



HAL
open science

Accompagner les chercheurs pour les aider à mieux gérer leurs données de recherche : le métier de data librarian

Laetitia Bracco

► To cite this version:

Laetitia Bracco. Accompagner les chercheurs pour les aider à mieux gérer leurs données de recherche : le métier de data librarian. Réalités industrielles. Annales des mines, 2022, Août 2022, pp.55-58. 10.3917/rindu1.223.0055 . hal-03763270

HAL Id: hal-03763270

<https://hal.univ-lorraine.fr/hal-03763270>

Submitted on 29 Aug 2022

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.

Accompagner les chercheurs pour les aider à mieux gérer leurs données de recherche : le métier de *data librarian*

Par Laetitia BRACCO

Conservatrice des bibliothèques à l'Université de Lorraine

Les données de la recherche sont progressivement considérées comme des productions scientifiques à part entière, dans toutes les communautés disciplinaires. Dans le contexte de la *Science ouverte*, de plus en plus d'exigences sont émises par les financeurs et par les politiques publiques en général pour mieux produire, structurer, conserver et ouvrir ces données. L'accès aux données sous-jacentes aux publications est, en outre, de plus en plus demandé par les revues scientifiques, dans une logique d'intégrité et de transparence. Les personnels de l'information scientifique et technique, dans les bibliothèques universitaires comme dans d'autres structures, sont ainsi amenés à accompagner et à former les chercheurs à toutes ces problématiques. Pour mener à bien cette mission, une montée en compétences est nécessaire, ainsi que le développement de nouvelles activités, pour celles et ceux qui occupent ce métier émergent qu'est celui de *data librarian*.

Les données de la recherche sont un produit d'étude complexe à définir. La définition de l'OCDE fait néanmoins consensus : ce sont « des enregistrements factuels (chiffres, textes, images et sons), qui sont utilisés comme sources principales pour la recherche scientifique et sont généralement reconnus par la communauté scientifique comme nécessaires pour valider des résultats de recherche »¹. On peut ainsi considérer que les données de la recherche constituent le matériau varié (allant de l'image au tableur, en passant par la vidéo) sur lequel s'appuient les chercheurs pour réaliser et publier leur thèse, leur article ou leur monographie. Quels rôles pourraient bien jouer les bibliothèques dans ce travail de recherche, par essence spécialisé et intimement lié à la discipline scientifique de chacun des chercheurs ?

Pourquoi accompagner les chercheurs dans l'exploitation des données ?

Le métier de *data librarian* est récent dans le paysage français des bibliothèques universitaires. S'il a commencé à émerger vers la moitié des années 2010², notamment avec la création de la liste de diffusion *data librarians* en 2017³, qui

¹ OCDE (2007), *Principes et lignes directrices pour l'accès aux données de la recherche financée sur fonds publics*, p. 18, <https://www.oecd.org/fr/science/inno/38500823.pdf>

² CALENGE B. (2015), « Entre *data librarians* et médiateurs du savoir », *Bertrand Calenge : carnet de notes*, <https://bccn.wordpress.com/2015/02/06/entre-data-librarians-et-mediateurs-du-savoir/>

³ Accès à la liste : <https://groupes.renater.fr/sympa/info/datalibrarians>

compte à ce jour 134 abonnés, c'est en 2019 que les exigences des financeurs en matière de données de la recherche ont émergé et, avec elles, le besoin d'accompagnement des chercheurs.

En effet, le programme de financement européen Horizon 2020 lança, en 2014, l'*Open Research Data Pilot*⁴, qui exigeait que les chercheurs rédigent un plan de gestion des données (ou *data management plan*). Ce pilote, qui ne concernait que certains projets, a été étendu à l'ensemble des projets financés par Horizon 2020 en 2017⁵. L'Agence nationale de la recherche (ANR), a emboîté le pas de l'Union européenne et a demandé à son tour, à partir de 2019, la fourniture d'un plan de gestion des données (PGD) pour tous les projets qu'elle finance⁶.

L'exigence de la production de ce document peut paraître anodine. En réalité, 2019 est une année charnière, qui a rendu le métier de *data librarian* incontournable pour l'accompagnement des chercheurs dans leurs travaux. En effet, le PGD n'est pas un document administratif, mais bien un outil de pilotage, qui nécessite de se poser de nombreuses questions et surtout d'y répondre : quels types de données vais-je produire ? Nécessitent-elles de prendre des précautions particulières ? Comment vais-je les partager après l'aboutissement de mon projet ?

Il est évident que les chercheurs se sont posé ces questions bien avant 2019 et qu'ils n'ont pas attendu les injonctions des financeurs pour organiser leurs données et les partager, notamment dans certaines disciplines comme l'astronomie, qui est pionnière en matière de partage des données⁷.

Mais la systématisation de ce document de planification a induit une prise de conscience parmi toutes les communautés scientifiques et a mis en lumière tant les besoins d'infrastructures techniques nécessaires à la bonne gestion de ces données (plateformes de stockage, entrepôts de données...) que les besoins en formation.

En effet, pour compléter un PGD, il est nécessaire de comprendre aussi bien les notions de métadonnées, de thésaurus ou de signalement que celles de données personnelles ou d'archivage pérenne ; autant de domaines qui peuvent sembler obscurs pour un public non averti.

⁴ OPENAIRE (2017), *What is the open research data pilot*, <https://www.openaire.eu/what-is-the-open-research-data-pilot>

⁵ EUROPEAN COMMISSION (2017), *H2020 Online Manual, Cross-cutting issues, Data management*, https://ec.europa.eu/research/participants/docs/h2020-funding-guide/cross-cutting-issues/open-access-data-management/data-management_en.htm

⁶ AGENCE NATIONALE DE LA RECHERCHE (2019), *L'ANR met en place un plan de gestion des données pour les projets financés dès 2019*, <https://anr.fr/fr/actualites-de-lanr/details/news/lanr-met-en-place-un-plan-de-gestion-des-donnees-pour-les-projets-finances-des-2019/>

⁷ GENOVA F. (2014), « Du nécessaire partage des données scientifiques : l'exemple de l'astronomie », *Arabesques*, n°73, pp. 12-13, <https://publications-prairial.fr/arabesques/index.php?id=999>

Quelles compétences à acquérir pour les *data librarians* ?

Former et accompagner dans l'élaboration des plans de gestion des données

Le PGD se présente sous la forme de questions, qui interrogent le chercheur sur sa production et sa gestion des données tout au long de son projet. S'il n'est pas question pour les bibliothécaires de se substituer au chercheur, lequel est seul à même d'explicitier les données qu'il produit et leur pertinence au regard du projet scientifique qu'il mène, ils peuvent néanmoins lui apporter leur expertise sur la plupart des questions posées dans le document précité.

En effet, la structuration de l'information, sa description par des métadonnées et des thésaurus, son organisation selon des règles de nommage précises ou encore l'ouverture des données constituent le cœur de métier des bibliothécaires, en particulier dans un contexte de *Science ouverte*.

Le PGD constitue ainsi une formidable occasion d'amener le chercheur à prendre en considération toutes les bonnes pratiques en matière de gestion des données ; cette sensibilisation est d'autant plus importante pour le futur des PGD, que ceux-ci ont vocation à être lus à l'avenir par des machines et devront donc être élaborés en respectant des standards précis.

Le métier de *data librarian* peut ainsi être défini au prisme des services qu'il apporte aux chercheurs. Outre l'accompagnement dans la rédaction du PGD, les *data librarians* peuvent en effet intervenir à chaque étape du cycle de vie de la donnée, lequel peut être défini comme l'ensemble des phases de transformation de la donnée (laquelle est collectée, décrite, traitée, analysée...) tout au long d'un projet de recherche⁸. Au début d'un projet, il peut notamment être intéressant de chercher des jeux de données préexistants pour compléter l'état de l'art. Ce service peut être apparenté à de la recherche documentaire sur ce nouvel objet que sont les données.

En cours du projet, les *data librarians* peuvent apporter leur expertise pour la plupart des questions qui se posent en matière de gestion des données : choix des formats, organisation des fichiers, solutions de stockage, accompagnement juridique de premier niveau... À la fin du projet, les chercheurs peuvent aussi être aidés dans la sauvegarde et l'ouverture de leurs données par le biais d'un entrepôt adapté, à choisir parmi les milliers qui s'offrent à eux. Les *data stewards*, ou curateurs de données, qui, en général, ont un profil d'ingénieur, peuvent, quant à eux, intervenir sur les données elles-mêmes, disposant des compétences disciplinaires nécessaires.

Induire une montée en compétences progressive des chercheurs et des *data librarians*

En outre, les *data librarians* font généralement preuve d'une forte activité en matière de formation des chercheurs, ingénieurs, doctorants et professionnels de l'information aux enjeux et bonnes pratiques en matière de données, et ce dans le cadre d'ateliers, de rendez-vous individuels ou encore par la production de supports

⁸ SCIENCE OUVERTE COUPERIN, *Données de la recherche, définitions*, <https://scienceouverte.couperin.org/donnees-recherche-definitions/>

et de guides d'accompagnement, comme le propose, par exemple, l'Université de Lorraine dans sa boîte à outils Science ouverte⁹.

Enfin, ce métier comporte une dimension Communication : le cadre réglementaire en matière de données de recherche et les exigences des financeurs évoluent rapidement, il est donc nécessaire de tenir les chercheurs régulièrement informés.

Si de nombreux services liés aux données, tels que la recherche de jeux de données, l'aide à la citation ou la description de données par des métadonnées, font partie des compétences traditionnelles des bibliothécaires intervenant dans le domaine de la Science ouverte, ce n'est pas forcément le cas de tous. Tous les *data librarians* doivent ainsi acquérir de nouvelles compétences.

D'une part, la mise en œuvre du plan de gestion des données amène les bibliothécaires à travailler en étroite collaboration avec d'autres services que ceux de la bibliothèque. En effet, dans le domaine de l'informatique et du juridique, seul un accompagnement de premier niveau peut être apporté : le *data librarian* n'est ni un architecte des systèmes d'information ni un juriste. Aussi, c'est avant tout une excellente connaissance du paysage de la recherche et de ses acteurs qui est attendue du *data librarian*, afin, le cas échéant, de rediriger les demandes vers le bon service. C'est toute la philosophie des guichets uniques qui se mettent en place pour accompagner les chercheurs dans leur recherche de données.

D'autre part, l'objet « donnée » est plus complexe à appréhender que celui de la publication scientifique. Il peut en effet prendre des formes très variées, tant en termes de format que de discipline. Le *data librarian* ne peut être un expert de tous ces domaines scientifiques. Mais l'accompagnement qu'il apporte nécessite pour lui de comprendre, ne serait-ce que de manière superficielle, les données générées par les chercheurs. C'est pourquoi avant tout accompagnement, un entretien individuel est nécessaire pour permettre une vulgarisation scientifique de la part du chercheur et une précision de ses apports méthodologiques de la part du *data librarian*. Ce moment d'échange est précieux pour forger une relation de confiance mutuelle entre ces deux acteurs et ainsi appréhender les données en mobilisant toutes les palettes de compétences scientifiques et documentaires nécessaires. La gestion des données relève donc bien d'un travail d'équipe.

Quelle est la structuration de l'accompagnement des chercheurs en France ?

Une communauté professionnelle française très développée

Ces nouvelles compétences à acquérir nécessitent la mise en place de formations professionnelles. L'ENSSIB¹⁰, comme les réseaux des CRFCB¹¹ et les Urfist¹² ont rapidement investi ce terrain et proposent aujourd'hui de nombreuses formations sur

⁹ UNIVERSITE DE LORRAINE, *Boîte à outils de la Science ouverte*, <https://scienceouverte.univ-lorraine.fr/boite-a-outils/>

¹⁰ École nationale supérieure des sciences de l'information et des bibliothèques.

¹¹ Centre régional de formation aux carrières des bibliothèques.

¹² Unité régionale de formation à l'information scientifique et technique.

ce sujet. Mais pour ce domaine en mutation, une veille professionnelle régulière est indispensable en complément de la formation continue. La communauté professionnelle française est très structurée et foisonnante sur ce point.

Le GTSO Données de Couperin – un groupe de travail composé de professionnels de l’information et de la documentation – propose de nombreux supports à destination des établissements et anime la communauté au travers de webinaires réguliers¹³. En 2021, ce groupe a publié les résultats de son « Enquête sur l’appui à la gestion des données de la recherche en service de documentation et d’information scientifique et technique »¹⁴. L’objectif de cette démarche était de comprendre la manière dont les services d’accompagnement s’étaient emparés des données de la recherche. Quelques conclusions principales peuvent en être tirées :

- en dépit des nouvelles exigences des financeurs, près de la moitié des répondants ne disposaient alors d’aucun service d’accompagnement ;
- pour les services déjà en place, la plupart étaient encore en phase de préfiguration ;
- les inégalités en matière d’accompagnement sont très fortes entre les petites et les grandes structures ;
- les services d’accompagnement sont davantage d’un niveau de sensibilisation que d’un niveau technique (préparation des données avant dépôt, description des données avec des métadonnées...) ;
- la gestion des données nécessite de travailler en coopération avec de nombreux services, y compris ceux externes à l’établissement.

Un besoin fort de soutien méthodologique a été exprimé par les 82 établissements ayant répondu à l’enquête.

Afin d’y répondre, le GTSO Données a commencé à produire des fiches pratiques thématiques, dont l’objectif est de répondre à une question précise que peuvent se poser les *data librarians* de toute la France : comment animer un atelier PGD ? Comment mettre en place un service d’accompagnement ? Comment former les doctorants ?

Pour animer la communauté, le GTSO Données propose également des webinaires¹⁵, une lettre de veille¹⁶ ou encore le partage de retours d’expérience. D’autres groupes de travail constitués autour des données sont très actifs en France, tels que RDA France, le groupe Données inter-réseaux du CNRS ou encore le Collège Données du Comité pour la Science ouverte. Tous ces groupes de travail

¹³ COUPERIN, GTSO Données, <https://www.couperin.org/science-ouverte/ressources-du-gtso/groupe-donnees>

¹⁴ FERET R., BRACCO L., LEHOUX E. & AUGOUVERNAIRE M. (2021), *Enquête sur l’appui à la gestion des données de la recherche en service de documentation et d’information scientifique et technique*, <https://doi.org/10.5281/zenodo.5078504>

¹⁵ GTSO Données, chaîne Youtube, <https://www.youtube.com/channel/UCswnouUXeMhHAsEzNpYGhMA>

¹⁶ BRACCO L., *Quoi de neuf dans les données de la recherche ?*, <https://www.getrevue.co/profile/laetitia-bracco>

sont complémentaires. Le métier de *data librarian* consiste donc également à se former et à s'informer en permanence, une démarche qui est facilitée par l'existence d'une communauté professionnelle nationale soudée ainsi que celle de réseaux internationaux dynamiques.

L'exemple de l'Université de Lorraine

L'accompagnement des chercheurs à la gestion des données de la recherche est une priorité fixée dans le deuxième plan national pour la Science ouverte : « pour accompagner et conseiller les chercheurs tout au long du cycle de vie des données, des "ateliers de la donnée" réunissant une large palette de métiers seront structurés sur tout le territoire »¹⁷. Une structuration nationale est donc en cours, par le biais d'appels à projets visant la labellisation en « ateliers de la donnée » de cellules existantes d'accompagnement à la gestion des données.

L'Université de Lorraine propose, depuis 2019, un accompagnement des chercheurs dans la gestion de leurs données. Initié par l'aide apportée pour élaborer le plan de gestion des données et la formation des doctorants, il s'est progressivement développé pour couvrir l'ensemble du cycle de vie de la donnée.

La palette de services proposés est large : outre les deux éléments précités, on y trouve le guichet unique de réponses, un accompagnement en matière de stockage, de dépôt dans des entrepôts, de respect du RGPD ou encore de publication d'un *data paper*¹⁸. Ces services ont vocation à se développer et à se massifier mais sont déjà reconnus par le label national « Atelier de la donnée ».

Conclusion

Le métier de *data librarian* connaît un fort élan depuis 2019. L'accompagnement des chercheurs dans la gestion de leurs données a pris son essor avec l'exigence de la fourniture d'un PGD pour les projets financés notamment par l'Union européenne et l'ANR ; un essor qui dépasse ce cadre aujourd'hui. La gestion des données ne peut s'envisager qu'au travers d'une combinaison de compétences scientifiques, documentaires et informatiques, un cadre au sein duquel les *data librarians* ont un rôle majeur à jouer.

¹⁷ MESRI (2021), *Deuxième plan national pour la Science ouverte*, p. 13, <https://www.ouvrirlescience.fr/deuxieme-plan-national-pour-la-science-ouverte/>

¹⁸ Voir le détail des services proposés sur la page dédiée : <http://scienceouverte.univ-lorraine.fr/donnees-de-la-recherche/>