



HAL
open science

Construire de nouveaux modes de connaissances autour des données de la recherche

Candice Fillaud

► **To cite this version:**

Candice Fillaud. Construire de nouveaux modes de connaissances autour des données de la recherche. domain_shs.info.medi. 2022. mem_04135415

HAL Id: mem_04135415

https://memic.ccsd.cnrs.fr/mem_04135415v1

Submitted on 20 Jun 2023

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



Distributed under a Creative Commons Attribution - NonCommercial - ShareAlike 4.0 International License

Master Information et Médiation Scientifique et Technique

Parcours Épistémologie et Ingénierie de la Science Ouvverte



Mémoire de Master

**Construire de nouveaux modes de
connaissances autour des données de la
recherche**

Candice Fillaud

Année universitaire 2021/2022

Sous la direction de Chérifa Boukacem-Zeghmouri

Professeure en Sciences de l'Information et de la Communication

DATACTIVIST

Tuteur professionnel : Antoine Blanchard, consultant senior



Remerciements

Je souhaite avant tout remercier mon tuteur professionnel, Antoine Blanchard, merci de m'avoir encadré tout au long de mon année d'alternance au sein de Dataactivst. Je le remercie pour sa grande pédagogie, sa disponibilité, sa rigueur et ses judicieux conseils. Ce fut une année très riche durant laquelle j'ai énormément appris professionnellement grâce à ses éclairages et ses critiques constructives. Merci à Antoine d'avoir partagé ses connaissances, son expérience professionnelle et de m'avoir apporté de nombreux outils méthodologiques.

Merci également à tous les membres de Dataactivst que j'ai pu côtoyer cette année, merci pour leur accueil chaleureux et les échanges très enrichissants que j'ai pu partager. Merci pour les partages d'expériences et de connaissances. Je remercie spécifiquement les personnes avec lesquelles j'ai travaillé sur la mission Sprint PGD : Arthur Sarazin, Julia Dumont et Mathieu Morey.

Je tiens à remercier grandement ma tutrice pédagogique Chérifa Boukacem-Zeghmouri. Je la remercie pour son accompagnement jusqu'au bout et les précieux échanges que nous avons eus. Merci également aux enseignants chercheurs du Master IMST EISO qui ont su nourrir les réflexions de ce mémoire par leurs enseignements.

Je voudrais remercier Amélie Barrio pour la relecture de mon mémoire et les échanges durant l'année, ça a toujours été un plaisir d'échanger, merci pour ses conseils.

Je remercie l'ensemble de ma promotion IMST EISO, en particulier Jérémie Buttard, Thomas Chavanon et Silvia Pampani pour l'entraide quotidienne et leur soutien.

Enfin, je souhaite remercier mes proches et ma famille de m'avoir accompagné durant ces deux années de Master et de m'avoir encouragé dans l'élaboration de ce mémoire.

Résumé

Les données de la recherche sont un des nouveaux enjeux des politiques de science ouverte. À travers le monde, nous assistons au développement de la culture de la donnée dans l'écosystème scientifique, propulsé par une demande d'ouverture des données de la recherche et le caractère obligatoire des plans de gestion de données (PGD). Ce mémoire s'intéresse à l'insertion de la culture de la donnée dans l'écosystème scientifique, avec les normes qui l'accompagne. Il prend appui sur la conceptualisation d'un atelier d'accompagnement à la rédaction de PGD afin de mettre en lumière les enjeux du développement d'une telle formation. Enfin, ce mémoire évoque les nouvelles mutations et les besoins liés à cette transition de la science vers la science ouverte en matière de données.

Mots-clés

Données de la recherche ; Principes FAIR ; Plan de gestion des données ; Gestion des données de la recherche ; Science ouverte ; Services d'accompagnement à la gestion des données

Abstract

Research data is one of the contemporary matters of open science policies. Propelled by a demand in open research data and the mandatory nature of data management plans (DMP), we are witnessing the development of data culture in the scientific ecosystem across the globe. This master's thesis focuses on the insertion of data culture—and all its associated standards—in the scientific ecosystem. It builds on the conceptualization of a support workshop with regard to the drafting of DMPs to highlight the challenges of developing such schooling. Finally, this thesis discusses the new mutations and the needs linked to the transition of science towards open science in terms of data.

Keywords

Research data ; FAIR data principles ; Data management plan ; research data management ; Open science ; Research data management services

SOMMAIRE

<i>Remerciements</i>	3
<i>Résumé</i>	4
<i>Mots-clés</i>	4
<i>Introduction</i>	6
I. Prolongement du champ scientifique par la culture de la donnée	9
1. Le champ « invisible » des données de la recherche	9
2. Culture de la donnée « en action ».....	15
3. Processus de normalisation des principes FAIR.....	19
II. Construction d'une formation d'accompagnement à la rédaction d'un PGD : méthodologie et outils	25
1. Les enjeux liés à la formation aux données de la recherche	25
2. Exemple d'action de formation : un Sprint PGD.....	29
3. Bilan et perspectives	36
III. L'écosystème scientifique en transition	41
1. Les pratiques de gestion face à une transition vers la science ouverte	41
2. Mise en œuvre de nouveaux dispositifs facilitant la transition.....	46
3. Transformation des services d'accompagnement à la gestion des données.....	49
<i>Conclusion</i>	52
<i>Bibliographie</i>	54
<i>Annexes</i>	63
<i>Annexe 1 : Déroulé d'animation de l'atelier Sprint PGD</i>	63
<i>Annexe 2 : Tableau récapitulatif des évolutions du Sprint PGD</i>	64
<i>Annexe 2 : Questionnaire d'évaluation du Sprint PGD</i>	66
<i>Annexe 3 : Tableau des réponses au questionnaire d'évaluation de l'atelier</i>	67
<i>Annexe 4 : Captures d'écran du nouveau parcours PGD Open Data Canvas</i>	68
<i>Annexe 5 : Logigramme de communicabilité d'une donnée</i>	71

Introduction

“C’est la puissance des données combinées à leur fragilité qui en fait un sujet d’étude fascinant dans la communication savante”

C.Borgman

Traduction de C.Matoussowsky

Date de l’entrée en vigueur de la Loi pour une République Numérique, le 9 octobre 2016 marque un nouveau tournant dans l’ouverture des savoirs scientifiques, notamment dans celui des données de la recherche :

« Art. L. 533-4.-II.-Dès lors que les données issues d’une activité de recherche financée au moins pour moitié par des dotations de l’État, des collectivités territoriales, des établissements publics, des subventions d’agences de financement nationales ou par des fonds de l’Union européenne ne sont pas protégées par un droit spécifique ou une réglementation particulière et qu’elles ont été rendues publiques par le chercheur, l’établissement ou l’organisme de recherche, leur réutilisation est libre.

« III. L’éditeur d’un écrit scientifique mentionné au I ne peut limiter la réutilisation des données de la recherche rendues publiques dans le cadre de sa publication.

Cette loi¹ permet aux chercheurs de reprendre leurs droits sur leurs publications ainsi que sur leurs données en dépit des éditeurs scientifiques (Robin 2017). Elle autorise les chercheurs à publier leurs articles dans des archives ouvertes, comme HAL, la plus couramment utilisée. Dans une logique d’ouverture des savoirs et dans un mouvement international de libre accès à la connaissance, cette loi promeut également l’ouverture des données de manière publique à des fins de valorisation et réutilisation. Cette nouveauté marque une transition dans le monde de la recherche scientifique. Il s’en est suivi une cascade d’évolutions dans les politiques de recherche, prenant exemple sur les instances européennes et internationales. Le 4 juillet 2018 pour la conférence européenne des bibliothèques de recherche (LIBER) à Lille, Frédérique Vidal annonce le premier Plan National pour la Science Ouverte (PNSO) définissant un nouveau cadre en matière de données de la recherche en rendant obligatoire l’ouverture des publications et des données pour tout projet de recherche financé sur fond public. Trois ans plus tard, le deuxième Plan National pour la Science Ouverte (PNSO) est annoncé, insufflant une nouvelle dynamique en incluant l’ouverture des codes

¹ Décret d’application de la loi :

<https://www.legifrance.gouv.fr/dossierlegislatif/JORFDOLE000031589829/?detailType=ECHEANCIER&detailId=>

source et protocoles de recherche. Le deuxième axe du deuxième PNSO, consacré aux données de la recherche, annonce notamment la création d'une nouvelle plateforme fédérée de dépôt des données : Recherche Data Gouv, inaugurée le 8 juillet 2022². Une généralisation de l'utilisation des Plans de Gestion des Données (PGD) est annoncée également. Le PGD, ce document de planification répondant à une structure précise, décrit les conditions de réutilisation et de valorisation des données d'un projet de recherche, nécessaires à leur ouverture conformément au principe « *aussi ouvert que possible, aussi fermé que nécessaire* ». Déjà implanté en Europe pour les appels à projets européens (H2020, Horizon Europe), qui disposent que tout projet de recherche financé s'accompagne d'un PGD, ce document l'est également en France. L'Agence Nationale de la Recherche (ANR) rend obligatoire le PGD en 2019. Un PGD permet de cadrer et d'aider le chercheur à définir une feuille de route pour gérer ses données produites ou collectées dans le cadre de son projet de recherche. Un peu plus tard, le 3 décembre 2021, le décret n° 2021-1572 relatif au respect des exigences de l'intégrité scientifique marque une nouvelle étape, propulsant les PGD comme « *l'outil central pour déterminer quelles données sont nécessaires à une recherche, quelles données peuvent être partagées et où, quel est leur devenir à moyen terme (conservation et archivage)* » (Blanchard, 2022). Depuis le premier PNSO il est également question de la « FAIRisation » des données des chercheurs, les principes FAIR apparaissent comme des règles à suivre pour rendre les données faciles à trouver, accessibles, interopérables et réutilisables.

Ces nouvelles politiques de gestion et diffusion des données ont amené à de nouvelles pratiques chez un bon nombre d'acteurs de la recherche scientifique, que ce soit les chercheurs, les doctorants, les ingénieurs de recherche et projet, les personnels d'accompagnement à la gestion des données... Toute personne prenant part au bon fonctionnement de l'écosystème scientifique a pu faire évoluer ses pratiques, apprendre ou réapprendre à gérer ses données en y instaurant de « bonnes pratiques ». Cette centralisation autour des données est donc le marqueur de transitions auxquelles le monde scientifique fait face, définissant de nouvelles modalités de construction de connaissances scientifiques.

L'analyse proposée dans ce mémoire est issue de ma formation et de mon expérience professionnelle acquises ces deux dernières années dans le cadre de mon Master. Durant ces deux dernières années j'ai pu m'intéresser spécifiquement aux données de la recherche, d'abord au cours de mon stage de M1³ où j'ai réalisé la première version du PGD de l'institut de recherche ANITI (Artificial and Natural Intelligence Toulouse Institute)⁴, fondé sur une approche qualitative, qui a permis de réaliser des entretiens semi-directifs avec des chercheurs et doctorants de la communauté

² <https://www.ouvrirlascience.fr/inauguration-de-recherche-data-gouv/>

³ Stage d'une durée de six mois

⁴ <https://aniti.univ-toulouse.fr/>

en intelligence artificielle. Ensuite, pour mon Master 2, dans le cadre de mon alternance qui a débuté en septembre 2021 au sein de la société coopérative Dataactivist⁵, j'ai pu aborder la notion des données de la recherche sous un nouveau prisme : celui du développement de nouveaux modes de connaissances, avec la conceptualisation d'un atelier d'accompagnement à la rédaction de PGD, mêlant des aspects de formation et d'acculturation aux bonnes pratiques de gestion des données pour les chercheurs et doctorants.

Ces expériences qui ont forgé le parcours de ces deux dernières années m'amènent à considérer les données de la recherche avec une dimension plus profonde comprenant les évolutions de la pluralité des enjeux liés à la question de la gestion des données. De cette expérience découle la problématique suivante : comment observer auprès des chercheurs, et comment révéler la place des données dans des modalités de préfiguration et de construction de connaissances ? Peut-on identifier ou comprendre une évolution dans la forme de constitution des savoirs ?

Ce mémoire s'inscrit donc au carrefour de deux enjeux : d'une part celui de l'ingénierie avec un aspect plus empirique, dont l'expérience de conceptualisation d'un atelier par méthodologie itérative et d'autre part, celui de l'épistémologie qui touche aux conditions d'élaboration de la connaissance. La première partie de ce mémoire traitera de la culture de la donnée, de son insertion dans le champ scientifique au processus de normalisation des principes FAIR. La deuxième partie, inspirée de mon expérience au sein de Dataactivist décrira la construction d'un nouveau mode de connaissance en lien avec l'atelier Sprint PGD que j'ai contribué à réaliser j'ai contribué à la réalisation ainsi que les perspectives associées. La troisième et dernière partie amorcera la notion de transition de l'écosystème scientifique actuel, en passant par les pratiques de gestion, les nouveaux dispositifs mis en place ainsi que l'évolution des services d'accompagnement à la gestion des données.

⁵ <https://dataactivist.coop/fr/>

I. Prolongement du champ scientifique par la culture de la donnée

L'écosystème scientifique est défini selon deux modèles incontournables et complémentaires par les sociologues Robert K. Merton et Pierre Bourdieu. Le premier modèle aborde dans les années 40, la formalisation des valeurs de la science par Merton suite à l'aboutissement d'un long travail d'analyse à son époque. Ensuite, et de manière complémentaire, Pierre Bourdieu apporte une vision plus proche de notre époque et formalise le champ scientifique (Bourdieu, 1976) au moment où la science a connu une évolution vers une « *Big science* ».

Dans l'univers scientifique, les chercheurs sont confrontés à de véritables enjeux de concurrence pour le financement de leurs recherches et pour la publication de leurs articles. Depuis peu, la culture de la donnée de la recherche se répand et tend à s'intégrer à ce champ qu'est la science. La notion de "culture de la donnée" reste très complexe à définir, Robert K. Merton propose une définition de la culture que nous utiliserons : « *La culture, c'est la compréhension des configurations historiques et singulières des idéaux, des valeurs et principes normatifs* ». Dans cet ensemble, la compréhension des configurations historiques reliée à la science, c'est que la science a ses pratiques et ses traditions qui représente les configurations historiques. La compréhension des configurations singulières représente la compréhension des spécificités de la science par les différentes disciplines scientifiques. La culture de la donnée prend en compte la compréhension de l'histoire des pratiques de gestion des données, les traditions scientifiques et les spécificités de cette culture. Dans un sens, lorsqu'on produit de la donnée, on ne produit pas seulement des « données » mais on produit tout le contexte et donc toute la culture qui lui est associée.

Cette partie permet de comprendre la position actuelle de la culture de la donnée dans le champ scientifique et comment elle se retrouve en tension entre les visions mertonienne et bourdieusienne de l'écosystème scientifique, ainsi que les politiques et les cadres gestionnaires et documentaires.

1. Le champ « invisible » des données de la recherche

Considéré comme fondateur de la sociologie des sciences, Robert K. Merton apporte une vision de la science dès les années 40 fondé sur ses valeurs, avec un monde scientifique idéal au travers d'un ensemble de caractères et de normes éthiques, appelés « Ethos de la science », qu'il a commencé à formaliser dès les années 1940 en faisant le lien entre la science et la démocratie : l'universalisme, le communalisme, le désintéressement et le scepticisme organisé (Merton, 1942). En 1976, le sociologue Pierre Bourdieu va venir éclairer une autre facette de la science et introduire sa vision plus proche de notre époque. Il définit le monde scientifique comme un champ social à part entière prenant en considération les rapports de force et les luttes qui constituent le système

scientifique, les monopoles, les stratégies, les intérêts et les profits. La vision bourdieusienne raisonne en complémentarité de celle de Merton, qui était surtout descriptive des valeurs de la science dans les années 40. Au sein du champ, Bourdieu y définit l'enjeu spécifique pour lequel luttent les agents du champ : le monopole de l'autorité scientifique. C'est-à-dire le prestige, la reconnaissance, la célébrité, autrement dit : asseoir sa légitimité scientifique. Basé sur une économie de capitalisation, le capital social du champ scientifique est donc le capital de reconnaissance symbolique. Bourdieu présente le champ scientifique comme un lieu marqué par de nombreuses luttes et inégalités avec une hiérarchisation des acteurs qui le compose : dominants et dominés (ou encore nouveaux entrants). Dès lors, des stratégies individuelles et politiques sont activement mises en place, d'un côté les dominants avec des positions plus hautes dans la structure de distribution du capital scientifique, de l'autre les nouveaux entrants. Cette opposition est marquée par des stratégies antagonistes ; d'un côté une stratégie de conservation de place pour les dominants et de l'autre, une stratégie subversive pour les nouveaux entrants. Sa vision met en lumière une réalité du monde scientifique d'aujourd'hui et de ses phénomènes de domination qu'il est tout à fait possible d'observer et d'analyser. La notion de champ scientifique employée par Bourdieu est constante et immuable mais à la fois modifiée par ceux qui font les règles, les agents les plus haut dans la structure de distribution : les dominants. Sur l'étude du champ scientifique, Yves Gingras précise :

« Les stratégies des chercheurs (choix des objets d'étude, des lieux de publication) se comprennent, dans ce modèle, comme des investissements en référence à une anticipation (consciente ou inconsciente) des chances probables de profit symbolique (reconnaissance, prix) » (Gingras, 2017)

Les luttes principales des chercheurs se déploient afin d'acquérir des financements pour leurs recherches : plus un chercheur publiera dans des revues prestigieuses, plus l'auteur sera cité, plus il pourra acquérir des financements. De ces luttes et de la pression à la publication apparaît l'expression *publish or perish* témoignant la pression à la publication dans le milieu académique. Le premier usage dans ce contexte apparaît en 1942 relevé par Eugène Garfield (Garfield, 1996), ingénieur connu pour être à l'origine de la bibliométrie et scientométrie. Les métriques de citations, la bibliométrie et la scientométrie s'intègrent dans le champ scientifique et témoignent à leur tour de la concurrence entre chercheurs pour la reconnaissance.

Bien que le modèle bourdieusien se rapproche le plus de ce qu'est la science à notre époque, les deux cadres ont symboliquement leur place dans la définition, l'observation et l'interprétation du champ scientifique contemporain. En effet, l'un et l'autre indépendamment ne sont pas représentatifs de la science dans sa globalité, en revanche ils forment une complémentarité qui permet de définir l'écosystème scientifique dans son ensemble. L'écosystème scientifique n'est pas uniquement un champ de lutte tel que Bourdieu le décrit mais un écosystème avec une structure sociale et une culture historique fondée sur des valeurs historiques comme le définit Merton. Les

données de la recherche témoignent de cette complémentarité puisque dans ce cas précis les deux visions sont valables. La volonté de partager les données correspond aux valeurs de Merton quant à la réalité de terrain, si nous prenons pour exemple les craintes des chercheurs ou la marchandisation des données, cela correspond davantage à la vision des luttes de Bourdieu.

De ces modèles, nous pouvons alors nous interroger sur la place de la culture de la donnée qui depuis la progression du numérique ne cesse de prendre en valeur en constituant une nouvelle dimension économique en plus d'être au cœur des politiques. La culture de la donnée fait partie intégrante de l'économie du numérique :

« L'économie numérique regroupe des champs différents selon les acteurs et observateurs du secteur. Elle peut se définir comme l'ensemble des activités relatives aux Technologies de l'Information et de la Communication (TIC), à la production et à la vente de produits et services numériques. »
(economie.gouv.fr)

L'essor des données massives ou « *Big Data* » s'est retrouvé au cœur des stratégies industrielles, les données ont trouvé une valeur significative par leur utilisation. La donnée en tant que telle n'a pas de valeur marchande, c'est sa gestion associée au contexte qui en est le capital. Les données numériques ont souvent été qualifiées « d'or noir » en raison de la marchandisation de celles-ci, comme l'enregistrement et le traitement de données personnelles⁶ pour le e-commerce. L'exploitation des données par les GAFAM⁷ et d'autres entreprises privées constituent un nouvel espace économique où les données personnelles deviennent une marchandise. Les informations personnelles des utilisateurs d'internet ou d'application voient leurs données utilisées de manière implicite en autorisant dès l'installation l'accès à des données comme leur historique de conversation, leur carnet d'adresse, leur historique de navigation, leur géolocalisation (Fontaine et al. 2017) etc...

Grâce à l'accumulation de données de milliers d'utilisateurs, des analyses statistiques permettent de définir des comportements d'utilisateurs ou de cibler des produits pour la publicité. En Europe, le Règlement général à la protection des données (RGPD) oblige les entreprises qui stockent les données utilisées sur internet à pouvoir les supprimer à la demande des utilisateurs à n'importe quel moment. Depuis le 18 mai 2018 (Vétois, 2018), les normes RGPD sont entrées en application en Europe, obligeant ainsi les GAFAM à se plier à ces règles et repenser leurs politiques. Afin de protéger les données des utilisateurs européens, les GAFAM seront soumis à des condamnations

⁶ « Une donnée personnelle est toute information se rapportant à une personne physique identifiée ou identifiable. Mais, parce qu'elles concernent des personnes, celles-ci doivent en conserver la maîtrise. » CNIL.

⁷ Google (Alphabet), Apple, Facebook (Meta), Amazon et Microsoft

en justice ainsi que de lourdes sanctions financières si elles ne les respectent pas (jusqu'à 4% du chiffre d'affaires de l'entreprise) (Forgeron, 2017).

Encore invisible il y a quelques années, c'est grâce à l'avènement de l'*open data*, et du boom du numérique que les données numériques se retrouvent au centre de nombreux débats et mutations comme en témoignent les stratégies politiques en faveur de l'ouverture des données. Les données prennent part à l'économie générale de la science, indiquant que nous sommes à un moment charnière :

« Cette nouvelle attention aux données s'inscrit dans un contexte d'évolution de l'économie générale de la science, marqué notamment par la promotion du financement de la recherche par projets. » (Galonnier et al. 2019)

Les travaux de Violaine Rebouillat (2019) montrent que les données de la recherche sont associées à une pluralité de valeurs : valeur normative, valeur économique, valeur d'usage... Les politiques publiques voient en l'ouverture des données un potentiel de réutilisation qui contribue à la transparence de la recherche et l'innovation et donc à la croissance économique. Les données sont considérées par les chercheurs qui les produisent comme des ressources « précieuses » ; les chercheurs évoquent même un lien de paternité pour leurs données. La valeur d'usage des données est associée aux publications des chercheurs, les données sont utiles et fonctionnent comme élément de preuve pour les résultats des scientifiques.

Selon la définition bourdieusienne du champ scientifique, entre luttes et rapports de force, la culture des données de la recherche peut-elle rejoindre les luttes du champ scientifique ? Dès lors que les données de la recherche sont soumises à des rapports de domination et de concurrence, elles vont s'insérer dans le champ scientifique conceptualisé par Bourdieu. Étant donné leurs valeurs et leur rôle central dans les politiques, les données de la recherche remplissent tous les critères du capital scientifique. Tout comme ils sont déjà soumis à une course, dorénavant les chercheurs sont poussés à ouvrir leurs données par les agences de financement et les revues sous couvert de l'argument « *public money, public data* ». Entre les chercheurs, le partage des données n'est pas réalisé de manière automatique, des réticences de la part des chercheurs perdurent, certains craignent de se faire « voler » leurs données, d'autres ne se rendent pas compte de la valeur de leurs données, et méconnaissent les manières d'ouvrir les données comme le montre cette étude (Charrier, 2020) sur les pratiques et les usages relatifs aux outils numériques.

La notion de paternité sur les données montre la concurrence à laquelle sont confrontés les chercheurs. Un chercheur qui partage ses données peut éprouver la crainte que d'autres chercheurs les utilisent, et publient un article qui sera plus cité ou reconnu dans une revue plus prestigieuse. Une autre crainte des chercheurs est le partage de leurs données associées aux résultats et l'influence

que cela peut avoir sur leur carrière. Par exemple, si lors de la publication d'un article, une erreur est relevée (sur la méthode, les instruments etc...) cela engendre la crainte de voir son article rétracté et donc de perdre en autorité scientifique, perdre en reconnaissance et donc en capital social comme le désigne Bourdieu. Le chercheur qui partage ses données dans des entrepôts, les met à disposition et permet de potentielles réutilisations. La réutilisation des données rejoint d'une certaine manière le principe des citations où les chercheurs vont citer les travaux d'autres scientifiques, alors si ses données sont disponibles, correctement documentées et réutilisées pour d'autres projets, le chercheur sera cité pour ses données et gagnera en capital social.

La culture de la donnée se nourrit également par d'une nouvelle forme de publication, les *data papers*. Tout comme les publications, ils sont un moyen de partager le travail des chercheurs, mais ne sont pas encore pris en compte dans les évaluations individuelles. Ces articles sont publiés dans des revues spécialisées et peuvent être cités par les pairs. Au-delà de la publication des données dans un entrepôt à des fins de réutilisation, les chercheurs trouvent aussi satisfaction à ce nouveau mode de publication comme en témoigne le chercheur Jérémie Garniaux dans un tweet publié le 25 juillet 2022 : « *En train de contribuer à mon premier data paper! Il y a quelque chose de vraiment satisfaisant à décrire avec une rigueur scientifique les processus & méthodes d'acquisition ou de structuration d'un jeu de données. Open science for the win!* » (Garniaux, 2022). Ce témoignage renforce aussi l'idée que la culture scientifique s'inscrit dans les deux visions de Merton et Bourdieu. D'un côté Jérémie Garniaux se satisfait d'un projet de publication, le *data paper*, donc d'une reconnaissance symbolique qui est le fait de publier, de l'autre le fait il y a la satisfaction d'ouvrir ses données par le biais du *data paper* et de pouvoir y décrire rigoureusement les processus et méthodes d'acquisition, soit participer au partage de la science, ce qui rejoint les valeurs de Merton. Dans un autre contexte, l'article de blog « L'ouverture des données de recherche : un retour aux sources de l'Ethos de la Science ? » écrit par Lionel Maurel (calimaq), juriste de formation et conservateur des bibliothèques, revient sur son intervention lors du colloque « Intégrité scientifique et Science Ouverte » organisé par l'Office Français de l'Intégrité Scientifique (OFIS) le 4 avril 2019. La session dans laquelle Lionel Maurel intervenait concernait l'ouverture des données de la recherche, pour rappeler que « les grands principes de la Science Ouverte offrent l'opportunité d'un retour aux sources même de l'Ethos de la Science » (Maurel, 2019).

Cette analyse vient précisément corroborer l'idée que j'ai avancée plus haut selon laquelle la culture de la donnée est un ensemble qui se construit entre la vision de Merton et celle de Bourdieu.

L'existence du Prix science ouverte « données de la recherche » montre que désormais les données s'inscrivent dans la dimension du capital social et de lutte pour l'autorité scientifique. Les récompenses ont eu lieu le 8 juillet 2022 par Sylvie Retailleau lors de l'inauguration de l'entrepôt

national fédéré Recherche Data Gouv, le lauréat pour le Prix science ouverte données de la recherche est Victor Gay, professeur adjoint à l'École d'économie de Toulouse et chercheur à l'Institut d'études avancées de Toulouse dans la catégorie « jeunes chercheurs pour la réutilisation des données ». Ces prix Il existe désormais une concurrence qui pousse les chercheurs à ouvrir leurs données. Dans le cadre des données le rapport dominant/dominé peut juste être moins prononcé que dans le champ scientifique, en matière de publication. En effet, dans la culture et dans les pratiques de la donnée, ceux qui manipulent de la donnée sont principalement les jeunes chercheurs donc les doctorants qui sont en première ligne sur les expérimentations particulièrement dans les domaines scientifiques comme les STM (*Science Technology and Medicine*).

Les jeunes chercheurs sont plus enclins à gérer les données puisqu'ils s'y attèlent dès le début de leur parcours en intégrant les savoirs faire techniques, les standards, les plateformes, les bonnes pratiques et le partage. À l'échelle d'un laboratoire, les jeunes chercheurs partagent les données et partagent aussi leur début de culture de la donnée au sein des laboratoires, ce sont souvent eux qui amènent aux seniors les bonnes pratiques. Cependant, les chercheurs expérimentés restent les décisionnaires sur la gestion des données, ce sont eux qui dominent sur la culture du partage (où partager ses données, quand les partager...). Les travaux montrent que si les jeunes chercheurs sont préparés à ouvrir les données, ce sont les seniors, qui vont les acculturer à la prudence, voire à la défiance et donc à freiner le partage des données, ceci pour ne pas se les faire voler et donc les alerter sur la paternité concernant leurs travaux (Nicholas et al. 2020). Même si les seniors restent les principaux décisionnaires, les jeunes chercheurs bénéficient d'expérience et de formations plus tôt dans leur carrière ce qui atténue le rapport dominant/dominé qu'il peut y avoir par rapport aux publications.

2. Culture de la donnée « en action »

L'insertion d'une culture de la donnée généralisée dans le champ scientifique résulte de multiples actions et évolutions législatives. Les motivations du partage et de l'ouverture des données se rassemblent dans quatre grands raisonnements d'ouverture (Borgman, 2020 p. 256) :

1. Pour reproduire la recherche ou vérifier la recherche
2. Pour mettre des ressources publiques à la disposition de la population
3. Pour permettre d'autres à poser de nouvelles questions sur les données existantes
4. Pour faire progresser la recherche et l'innovation

À l'échelle mondiale, la politique de la science ouverte se développe incluant la culture de la donnée directement associée puisque l'ouverture de la science passe par l'ouverture des données de la recherche. Ce nouvel environnement ouvert s'est en fait formé sous les nouvelles contraintes et obligations déterminées par les agences de financement, les éditeurs, les comités d'éthique et les délégués à la protection des données (DPD ou *DPO* en anglais).

Les raisonnements d'ouverture assimilés par les établissements publics et privés ont laissé place à de nouveaux moyens de formation et de nouvelles infrastructures. Nous nous demanderons dans cette partie, comment la culture de la donnée s'est installée au sein de l'écosystème scientifique. Sa mise en place a-t-elle modifié les formes de production des savoirs ?

Les données de la recherche sont mentionnées dans la plupart des feuilles de route des établissements, le Centre National de la Recherche Scientifique (CNRS) a engagé un plan d'action en 2020 pour « booster » l'ouverture des données des chercheurs, notamment dans les pratiques, en formant les chercheurs et les personnels à la gestion des données. Le cadre législatif actuel participe au prolongement de la culture de la donnée dans le champ scientifique puisqu'il pousse les chercheurs à ouvrir leurs données autant que faire se peut. L'adoption des nouvelles politiques prévoit des recommandations sur les pratiques de gestion, comme l'application des principes FAIR (cf. partie I.3). L'application des politiques peut différer selon les établissements de recherche, et institutions puisqu'ils établissent leurs propres politiques de partage et dépôt des données ainsi que leurs propres chartes en matière de science ouverte. Les initiatives européennes telles que l'European Open Science Cloud (EOSC), GO FAIR, Research Data Alliance (RDA) sont en train de dessiner l'histoire de la culture de la donnée en y contribuant et en répandant celle-ci.

L'EOSC est un projet de la Commission européenne, officiellement lancé en 2018 visant à développer un dispositif numérique pluridisciplinaire permettant aux utilisateurs (chercheurs, ingénieurs, entreprises, citoyens européens) d'échanger les pratiques de science ouverte, de publier et réutiliser des données de la recherche en soutenant trois objectifs :

« 1. Accroître la valeur des données scientifiques en les rendant facilement accessibles à 1,7 million de chercheurs européens et à 70 millions de professionnels des sciences, de la technologie, des sciences humaines et sociales grâce à un environnement virtuel doté de services ouverts et intégrés pour le stockage, la gestion, l'analyse et la réutilisation des données de recherche, par-delà des disciplines scientifiques (interdisciplinarité) et les frontières,

2. Réduire les coûts de la gestion des données scientifiques en fédérant les infrastructures de données scientifiques existantes, actuellement dispersées entre disciplines et États membres,

3. Tout en assurant une protection adéquate des informations et données personnelles conformément aux règles communautaires applicables. » (Ouvrirlascience.fr)

L'EOSC a pu développer un portail Web, ce portail permet aux utilisateurs d'accéder aux services de gestion des données et aux autres ressources des infrastructures de recherche publiques nationales, régionales et institutionnelles en Europe, afin que les utilisateurs puissent bénéficier des avantages d'une coalition européenne axée sur la science sur les données. En 2017, la France a rejoint l'initiative GO FAIR en tant que co-fondatrice, afin de développer les principes FAIR et notamment « ouvrir progressivement les données de la recherche existantes au sein des institutions scientifiques et académiques dans tous les domaines de la recherche et au-delà des frontières nationales. »

Ce partenariat témoigne de l'implication de la France dans une dynamique internationale axée sur les données de la recherche et contribuant également aux développements de l'EOSC. L'adhésion à ces initiatives permet à la France de s'assembler avec l'Europe et permet ainsi de faciliter les échanges et l'interopérabilité des données de la recherche tout en perpétuant la démarche FAIR.

À l'échelle internationale, l'organisation Research Data Alliance - RDA est une initiative créée en mars 2013 par la Commission européenne, la National Science Foundation et le National Institute of Standards and Technology du gouvernement des États-Unis, et le ministère de l'Innovation du gouvernement australien. Cette organisation rassemble des membres de 137 pays alliant des communautés, des groupes d'intérêt groupes de travail au sein des pays. La création de cette organisation vise à travailler à échelle mondiale sur les questions de données notamment pour faciliter le partage des données de recherche, de la gestion et de la diffusion. La dimension internationale de la RDA permet également de construire des passerelles technologiques et sociologiques sur le partage des données⁸. En France, c'est le CNRS qui est chargé du développement du Noeud RDA France mis en place dans le cadre du projet RDA Europe 4.0 (2018-2020) dont les coordinatrices sont Françoise Genova astronome et directrice de recherche

⁸ Research Data Alliance : présentation générale. [Vidéo]. Canal-U. <https://www.canal-u.tv/115980>. (Consultée le 12 juillet 2022)

au CNRS et Laurence El Khouri Ingénieure de recherche à la Direction des Données Ouvertes de la Recherche. En France, les membres du Noeud participent à construire une base de travail en formant des groupes de travail et des groupes d'intérêt abordant tous les sujets possibles se rapportant au partage des données scientifiques. Les groupes de travail proposent des recommandations implémentables et les groupes d'intérêt peuvent travailler sur des enquêtes ou proposer des guides de bonnes pratiques sur des thèmes plus spécialisés. Des journées annuelles, des ateliers et des webinaires sont organisés sur différentes thématiques, ouverts à toutes et tous. Il est également possible de s'inscrire aux listes de diffusion et de discussion du Noeud RDA.

La société coopérative et participative Dataactivist dans laquelle j'ai effectué mon alternance cette année participe ainsi à accroître la culture de la donnée. Cette société participe plus précisément à ouvrir les données, les rendre utiles et utilisables⁹ en travaillant avec des collectivités territoriales, des administrations, des entreprises ou encore des universités. Depuis l'année 2020, Dataactivist accompagne un nombre croissant d'acteurs de l'enseignement supérieur, de la recherche et de l'innovation (ESRI) qui élaborent ou mettent en œuvre leur stratégie science ouverte et se mettent en conformité avec les exigences de leurs financeurs en matière d'*open data* et de gestion des données de recherche.

La culture de la donnée se construit dans un contexte international, grâce aux initiatives menées sur les différentes échelles (nationale, européenne, mondiale) et les passerelles qui se développent. En créant des passerelles internationales, il y a une volonté de construire une culture de la donnée commune. L'insertion de la culture de la donnée est aussi propulsée par les PGD, en effet le PGD est un document décrivant l'ensemble du cycle de vie de la gestion des données : la collecte/réutilisation des données, le traitement des données, l'analyse des données, le stockage et la sauvegarde des données, le partage des données, leur conservation et archivage, ainsi que la réutilisation des données. Il peut être considéré comme un levier pour développer cette culture et un élément clé pour répandre les bonnes pratiques en matière de gestion des données (Schöpfel, 2018).

Les PGD sont des documents demandant aux chercheurs d'appliquer les principes FAIR sur leurs données. L'hétérogénéité disciplinaire ne facilitant pas la généralisation, la gestion des données est propre à chaque discipline selon ses normes et bonnes pratiques. Dans le bilan du deuxième axe concernant les données de la recherche du PNSO¹⁰, il est écrit : « Les communautés disciplinaires sont inégalement familiarisées aux enjeux de la gestion et de l'ouverture des données de recherche et inégalement pourvues en infrastructures et services leur permettant d'y répondre ». Là où les

⁹ Le slogan de Dataactivist est « We open data, we make them useful »

¹⁰ <https://www.ouvrirlascience.fr/wp-content/uploads/2021/07/Bilan-PNSO-2018-2021.pdf>

Construire de nouveaux modes de connaissances autour des données de la recherche Introduction

PGD ont un rôle important à jouer en tant que levier dans la culture des données et permettraient de pallier les disparités de diverses communautés disciplinaires face aux enjeux de structuration, d'accessibilité, de réutilisation, d'interopérabilité, de citation, de partage et d'ouverture des données de la recherche : « La généralisation de l'usage des plans de gestion des données impulsée par les agences de financement de la recherche contribuera également à cette maturation ». Les PGD sont considérés comme norme faisant le lien entre le processus scientifique, l'environnement réglementaire et les données de la recherche.

3. Processus de normalisation des principes FAIR

Les principes FAIR sont issus d'une logique d'amélioration des infrastructures des données de la recherche afin d'aider à standardiser et généraliser la gestion des données. En 2014, à l'initiative du *Data FAIRPort*, un atelier réunit 25 participants comprenant des membres d'infrastructures de recherche et des instituts politiques, des éditeurs, des spécialistes du web sémantique, des innovateurs, des informaticiens et des scientifiques expérimentaux, pour définir une infrastructure mondiale et construire des principes. L'objectif des principes FAIR est présenté comme le moyen de faciliter la découverte des données, de les rendre accessibles, interopérables et réutilisables pour répondre aux demandes des financeurs en matière de données.

Les lignes directrices, appelées *FAIR data principles* ont été publiées sur le site du groupe FORCE11, chaque principe FAIR se déclinant en un ensemble de caractéristiques de référence dont doivent se rapprocher les données et les métadonnées¹¹ pour faciliter leur découverte et leur utilisation par les chercheurs mais aussi par les ordinateurs (machines) afin de faciliter le traitement automatique :

« Pour être faciles à trouver :

- F1. Les (méta)données se voient attribuer un identifiant unique au monde et éternellement persistant.
- F2. Les données sont décrites avec des métadonnées riches.
- F3. Les (méta)données sont enregistrées ou indexées dans une ressource consultable.
- F4. Les métadonnées spécifient l'identifiant des données.

Pour être accessibles :

- A1 Les (méta)données sont récupérables par leur identifiant à l'aide d'un protocole de communication standardisé.
 - A1.1 Le protocole est ouvert, gratuit et universellement implémentable.
 - A1.2 Le protocole permet une procédure d'authentification et d'autorisation, si nécessaire.
- A2 Les métadonnées sont accessibles, même lorsque les données ne sont plus disponibles.

Pour être interopérables :

- I1. Les (méta)données utilisent un langage formel, accessible, partagé et largement applicable pour la représentation des connaissances.
- I2. Les (méta)données utilisent des vocabulaires qui suivent les principes FAIR.
- I3. Les (méta)données incluent des références qualifiées à d'autres (méta)données.

¹¹ Les métadonnées sont des données sur les données.

Pour être réutilisables :

R1. Les (méta)données ont une pluralité d'attributs précis et pertinents.

R1.1. Les (méta)données sont publiées avec une licence d'utilisation des données claire et accessible.

R1.2. Les (méta)données sont associées à leur provenance.

R1.3. Les (méta)données respectent les normes communautaires applicables au domaine. »

Alors que la promotion de l'ouverture des données se généralise et est présente dans toutes les politiques, il ne s'agit donc plus de seulement diffuser et ouvrir ses données mais de les rendre FAIR. Ces principes servent à répondre à la critique habituelle du mouvement d'*open data* en général où les données sont certes ouvertes mais difficiles à trouver, peu accessibles et interopérables et donc généralement non réutilisables. Des chercheurs ont pu le constater au moment de constituer des corpus de recherche, ce à quoi essaient de répondre certains acteurs comme PROGEDO¹². D'autre part, la diffusion de ces principes et leur promotion dans divers documents de politiques nationales répondent aux besoins d'améliorer la réutilisation des données de recherche pour plus d'efficacité, de transparence et d'intégrité. Dans une étude menée par la Commission Européenne (Directorate-General for Research and Innovation (European Commission), PwC EU Services, 2018), les coûts de l'absence d'une bonne gestion des données, soit des données non « FAIRisées », s'élèvent 3 milliards d'euros juste pour la France. Cette perte est liée aux pertes de temps, à la non-optimisation des coûts de stockage, aux frais de licence, aux problèmes de duplication de la recherche.

L'adhésion à ces principes est internationale et validée dans les plus hauts comités comme lors du G7 en 2016 et G20 (Mons et al. 2017). Par exemple la Commission européenne les adopte dans le cadre des projets financés *Horizon 2020* dans le texte *Guidelines on FAIR Data Management in Horizon 2020* où ils sont érigés comme une base de partage des données :

« This document helps Horizon 2020 beneficiaries make their research data findable, accessible, interoperable and reusable (FAIR), to ensure it is soundly managed. Good research data management is not a goal in itself, but rather the key conduit leading to knowledge discovery and innovation, and to subsequent data and knowledge integration and reuse. »

La rapidité de leur mise en œuvre dans les politiques amène à se demander pour quelles raisons ces principes spécifiques ont-ils fait adhérer le monde scientifique et pourquoi sont-ils devenus universels ?

¹² « Très grande infrastructure de recherche (IGIR), PROGEDO a pour but de développer la culture des données, d'impulser et structurer une politique des données d'enquêtes pour la recherche en sciences sociales. » <https://www.progedo.fr/donnees/quetelet-progedo-diffusion/>

L'incorporation de ces principes en 2018 comme une aspiration du Plan National pour la Science Ouverte, marque leur utilisation en tant que référence en matière de gestion des données : « Notre ambition est de faire en sorte que les données produites par la recherche publique française soient progressivement structurées en conformité avec les principes FAIR (Facile à trouver, Accessible, Interopérable, Réutilisable), préservées et, quand cela est possible, ouvertes ». Afin de promouvoir les principes FAIR et de développer l'ouverture des données, un appel ANR flash « Science ouverte : pratiques de recherche et données ouvertes », est annoncé dans le PNSO1. Dans le Deuxième Plan National pour la science ouverte (PNSO2), ils représentent la sixième mesure : « promouvoir l'adoption d'une politique de données sur l'ensemble du cycle des données de la recherche, pour les rendre faciles à trouver, accessibles, interopérables et réutilisables (FAIR) ». Il est précisé dans le Plan National que ces principes ne s'appliquent pas qu'aux données mais englobent toutes les productions scientifiques en lien avec la gestion des données, que ce soit les outils de collecte, de réutilisation, également les algorithmes, les logiciels. Toutes les données de recherche numériques sont concernées par les principes FAIR ainsi que leurs métadonnées.

Ainsi les principes FAIR ont été associés directement dans les politiques pour définir le cadre d'ouverture des données de la recherche : les données ne doivent pas forcément être ouvertes mais se rapprocher le plus possible d'une gestion FAIR.

Dans l'article de Mons, (Mons et al., 2017) proposant une révision et une clarification des principes FAIR, il est également précisé que ces principes ne constituent en aucun cas une « norme » mais un comportement à avoir pour rendre les données réutilisables. Selon Mons et les autres auteurs, la qualification de « norme » est utilisée à tort puisque ces principes sont permissifs avec de nombreuses approches différentes pour rendre les données FAIR. Les principes FAIR ne spécifient délibérément pas d'exigences techniques, leur exécution est plus ou moins libre et les établissements de recherche qui régissent leurs politiques d'établissements peuvent les suivre et de les moduler selon leurs spécificités. Par exemple, pour rendre ses données faciles à trouver, il est prévu par les principes FAIR que les données soient accompagnées d'un identifiant pérenne. L'identifiant pérenne le plus généralement utilisé est le DOI (*Digital object identifier*), celui-ci est attribué lorsque les données sont déposées sur un entrepôt et il est, en partie, du rôle de l'institution de guider ses chercheurs dans le dépôt de ses données dans un entrepôt spécialisé, qu'il soit disciplinaire, institutionnel, national... En soi, les principes FAIR orientent les politiques mais ne prescrivent pas exactement les moyens et approches pour rendre les données FAIR. Hormis cet exemple et les précisions mentionnées, l'acronyme FAIR est inscrit sur toutes les politiques en gestion des données et pointé comme « la » référence, même si la volonté des auteurs des principes n'était pas de constituer une norme pour le monde scientifique. En s'implémentant dans toutes les ressources liées à la gestion des données, les principes FAIR sont devenus la référence habituelle et une conformité à laquelle les chercheurs devront tendre.

Parmi les différents acteurs de la donnée, comprenant les parties prenantes, les chercheurs sont directement concernés par l'application de ces principes, notamment s'ils veulent réutiliser les données d'autres chercheurs ou améliorer leur gestion des données afin d'acquies des financements. Les agences de financement sont de plus en plus soucieuses des données des chercheurs en demandant l'ouverture de celles-ci dans la mesure du possible et une gestion FAIR. L'ouverture des données passe donc par de bonnes pratiques de gestion, puisque l'ouverture des données n'est utile que si les données sont de bonne qualité, documentées et réutilisables. L'application des principes FAIR dans la gestion des données est donc une compétence importante.

Fabrique de la normalisation *versus* la réalité

Cependant, l'application des principes FAIR se heurte à certains obstacles, que ce soit la connaissance des chercheurs vis-à-vis de ces principes ou l'impact de leur application en matière de standards et autres normes qui s'y ajoutent. Dans les études portant sur la gestion des données dans le milieu de la recherche, les articles ont généralement un périmètre disciplinaire (Inau et al., 2021 ; Persaud et al., 2021 ; Löffler et al., 2021). Par exemple, l'étude de Persaud discute l'importance d'intégrer des bonnes pratiques de gestion dès le début du projet de recherche. Une étude sur la recherche biomédicale (Inau et al., 2021) conclut que la « FAIRisation » des données cliniques est primordiale pour soutenir de bonnes pratiques de recherche basées sur la transparence. Chaque discipline ayant ses propres enjeux, il n'existe que peu d'analyses transversales permettant d'avoir une vision globale sur l'intégration de la culture des données au sein des communautés et sur l'évolution de la gestion des données dans la recherche scientifique.

Dans l'enquête menée par Samuel datant de 2021 sur la « crise de la reproductibilité » (Samuel, König-Ries 2021), une question portant sur l'influence des principes FAIR concernant les pratiques des chercheurs montre que 38% des répondants n'ont jamais entendu parler des principes FAIR ou ne savent pas ce qu'ils signifient. En d'autres termes, malgré la généralisation de ces principes ces dernières années, une acculturation des chercheurs est encore nécessaire. À propos de l'application des principes FAIR, les résultats représentent que parmi les quatre principes, l'interopérabilité des données reste le principe le plus difficile à appliquer. Il est également mentionné que des formations autour des principes FAIR sont nécessaires à leur développement et aideraient à les acculturer les chercheurs.

L'étude de Bloemers présente aussi la situation actuelle et confirme le manque de connaissances des chercheurs vis-à-vis des principes FAIR :

« *Researchers are not well prepared or supported :*

- *Many researchers are not yet familiar with the FAIR principles, or what they mean. Also the concept of machine readable (actionable) information is not known, nor are the practices and tool available ;*
- *Researchers fear that the RFOs requirements lead to additional costs and therefore reduction of research budget ;* » (Bloemers, 2020)

Dans un article de 2018, Joachim Schöpfel propose une réflexion sur le rôle des normes et des standards dans la gestion des données de la recherche (Schöpfel, 2018). En effet, les principes FAIR sont en train de devenir les nouveaux standards de référence pour la gestion des données. L'application de ces principes dépend néanmoins d'autres normes et standards ; la normalisation joue donc un rôle important dans la démarche. Dans l'application des quatre principes, la littérature et les documents officiels demandent de suivre des normes ou standards. Par exemple, si nous détaillons les recommandations de l'axe F ou encore la partie I :

« *Findable*

Pour faciliter la recherche des données, il faut privilégier des standards (normes) à trois niveaux :

- Identifiants standards (DOI, URI, handle...)
- Formats standards pour les métadonnées (DataCite Metadata Schema), enrichis de formats standards disciplinaires
- Mécanismes d'interrogation standards (API, SPARQL, SQL etc.)

Interoperable

Pour augmenter l'interopérabilité des dispositifs et infrastructures des données, la normalisation est requise à deux niveaux :

- Utilisation d'un langage formel et standard pour la représentation des connaissances, comme Web Ontology Language (OWL), Resource Description Framework (RDF) et Simple Knowledge Organization System (SKOS)
- Utilisation de terminologies largement partagées et reconnues comme standards disciplinaires »

Dans les faits, en plus de s'acculturer aux enjeux des principes FAIR, les chercheurs sont donc amenés à devoir respecter d'autres normes et standards sur leurs données pour les appliquer.

Construire de nouveaux modes de connaissances autour des données de la recherche Introduction

En ce qui concerne la raison du succès de ces principes, Schöpfel mentionne :

« [...] il faut garder en tête qu'il s'agit d'abord d'un ensemble de principes et de règles pour interconnecter des machines, une condition nécessaire pour l'exploitation massive du big data de la recherche. Autrement dit, le moteur et principal intéressé d'une telle démarche est l'industrie de l'information, avec les grandes infrastructures, qui a besoin de ces données comme « fuel of economy » (Neelie Kroes, EC) pour créer de la valeur. »

Si les principes FAIR ont autant séduit les éditeurs, financeurs et politiques, il se pourrait que ce soit pour faciliter le traitement des données à des fins économiques notamment concernant l'industrie de l'information. Les principes FAIR représentent donc un témoin de l'inclusion de la culture de la donnée dans le domaine scientifique - et donc le champ scientifique- mené par les données.

II. Construction d'une formation d'accompagnement à la rédaction d'un PGD : méthodologie et outils

Cette partie, plus empirique, se concentre sur les enjeux scientifiques et techniques auxquels peuvent être confrontés la construction de nouveaux modes de connaissances concernant les formations à la gestion des données de la recherche. Nous prendrons en étude de cas la conceptualisation de l'atelier Sprint PGD réalisé lors de mon alternance chez Dataactivist, un atelier d'accompagnement des chercheurs et des doctorants à la rédaction de PGD permettant également d'acculturer aux principes FAIR. De la conceptualisation à la réalisation du Sprint PGD, nous avons développé une méthodologie itérative innovante. Nous discuterons de la réalisation de ce premier Sprint PGD et de l'évaluation réalisée par les participants afin d'analyser les perspectives et les évolutions potentielles de ce type d'atelier.

1. Les enjeux liés à la formation aux données de la recherche

Enjeux relatifs à l'introduction de nouvelles pratiques de gestion de données pour les chercheurs

La publication des principes FAIR reste récente¹³, avec leur incorporation en 2018 comme une aspiration du Plan National pour la Science ouverte français : « *Notre ambition est de faire en sorte que les données produites par la recherche publique française soient progressivement structurées en conformité avec les principes FAIR (Facile à trouver, Accessible, Interopérable, Réutilisable), préservées et, quand cela est possible, ouvertes* ». Alors que les principes FAIR sont en voie de devenir une norme, il est important de sensibiliser les chercheurs à ces principes, et de les aider à les mettre en œuvre dans leur gestion des données. Des recherches fondées sur des études quantitatives et qualitatives montrent que les chercheurs accordent peu de temps à la gestion des données, qui s'ajoutent à leur charge de travail (Rebouillat, 2021). Des entretiens montrent que les scientifiques travaillent à « flux tendu » privilégiant la valorisation de leurs résultats par la publication plutôt que le partage des données : le temps consacré à la gestion des données est perçu comme un obstacle à la diffusion des données (Le Béhec et al., 2022, p-90).

Au niveau des formations au PGD existantes, il n'existe que très peu de retours d'expérience de la part des formateurs, alors qu'ils sont considérés comme très utiles (Perrier & Barnes, 2018). Nous pouvons citer des webinaires qui donnent aux chercheurs, doctorants et toute personne intéressée, les clés pour rédiger son PGD avec l'outil DMP OPIDoR et présenter des bonnes pratiques de gestion. D'autres ateliers entre professionnels du soutien à la recherche portent plutôt sur l'offre d'accompagnement des chercheurs à l'élaboration des PGD et la gestion des données. Les

¹³ Voir partie I

nouvelles pratiques de gestion des données avec pour élément central l'ouverture des données selon les principes FAIR donnent de plus en plus de matière à emmagasiner pour les acteurs des données. Il est important de mutualiser les connaissances et les innovations pour être en mesure d'accompagner et former les chercheurs le plus tôt et efficacement possible.

Enjeux relatifs aux disparités entre les communautés scientifiques et les personnels d'accompagnement à la gestion des données

L'hétérogénéité des communautés de recherche implique des cultures de la donnée très hétérogènes du fait de pratiques diverses et de spécificités propres à chaque discipline (Amiel et al., 2020). L'accompagnement à la rédaction des PGD relève de chaque établissement de recherche, et va différer selon le cadre administratif et le soutien institutionnel de chacun (Kim, 2021). Le soutien à la gestion des données est assuré en grande partie par les bibliothèques universitaires, qui font face à l'augmentation du besoin en gestion des données des étudiants et enseignants-chercheurs (Tenopir et al., 2014). L'évolution de ces métiers vers la gestion des données et l'élaboration de PGD (avec notamment l'avènement du *data librarian*) n'a pas forcément été anticipée, et nécessite de développer leur profil de compétences.

Un rapport du corps d'inspection de l'enseignement supérieur et de la recherche (Letrouit et al., 2021) a mis en évidence que les bibliothèques universitaires « *n'ont pas les moyens humains de répliquer pour chaque discipline l'assistance proposée à un niveau générique* » : même si « *l'approche disciplinaire n'a rien de nouveau pour les bibliothèques* », « *le degré de spécialisation nécessaire et les moyens humains à investir* » entravent leur action. Concrètement, le webinaire RDA Alliance de mars 2022 (Deboin et al., 2022) sur l'accompagnement des scientifiques à la rédaction des PGD, lequel rassemblait de nombreux formateurs, *data managers* et *data librarians*, a mis en lumière ces disparités et la nécessité de mettre en place des formations pour les formateurs dont certaines sont déjà mises en place¹⁴.

La difficulté de créer un continuum entre les personnels d'accompagnement, leur profil de compétences, et les besoins spécifiques des communautés scientifiques conduisent à ce deuxième enjeu. Afin d'éviter d'émietter leur accompagnement face à des besoins croissants, et pour contribuer à construire un socle interdisciplinaire de la science ouverte, les services auraient intérêt à développer une formation pouvant rassembler des chercheurs de disciplines distinctes dans leur démarche de rédaction de PGD. Pour cela, ils devront combiner des contenus généralistes, disciplinaires, institutionnels (selon la politique de gestion des données propre à chaque établissement).

¹⁴ Formation « Accompagner la rédaction d'un plan de gestion de données (PGD) » <https://sygefor.reseau-urfist.fr/#/training/8775/10054?from=true>

Enjeux relatifs aux modalités pédagogiques utilisées aujourd'hui pour aider à la conception d'un PGD

Un enjeu majeur des formations à la gestion des données est d'anticiper la réaction négative des chercheurs qui considèrent la rédaction de PGD comme un exercice ennuyeux, et le PGD lui-même comme un document administratif plutôt qu'un outil qui leur serait directement utile (Vitale & Moulaison Sandy, 2019). Même s'ils reconnaissent l'importance d'une bonne gestion des données, or la plupart n'est pas forcément disposée à y consacrer du temps. Or, des retours d'expérience montrent au contraire que d'autres considèrent l'exercice facile (Dällenbach, 2019.)

La revue systématique des formations à la gestion de données d'Oo et al, 2021, donne quelques conditions de succès notamment la conception centrée sur l'utilisateur, la pédagogie basée sur la pratique et le travail collaboratif. Cette étude dresse une analyse et un bilan basés sur 28 articles en lien avec la formation à la gestion des données. Un classement des formations a été réalisé selon quatre critères de succès pour aider les institutions à s'emparer de points de référence dans le but de développer des formations :

- Une participation active pendant la formation ;
- Une demande de formation importante (dans notre cas actuel on remarque une augmentation de la demande de formation à la gestion des données) ;
- Une amélioration de la connaissance et de la compréhension de la gestion des données par les participants ainsi que de leur confiance dans la mise en œuvre des pratiques de gestion des données ;
- Un retour positif après la formation.

À la suite de cette analyse, les auteurs proposent trois thèmes clés d'une formation réussie à la gestion des données : un processus de conception de formation centré sur l'utilisateur, une formation qui met l'accent sur l'aspect pratique et enfin une approche collaborative de la conception et de la mise en œuvre de la formation.

Clé 1 : Établir un processus de conception centré sur l'utilisateur

Tous les articles de l'enquête montrent que les formations ont été fondées sur la centralité de l'utilisateur – c'est -à-dire, une formation axée sur le profil et les besoins du public cible donc plutôt sur une approche de formation « *bottom up* ». Une analyse des besoins du public cible permet d'éclairer la conception de la formation, d'éclairer les décisions quant aux domaines à privilégier. Le fait de collecter des informations de manière collaborative avec le corps enseignant et d'autres acteurs de la gestion des données, de concevoir des enquêtes sur les pratiques en termes de gestion des données fait partie des bonnes pratiques pour concevoir une formation. Pour répondre aux niveaux d'exigences des participants, pour tenir compte des différentes disciplines et domaines de recherches, adapter sa formation est un point auquel il faut porter une attention toute particulière.

Construire de nouveaux modes de connaissances autour des données de la recherche Introduction

Certaines personnes peuvent déjà être familière avec les notions de gestion des données, d'autres ne les connaissent pas. Adapter sa formation pour satisfaire tous les participants est donc judicieux.

Clé 2 : Mettre l'accent sur l'aspect pratique

La contextualisation de la formation prend tout autant une importance en définissant le format et les activités d'apprentissage avec des objectifs précis. Dans la conception, il est révélé que les tests pilotes permettent de mieux comprendre les besoins du public cible afin de concevoir une formation pertinente et utile. Une formation à la gestion des données engageante doit comprendre des activités d'apprentissage interactives telles que des méthodes d'enseignement actives, des présentations, des travaux de groupe et des discussions (partage d'expérience, discussions actives entre les participants). De manière générale, la formation doit être étayée par une approche pédagogique, éducative et théorique.

La participation active des chercheurs lors de la formation favorise l'apprentissage. L'incorporation d'exemples de la vie réelle (scénarios utilisés dans la formation en lien avec ce que les participants sont susceptibles de rencontrer dans leur travail et leurs recherches) et activités d'apprentissage par la pratique, particulièrement important pour les doctorants (novices en matière de recherche et ont besoin d'accompagnement à la gestion des données).

Clé 3 : S'appuyer sur une approche collaborative de la conception et de la mise en œuvre de la formation

Dans l'étude, 79% des articles soulignent avoir développé leur formation en collaboration avec des professeurs, étudiants, bibliothécaires, *data management services*... Une approche collaborative est essentielle pour une formation réussie, l'aspect collaboratif de la conception contribue à l'efficacité de la formation. La collaboration avec les parties prenantes internes est nécessaire pour une formation à la gestion des données efficace. Cette clé reconnaît que la gestion des données s'étend à plusieurs institutions : facultés, départements et unités etc... D'autres points d'amélioration sont identifiés tels que l'instauration d'une formation au plus tôt pour les doctorants et chercheurs (Wiljes & Cimiano, 2019). En France, cependant, les données probantes manquent.

2. Exemple d'action de formation : un Sprint PGD

Depuis l'année 2020, Dataactivist accompagne un nombre croissant d'acteurs de l'enseignement supérieur, de la recherche et de l'innovation (ESRI) qui élaborent ou mettent en œuvre leur stratégie science ouverte et se mettent en conformité avec les exigences de leurs financeurs en matière d'open data et de gestion des données de recherche. Dataactivist se positionne également dans des activités de médiation et de pédagogie de la donnée, en organisant par exemple depuis 2019 le Challenge Data (Sarazin, 2021) à destination des étudiants de Master. Cet événement d'une semaine de pédagogie active est associé à des objectifs précis : une mise en pratique concrète des étudiants pour l'acquisition d'une méthodologie éprouvée et confrontée aux réalités du terrain. Avec, au bout d'une semaine, la publication et l'exploitation (visualisation de données, cartographie...) de jeux de données publiques. Inspiré par le Challenge Data, le service de documentation d'une université française a proposé à Dataactivist de développer nouveau format de médiation, visant cette fois la production d'un PGD plutôt que la publication de données. Ce format appelé « Sprint PGD » devait permettre aux chercheurs de rédiger un PGD et de s'acculturer aux bonnes pratiques de gestion des données.

Développement d'une méthodologie itérative

La rédaction du PGD par les chercheurs et son accompagnement par les services support se confrontent à des enjeux techniques cités précédemment partie II, 1. Pour les résoudre, nous avons développé un atelier s'inscrivant dans une démarche de recherche action (Lewin, 1946) basé sur la primauté de l'expérience et le changement de stratégie. Une démarche de recherche dans laquelle s'installe une spirale de cycles de planification, d'action, d'observation et de réflexion (McTaggart, 1988 citée par MacIsaac, 1996). La méthodologie itérative appliquée a été documentée dans un journal de bord expliquant les pistes de réflexions que nous avons suivies depuis le début du projet ainsi que la justification des choix et du tournant qu'a pris le Sprint PGD. Les étapes décisives sont reproduites dans le tableau en Annexe 1.

La figure n°1 ci-dessous représente le modèle de Kemmis et McTaggart :

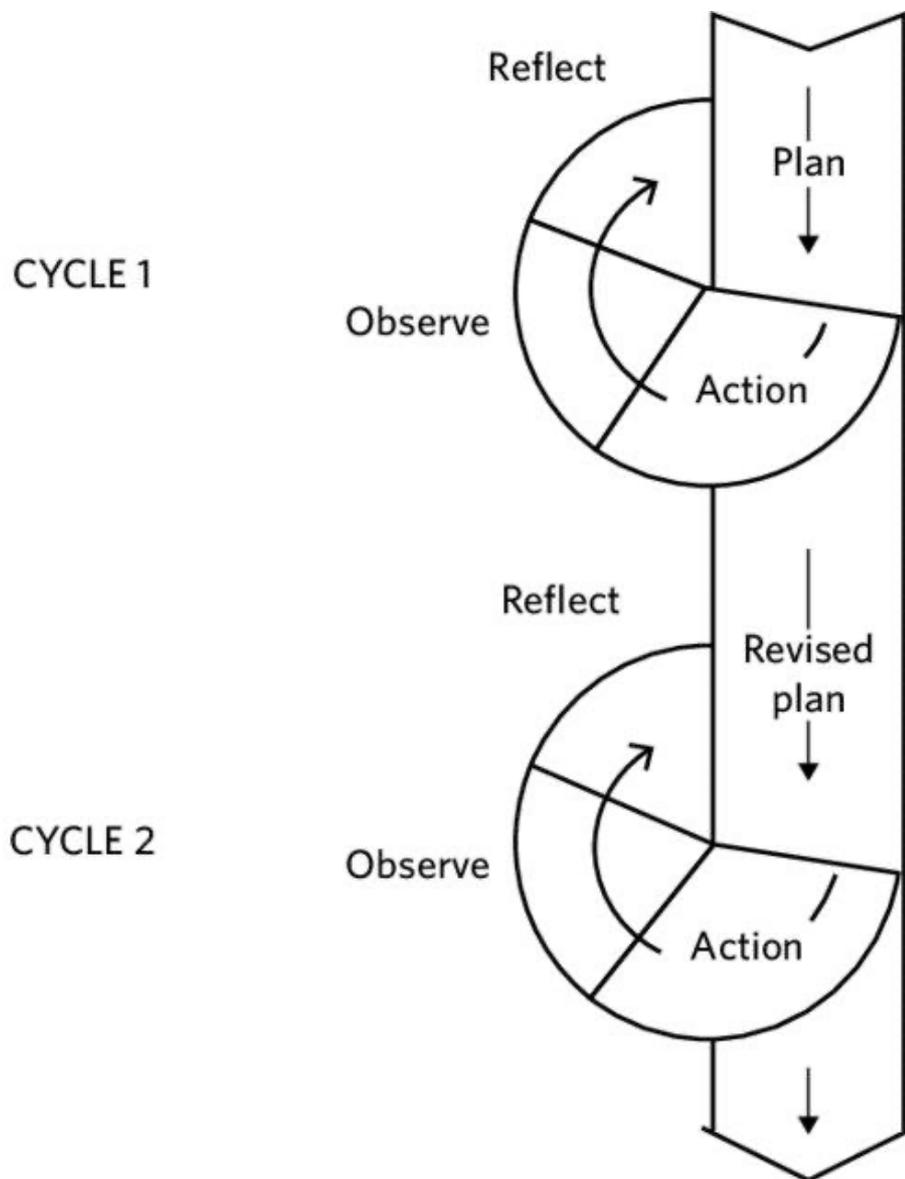


Figure n°1 : Modèle Kemmis et McTaggart (1988:11-14, cité dans Burns 2010)

La méthodologie développée pour la conception s'est également appuyée sur le modèle ADDIE (Branson, 1978) un processus de conception pédagogique basé sur une analyse du matériel existant (A) (ici, des formations existantes) ; la conception de la nouvelle formation (D) et sa plus-value. Le développement du matériel nécessaire (D), ici le développement et la mise à disposition de ressources dans un outil numérique dédié ; la mise en œuvre de formations pilotes (I), première version de l'atelier qui a fait l'objet d'une évaluation (E). La méthodologie prend en compte le

contexte lié aux activités des chercheurs comme le manque de temps. Pour appliquer cette méthodologie itérative l'idée était de produire plusieurs versions du Sprint PGD, suivies d'évaluations identifiant les manques et les perspectives. La figure ci-dessous représente le modèle ADDIE :

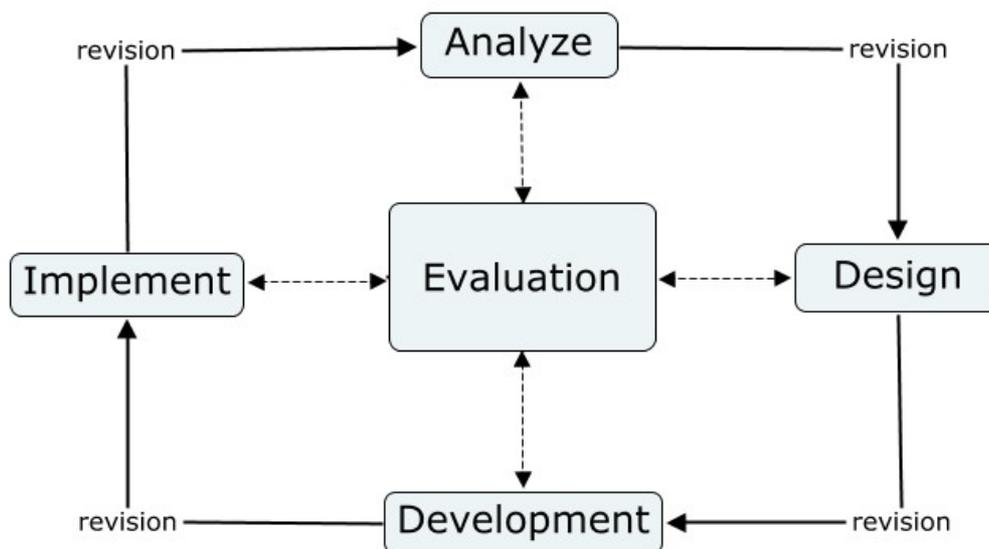


Figure n°2 : Modèle ADDIE (Branson, 1978)

L'objectif initial du Sprint PGD est de permettre renseigner tout modèle de PGD (ANR, Science Europe, Horizon 2020 etc...) selon de bonnes pratiques pour pouvoir obtenir des PGD de qualité, suffisamment documentés. La première piste pour la conception consistait à ne pas coller au modèle de PGD actuel et prendre un peu de distance en axant surtout l'atelier autour de la « gestion des données », afin d'en faire une formation véritable. Étant donné que les principes FAIR se retrouvent au cœur des politiques de gestion des données, nous avons pensé acculturer aux principes FAIR et monter en généralités autour des notions essentielles du PGD pour les chercheurs.

Le premier format du Sprint a été conçu pour aider les chercheurs à rédiger leur PGD, anticiper la gestion de leurs données tout au long du projet de recherche, comprendre quels produits de recherche sont concernés par le PGD et comprendre les enjeux que posent les principes FAIR. Le tout dans un format resserré afin de s'adapter aux contraintes de temps et aux disponibilités des équipes de recherche, très occupées par de multiples obligations et sollicitations. Les modalités pédagogiques de notre atelier associaient des outils numériques, un accompagnement composé de plusieurs animateurs, ce qui permettait d'apporter des conseils et bonnes pratiques.

Pour ce premier Sprint nous nous sommes fixé les objectifs suivants :

- Aider les personnels à préparer les données, répondre à toutes leurs questions et leurs présenter les bonnes pratiques ;
- Amorcer un suivi personnalisé tout au long de la vie des projets ;
- Permettre aux personnels de rédiger des PGD (modèles ANR / Europe, ou autre selon les besoins) ;
- Acculturer les participants aux principes FAIR ;
- Fournir des recommandations adaptées, quelle que soit la discipline et l'objet de recherche.

Conception et évaluation d'un patchwork PGD sur l'outil numérique : l'Open Data Canvas (ODC)

L'atelier s'appuie sur un outil numérique : l'Open Data Canvas, conceptualisé et réalisé par Arthur Sarazin, chercheur en systèmes d'information, spécialisé en design science au sein de Dataactivist. Un outil encourageant l'utilisation des données ouvertes développé dans le cadre de sa thèse soutenue en 2021 (Sarazin, 2021). L'Open Data Canvas est une projet collaboratif numérique avec pour volonté de créer un commun numérique recensant et caractérisant des ressources relatives à l'*open data*, par exemple :

« La ressource est-elle d'accès gratuit ? La ressource est-elle sous licence libre ? Les règles de gestion sont-elles définies collectivement ? Ces règles permettent-elles la contribution de tous ? La communauté est-elle ouverte à d'autres ? est-elle motivée par des intérêts marchands ? »

On comprend de sa définition (Guerry, 2021) qu'un commun numérique repose sur trois éléments : une ressource numérique, une communauté qui la gère, des règles de gestion explicite, ici un manifeste¹⁵. Initialement créé pour le développement de l'*open data*, l'outil permet aux utilisateurs d'accéder librement à une liste de ressources appelées composants (françaises et internationales) sur l'*open data* : de la collecte au traitement des données, et de leur publication à leur réutilisation. L'ODC fournit également un manifeste de l'*open data*, qui explique les principes essentiels pour qu'une stratégie *open data* ait de l'impact. Grâce à cet outil, nous avons pu développer un parcours numérique type patchwork¹⁶, en accès libre, associé à notre atelier. Ce modèle patchwork est une combinaison de composants qui permettent de développer de A à Z un dispositif de manière générique. Cette appellation « patchwork » est une métaphore renvoyant à la technique de couture, permettant aux composants d'être reliés les uns aux autres dans le but de constituer un tout, plus

¹⁵ <https://opendatacanvas.org/manifeste>

¹⁶ Ici la version du parcours est basée sur le cycle de vie de la donnée, non sur les principes FAIR, (2022) (<https://opendatacanvas.org/sprint-pgd>)

grand et cohérent. Pour créer ce parcours, nous avons d'abord travaillé avec les modèles de PGD les plus utilisés : le modèle ANR, le modèle H2020, le modèle Science Europe.

De cette réflexion, une analyse des différentes sections des modèles de PGD (ANR, H2020, Europe) a été réalisée permettant de savoir si les modèles étaient transposables en atelier autour des principes FAIR. Nous avons rassemblé toutes les questions et recommandations du modèle que nous avons ensuite transposé aux principes FAIR afin de créer un parcours organisé en quatre parties : F pour rendre ses données faciles à trouver, A pour rendre ses données accessibles, I pour rendre ses données interopérables et R pour rendre ses données réutilisables. Une fois que nous avons associé les questions du PGD à chacune des parties, nous avons pu construire le parcours sur notre outil numérique. Les étapes étant chacune organisées en différents pas-à-pas, l'utilisateur allait en parallèle pouvoir rédiger son PGD sur DMP OPIDoR. Pour cette première version le parcours s'organisait en quatre parties qui traitaient des notions suivantes :

Étape F : Rendre ses données faciles à trouver :

- F.1 Identifier l'ensemble des produits de recherche d'un projet
- F.2 Préparer les métadonnées et la documentation associées à chaque produit de recherche
- F.3 Choisir un entrepôt pour stocker les produits de recherche identifiés

Étape A : Rendre ses données accessibles :

- A.1. Choisir les niveaux d'accès adaptés à chaque produit de recherche
- A.2 Décrire le chemin d'accès aux produits de recherche pour chaque niveau
- A.3 Budgétiser le coût global de gestion des produits de recherche

Étape I : Rendre ses données interopérables :

- I.1 Vérifier que le format de données choisi est ouvert et partagé
- I.2 S'assurer les métadonnées répondent à un standard
- I.3 Vérifier que le vocabulaire utilisé respecte les principes FAIR

Étape R : Rendre ses données réutilisables :

- R.1 Choisir le type de licence adapté aux besoins
- R.2 Vérifier la conformité juridique et éthique de la publication des données

Construire de nouveaux modes de connaissances autour des données de la recherche Introduction

Chaque partie (exemple F.1 ou A.1) correspond à un pas-à-pas, par exemple si on prend le pas-à-pas F.1 afin que l'utilisateur identifie l'ensemble de ses produits de recherche, il devait suivre l'ensemble des pas :

- Avant de commencer (consignes et recommandations)
- Premier pas : se familiariser avec la notion de produit de recherche
- Deuxième pas : détecter tous les produits de recherche du projet
- Troisième pas : lister tous les produits de recherche du projet
- S'inspirer d'exemples de PGD publics

Pour chaque pas il était précisé dans la partie « avant de commencer » à quelle question du modèle du PGD correspondait le pas et le temps maximum à passer sur ce pas.

Des ressources étaient également mises à disposition pour chaque pas que ce soient des vidéos pédagogiques DoRANum (« Données de la recherche et apprentissage numérique ») ou des ressources que nous avons implémentées dans un gitbook afin d'aider les participants s'ils ne connaissaient pas les notions abordées en leur proposant des recommandations. En fin de pas-à-pas, nous avons mis à disposition des exemples de réponses aux questions du PGD du modèle de l'ANR que nous avons jugé complètes et en accord avec les principes FAIR. Ces exemples sont issus des PGD publics mis à disposition sur DMP OPIDoR.

Réalisation du premier Sprint PGD

Le premier Sprint a été co-construit avec d'autres acteurs, notamment des professionnels de la donnée et des personnels d'accompagnement à la gestion des données, avec lesquels nous avons pu concevoir et animer la première version du Sprint PGD le 10 mai 2022 à Toulouse. Pour cette version prototype nous avons travaillé en partenariat avec l'Unité Régionale de Formation à l'Information Scientifique et Technique Occitanie et le Comité de Réflexion pour la Science Ouverte (CéSO) de Toulouse. Le but était de pouvoir réaliser une V0 du Sprint PGD sous forme d'un prototype avec des chercheurs et doctorants d'une même communauté disciplinaire et de poursuivre la conceptualisation de la formation. À l'issue de notre première réunion de présentation de l'atelier, nous avons modifié le format de la journée d'animation, et validé ensemble l'articulation du parcours autour des principes FAIR. En concertation avec les personnels d'accompagnement à la rédaction des PGD de Toulouse, la journée d'animation a été raccourcie en une demi-journée afin de s'adapter à la contrainte de temps des chercheurs.

Pour ce premier Sprint, nous avons voulu nous concentrer sur un périmètre de chercheur resserré, nous nous sommes orientés vers la communauté des sciences du numérique. Le choix de la communauté des sciences du numériques pour tester l'atelier en V0 permettait un premier état des lieux de l'adéquation entre l'animation et les ressources numériques. L'objectif ultime étant

d'orienter le Sprint PGD vers un public pluridisciplinaire avec intégration de nouvelles ressources spécifiques permettant une personnalisation des parcours. Un des objectifs du parcours ODC était également de pouvoir injecter les recommandations spécifiques d'un établissement ou des recommandations dans les ressources pour personnaliser le parcours et aider davantage les participants dans la rédaction du PGD (choix des entrepôts, politique institutionnelle d'archivage ou de gestion des données personnelles, etc). Ce test n'a pas pu avoir lieu car nos interlocuteurs toulousains n'avaient pas de recommandations spécifiques à proposer.

Cette demi-journée était animée par trois consultants Dataactivist, et trois experts toulousains (Michelle Sibillat, Amélie Barrio, Soraya Demaye). Parmi les 18 inscrits figuraient d'autres personnels d'accompagnement comme Emilie Marchand du Laboratoire d'analyse et d'architecture des systèmes (LAAS), qui participait avec un doctorant du laboratoire. Le déroulé de la demi-journée (Annexe 1) se composait de l'accueil des participants suivie d'un temps plénière de quinze minutes détaillant le déroulement de la demi-journée, les enjeux du PGD, un rappel de ce que sont les principes FAIR ainsi qu'une explication de l'articulation entre le parcours ODC et la rédaction sur DMP OPIDoR.

Au préalable nous avons envoyé aux participants inscrits une fiche de préconisations pour les guider dans leur inscription sur DMP OPIDoR et leur permettre de venir à l'atelier avec un compte et un modèle de PGD déjà définis afin de ne pas perdre de temps le jour J étant donné le format resserré. À la suite de la plénière, le parcours pouvait commencer en autonomie où les chercheurs et doctorants ont pu suivre le parcours en pas-à-pas et rédiger en parallèle sur DMP OPIDoR. Nous avons réparti les participants par équipe projet dans la mesure du possible, afin de partager le travail (suivre le parcours et rédiger le PGD) mais aussi de construire une vision collective de la gestion des données du projet. De notre côté nous sommes restés à disposition afin d'échanger avec les participants et répondre à leurs questions. À la fin de la demi-journée nous avons pu faire un bilan avec les participants dans un format plénière, de manière à leur soumettre les recommandations pour la suite de la rédaction de leur PGD ainsi qu'un questionnaire qualitatif d'évaluation de la demi-journée d'atelier.

3. Bilan et perspectives

Cette partie s'organise en plusieurs axes, d'abord nous aborderons les premiers retours que nous avons eu en tant qu'animateurs à l'issue de l'atelier Sprint PGD que ce soit sur l'ergonomie des outils ou l'animation de l'atelier. Ensuite, nous analyserons les réponses des participants au questionnaire d'évaluation et nous verrons en quoi leurs retours permettent de cibler les points faibles et forts de ce premier atelier. Nous évoquerons également la reconceptualisation du Sprint PGD permettant d'ouvrir de nouvelles perspectives.

À l'issue de l'atelier nous avons discuté de nos premiers ressentis concernant le déroulement en tant qu'animateurs. Nous avons trouvé que plusieurs points méritaient une amélioration : l'ergonomie du parcours et des différents outils, la linéarité entre les modèles de PGD et l'articulation du parcours autour des principes FAIR. De même, concernant le déroulement de l'animation, nous avons considéré que le temps de plénière était plutôt court. Il aurait été judicieux de prévoir un temps plus long afin d'expliquer plus exactement le fonctionnement des pas-à-pas du parcours sur l'ODC, expliquer plus largement les différents outils (gitbook, DMP OPIDoR) pour ne pas créer de confusion lors de l'atelier. En effet, au cours de l'atelier les chercheurs étaient parfois perdus entre la linéarité du modèle de leur PGD et celle du parcours, le parcours ne permettant pas de rédiger les PGD dans l'ordre des questions du modèle. De plus, certains participants suivaient exclusivement les ressources que nous avons mises à disposition sur le gitbook sans suivre les pas-à-pas de l'ODC. Ces personnes-là arrivaient quand même à rédiger leur PGD puisque les ressources permettaient de répondre aux questions, néanmoins ils ne suivaient pas les pas-à-pas et n'avaient pas le fil conducteur de l'atelier lié aux principes FAIR. Une explication plus claire sur les outils aurait facilité l'utilisation des outils et donc la rédaction du modèle.

Concernant les retours des participants, sur les 18 présents lors de l'atelier, 11 personnes ont pu remplir le questionnaire d'évaluation (Annexe 2) mettant en lumière les points forts et les points faibles de l'atelier (réponses des participants Annexe 3). Dans ce questionnaire plusieurs dimensions nous intéressaient. D'abord, savoir si durant l'atelier les participants ont considéré que l'atelier leur a été utile, améliorant leur capacité de rédaction du PGD. Ensuite, il était important de savoir si une notion en particulier leur semblait plus difficile qu'une autre. L'utilisation du canvas, du gitbook et de DMP OPIDoR était aussi un point sur lequel nous attendions des retours, pour cela une question portait sur les éléments perçus comme moins abordables selon les participants. Afin de nous éclairer sur ce qui leur était le plus utile durant l'atelier, le questionnaire comportait une section sur le facteur facilitant le plus la rédaction (ressources, parcours ODC, accompagnement de l'équipe...). Dans l'intérêt d'améliorer le Sprint PGD nous leur avons demandé si un élément leur avait manqué au cours de l'atelier, et enfin si l'atelier leur avait permis de s'acculturer des principes FAIR.

En moyenne, les participants ont évalué leur progrès dans la rédaction du PGD par deux points sur cinq maximum, l'atelier leur a donc été plutôt utile. Une grande majorité des participants estime que le temps passé sur l'atelier était utile, un seul participant estimant n'avoir ni gagné ni perdu du temps. Parmi les étapes qui ont donné le plus de difficultés aux participants revient souvent l'étape R qui traitait des questions liées aux aspects juridiques et éthiques (RGPD, licences...) (participants 1, 4, 6 et 11), la partie concernant les métadonnées et standards de métadonnées (participants 5 et 6), ou encore la non-linéarité du parcours et des modèles de PGD (participants 3). Les aspects juridiques concernant les données de la recherche sont un point important à relever. Que ce soit lors des entretiens semi-directifs lors mon stage de M1, ou cette année, la notion de licences – enjeu pourtant crucial dans l'ouverture des données de la recherche - a toujours été une notion floue avec un manque de connaissance évident de la part des chercheurs. Les réponses au questionnaire confirme les difficultés qu'éprouvent les chercheurs sur le sujet, sous un autre contexte, il peut être intéressant de passer plus de temps sur les aspects juridiques dans les formations.

Globalement, ce qui a fonctionné dans l'atelier étaient la richesse des ressources présentes dans le gitbook, la formation d'équipes de travail durant l'atelier ainsi que l'accompagnement et l'animation de l'atelier. Comme vu en partie II.2, les trois clés d'une bonne formation mis en lumière par l'article de Oo étaient le fait d'établir un processus de conception centré sur l'utilisateur, de mettre l'accent sur l'aspect pratique et de s'appuyer sur une approche collaborative de la conception et de la mise en œuvre de la formation. Le Sprint PGD respecte ces trois clés, d'une part les participants sont actifs durant la formation puisque le parcours se réalise en semi-autonomie avec l'accompagnement des animateurs. Ce qui rend les participants actifs en plus de rédiger simplement le PGD, est le fait qu'ils suivent un parcours permettant de répondre à leurs questions en trouvant les ressources nécessaires. La demande de formation concernant les données de la recherche ne cesse de croître avec des formations spécifiques aux établissements ou plus générales. Enfin dans le retour nous constatons que dans la majorité des réponses, les participants étaient satisfaits et ont trouvé la compilation des ressources complètes et utiles. Ils ont également bien fait ressortir, grâce au questionnaire, que l'accompagnement mis en place leur a été très bénéfique pour la rédaction. Le fait que les animateurs soient disponibles pour répondre à leurs questions et apporter des précisions a bien fonctionné. En tant qu'animatrice de l'atelier ce format et aussi agréable puisqu'il offre la possibilité d'échanger individuellement avec les participants, de découvrir leur projet et de les guider plus finement.

Le questionnaire et le développement de la méthodologie itérative participe à la création d'une formation de manière ascendante « *bottom up* » puisque qu'à la suite des évaluations nous pouvons identifier les points sur lesquels les participants, donc les chercheurs et doctorants éprouvent le plus de difficultés pour paramétrer ensuite l'ensemble de l'atelier et les ressources afin de faciliter les points plus compliqués.

Dans les retours que nous avons eus avec les partenaires de Toulouse, il y avait une satisfaction générale concernant le déroulement de l'atelier. L'atelier crédibilisait l'accompagnement et les ressources validaient l'accompagnement. Dans une optique de réitération de l'atelier, le format du Sprint PGD était intéressant avec un parcours et une animation pour des chercheurs et doctorants. Il est ressorti qu'un parcours clé en main ou une formation aux formateurs sur ce genre de parcours ODC serait intéressante pour des structures comme l'URFIST Occitanie. Notamment dans l'idée de former des personnels d'accompagnement pour avoir la capacité d'assurer un Sprint PGD à leur tour pour multiplier les possibilités et qu'eux aussi puissent en bénéficier. À la suite des bilans en interne et avec les partenaires (URFIST, CÉSO), nous avons retravaillé le contenu du parcours ainsi que l'ergonomie des outils numériques. Une question se posait : était-il judicieux de rester sur une articulation autour des principes FAIR étant donné que le lien entre la rédaction et le parcours n'était pas linéaire ? Si nous avons très tôt pris la décision d'axer l'atelier sur les principes FAIR pour en faire un parcours original et novateur, l'idée était satisfaisante mais la conception a demandé une certaine gymnastique entre l'adéquation des modèles de PGD, les principes FAIR et les pas-à-pas du parcours.

Du fait que les chercheurs ont une vision plus intuitive du cycle de vie de la donnée plutôt que des principes FAIR, l'articulation autour des principes FAIR n'était pas la bonne et complexifiait l'exercice au lieu de le simplifier. Nous avons choisi d'articuler l'atelier autour du cycle de vie de la donnée. Par souci de linéarité, nous avons reconceptualisé l'atelier autour du cycle de la donnée en 5 étapes (cf. Annexe 4) :

Étape 1 - Acquisition des données

Étape 2 - Traitement des données

Étape 3 - Accès et partage des données

Étape 4 - Conservation et archivage des données

Étape 5 - Réutilisation des données

Nous avons également retravaillé les pas-à-pas de sorte qu'ils soient plus prescriptifs que ceux du parcours précédent et guident davantage les utilisateurs dans la rédaction du PGD. Une question demeurait toujours concernant les modèles de PGD. Comment construire un parcours d'accompagnement qui permettrait de rédiger n'importe quel modèle de PGD de manière linéaire ? Durant la réalisation de notre atelier prototype Sprint PGD, certains participants suivaient le modèle ANR, d'autres H2020 ou encore Science Europe (désormais modèle ANR par défaut). Tous ces modèles traitent des mêmes notions de gestion des données et nécessitent, à quelques notions près, les mêmes informations ; néanmoins les sections ne sont pas dans le même ordre et les questions n'ont pas toujours la même granularité. La reconception du parcours ODC s'est

heurtée à des modèles de PGD trop différents, des incohérences dans les questions et les recommandations, des manques de précisions ou des questions trop précises, qui rendent la conception d'un parcours universel difficile. Pour cette raison nous avons préféré nous concentrer sur la conception d'un parcours par modèle de PGD. À l'aide de l'outil *Airtable*, nous avons pu concevoir des parcours avec des pas-à-pas uniques pour rédiger chaque modèle de PGD. Nous avons créé un nouveau parcours autour du cycle de vie de la donnée afin de pouvoir remplir le modèle ANR dont les captures d'écran de l'ODC se trouvent dans l'Annexe 4. L'objectif et l'aboutissement de l'ODC est de pouvoir construire plusieurs parcours en patchwork, associé à un modèle unique : un parcours pour le modèle ANR, un parcours pour le modèle Science Europe, un parcours pour le modèle H2020, afin d'avoir une rédaction plus linéaire qui ne demande pas au participant une gymnastique entre le canvas et la rédaction sur DMP OPIDoR.

Enfin, lors de l'atelier, nous avons constaté que les participants ont passé plus de temps sur la prise en main des outils et DMP OPIDoR que le temps prévu à cette étape. De même, ils ont eu tendance à rester sur les ressources du gitbook au lieu de suivre le parcours ODC malgré l'explication du modèle en pas-à-pas. Afin de pallier ces soucis, nous avons décidé de revoir toutes les ressources et de ne plus les constituer autour d'un gitbook mais plutôt sous forme de fiches numériques unitaires associées exclusivement à un pas, ce qui permettrait de ne pas perdre les utilisateurs. Nous avons utilisé d'autres outils que le gitbook pour créer ces fiches, une de mes missions étant notamment de créer les nouvelles fiches ressources en format *rmarkdown*¹⁷ sous R Studio avec un *repository* github dédié. En utilisant ces formats les fiches se retrouvent dans un format plus ouvert que le gitbook initial. Nous avons également choisi de diversifier les ressources du parcours en incluant de courtes vidéos explicatives sur des notions comme les standards de métadonnées, l'anonymisation et la pseudonymisation etc... Nous avons reproduit des ressources existantes (en licence libre) dans des formats réutilisables et plus ouverts que le format initial. C'est le cas du logigramme de communicabilité d'une donnée (Annexe 5) conçu par Nicolas Becard, Céline Castets-Renard, Gauthier Chassang, Martin Dantant, Laurence Freyt-Caffin, Nathalie Gandon, Caroline Martin, Andrea Martelletti, Alexandra Mendoza-Caminade, Nathalie Morcrette et Claire Neirac dans le Guide d'analyse du cadre juridique en France, 2017 qui était intégré dans un long document .pdf et présentait plusieurs inconvénients : image en basse qualité, image non interactive, texte non copiable, fichier lourd à charger, pas de lien direct vers la page de la figure... J'ai pu le recréer sous forme d'une *mindmap* interactive en format *markdown* grâce à l'application *mymarkmap*¹⁸.

¹⁷ « L'extension *rmarkdown* permet de générer des documents de manière dynamique en mélangeant texte mis en forme et résultats produits par du code R. » <https://larmarange.github.io/analyse-R/rmarkdown-les-rapports-automatises.html>

¹⁸ <https://mymarkmap.vercel.app/>

Grâce à l'adoption d'une méthodologie itérative, nous avons pu tester la conception d'une première version du parcours ODC et réaliser un atelier Sprint PGD avec des chercheurs et doctorants à Toulouse. Ce premier Sprint a mis en lumière certains aspects encore fragiles de la conception que nous avons pu revoir et repenser afin d'améliorer le parcours. Dans le cas précis d'une formation d'accompagnement à la rédaction d'un PGD, créer une formation universelle pouvant permettre à des chercheurs n'importe quel modèle de PGD peut être difficile et demander du temps. En revanche le fait de créer un parcours participant à un commun numérique et pouvant être implémenter de diverses ressources permet de d'innover en créant des parcours spécifiques pour une discipline ou un établissement, ou encore des parcours spécifiques permettant de rédiger n'importe quel type de modèle de PGD.

Un des objectifs de l'atelier Sprint PGD sur le long terme serait d'acquérir une dimension pluridisciplinaire qui permettrait de le décliner pour chaque discipline et pour chaque modèle de PGD ce qui ferait la force de ce type de parcours. De mon expérience, la conceptualisation d'un tel atelier avec la création de ressources numériques est très longue que ce soit la veille à effectuer sur les ressources ou encore les processus de conception de nouvelles ressources. C'est un travail louable qui serait utile sans aucun doute mais qui nécessiterait plus de temps et de moyens que ce nous avons réalisé avec Dataactivist en quelques mois. De mon point de vue, conceptualiser une formation avec des chercheurs ou des experts pourrait s'avérer plus efficace, dans le sens où leur implication permettrait de cibler les aspects de la formation qui mériteraient plus de temps, d'autres moins avec des ressources percutantes. Les chercheurs et les doctorants sont les plus à même de connaître leurs lacunes concernant leur gestion des données. Même si l'expérience et les connaissances de chaque chercheur sont hétérogènes, ce dispositif faciliterait la conceptualisation. Par exemple, il aurait peut-être été utile de demander dès le départ à des chercheurs si l'orientation du parcours autour des principes FAIR ou autour du cycle de vie de la donnée.

III. L'écosystème scientifique en transition

Le mouvement science ouverte s'est propagé à travers tout l'écosystème scientifique dans une dimension internationale. Comme vu précédemment, ce sont tous les acteurs de la science qui sont concernés par ce mouvement, nous assistons à une transition de la science vers la « science ouverte ». Cette transition implique des évolutions, qui se retrouvent à la fois dans les pratiques de gestion, dans les formations pour chercheurs et étudiants qui s'axent de plus en plus sur les données de la recherche et particulièrement dans les corps de métier. La transition vers le numérique amplifie et accélère la transition vers la science ouverte, incluant des innovations, de nouveaux outils numériques, de nouvelles plateformes et le développement des services automatisés.

Cette transition enclenche une agitation dans l'écosystème scientifique, à savoir quelles décisions sont positives, négatives, qu'est-ce qui réussit et échoue... Nous nous intéresserons à savoir comment cette transition, déjà engagée, influe-t-elle sur les pratiques de la recherche ?

1. Les pratiques de gestion face à une transition vers la science ouverte

Au regard d'enquêtes sur les pratiques de gestion des données menées dans deux sites français, nous allons nous intéresser aux mutations dans ces pratiques liées à la science ouverte, ainsi qu'aux besoins des chercheurs et professionnels de l'IST¹⁹ pour faire face à cette transition. La première enquête « Données de la recherche : Quelles pratiques ? Quels besoins ? Enquête à Aix-Marseille Université » a été réalisée par Caroline Sophia Donati, chargée de mission données de la recherche à l'Université Aix-Marseille (AMU). Cette synthèse publiée en février 2020 propose un état des lieux des pratiques en termes de perception des données de la recherche, de compréhension du mouvement de la science ouverte et des pratiques de la gestion des données de la recherche pour la communauté scientifique AMU. L'enquête rassemblait des enseignants chercheurs (EC) et des personnels de l'IST (IST) de l'AMU, les deux communautés ont dû répondre à un formulaire distinct. Nous nous focaliserons essentiellement aux pratiques des EC concernant leur gestion des données, notamment le biais de l'utilisation du PGD et sur les besoins et attentes des EC et IST.

Dans le formulaire des EC, la première question relative aux PGD montre que 98 répondants (49,75%), ne savent pas de quoi il est question, 39 répondants (19,80 %) sont favorables à cette pratique, et 23 (11,68 %) ne savent pas comment faire mais le font. Comme vu dans les parties précédentes, les PGD constituent un levier pour sensibiliser les chercheurs à la culture des données de la recherche. Ils permettent d'assimiler les bonnes pratiques de gestion des données, et le fait

¹⁹ « L'I.S.T. regroupe ainsi l'ensemble des informations produites par la recherche et nécessaires à l'activité scientifique et intervient donc en amont et en aval dans tout le cycle de production de nouveaux contenus scientifiques quelle que soit leur forme : articles, données, ouvrages, archives ouvertes, etc. » <https://www.enseignementsup-recherche.gouv.fr/fr/information-scientifique-et-technique-51161>

Construire de nouveaux modes de connaissances autour des données de la recherche Introduction

que les chercheurs ne connaissent pas cette notion contribue au manque d'acculturation de gestion des données de la recherche.

Dans les réponses concernant les risques liés à la mise à disposition de leurs données, le premier argument des chercheurs est : la perte de priorité scientifique. En 2022, c'est encore un argument principal. Les chercheurs sont toujours dans une dimension de lutte de légitimité et de capital scientifique. Les chercheurs n'identifient pas les professionnels IST disponibles pour leur venir en aide, en conséquent, ils n'identifient pas non plus les ressources disponibles de leur unité.

Concernant les réponses des professionnels de l'IST, la première question était sur l'appartenance à un réseau professionnel ou non et seulement 19 répondants (soit 31,5%) ont répondu oui. Le fait d'appartenir à un réseau professionnel permet à la fois de faciliter la veille par le partage au sein du réseau. Un réseau est constitué apporter des réponses à des questions précises sur des sujets. Dans le cas de l'AMU, cette question permettait de savoir si par la suite les professionnels de l'IST voudraient faire partie d'une communauté AMU sur les questions de science ouverte.

En termes de besoin, les répondants au questionnaire IST sont majoritairement intéressés par de l'assistance technique (45,90 %), donc -le développement de ressources humaines permettant d'aider les chercheurs dans leur gestion des données. Les mutations actuelles dans les pratiques de gestion, nécessitent de nouveaux besoins pour les chercheurs et les personnels IST afin de répondre à leurs questions et les guider.

Si nous voulons poursuivre la transition vers une science ouverte, appliquer une bonne gestion des données -donc appliquer les principes FAIR et aider à la rédaction de PGD documentés est indispensable.

Les personnels des établissements doivent pouvoir trouver les informations nécessaires afin de les accompagner et de les aider. D'où les points essentiels à retenir de cette enquête :

- Le manque de temps ;
- L'absence de support logistique et/ou technique ;
- Le manque de légitimité et de reconnaissance de cette mission comme importante, nécessaire, prioritaire au sein des laboratoires.

La deuxième enquête a été réalisée par la Cellule Data Rhône Alpes, donnant lieu au rapport publié en mai 2022 sur les usages et les besoins pour la gestion des données de la recherche sur le site de l'Université Grenoble Alpes (UGA). Plus récente, cette enquête est traversée par les changements de stratégie en matière de politique science ouverte. Nous nous intéresserons particulièrement aux différentes pratiques des communautés, en particulier autour l'utilisation du PGD qui est le témoin de l'insertion de la culture de la donnée et qui nous permettra d'identifier ou non le développement d'une stratégie de gestion des données. Nous nous intéresserons également aux besoins identifiés des chercheurs et personnels IST pour améliorer leurs pratiques de gestion.

« D'une façon générale, les Plans de Gestion de Données, documents qui décrivent la façon dont les données seront produites ou obtenues, (ré)utilisées, traitées, organisées, stockées, sécurisées, préservées, documentées, partagées, au cours et à l'issue d'un projet de recherche, sont assez méconnus. »

Nous pouvons faire le même constat pour cette enquête : les PGD restent encore des notions méconnues des chercheurs malgré le caractère obligatoire. Dans ce cas l'acculturation au PGD ne peut qu'évoluer favorablement et ainsi permettre aux chercheurs d'assimiler une bonne gestion des données et faire évoluer leurs pratiques, tout comme l'enquête précédente.

« Par ailleurs, sur les habitudes de stockage, les volumes de données en jeu influent sur les pratiques, et certaines communautés (par exemple la physique des particules, l'astrophysique, les géosciences, l'environnement et l'écologie) sont mieux structurées que d'autres sur ce sujet. »

Dans les pratiques de gestion et habitudes pour les chercheurs de gérer leurs données, comme il est écrit ci-dessus, des disparités existent. Certaines disciplines (cf. exemples ci-dessus) ont des habitudes plus structurées, donc une culture de la donnée plus imprégner tandis que dans d'autres disciplines où le volume est moins important, les pratiques ne sont pas forcément structurées et n'ont pas d'habitudes de gestion des données. Afin de créer une base commune autour de la culture de la donnée, le développement de ressources disciplinaires spécifiques permettraient d'aider les communautés qui n'ont pas les mêmes pratiques ; soit de développer une culture de la donnée disciplinaire.

« De même, le degré de complexité, tant réglementaire que technique, lié au travail sur des données sensibles, impacte plus certaines communautés (sciences humaines et sociales, santé) que d'autres. »

Dans la partie II.3 Bilan et perspectives, nous avons vu que les chercheurs éprouvaient plus de difficultés à réaliser l'étape sur les parties juridiques du PGD. La partie bilan, les données sensibles sont sujettes à des questions juridiques et éthiques, là où les chercheurs éprouvent plus difficultés par manque de connaissance sur le sujet. L'accompagnement par des personnels spécialisés sur ces questions faciliteraient la gestion et l'ouverture des données par des processus d'anonymisation,

pseudonymisation etc... pour les chercheurs des communautés les plus à même de générer ce genre de données.

« De manière générale, la réutilisation de données existantes n'est pas simple, car bien souvent les principes FAIR c'est-à-dire trouvables, accessibles, interopérables et réutilisables, ne sont pas complètement respectés. Enfin, l'action de diffuser ses données comme un produit de recherche à part entière n'est pas encore bien intégré dans le processus de recherche. Ces constats encouragent la poursuite des efforts réalisés en ce qui concerne l'information autour des outils et services déjà existants. Ils confortent la nécessité de renforcer l'accompagnement qui doit être au cœur des dispositifs de soutien. Les actions de formation doivent être consolidés et développés, à tous les niveaux. Enfin, il est essentiel d'aller vers une simplification de tout le dispositif pour une meilleure compréhension et une bonne adhésion de toutes les communautés scientifiques, quel que soit leur niveau de maturité dans le domaine de la gestion des données de la recherche. »

Ces constats sont intéressants sur plusieurs niveaux, d'abord sur l'intégration de la diffusion des données dans les processus de recherche : malgré les nouvelles politiques, les chercheurs ne partagent pas leurs données. L'accompagnement et la consolidation des formations et la simplification des dispositifs font partie des enjeux principaux de la science ouverte concernant les données, ce qui renforce les idées développées dans ce mémoire.

Presque la moitié des répondants (45%) n'identifient pas de personne ressource pour les aider dans la gestion des données. Les répondants ont donc recours à des personnes de proximité, soit directeur de thèse, personnel interne au laboratoire, parfois documentaliste. Il y a également une forme d'entraide puisque parfois des répondants identifient des personnes extérieures à l'établissement, par la proximité disciplinaire.

Globalement, nous constatons un manque important d'information sur les possibilités d'accompagnement et d'aide sur la gestion des données. Il y a très clairement un besoin important d'accompagnement et de formations sur la gestion des données en général dans toutes ses ramifications. Le stockage, dans toute sa complexité (stockage, sauvegarde, archivage, entrepôt de données mais aussi partage de données) représente également un des points sur lesquels il est indispensable de proposer à la fois du conseil mais aussi des infrastructures adaptées. Dans les pratiques en elles-mêmes ce sont souvent les étapes de stockage, sécurisation et archivage qui sont inorganisées, une autre enquête confirme cette idée :

« Il en est ainsi en particulier à propos des pratiques de stockage, d'archivage, ou de diffusion des données (hors de toute publication scientifique), qui restent globalement inorganisées, alors même que le sentiment de propriété développé autour des données est réel. » (Amiel et al. 2020)

Les besoins mentionnés et les différents manques sont en train de se solutionner par le biais de réunions, mises en communs, échanges de bonnes pratiques et surtout d'un début de publications

sur les pratiques de gestion. Des établissements de recherche sont parfois plus avancés que d'autres sur le sujet. Par exemple, dans l'accompagnement à la rédaction des PGD certains établissements préremplissent ou inscrivent directement leurs recommandations spécifiques en matière de stockage ou autre, dans des modèles DMP OPIDoR dédiés. Ce qui permet aux chercheurs des établissements de trouver des recommandations spécifiques.

Dans le bilan du PNSO 1, sur la partie « Coordonner : Construire autour de l'administrateur des données un réseau de correspondants dans les établissements, pour répondre aux questions que se posent les chercheurs sur les données de la recherche », nous remarquons que le niveau n'est qu'à une étoile sur cinq. Ce constat contribue au fait que les demandes en formations et en personnels ne sont pas encore soulagées malgré la généralisation des PGD, le développement d'entrepôts de données ou de prix des données de la recherche. L'enjeu reste donc la coordination entre les professionnels de la donnée et les chercheurs afin de développer des offres de service structurées et efficace pour atteindre les objectifs fixés par les politiques nationales.

2. Mise en œuvre de nouveaux dispositifs facilitant la transition

Cette partie illustre les initiatives et les moyens mis en œuvre facilitant la transition vers la science ouverte et permettre ainsi la disposition de moyens technologiques et humains. L'insertion de la culture de la donnée dans l'univers scientifique, nécessite la construction commune de nouvelles formations ou des modalités pédagogiques sur la gestion des données de la recherche. À titre d'exemple, le webinaire, très complet, organisé par le Nœud RDA France le 18 mars 2022 « Amener les scientifiques vers la démarche d'élaboration des plans de gestion des données ». Ce webinaire rassemblait une grande hétérogénéité d'acteurs du monde de la recherche : des chercheurs, ingénieurs de projet, personnels d'accompagnement aux gestions des données, des chargés de missions, des animateurs d'atelier, des documentalistes etc... Une synthèse a été rédigée rassemblant les contributions des participants lors des deux webinaraires organisés sur le même sujet.

Une synthèse des échanges autour des questions liés au PGD a été réalisé, autour de la question : comment amener les chercheurs à la rédaction de PGD ? La synthétisation des écrits des webinaraires permet de mettre en commun, dans une initiative nationale et d'essayer de trouver des solutions de formation afin d'essayer de former au mieux les chercheurs et leur donner envie de gérer leurs données en associant des bonnes pratiques. Voici dans les grandes lignes, les réponses aux questions demandées lors de ce webinaire :

Qu'est ce qui marche ?

- Avoir un contact direct, discuter, avoir le témoignage de scientifiques
- Sensibiliser, répéter l'information, argumenter, positiver, prioriser
- Faire œuvre de pédagogie, utiliser les ressources, des cas d'usage
- Disposer de compétences, s'organiser
- Accompagner, aider, proposer des contreparties, réfléchir collectivement

Quelles difficultés rencontrées vous ?

- Manque d'exemples concrets et de sensibilisation dans certains domaines
- Manque de temps, lourdeur administrative, complexité du PGD
- Manque d'acculturation, de compétences, difficulté à convaincre
- Manque d'appui des tutelles, d'association d'acteurs, de valorisation du travail

Dans les difficultés rencontrées, à plusieurs reprises vient le manque de connaissance sur la question des licences, que j'ai aussi pu rencontrer lors de l'organisation du Sprint, la plupart des chercheurs ne sont pas familiers avec les questions juridiques et éthiques c'est ce qui ressort le plus souvent. Les plans de gestion des données structurés et les plans des institutions permettent d'associer directement les recommandations et bonnes pratiques dans le PGD, facilitent la vie des chercheurs.

Lors du Sprint PGD certains chercheurs ont fini leur rédaction plus tôt que d'autres car ils avaient utilisé un modèle avec les recommandations pour leur établissement.

Qu'est-ce qui pourrait améliorer la situation ?

- Mettre en place de formations pour les doctorants, les chefs de projet
- Disposer de supports, fiches synthétiques, questionnaires-types
- Avoir un cadre commun, des outils plus intégrés
- Avoir des modèles et des exemples, intégrer le PGD dès le début de projet
- Reconnaître cette activité, avoir des moyens humains et financiers

Ces échanges ont permis de partager les expériences et les différentes visions des acteurs présents lors du webinaire afin de créer un cadre commun. La compilation des écrits offre des solutions. Il y a un certain nombre de fois où le besoin de créer des groupes de travail autour de la gestion des données est revenu, comme la cellule data Grenoble, le GTSO données de la recherche Couperin, ou le Nœud France RDA. Ces initiatives et groupes de travail font partie des nouveaux dispositifs qui créent de l'échange et le débat et aident l'écosystème scientifique dans sa transition. Une cartographie nationale de ce qui est développé individuellement dans les établissements et laboratoire avec des retours d'expérience permettraient de définir ce qui fonctionne ou non, et permettrait d'aider les établissements moins avancés à trouver des solutions.

Les pratiques de gestion des données commencent à être documentées dans la littérature ou par le biais de formation mise en libre accès. Le développement de services mutualisés à la sensibilisation et la formation est en plein essor, des plateformes comme le portail Callisto, *open classrooms*, etc. rassemblant des ressources et outils numériques. La collection de ressources pédagogiques DoRANum développé par l'INIST-CNRS, associé aux Unités régionales de formation à l'information scientifique et technique (URFIST), permet aux chercheurs et doctorants de se former en autonomie grâce à des ressources pédagogiques publiées sous licence Creative Commons Attribution (CC-BY). Le Comité pour la Science ouverte soutient également le projet DoRANum disciplinaire afin de développer des contenus en fonction des pratiques disciplinaires de communautés scientifiques. Un projet rassemblant des experts d'une dizaine de disciplines.

Le matériel pédagogique s'ouvre aussi d'un côté et permet aux chercheurs d'avoir plus d'accès à l'information pour se former en autonomie. Le réseau des URFIST fournit aussi des articles ou billet de blog répondant à des questions concernant les pratiques de gestion des données ou encore le PGD avec les nouveautés. Ces articles²⁰ sont accessibles pour tous.

Les outils pour aider les pratiques de gestion évoluent évolution des outils. Désormais l'importation des plans dans DMP OPIDoR permet à chaque établissement d'ajouter ses recommandations

²⁰ <https://urfistinfo.hypotheses.org/tag/pgd>

Construire de nouveaux modes de connaissances autour des données de la recherche Introduction

directement dans les modèles de PGD, offrant la possibilité aux chercheurs d'accéder aux recommandations de gestion de leur unité et de les appliquer. Le nouveau modèle par défaut de DMP OPIDoR « Science Europe » est devenu le modèle par défaut afin de tendre vers des PGD « *machine actionnable* » et de répondre à plusieurs objectifs de facilitation des pratiques et des usages. Ce nouveau modèle structuré permet :

- d'importer en un clic des informations concernant les projets financés par l'ANR,
- de promouvoir la reconnaissance de l'ensemble des contributeurs et promouvoir l'usage d'identifiants pérennes,
- de calculer des coûts et d'établir la liste des coûts associés à la gestion,
- de bénéficier de suggestions de services référencés dans Cat OPIDoR et référencer les services utilisés,
- d'exploiter les informations du DMP représentées en un format exploitable par les machines pour demander des ressources extérieures, pour alimenter d'autres systèmes d'informations,
- de produire le DMP dans un format conforme aux recommandations du groupe RDA DMP Common Standards et faciliter ainsi les échanges entre outils de DMP.

3. Transformation des services d'accompagnement à la gestion des données

Afin d'acquérir une bonne gestion des données, les chercheurs doivent effectuer un travail commun entre l'équipe de recherche et les services d'accompagnement à la gestion des données (informaticiens, archivistes, documentalistes etc...). Les chercheurs n'ont pas besoin d'accompagnement durant toute la durée de leur projet de recherche ; par exemple, dans les phases de collecte et de création de données les chercheurs connaissent très bien les processus. En revanche, c'est au moment de diffuser et valoriser l'information que les chercheurs peuvent faire appel à des personnels qualifiés comme les bibliothécaires (Délémontez, Boukacem-Zeghmouri 2015). Les réponses aux questionnaires d'évaluation de notre atelier Sprint PGD et les enquêtes précédentes confirment que la plupart des chercheurs éprouvent des difficultés des étapes particulières de la gestion des données comme l'application des licences qui est un aspect juridique. Les chercheurs et doctorants ne sont pas nécessairement formés aux aspects juridiques d'un projet de recherche, ces connaissances ne font pas partie de leurs compétences. C'est en ça que les pratiques de gestion des données ne peuvent se rapporter qu'à une seule personne. Dans une optique d'amélioration de la gestion des données à des fins d'ouverture et de réutilisation, de la collecte à l'archivage, il est nécessaire de partager le travail et de solliciter les compétences de spécialistes notamment des délégués à la protection des données (DPD, DPO) si des jeux de données soulèvent des enjeux juridiques, des archivistes pour les questions d'archivage etc... Le tout est de savoir comment impliquer et intégrer tous les acteurs au cycle de la recherche ?

Les services d'accompagnement à la gestion des données ont dû assumer de nouveaux rôles et de nouvelles responsabilités afin d'améliorer leur capacité d'expertise et leur offre de service, malgré les moyens contraints des établissements et les contraintes de temps des chercheurs (Rolando et al., 2015). En outre, pour aider les chercheurs à mener des recherches reproductibles et répondre aux besoins liés à leurs données comme la rédaction de PGD, les services et formations proposés ont dû se renforcer sur le cadre légal du partage et de la publication des données, les questions juridiques et déontologiques comme les licences et la confidentialité, ou encore la protection des données personnelles (Kim, 2021). Afin de pallier ce problème, les services d'accompagnement ont fait appel à d'autres expertises (juristes, professionnels de l'information, ingénieurs informatiques...) pour permettre d'adapter leur offre et de répondre aux besoins des communautés scientifiques.

Actuellement, des investissements stratégiques en ressources humaines permettent à de nouveaux métiers liés aux données de voir le jour et plus. Ces métiers sont référencés sur la plateforme nationale REFERENS²¹. Même si les métiers initiaux connus comme chargé de ressources documentaires, pour acquérir et mettre à disposition l'information scientifique ; juriste, pour gérer les droits de propriété intellectuelle liée aux données, ont toujours été impliqué dans la gestion des données de la recherche ; des nouveaux métiers sont apparus comme architecte de données, qui aide à définir les standards relatifs à la nature des données, leur collecte et leur mode de stockage ou encore chargé du traitement des données scientifiques qui assure et maintient le contenu de bases de données scientifiques, terminologiques, bibliométriques et contribuer à l'étude statistique des données (REFERENS).

Dans les services d'accompagnement à la gestion des données, initialement les métiers de professionnels de l'information, documentaliste ou bibliothécaires pouvaient être des appuis à la gestion des données de la recherche. La plupart des postes associés à la gestion des données sollicitent des diplômes en sciences de l'information et de la communication (Odeh, 2017). Le métier de documentaliste de la donnée associe de nouvelles compétences pour faciliter l'accès aux données, sensibiliser et former à la gestion des données et l'archivage de données à long terme (Odeh, 2017). Les postes de documentaliste, bibliothécaire, professionnel de l'IST demandent de plus en plus de polyvalence. De nouvelles missions apparaissent qui ne sont pas forcément présentes sur les fiches de poste initiales. Dans l'ensemble, les compétences ne changent pas mais il y a un élargissement des connaissances à prendre en compte. Néanmoins, il existe une réelle différence de culture professionnelle entre les personnels d'accompagnement à la gestion des données et les chercheurs. Cette différence relève d'une formation initiale entre SHS et STS ne présentant pas les mêmes enjeux de l'univers scientifique et ces acteurs ne connaissant pas forcément la réalité du travail des chercheurs. C'est pourquoi des enquêtes, comme celles évoquées dans la partie précédente permettent de mettre en parallèle les connaissances et les difficultés de chacun afin d'instaurer par la suite une relation plus homogène entre les personnels d'accompagnement et les chercheurs.

De nouvelles possibilités de carrières dues à l'ouverture des données de la recherche on voit ouvrir de nouveaux postes et des besoins spécifiques comme des chargés de mission à la gestion des données ou chargé de mission science ouverte dans certains établissements. Au sein de Dataactivist et avec l'expérience du Sprint PGD, l'intérêt de l'URFIST Occitanie concernant le développement d'une formation des formateurs plutôt que de la formation classique proposée, permet d'identifier les évolutions des prochains métiers. Les services d'accompagnement devront se forger une culture

²¹https://data.enseignementsup-recherche.gouv.fr/pages/fiche_emploi_type_referens_iii_itrf/?refine.referens_id=F2A43

de la donnée de la recherche et du PGD afin de pouvoir accompagner les chercheurs des établissements sur toutes les étapes du cycle de vie de la donnée. La demande en formation de formateur témoigne de cette transformation des services d'accompagnement qui voient acquérir de nouvelles compétences indispensables pour pouvoir participer et aider les chercheurs dans cette transition vers la science ouverte. Le développement d'une nouvelle offre de service s'inscrit sur la durée pour permettre une coordination efficace.

Conclusion

Les données de la recherche sont reconnues pour être un sujet complexe abordant un ensemble de notions sous-jacentes. En plaçant les données comme un élément central de la transition de la science vers la science ouverte, les pratiques des chercheurs évoluent pour répondre aux obligations des politiques internationales et aux exigences des financeurs. Dans ce mémoire, nous avons abordé le sujet des données de la recherche comme représentant une culture à part entière avec ses propres configurations et spécificités. Nous avons essayé de révéler leur place à la fois en tension entre une culture partageant des valeurs de la science, et faisant partie intégrante du champ scientifique défini par Bourdieu. Le constat étant, dans l'analyse des pratiques, les chercheurs et doctorants sont en train de se forger une culture de la donnée par le biais des plans de gestion des données et des recommandations de gestion des établissements. Néanmoins, une majorité de la communauté scientifique ne sait pas ce qu'est le PGD ou a encore du mal à s'en emparer. De plus, les pratiques de gestion des données des chercheurs se tissent entre les modèles bourdieusien et mertonien. D'un côté, la science ouverte notamment l'ouverture des données de la recherche nous fait revenir aux valeurs de l'Ethos de la science, de l'autre côté il reste une part de conflit et d'intérêt : des chercheurs restent tiraillés entre la valorisation de l'ouverture des données et la perte de propriété sur leurs données à travers leur ouverture.

Les politiques d'ouverture donnent un cadre à suivre en recommandant l'application des principes FAIR, repris dans toutes les politiques science ouverte des établissements, leur adoption à échelle internationale les érige en norme puisqu'ils fournissent un cadre bonnes pratiques en matière de gestion des données. La place centrale des données, bouscule les chercheurs et le monde de l'IST en créant de nouveaux enjeux sur lesquels il faut former, aider et guider les chercheurs. De nouvelles modalités pédagogiques se développent dû à la demande en formation très forte donnant place à de nouvelles initiatives, le déploiement de groupes de travail pour répondre sur ces questions, innover, adapter la matière pédagogique actuelle en alliant le pratique et le théorique. Grâce au Sprint PGD nous avons pu déterminer les éléments clés de conceptualisation d'une formation également tout ce que la conceptualisation d'un nouveau

Dans ce travail, nous avons vu que pour se spécialiser dans la gestion des données de la recherche, devenir un expert du sujet dans son intégralité, il faudrait avoir les connaissances sur les spécificités disciplinaires, ainsi que dans toutes les étapes du cycle de vie de la donnée (archivistique, juridique, informatique etc.). Or, malgré les nouveaux métiers qui se développent, l'accompagnement des chercheurs dans la gestion des données est effectué par plusieurs experts avec des combinaison de compétences scientifiques. Dans ce sens, il est important de créer un socle commun autour des pratiques de gestion c'est en ça que les principes FAIR participent afin de définir des normes communes dont chaque acteur peut s'emparer pour tendre à aller vers ces « bonnes pratiques ». Là où il y a également un grand enjeu c'est dans la construction de nouveaux modes de connaissance,

les besoins en information, en formation sont très grands. Encore en 2022, malgré les nouvelles politiques depuis 2018, ces notions sont récentes, la transition est encore « jeune » même si de plus en plus de moyens sont développés notamment par le MESRI. Les initiatives nationales et mondiales autour des questions des données de la recherche, le développement des plateformes pédagogiques, les formations sont des éléments indispensables afin de mettre tous les acteurs du monde scientifique sur un pied d'égalité dans la culture de la donnée. Afin de maintenir également ce développement, des moyens humains sont attendus par les établissements de recherche. Avec la normalisation mondiale des principes FAIR en matière de gestion des données, les initiatives nationales et internationales qui établissent des passerelles et ponts communs pour partager les connaissances. Il y a une tentative de faciliter les notions complexes liées aux données et à leur par le biais de représentations communes, normes, standards. Néanmoins, il ne faut pas oublier que d'un point de vue historique, les disciplines ont des cultures différentes et des disparités dans leur manière de gérer les données. Même si le mouvement de la science ouverte essaie de construire des représentations et valeurs communes, il reste encore des frontières entre les communautés scientifiques. À mon sens c'est un des plus grands défis de la science ouverte car chaque communauté scientifique possède sa propre culture de la donnée avec ses pratiques et ses habitudes de gestion, les établissements ont également leurs propres pratiques (stockage, sécurisation etc...) même s'il y a une réelle volonté d'y parvenir.

Bibliographie

AMIEL, Philippe, FRONTINI, Francesca, LACOUR, Pierre-Yves et ROBIN, Agnès, 2020. Pratiques de gestion des données de la recherche : une nécessaire acculturation des chercheurs aux enjeux de la science ouverte ? Cahiers Droit, Sciences & Technologies [en ligne]. 11 mai 2020. N° 10, pp. 147-168. [Consulté le 19 février 2022]. DOI 10.4000/cdst.2061. Disponible à l'adresse : <https://journals.openedition.org/cdst/2061>

BLANCHARD, Antoine, 2022. Intégrité scientifique et données de la recherche : comment se mettre en conformité avec le décret.... Dataactivist [en ligne]. 22 mars 2022. [Consulté le 2 juillet 2022]. Disponible à l'adresse : <https://medium.com/dataactivist/int%C3%A9grit%C3%A9-scientifique-et-donn%C3%A9es-de-la-recherche-comment-se-mettre-en-conformit%C3%A9-avec-le-d%C3%A9cret-278fb4f1f0be>

BLOEMERS, Margreet et MONTESANTI, Annalisa, 2020. The FAIR Funding Model: Providing a Framework for Research Funders to Drive the Transition toward FAIR Data Management and Stewardship Practices. *Data Intelligence* [en ligne]. 1 janvier 2020. Vol. 2, n° 1-2, pp. 171-180. [Consulté le 15 juin 2022]. DOI [10.1162/dint.a.00039](https://doi.org/10.1162/dint.a.00039). Disponible à l'adresse : <https://doi.org/10.1162/dint.a.00039>

BORGMAN, Christine L., 2015. Big Data, Little Data, No Data: Scholarship in the Networked World. Cambridge, MA, USA : MIT Press. ISBN 978-0-262-02856-1.

BORGMAN, Christine L., 2020. Qu'est-ce que le travail scientifique des données ? : Big data, little data, no data [en ligne]. Marseille : OpenEdition Press. [Consulté le 20 janvier 2022]. Encyclopédie numérique. ISBN 979-10-365-6541-0. Disponible à l'adresse : <http://books.openedition.org/oep/14692>

BOURDIEU, Pierre, 1975. La spécificité du champ scientifique et les conditions sociales du progrès de la raison. *Sociologie et sociétés* [en ligne]. 1975. Vol. 7, n° 1, pp. 91-118. [Consulté le 17 juin 2022]. DOI 10.7202/001089ar. Disponible à l'adresse : <https://www.erudit.org/fr/revues/socsoc/1975-v7-n1-socsoc122/001089ar/>

BOURDIEU, Pierre, 1976. Le champ scientifique. *Actes de la Recherche en Sciences Sociales* [en ligne]. 1976. Vol. 2, n° 2, pp. 88-104. [Consulté le 10 juillet 2022]. DOI 10.3406/arss.1976.3454. Disponible à l'adresse : https://www.persee.fr/doc/arss_0335-5322_1976_num_2_2_3454

BRANSON, Robert K., 1978. The Interservice Procedures for Instructional Systems Development. Educational Technology [en ligne]. 1978. Vol. 18, n° 3, pp. 11-14. [Consulté le 31 mars 2022]. Disponible à l'adresse : <http://www.jstor.org/stable/44418942>

MAUREL, 2019. L'ouverture des données de recherche : un retour aux sources de l'Ethos de la Science ? - S.I.Lex - [en ligne]. 5 juin 2019. [Consulté le 20 juillet 2022]. Disponible à l'adresse : <https://scinfolex.com/2019/06/05/louverture-des-donnees-de-recherche-un-retour-aux-sources-de-lethos-de-la-science/>

CHARRIER, Philippe, 2020. Objectifs et démarche de l'enquête SOSP-FR (State of Open Science Practices in France). SOSP-FR [en ligne]. [Consulté le 5 août 2022]. Disponible à l'adresse : <https://ecosysteme.hypotheses.org/42>

DÄLLENBACH, David et THONNEY, Delphine, 2019. Data Management Plan : retour d'expérience. Recherche d'idées [en ligne]. 29 octobre 2019. [Consulté le 24 février 2022]. Disponible à l'adresse : <https://campus.hesge.ch/blog-master-is/data-management-plan-retour-dexperience/>

DEBOIN, Marie-Claude, GENOVA, Françoise, BAGE, Anne-Sophie et LAÏ, Paolo, 2022. Amener les scientifiques vers la démarche d'élaboration de plan de gestion de données (PGD) : compilation des contributions des participants du webinaire RDA France, 18 et 25 mars 2022. [en ligne]. 19 juillet 2022. [Consulté le 29 juillet 2022]. DOI 10.18167/agritrop/00587. Disponible à l'adresse : <https://hal-cnrs.archives-ouvertes.fr/hal-03739453>

DEBOIN, Marie-Claude, 2018. Découvrir de nouveaux métiers liés aux données de la recherche [en ligne]. CIRAD. [Consulté le 20 juillet 2022]. Disponible à l'adresse : <http://coop-ist.cirad.fr/metiers>

DÉLÉMONTEZ, Renaud et BOUKACEM-ZEGHMOURI, Chérifa, 2015. Données de la recherche : entre discours, réalités et valeur. *I2D - Information, données & documents* [en ligne]. 2015. Vol. 52, n° 4, pp. 56-57. [Consulté le 10 août 2022]. DOI [10.3917/i2d.154.0056](https://doi.org/10.3917/i2d.154.0056). Disponible à l'adresse : <https://www.cairn.info/revue-i2d-information-donnees-et-documents-2015-4-page-56.htm>

DONATI, Caroline Sophie, 2019. Données de la recherche : Quelles pratiques ? Quels besoins ? Enquête à Aix-Marseille Université. . pp. 34.

EUROPEAN COMMISSION. DIRECTORATE GENERAL FOR RESEARCH AND INNOVATION. et PWC EU SERVICES., 2018. Cost-benefit analysis for FAIR research data : cost of not having FAIR research data. [en ligne]. LU : Publications Office. [Consulté le 10 août 2022]. Disponible à l'adresse : <https://data.europa.eu/doi/10.2777/02999>

FONTAINE, Mathieu, JUILLET, Sylvain et FROGER, Didier, 2017. La donnée numérique : l'or noir du XXI^e siècle? Actu-Juridique [en ligne]. 7 septembre 2017. [Consulté le 6 août 2022]. Disponible à l'adresse : <https://www.actu-juridique.fr/civil/la-donnee-numerique-lor-noir-du-xxie-siecle/>

FORGERON, Jean-François, 2017. Dark Analytics : les « données sombres », un enjeu majeur. Lexing Alain Bensoussan Avocats [en ligne]. 10 juillet 2017. [Consulté le 3 août 2022]. Disponible à l'adresse : <https://www.alain-bensoussan.com/avocats/dark-analytics-donnees-sombres-enjeu/2017/07/10/>

GARFIELD, Eugene, 1996. What is the Primordial Reference for the phrase « Publish or Perish »? . pp. 2.

GINGRAS, Yves, 2017. Champ. In : BOUCHARD, Frédéric, DORAY, Pierre et PRUD'HOMME, Julien (éd.), *Sciences, technologies et sociétés de A à Z* [en ligne]. Montréal : Presses de l'Université de Montréal. pp. 37-40. Thématique Sciences sociales. [Consulté le 6 juillet 2022]. ISBN 978-2-8218-9562-1. Disponible à l'adresse : <http://books.openedition.org/pum/4267>

GUERRY, 2021. Points de repère sur les communs numériques. [en ligne]. 20 octobre 2021. [Consulté le 3 août 2022]. Disponible à l'adresse : <https://bzg.fr/points-de-repere-sur-les-communs-numeriques/>

INAU, Esther Thea, SACK, Jean, WALTEMATH, Dagmar et ZELEKE, Atinkut Alamirrew, 2021. Initiatives, Concepts, and Implementation Practices of FAIR (Findable, Accessible, Interoperable, and Reusable) Data Principles in Health Data Stewardship Practice: Protocol for a Scoping Review. *JMIR Research Protocols* [en ligne]. 2 février 2021. Vol. 10, n° 2, pp. e22505. [Consulté le 25 février 2022]. DOI [10.2196/22505](https://doi.org/10.2196/22505). Disponible à l'adresse : <https://www.researchprotocols.org/2021/2/e22505>

JETTEN, Mijke, GROOTVELD, Marjan, MORDANT, Annemie, JANSSEN, Mascha, BLOEMERS, Margreet, MIEDEMA, Margriet et GELDER, Celia W. G. van, 2021. Professionalising data stewardship in the Netherlands. Competences, training and education.

Dutch roadmap towards national implementation of FAIR data stewardship [en ligne]. Zenodo. [Consulté le 12 juillet 2022]. Disponible à l'adresse : <https://zenodo.org/record/4623713>

KIM, Jeonghyun, 2021. Determining research data services maturity: The role of library leadership and stakeholder involvement. *Library & Information Science Research* [en ligne]. avril 2021. Vol. 43, n° 2, pp. 101092. [Consulté le 25 février 2022]. DOI [10.1016/j.lisr.2021.101092](https://doi.org/10.1016/j.lisr.2021.101092). Disponible à l'adresse : <https://www.sciencedirect.com/science/article/pii/S0740818821000220>

LE BÉCHEC, Mariannig, BOUCHARD, Aline, CHARRIER, Philippe, DENECKER, Claire, GALLEZOT, Gabriel et RENNES, Stéphanie, 2022. Pratiques et usages des outils numériques dans les communautés scientifiques en France [en ligne]. report. Comité pour la science ouverte. [Consulté le 18 mars 2022]. Disponible à l'adresse : <https://hal-lara.archives-ouvertes.fr/hal-03545512>

LETROUT, Carole, CACHARD, Pierre-Yves, DUPUIS, Monique et FROMENT, Bernard, 2021. La place des bibliothèques universitaires dans le développement de la science ouverte. *enseignementsup-recherche.gouv.fr* [en ligne]. février 2021. [Consulté le 31 mars 2022]. Disponible à l'adresse : <https://www.enseignementsup-recherche.gouv.fr/fr/la-place-des-bibliotheques-universitaires-dans-le-developpement-de-la-science-ouverte-47671>

LÖFFLER, Felicitas, WESP, Valentin, KÖNIG-RIES, Birgitta et KLAN, Friederike, 2021. Dataset search in biodiversity research: Do metadata in data repositories reflect scholarly information needs? *PLOS ONE* [en ligne]. 24 mars 2021. Vol. 16, n° 3, pp. e0246099. [Consulté le 25 février 2022]. DOI [10.1371/journal.pone.0246099](https://doi.org/10.1371/journal.pone.0246099). Disponible à l'adresse : <https://journals.plos.org/plosone/article?id=10.1371/journal.pone.0246099>

MERTON, Robert K., 1942. The Normative Structure of Science. [en ligne]. 1942. [Consulté le 29 juillet 2022]. Disponible à l'adresse : <https://www.panarchy.org/merton/science.html>

MONS, Barend, NEYLON, Cameron, VELTEROP, Jan, DUMONTIER, Michel, DA SILVA SANTOS, Luiz Olavo Bonino et WILKINSON, Mark D., 2017. Cloudy, increasingly FAIR; revisiting the FAIR Data guiding principles for the European Open Science Cloud. *Information Services & Use* [en ligne]. 1 janvier 2017. Vol. 37, n° 1, pp. 49-56. [Consulté le 07 juillet 2022]. DOI [10.3233/ISU-170824](https://doi.org/10.3233/ISU-170824). Disponible à l'adresse : <https://content.iospress.com/articles/information-services-and-use/isu824>

NICHOLAS, David, WATKINSON, Anthony, ABRIZAH, Abdullah, RODRÍGUEZ-BRAVO, Blanca, BOUKACEM-ZEGHMOURI, Cherifa, XU, Jie, ŚWIGOŃ, Marzena et HERMAN, Eti,

2020. Does the scholarly communication system satisfy the beliefs and aspirations of new researchers? Summarizing the Harbingers research. *Learned Publishing* [en ligne]. 2020. Vol. 33, n° 2, pp. 132-141. [Consulté le 12 août 2022]. DOI 10.1002/leap.1284. Disponible à l'adresse : <http://onlinelibrary.wiley.com/doi/abs/10.1002/leap.1284>

ODEH, Souad, 2017. Les données de la recherche : transformation ou transmission du métier de documentaliste. *I2D - Information, données documents* [en ligne]. 2017. Vol. 54, n° 4, pp. 4-7. [Consulté le 23 juin 2022]. Disponible à l'adresse : <http://www.cairn.info/revue-i2d-information-donnees-et-documents-2017-4-page-4.htm>

OO, Cherry Zin, CHEW, Adrian W., WONG, Adeline L. H., GLADDING, Joanne et STENSTROM, Cecilia, 2021. Delineating the successful features of research data management training: a systematic review. *International Journal for Academic Development* [en ligne]. 10 mars 2021. Vol. 0, n° 0, pp. 1-16. [Consulté le 24 février 2022]. DOI [10.1080/1360144X.2021.1898399](https://doi.org/10.1080/1360144X.2021.1898399). Disponible à l'adresse : <https://doi.org/10.1080/1360144X.2021.1898399>

PERSAUD, Bhaleka D., DUKACZ, Krysha A., SAHA, Gopal C., PETERSON, Amber, MORADI, Laleh, O'HEARN, Stephen, CLARY, Erin, MAI, Juliane, STEELEWORTHY, Michael, VENKITESWARAN, Jason J., KHEYROLLAH POUR, Homa, WOLFE, Brent B., CAREY, Sean K., POMEROY, John W., DEBEER, Chris M., WADDINGTON, James M., VAN CAPPELLEN, Philippe et LIN, Jimmy, 2021. Ten best practices to strengthen stewardship and sharing of water science data in Canada. *Hydrological Processes* [en ligne]. 2021. Vol. 35, n° 11, pp. e14385. [Consulté le 25 février 2022]. DOI [10.1002/hyp.14385](https://doi.org/10.1002/hyp.14385). Disponible à l'adresse : <http://onlinelibrary.wiley.com/doi/abs/10.1002/hyp.14385>

PERRIER, Laure et BARNES, Leslie, 2018. Developing Research Data Management Services and Support for Researchers: A Mixed Methods Study. *Partnership: The Canadian Journal of Library and Information Practice and Research* [en ligne]. 8 mai 2018. Vol. 13, n° 1. [Consulté le 22 février 2022]. DOI [10.21083/partnership.v13i1.4115](https://doi.org/10.21083/partnership.v13i1.4115). Disponible à l'adresse : <https://journal.lib.uoguelph.ca/index.php/perj/article/view/4115>

ROBIN, Agnès, 2017. Les données scientifiques au prisme du dispositif open data. *Communication Commerce Électronique* [en ligne]. septembre 2017. Vol. 9, n° étude 14, pp. 7-14. [Consulté le 12 décembre 2021]. Disponible à l'adresse : <https://hal.umontpellier.fr/hal-01845205>

REBOUILLAT, Violaine, 2019. Les données scientifiques face aux enjeux de la recherche en Sciences, Technologie et Médecine : enquête exploratoire à l'Université de Strasbourg. *Études de communication. langages, information, médiations* [en ligne]. 1 juin 2019. N° 52, pp. 99-116. [Consulté le

10 juillet 2022]. DOI [10.4000/edc.8681](https://doi.org/10.4000/edc.8681). Disponible à l'adresse : http://journals.openedition.org/edc/8681#xd_co_f=NTk3N2QwZjktOGMzNS00MmE1LWFjNGYtNzM0N2FIYTIIYjg0~

REBOUILLAT, Violaine, 2021. Le partage des données vu par les chercheurs : une approche par la valeur. *Les Enjeux de l'information et de la communication* [en ligne]. 2021. Vol. 221, n° 1, pp. 35-53. [Consulté le 25 mars 2022]. Disponible à l'adresse : <http://www.cairn.info/revue-les-enjeux-de-l-information-et-de-la-communication-2021-1-page-35.htm?ref=doi>

ROLANDO, Lizzy, CARLSON, Jake, HSWE, Patricia et WELLS, Susan, 2015. Data Management Plans as a Research Tool. *Bulletin of the Association for Information Science and Technology* [en ligne]. 2015. Vol. 41, n° 5, pp. 43-45. [Consulté le 24 février 2022]. DOI [10.1002/bult.2015.1720410510](https://doi.org/10.1002/bult.2015.1720410510). Disponible à l'adresse : <http://onlinelibrary.wiley.com/doi/abs/10.1002/bult.2015.1720410510>

SAMUEL, Sheeba et KÖNIG-RIES, Birgitta, 2021. Understanding experiments and research practices for reproducibility: an exploratory study. PeerJ [en ligne]. 21 avril 2021. Vol. 9, pp. e11140. [Consulté le 28 février 2022]. DOI [10.7717/peerj.11140](https://doi.org/10.7717/peerj.11140). Disponible à l'adresse : <https://peerj.com/articles/11140>

SARAZIN, Arthur, 2021. 5 jours pour initier la démarche open data de 19 collectivités. *Dataactivist* [en ligne]. 21 mai 2021. [Consulté le 29 août 2022]. Disponible à l'adresse : <https://medium.com/dataactivist/5-jours-pour-initier-la-d%C3%A9marche-open-data-de-19-collectivit%C3%A9s-1b635a0b1645>

SARAZIN, Arthur, 2021. *Construire les usages des données ouvertes en France : proposition d'un Open Data Canvas par l'approche du design science* [en ligne]. Thèse de doctorat. Université Grenoble Alpes. [Consulté le 3 mai 2022]. Disponible à l'adresse : <http://www.theses.fr/2021GRALG007>

SCHÖPFEL, Joachim, 2018. Hors norme ? Une approche normative des données de la recherche. In : *In COSSI 2018, Processus de normalisation et durabilité de l'information* [en ligne]. Bordeaux, France. mai 2018. [Consulté le 14 mai 2022]. Disponible à l'adresse : <https://hal.archives-ouvertes.fr/hal-01800597>

SCHÖPFEL, Joachim, 2018. Vers une culture de la donnée en SHS [en ligne]. Research Report. Université de Lille. [Consulté le 22 juin 2022]. Disponible à l'adresse : <https://hal.archives-ouvertes.fr/hal-01846849>

TENOPIR, Carol, SANDUSKY, Robert J., ALLARD, Suzie et BIRCH, Ben, 2014. Research data management services in academic research libraries and perceptions of librarians. *Library & Information Science Research* [en ligne]. avril 2014. Vol. 36, n° 2, pp. 84-90. [Consulté le 2 mars 2022]. DOI [10.1016/j.lisr.2013.11.003](https://doi.org/10.1016/j.lisr.2013.11.003). Disponible à l'adresse : <https://www.sciencedirect.com/science/article/pii/S0740818814000255>

VÉTOIS, Jacques, 2018. RGPD et loi sur les données personnelles : nouvelles contraintes, nouvelles avancées ? Terminal. *Technologie de l'information, culture & société* [en ligne]. 17 juin 2018. N° 122. [Consulté le 27 août 2022]. Disponible à l'adresse : <http://journals.openedition.org/terminal/2182?lang=en>

VITALE, Cynthia et MOULAISON SANDY, Heather, 2019. Data Management Plans A Review. *DESIDOC Journal of Library & Information Technology*. décembre 2019. Vol. 39, pp. 322-328. DOI [10.14429/djlit.39.06.15086](https://doi.org/10.14429/djlit.39.06.15086).

WILJES, Cord et CIMIANO, Philipp, 2019. Teaching Research Data Management for Students. [en ligne]. [Consulté le 25 février 2022]. Disponible à l'adresse : <https://datascience.codata.org/articles/10.5334/dsj-2019-038>

WILKINSON, Mark D., DUMONTIER, Michel, AALBERSBERG, IJsbrand Jan, APPLETON, Gabrielle, AXTON, Myles, BAAK, Arie, BLOMBERG, Niklas, BOITEN, Jan-Willem, DA SILVA SANTOS, Luiz Bonino, BOURNE, Philip E., BOUWMAN, Jildau, BROOKES, Anthony J., CLARK, Tim, CROSAS, Mercè, DILLO, Ingrid, DUMON, Olivier, EDMUNDS, Scott, EVELO, Chris T., FINKERS, Richard, GONZALEZ-BELTRAN, Alejandra, GRAY, Alasdair J. G., GROTH, Paul, GOBLE, Carole, GRETHE, Jeffrey S., HERINGA, Jaap, 'T HOEN, Peter A. C., HOOFT, Rob, KUHN, Tobias, KOK, Ruben, KOK, Joost, LUSHER, Scott J., MARTONE, Maryann E., MONS, Albert, PACKER, Abel L., PERSSON, Bengt, ROCCA-SERRA, Philippe, ROOS, Marco, VAN SCHAIK, Rene, SANSONE, Susanna-Assunta, SCHULTES, Erik, SENGSTAG, Thierry, SLATER, Ted, STRAWN, George, SWERTZ, Morris A., THOMPSON, Mark, VAN DER LEI, Johan, VAN MULLIGEN, Erik, VELTEROP, Jan, WAAGMEESTER, Andra, WITTENBURG, Peter, WOLSTENCROFT, Katherine, ZHAO, Jun et MONS, Barend, 2016. The FAIR Guiding Principles for scientific data management and stewardship. *Scientific Data* [en ligne]. 15 mars 2016. Vol. 3, n° 1, pp. 160018. [Consulté le 20 décembre 2021]. DOI [10.1038/sdata.2016.18](https://doi.org/10.1038/sdata.2016.18). Disponible à l'adresse : <https://www.nature.com/articles/sdata201618>

Sitographie

Accompagner la rédaction d'un plan de gestion de données (PGD) ou Data Management Plan (DMP) [Formation de formateurs]. [en ligne]. [Consulté le 20 août 2022]. Disponible à l'adresse : <https://sygefor.reseau-urfist.fr/#/training/8775/10054?from=true>

Bilan du Plan national pour la science ouverte : des engagements tenus, des avancées majeures réalisées en 3 ans. *enseignementsup-recherche.gouv.fr* [en ligne]. [Consulté le 30 août 2022]. Disponible à l'adresse : <https://www.enseignementsup-recherche.gouv.fr/fr/bilan-du-plan-national-pour-la-science-ouverte-des-engagements-tenus-des-avancees-majeures-realisees-48296>

DMP publics. [en ligne]. [Consulté le 27 août 2022]. Disponible à l'adresse : https://dmp.opidor.fr/public_plans

DoRANum – Données de la recherche : Apprentissage Numérique. [en ligne]. [Consulté le 30 août 2022]. Disponible à l'adresse : <https://doranum.fr/>

Fiche d'un emploi type de RéFérens III — Plateforme open data (données ouvertes). [en ligne]. [Consulté le 20 août 2022]. Disponible à l'adresse : https://data.enseignementsup-recherche.gouv.fr/pages/fiche_emploi_type_referens_iii_itrf/?refine.referens_id=F2A43

FORCE11. [en ligne]. [Consulté le 20 août 2022]. Disponible à l'adresse : <https://force11.org/>

Formations en ligne et cours en accès libre. [en ligne]. [Consulté le 20 août 2022]. Disponible à l'adresse : <https://openclassrooms.com/fr/>

gt-dmp - Groupe de travail sur les formations aux DMP - info. [en ligne]. [Consulté le 20 août 2022]. Disponible à l'adresse : <https://groupes.renater.fr/sympa/info/gt-dmp>

Information scientifique et technique. *enseignementsup-recherche.gouv.fr* [en ligne]. [Consulté le 20 août 2022]. Disponible à l'adresse : <https://www.enseignementsup-recherche.gouv.fr/fr/information-scientifique-et-technique-51161>

Jérémy Garniaux, [@mapperfr] (2022) “En train de contribuer à mon premier data paper! Il y a quelque chose de vraiment satisfaisant à décrire avec ” Twitter. 25 juillet 2022, 12h47. https://twitter.com/mapperfr/status/1551519591574151168?s=20&t=g5Z9eyUhTJ_DyFqzQ6H6Mw

Construire de nouveaux modes de connaissances autour des données de la recherche Bibliographie

L'économie numérique, quelles opportunités de croissance ?. [en ligne]. [Consulté le 3 août 2022]. Disponible à l'adresse : <https://www.economie.gouv.fr/facileco/leconomie-numerique-queelles-opportunités-croissance>

Les données de la Recherche - CALLISTO | Canal U. [en ligne]. [Consulté le 4 août 2022]. Disponible à l'adresse : <https://www.canal-u.tv/chaines/capitoul/les-donnees-de-la-recherche/les-donnees-de-la-recherche-callisto>

myMarkmap. [en ligne]. [Consulté le 17 août 2022]. Disponible à l'adresse : <https://mymarkmap.vercel.app/>

PGD | UrfistInfo. [en ligne]. [Consulté le 4 août 2022]. Disponible à l'adresse : <https://urfistinfo.hypotheses.org/tag/pgd>

Portail Web de l'European open science cloud (EOSC). [en ligne]. [Consulté le 20 août 2022]. Disponible à l'adresse : <https://www.ouvrirlascience.fr/portail-web-de-leosc>

Recherche - Frédérique Vidal dévoile à l'Université de Lille le plan pour la science ouverte / Actualités / Actualités / Accueil - Les services de l'État dans le Nord. [en ligne]. [Consulté le 30 août 2022]. Disponible à l'adresse : <https://www.nord.gouv.fr/index.php/Actualites/Actualites/Recherche-Frederique-Vidal-devoile-a-l-Universite-de-Lille-le-plan-pour-la-science-ouverte>

Science ouverte : la France rejoint GO FAIR en tant que co-fondatrice. enseignementsup-recherche.gouv.fr [en ligne]. [Consulté le 30 août 2022]. Disponible à l'adresse : <https://www.enseignementsup-recherche.gouv.fr/fr/science-ouverte-la-france-rejoint-go-fair-en-tant-que-co-fondatrice-46198>

Victor Gay récompensé par le prix Science Ouverte pour les données de la recherche, 2022. TSE [en ligne]. [Consulté le 3 août 2022]. Disponible à l'adresse : <https://www.tse-fr.eu/fr/victor-gay-recompense-par-le-prix-science-ouverte-pour-les-donnees-de-la-recherche>

Annexes

Annexe 1 : Déroulé d'animation de l'atelier Sprint PGD

14:00 ⌚ 10	Accueil des participants Placement des participants dans les salles par équipes projet	Accueil
14:10 ⌚ 15	Présentation générale Présentation de l'atelier	Présentation
14:25 ⌚ 45	F - Facile à trouver Pas-à-pas de la partie F et rédaction dans DMP OPIDoR	Production
15:10 ⌚ 40	A - Accessible Pas-à-pas de la partie A et rédaction dans DMP OPIDoR	Production
15:50 ⌚ 5	Pause Description	
15:55 ⌚ 40	I - Interopérable Pas-à-pas de la partie I et rédaction dans DMP OPIDoR	Production
16:35 ⌚ 35	R - Réutilisable Pas-à-pas de la partie R et rédaction dans DMP OPIDoR	Production
17:10 ⌚ 10	Bilan de la journée Vous avez rédigé votre PGD, et après ?	Synthèse
17:20 ⌚ 10	Questionnaire d'évaluation À la suite de l'atelier les participants remplissent le questionnaire d'évaluation de l'atelier	Production
17:30 ⌚ 3:30		

Annexe 2 : Tableau récapitulatif des évolutions du Sprint PGD

Version du Sprint	Les moins	Solution(s) proposée(s)	TO-DO
Sprint PGD V0 : Journée d'atelier autour des principes FAIR	Réunion avec les équipes de Toulouse (17/02/22) : limite de temps pour les chercheurs Le format d'une journée est trop long	Modification du temps de l'atelier sur une demi-journée	Articulation de l'atelier autour des principes FAIR (plus d'enjeux pédagogiques que la simple écriture du PGD)
Sprint PGD V1 : Développement de l'Open Data Canvas	Brainstorming avec Antoine et Arthur - Dataactivist (22/03/22) : discussion autour des limites de l'outil ODC Besoin de réorganiser les ressources : recherche en webdesign pour un nouveau prototypage, option bubble.io, mais limite de temps	Rassembler les ressources du parcours DoRANum en utilisant un Gitbook	Construire un parcours en pas-à-pas suivant les principes FAIR, fluide et ergonomique comprenant des recommandations et ressources pour guider la rédaction du PGD
Sprint PGD V3 : Développement du Gitbook & du pas-à-pas du parcours ODC	Test en interne du parcours - Dataactivist (22/04/22) : Utilisateur peut se perdre sur le parcours entre les ressources et le pas-à-pas	Restructuration du Gitbook et de l'ODC selon le pas-à-pas, découpage plus fin des parties du Gitbook	Concevoir le Gitbook avec des pages représentant les sous-étapes du parcours (F1, F2, F3...) Concevoir le pas-à-pas de manière à remplir toutes les questions du PGD
Sprint PGD V3.1 : Modifications du Gitbook & du parcours ODC - précisions des sections DMP OPIDoR	Test en interne du parcours - Dataactivist (22/04/22) : test basé sur l'utilisateur, ici Samuel (Dataactivist) qui ne connaissant pas le parcours	Préciser les sections du PGD que l'utilisateur peut remplir dès le début d'un pas-à-pas Précision du temps également dans l'onglet "avant de commencer"	
Sprint PGD V4 : Réalisation de la version prototype du Sprint (bilan)	Premier atelier Sprint PGD (10/05/22) : 1- Rappel plus concret sur ce qu'est un PGD (en quoi est-ce nécessaire, pourquoi faut-il en faire un ?)	1- Au niveau de la présentation générale ou dans le pas-à-pas : rappeler ce qu'est un PGD et à quoi cela sert 2 - Axer le parcours plutôt	Garder le fonctionnement de l'atelier par équipe pour la rédaction À la suite de la présentation

Construire de nouveaux modes de connaissances autour des données de la recherche Annexes

	<p>2- Rédaction non linéaire entre le parcours ODC et DMP OPIDoR</p> <p>3 - Les exemples n'incluent pas les PGD H2020 et Science Europe</p> <p>4 - Le pas-à-pas n'est pas assez directif, il se cache trop derrière les questions du PGD et les ressources proposées, en laissant les chercheurs décider comment ils veulent traiter chaque point</p> <p>5 - Problème d'ergonomie pour l'utilisateur qui doit s'y retrouver entre ODC, Gitbook et DMP OPIDoR</p>	<p>sur le cycle de vie de la donnée, ce qui permettra d'avoir une rédaction linéaire (surtout pour le modèle ANR)</p> <p>3 - Prévoir plusieurs types d'exemples en fonction des différents modèles au début des parcours, et les remonter dans le pas-à-pas</p> <p>4 - Donner des consignes beaucoup plus précises, avec un objectif concret à chaque pas-à-pas</p> <p>5 - Intégrer tout le contenu du Gitbook dans l'ODC</p>	<p>générale, prévoir un court temps d'échanges sur les différents projets</p> <p>En repensant le pas-à-pas, trier les ressources pour faire ressortir les plus pertinentes comme le logigramme ou encore détailler plus au niveau des coûts de gestion...</p>
<p>Sprint PGD - Version Cycle de vie de la donnée</p>	<p>Nouvelle version autour du cycle de vie</p> <p>1 - Reconstruire les pas-à-pas en suivant le PGD ou le cycle de vie</p> <p>2 - Pas-à-pas exhaustifs qui répondent à toutes les recommandations formulées par DMP OPIDoR</p> <p>3 - Pas-à-pas personnalisés selon le modèle de PGD (H2020, SE, ANR)</p>	<p>1 - Nouvelle base Airtable composée de toutes les questions du PGD, questions des modèles, bibliothèque de pas-à-pas, associés à des ressources (Gitbook, lien, vidéo YT, autre...)</p> <p>2 - Association des pas-à-pas selon les questions du modèle</p>	

Annexe 2 : Questionnaire d'évaluation du Sprint PGD



Questionnaire d'évaluation de l'atelier Sprint PGD

Nous voici à la fin de l'atelier, merci de prendre encore quelques minutes pour nous faire part de votre expérience afin de nous aider à améliorer notre accompagnement.

Comment évaluez-vous votre capacité à rédiger un PGD avant cet atelier ? *

Comment évaluez-vous votre capacité à rédiger un PGD après cet atelier ? *

Quel est votre retour sur le temps investi dans l'atelier sur une échelle de 1 à 5 ? *

- 1 : Inutile : je n'ai rien gagné, rien appris
- 2 : Utile mais ça ne valait pas à 100% le temps que j'y ai passé
- 3 : Moyen : je n'ai pas perdu mon temps, sans plus
- 4 : Bon : j'ai gagné plus que le temps que j'y ai passé
- 5 : Excellent : ça valait bien plus que le temps que j'y ai passé

Quelle étape vous a semblé la plus difficile ? *

Mentionner ce qui vous a posé le plus de problèmes pendant la rédaction de votre PGD.

Annexe 3 : Tableau des réponses au questionnaire d'évaluation de l'atelier

		Questionnaire évaluation-Grid view					
Participants	Progression	Quel est votre retour sur le temps investi dans l'atelier sur une échelle de 1 à 5 ?	Qu'est-ce qui vous a manqué durant le Sprint ?	Quelle étape vous a semblé la plus difficile ?	Qu'est-ce qui vous a le plus aidé ?	Comment évaluez-vous vos connaissances sur les principes FAIR suite à l'atelier ?	Autre remarque
1	1	4 : Bon : j'ai gagné plus que le temps que j'y ai passé	rien!	Les questions RQPD complexifient les décisions de partage des données. Quelles données partager pour des entretiens qualitatifs très identifiants?	Le matériel était bien présenté pour faciliter le remplissage des rubriques les plus délicates.	Le parcours m'a permis de mieux les maîtriser	
2	1	5 : Excellent : ça valait bien plus que le temps que j'y ai passé	. du temps . des infos sur les données du projet, à chercher plus tard	. savoir quels jeux de données mentionner . savoir à l'avance ce qui va être fait des données, où elles vont être installées . pour le projet presque terminé, faut-il mentionner ou non les jeux de données qu'on n'a finalement pas utilisé ?	tout est très utile et complémentaire : l'accompagnement de l'équipe, le fait de découper la rédaction en étapes, les ressources en ligne	Le parcours m'a permis de mieux les maîtriser	En fait nous faisons de la recherche sur les métadonnées et les principes FAIR donc je n'ai pas trop appris. Mais vos outils et informations sont bien faits je trouve.
3	2	4 : Bon : j'ai gagné plus que le temps que j'y ai passé	Un petit document introductoire à lire avant aurait été le bienvenu : rédiger un PGD ça consiste à faire telle et telle chose, et de manière transversale nous allons découvrir la démarche FAIR qui consiste à détailler, pour chaque point, potentiellement 4 aspects : etc. et un petit exemple pour rendre les choses concrètes, notamment pour les personnes qui font ça pour la première fois.	Utiliser la trame non linéaire de rédaction des différents items du PGD (trame FAIR nécessitant de revenir plusieurs fois sur un item donné du PGD). Je me serais senti plus à l'aise d'aborder les items de manière linéaire par rapport à la trame du PGD, quitte à ce qu'on me guide pour expliciter chacun des points liés à la démarche FAIR pour un item donné	Accompagnement de l'équipe pour comprendre la démarche globale. Le site web ODC semble contenir beaucoup de choses (tout ?) mais du fait que le remplissage ne se fait pas de manière linéaire par rapport au est un peu perturbant.	Le parcours m'a permis de découvrir	Equipe très à l'écoute, peut-être faire des groupes de niveaux avec des objectifs différents (découverte, perfectionnement, approfondissement par exemple).
4	2	4 : Bon : j'ai gagné plus que le temps que j'y ai passé	Les autres membres de mon équipe auraient été un plus pour rendre plus facile la rédaction	La partie concernant les aspects juridiques	Le pas à pas Open Data Canvas mais surtout l'accompagnement de l'équipe qui était incroyable	Le parcours m'a permis de les découvrir	
5	2	4 : Bon : j'ai gagné plus que le temps que j'y ai passé	RAS	Bien définir les données, métadonnées et les aspects logiciels, juridique. Le PGD doit se construire avec le chercheur concerné et ne rien oublier.	L'ensemble des ressources mises à disposition. Je les garde précieusement.	Le parcours m'a permis de mieux les maîtriser	
6	2	4 : Bon : j'ai gagné plus que le temps que j'y ai passé	ras	les licences et les standards de métadonnées	L'accompagnement de l'équipe qui aide à y voir plus clair (les ressources sont très denses) et à débloquer des questions	Le parcours m'a permis de mieux les maîtriser	
7	2	5 : Excellent : ça valait bien plus que le temps que j'y ai passé	Pas identifiés pour le moment	Description des produits de recherche	Accompagnement équipe	Le parcours m'a permis de mieux les maîtriser	
8	2	4 : Bon : j'ai gagné plus que le temps que j'y ai passé	Une explication d'où un pgd est nécessaire. Pourquoi faut il en faire un.	Comprendre tous les concepts liés au PGD et pouvoir caractériser les besoins encore flou du projet.	Le personnel et les autres participants.	Le parcours m'a permis de les découvrir	
9	2	3 : Moyen : je n'ai pas perdu mon temps, sans plus	Je ne vois pas, désolé	Répondre aux questions du format H2020 sur les aspects que je ne maîtrise pas (pout et sécurité)	Les animateurs et leur expérience dans l'échange	Le parcours ne m'a rien apporté	Je ne mesurais absolument pas la difficulté de l'exercice consistant à créer un PGD avant cet atelier. Cela m'aura permis de me préparer pour les cas futurs
10	2	4 : Bon : j'ai gagné plus que le temps que j'y ai passé	Il aurait été bon d'avoir une version alternative du guide dans le cas où notre projet est d'un autre type qu'ANR (H2020 par exemple)	Aucune en particulier	Le retour de l'équipe présente sur place pour aider à répondre à certaines questions lorsque le projet sur lequel on doit réaliser un PGD présente de nombreuses particularités	Le parcours m'a permis de mieux les maîtriser	
11	3	4 : Bon : j'ai gagné plus que le temps que j'y ai passé	/	Étape R	Les ressources mises à disposition, le pas-à-pas Open Data Canvas, l'accompagnement de l'équipe	Le parcours m'a permis de mieux les maîtriser	

Annexe 4 : Captures d'écran du nouveau parcours PGD Open Data Canvas

Bienvenue dans le patchwork de construction d'un PGD

Un parcours d'accompagnement de bout-en-bout pour construire son Plan de Gestion des Données

Bâti autour du cycle de vie de la donnée, vous y trouverez un **patchwork de savoir-faire, d'outils et de documentations** tissé pour construire votre PGD.

5 étapes pour construire son PGD

Un parcours d'accompagnement pas-à-pas et de bout-en-bout pour construire son PGD



Étape 1 - Acquisition des données

Première étape du cycle de vie : acquisition des données, collecte et réutilisation de données existantes.



Étape 2 - Traitement des données

Deuxième étape du cycle de vie : le traitement des données, analyse, documentation...



Étape 3 - Accès et partage des données

Troisième étape du cycle de vie : accessibilité et stockage des données durant le projet de recherche.



Étape 4 - Conservation et archivage des données

Quatrième étape du cycle de vie : tri et préparation des données pour archivage.



Étape 5 - Réutilisation des données

Cinquième et dernière étape du cycle de vie : vérification des moyens mis en œuvre pour la réutilisation de vos données.



[Le Manifeste](#)

[Les composants](#)

[Patchwork #1 Open Data](#)

[Patchwork #2 PGD](#)

[Patchwork #3 Algorithmes](#)

[Contribuer](#)

Étape 1 - Acquisition des données

Une étape pour décrire l'acquisition de vos données

Cette étape se réalise en 6 pas-à-pas

Construisez votre PGD en suivant le parcours

Un parcours d'accompagnement pas-à-pas et de bout-en-bout pour construire votre PGD



1.1 Identifier les produits de recherche du projet



1.2 Expliquer si des données sont réutilisées ou collectées



1.3 Choisir le meilleur format de données



1.4 Respecter la législation sur les données à caractère personnel



1.5 Identifier les enjeux de propriété intellectuelle



1.6 Respecter les principes éthiques

1.5 Identifier les enjeux de propriété intellectuelle

Le pas-à-pas

Déterminer les droits de propriété intellectuelle qui s'appliquent aux données

^  Selon la nature des données, différents régimes juridiques peuvent s'appliquer, et les droits attachés à ces données peuvent engendrer des obligations d'ouverture, de fermeture, ou encore des règles d'accès spécifiques.

Le code de la propriété intellectuelle n'accorde de droit d'auteur que sur les œuvres (textes, dessins, photographies...) et les bases de données (leur contenu comme leur structure).

[Ressource clé](#)

Déterminer les titulaires des droits sur les données

^  Les titulaires des droits sur les œuvres sont les personnes alors que les titulaires des droits sur les bases de données sont l'organisme (producteur de la base de données), seul ou avec ses partenaires.

On ne dit pas d'un organisme de recherche qu'il est « propriétaire » des données : on dira plutôt que les données sont des informations publiques dont l'organisme est le producteur.

Encadrer entre partenaires les droits sur les données

 L'accord de consortium permet de clarifier le statut juridique des données et des résultats, tout en prévoyant à l'avance les modes de partage futurs.

Négocier entre partenaires le partage des données

^  Plus les modes de partage souhaités sont négociés tôt par les partenaires d'un projet, meilleures sont les chances d'une réutilisation sereine.

Le partage des données peut être couvert par l'accord de consortium ou des contrats de prestation, de cession, de transfert des données, etc.

Annexe 5 : Logigramme de communicabilité d'une donnée

