partenaires du projet auront-ils accès aux données pendant la recherche ? • Sécurité – Intégrité – traçabilité : quelles sont les mesures de protection mises en œuvre pour suivre la production et l'analyse des données au cours du projet ? <p>Archivage et conservation des données après la fin du projet</p> <ul style="list-style-type: none"> • Quelles sont les données à conserver à long terme et quelles sont les données à détruire ? • sur quelle plateforme d'archivage pérenne seront archivées les données à conserver sur le long terme ? Sinon, quelles procédures seront mises en place pour la conservation à long terme ? • Quelle est la durée de conservation des données ? • Qui sera responsable de la conservation à long terme ? • Quel sera le volume de ces données ? • Quelles garanties de financement couvriront les coûts associés à la conservation à long terme ?
INRAE – trame structure	9	<p>Information sur la structure</p> <ul style="list-style-type: none"> • Nom de la structure • Type de structure • ID de la structure

	<ul style="list-style-type: none"> • Responsabilités de la structure • Établissement(s) tutelle(s) • Département de rattachement INRAE • Financeur(s)
	<p>Informations sur le plan de gestion</p> <ul style="list-style-type: none"> • DOI • historique des versions
	<p>Présentation générale des données</p> <ul style="list-style-type: none"> • Mode d'obtention des données • Origine • type de données • nature des données • Format des données • Périmètre thématique des données
	<p>Droits de propriété intellectuelle</p> <ul style="list-style-type: none"> • Qui détiendra les droits sur les données et les autres informations créées ?
	<p>Confidentialité</p> <ul style="list-style-type: none"> • Identification des jeux de données contenant des données confidentielles • Quelles sont les mesures prises et les normes auxquelles il est nécessaire de se conformer pour

		<p>garantir cette confidentialité ?</p> <ul style="list-style-type: none"> • Le cas échéant, comme la confidentialité des données fournies par des personnes sera garantie lorsque les données seront partagées ou rendues disponibles pour une analyse de second niveau ?
		<p>Partage des données</p> <ul style="list-style-type: none"> • Y a-t-il une obligation de partage (ou à l'inverse une interdiction ou une restriction) ? • Quelles sont les réutilisations potentielles de ces données ? • La lecture des données nécessite-t-elle le recours à un logiciel ou un outil spécifique ? Si oui, lequel ? • Comment les données seront-elles partagées ? • Avec qui ? • Sous quelle licence ?
		<p>Organisation et documentation des données</p> <ul style="list-style-type: none"> • Quels méthodes et outils sont utilisés pour acquérir et traiter les données, depuis leur acquisition jusqu'à leur mise à disposition, leur archivage ou leur destruction ? • Quelles métadonnées seront utilisées pour accompagner le jeu de données ? Quels seront les standards, vocabulaires, taxonomies... utilisés pour décrire et représenter les données et éléments de métadonnées ? Comment les métadonnées seront-elles produites et mises à jour ? • Une documentation complémentaire aux métadonnées est-elle nécessaire pour décrire les données et assurer leur réutilisabilité sur le long terme • Comment les fichiers de données sont-ils gérés et organisés : contrôle des versions, conventions de nommage des fichiers, organisation des fichiers • Quel est le processus de contrôle qualité des données ?

		<p>Stockage et sécurité des données</p> <ul style="list-style-type: none">• Quels sont les types de flux empruntés par les données et les supports utilisés pour les stocker ?• Quelle est la volumétrie actuelle et prévisionnelle ?• L'entité hébergeant physiquement les données a-t-elle une politique de sécurité pour son système d'information ?• Sécurité – Confidentialité : les données font-elles l'objet d'échange ou de partage avec de tiers acteurs et selon quelles modalités ? comment sont déterminés les droits d'accès aux données avant leur publication ?• Sécurité – Intégrité – Traçabilité : Quelles sont les mesures de protection mises en œuvre pour suivre la production et l'analyse des données ?
		<p>Archivage et conservation des données</p> <ul style="list-style-type: none">• Quelles sont les données à conserver sur le moyen ou le long terme et quelles sont les données à détruire ?• sur quelle plateforme d'archivage pérenne seront archivées les données à conserver sur le long terme ? Sinon, quelles procédures seront mises en place pour la conservation à long terme ?• Quelle est la durée de conservation des données ?• Quelles garanties de financement couvriront les coûts associés à la conservation à long terme ?

- **Analyse des contenus**

Le remplissage des PGD, par modèle

[/!\ tous ces plans inclus entre une et trois pages de « garde »]

ANR

DMP	Pages	Niveau de complétion	Informations au bon endroit	Spécificité
	3 p.	Ex : rien / incomplet / complet	oui/non	« sans informations »
IFB_Training_salivarius	3	incomplet	Oui	Réponses par phrases courtes. Parfois « question sans réponse ». Tableaux
Bridge Research through Interoperable Data Governance and Environments	8	Complet	Oui	Tableaux
WISPER	4	Complet	Oui	Partie 4. : « non applicable »
PGD 1 : Suivi (fictif) de population de poissons dans le lac du Bourget	5	Complet	Oui	Projet fictif, « PGD destiné à être utilisé à des fins pédagogiques »
BIOSWOT-Med cruise	5	Rien	-	« Question sans réponse » à toutes les rubriques
TRANSFUNERAIRE: approche comparative des rituels collectifs de ré-inhumation en Europe et en Amérique-Latine XXe-XXIe	8	Complet	Oui	Quelques points « question sans réponse »

siècles				
Hospitam	5	Complet	Oui	Assez détaillé « / » pour les parties 5.3 et 4.3
IMPRINT	7	Complet	Oui	
LipInTB	5	Complet	Oui	Rubriques avec « non applicable »
PGD HyperOtlet	9	Complet	Oui	Mais, du point 2 au point 6 : 3 cadres à chaque fois, dont seulement un de remplis (pourtant, il y a toutes les informations nécessaires)

DCC Template (français)

DMP	Pages	Niveau de complétion	Informations au bon endroit	Spécificité
	3 p.	Ex : rien / incomplet / complet	oui/non	« sans informations »
Étude sur l'utilisation d'œuvres relevant des arts visuels dans les publications scientifiques	8	Incomplet	Oui	« question sans réponse »
ATLAS (volet recherche)	7	Complet	Oui	Assez détaillé

Horizon 2020 FAIR DMP (français)

DMP	Pages	Niveau de complétion	Informations au bon endroit	Spécificité
	3 p.	Ex : rien / incomplet / complet	oui/non	« sans informations »
PCR Confluence : Plan de Gestion de Données initial	11	Complet	Oui (le point Autres attend des procédures : ici glossaire)	Tableaux, glossaire

INRAE – Trame générique projet

DMP	Pages	Niveau de complétion	Informations au bon endroit	Spécificité
	3 p.	Ex : rien / incomplet / complet	oui/non	« sans informations »
IFB_Training_plant	12	Incomplet	Oui	« question sans réponse » Réponses succinctes PGD test
FLOWER-LAYER	10	Incomplet	Oui	Anglais. « Question sans réponse » « not applicable » réponses succinctes

G2WAS - Grape Genes for Water Scarcity	5	Complet	Oui	Anglais / réponse assez concises
--	---	---------	-----	----------------------------------

INRAE – Trame structure

DMP	Pages	Niveau de complétion	Informations au bon endroit	Spécificité
	3 p.	Ex : rien / incomplet / complet	oui/non	« sans informations »
Data_DiaPhen	14	Rien	-	« question sans réponse »
URGI Plant Bioinformatics Facility	25	Complet	Oui	3 parties sur 5 sont vides. Les 2 remplies sont complètes

I. Tableaux d'aide à l'analyse

- Taux de PGD complets sur l'ensemble du corpus, selon les modèles :

	ANR	DCC Template	H2020	INRAE générique	INRAE structure	Total
Nombre de PGD	10	2	1	3	2	18
Nombre de PGD complets	8	1	1	1	1	12
%	80	50	100	33	50	67

Table des matières

Remerciements.....	5
Glossaire.....	7
Introduction.....	11
Chapitre 1 – État de l’art.....	15
1. Le contexte : la science ouverte et les données de la recherche.....	15
1.1. La science ouverte, un mouvement récent.....	15
1.1.1. Du libre accès à la science ouverte.....	16
1.1.2. <i>L’environnement</i> juridique de la science ouverte.....	19
1.1.3. Les principes FAIR.....	20
1.2. Des données aux données de la recherche.....	22
1.2.1. Qu’est-ce que la « donnée » ?.....	22
1.2.2. Quelle définition pour les données de la recherche ?.....	23
1.3. Présentation des structures d’appui qui accompagnent le mouvement et participent à l’acculturation de la communauté universitaire.....	25
1.3.1. CoopIST-Cirad, <i>spécialisé en Information Scientifique et Technique</i>	26
1.3.2. Le CoSO, <i>concernant</i> la science ouverte en France.....	26
1.3.3. Science Ouverte France, Couperin : pour promouvoir l’Open Access et les données de la recherche.....	26
1.3.4. DORANum, outil proposant des formations aux données de la recherche.....	27
2. Les données de la recherche au travers du Plan de Gestion de données.....	28
2.1. Les enjeux des données de la recherche.....	28
2.1.1. L’ouverture des données de la recherche.....	28
2.1.2. La gestion des données de la recherche.....	29
2.2. Le Plan de Gestion de Données : la formalisation d’un processus.....	31
2.2.1. Le plan de gestion de données en détail.....	31

2.2.2. Les modèles de plans de gestion de données sur DMP OPIDoR.....	34
2.3. L'appropriation des plans de gestion de données par les chercheurs.....	35
Chapitre 2 – Méthodologie.....	40
1. Présentation du terrain.....	40
2. Le corpus.....	41
2.1. La sélection des PGD.....	41
2.2. La grille d'analyse.....	42
2.2.1. Un état des lieux.....	42
2.2.2. La structure des PGD.....	42
2.2.3. Le contenu du corpus.....	43
3. Les entretiens.....	44
3.1. La construction de l'échantillon.....	44
3.2. Le déroulement.....	45
3.3. La grille d'entretien.....	45
Chapitre 3 – Les plans de gestion de données et leurs usages : une analyse des observations.....	48
1. La mise en œuvre d'un PGD.....	48
1.1. Le modèle : un choix ou une obligation.....	48
1.2. Le rôle potentiel du parcours des intervenants.....	49
1.3. Les raisons de la rédaction d'un PGD.....	50
2. La rédaction du PGD.....	51
2.1. La manière de remplir le PGD comme indicateur de la compréhension des enjeux.....	51
2.2. Difficultés rencontrées et solutions appliquées.....	53
2.3. Un soutien nécessaire à la bonne compréhension.....	55
3. Les apports de l'expérience de rédaction d'un PGD.....	55
3.1. Compréhension des enjeux véhiculés par les PGD.....	55
3.2. La sensibilité aux enjeux d'ouverture.....	56

3.3. La perception du PGD par les chercheurs.....	57
Conclusion.....	59
Bibliographie.....	62
Table des annexes.....	69
Annexes.....	70
A. Grille d’entretiens : chercheurs.....	70
B. Grille d’entretiens : doctorants.....	72
C. Grille d’entretiens : ingénieurs.....	74
D. Grilles d’analyse.....	76
E. Retranscriptions des entretiens avec les chercheurs.....	78
F. Retranscriptions des entretiens avec les doctorants/post-doctorants.....	82
G. Retranscriptions des entretiens avec les ingénieurs.....	87
H. Grilles d’analyse complétées.....	93
I. Tableaux d’aide à l’analyse.....	111

Les Plans de Gestion de Données entre les mains des chercheurs, dans le contexte de l'ouverture des données de la recherche

Présenté par Alicia Maubailly

Le 5 juillet 2021

Résumé français

Le récent mouvement de la science ouverte en France a contribué à la mise en place de questionnements autour des données de la recherche et de leur ouverture. Suite à cela, on a vu se développer ces dernières années les plans de gestion de données, un document devenu obligatoire dans bon nombre de projets financés. Or, avec cette obligation progressive de produire des PGD, de nouvelles pratiques sont amenées à apparaître au sein de la communauté scientifique. L'enjeu est de savoir si ces PGD peuvent effectivement jouer un rôle dans la transmission des enjeux des données de la recherche auprès des chercheurs. Pour cela, l'étude s'appuie sur un échantillon d'entretiens et sur un corpus de PGD.

Mots-clés : Plan de Gestion de Données (PGD), données de la recherche, science ouverte, Horizon 2020, Recherche scientifique

Résumé anglais

Open science in France is a recent movement who has contributed to the development of questions around research data and their opening. Has a result, data management plans have been developed in recent years and have become mandatory to many funded projects. However, with the progressive obligation to produce DMPs, new practices are bound to appear within the scientific community. Now, the issue is to know if these DMPs can effectively play a rôle in the transmission of research data issues to researchers. For this, the study relies on a sample of interviews and a corpus of DMPs.

Key-words : Data Management Plan (DMP), research data, open science, Horizon 2020, research