



HAL
open science

Services de gestion et de partage des données de recherche : ce qu'en pensent les chercheurs

Violaine Rebouillat, Ghislaine Chartron

► To cite this version:

Violaine Rebouillat, Ghislaine Chartron. Services de gestion et de partage des données de recherche : ce qu'en pensent les chercheurs. 12ème Colloque international d'ISKO-France : Données et mégadonnées ouvertes en SHS : de nouveaux enjeux pour l'état et l'organisation des connaissances?, Oct 2019, Montpellier, France. hal-02307085

HAL Id: hal-02307085

<https://hal.science/hal-02307085>

Submitted on 7 Oct 2019

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.

Services de gestion et de partage des données de recherche : ce qu'en pensent les chercheurs

Violaine Rebouillat

Doctorante en Sciences de l'information et de la communication

EA 7339 DICEN-IdF – Cnam

violaine.rebouillat.auditeur@lecnam.net

Ghislaine Chartron

Professeure en Sciences de l'information et de la communication

EA 7339 DICEN-IdF – Cnam

ghislaine.chartron@lecnam.net

Résumé

En France, les professionnels de l'information scientifique et technique (IST) se positionnent sur le développement de services pour la gestion et la valorisation des données de recherche. L'article interroge l'utilisation de ces services par les chercheurs. Il s'appuie sur 46 entretiens, réalisés auprès de chercheurs de l'Université de Strasbourg. Le catalogue Cat OPIDoR, référençant les services de données français, a servi de base d'étude pour l'enquête. Les résultats montrent que les services développés par les professionnels de l'IST correspondent pour une faible partie à ceux qu'utilisent les répondants. Une des explications esquissées est qu'en matière de données les chercheurs sont davantage influencés par les recommandations des éditeurs que par celles des professionnels de l'IST.

Mots clés

Données de la recherche, Science ouverte, Services de données, Pratiques de recherche

Title

French Research Data Management and Sharing Services: are they used by scientists?

Abstract

French scientific and technical information (STI) staffs develop research data management and sharing services. In this paper, scientists' use of these services is questioned. Forty-six interviews were conducted with researchers from the University of Strasbourg. They were asked to explore the online catalog Cat OPIDoR, which lists French research data services. The results show that few services developed by STI staffs were used by respondents. One possible explanation is that, in terms of research data, scientists are more influenced by scientific publishers than by STI staffs.

Keywords

Research Data, Open Science, Research Data Services, Research Practices

INTRODUCTION

Depuis la Déclaration de Berlin en 2003 (Max Planck Gesellschaft, 2003) et dans un contexte d'engouement croissant pour les « *data* » (Cukier, 2010) et l'ouverture de la science (Leonelli, 2013), de nombreux services ont vu le jour, dans la perspective de proposer aux chercheurs un appui pour gérer et partager les données scientifiques.

La notion de donnée de recherche n'est pas une notion simple à définir, car elle rassemble des réalités très différentes. Christine L. Borgman (2015, p.28) considère qu'« *une observation, un objet, un document ou toute autre entité devient une donnée de recherche, dès lors qu'elle est utilisée comme preuve d'un phénomène, c'est-à-dire qu'elle est collectée, analysée et interprétée* ». Des travaux comme ceux de Sabina Leonelli et de Joachim Schöpfel ont mis en évidence le caractère contextuel des données de recherche. Selon Leonelli (2015), il n'existe pas de données en elles-mêmes. Ce qu'un scientifique considère comme « donnée » est toujours relatif à une question de recherche spécifique. Les données ne sont pas définies selon leurs propriétés intrinsèques mais selon leur fonction au sein de processus de recherche particuliers. On ne peut répondre à la question « qu'est-ce qu'une donnée » qu'en faisant référence à des situations de recherche concrètes. Schöpfel et al. (2017) ont ainsi isolé quatre paramètres de variation d'une donnée scientifique : sa nature factuelle, son enregistrement, la communauté qui la génère et/ou l'utilise et sa finalité. Dans cette communication, seront employés indistinctement les termes de « données de la recherche », de « données de recherche » et de « données scientifiques ».

A cette première complexité sémantique s'ajoute celle des pratiques de gestion et de partage des données, qui varient en fonction des individus et des communautés. Diverses études se sont attachées à comprendre comment et avec qui les chercheurs gèrent et partagent leurs données (Stuart et al., 2018 ; Schöpfel, 2018 ; Tenopir et al., 2015). Ces études ont contribué à mieux connaître les pratiques et ont servi de légitimation au développement d'offres de services.

En France, les professionnels de l'information scientifique et technique (IST) se sont très vite positionnés sur le développement de services pour la gestion et la valorisation des données de recherche. La mission première de ces professionnels, qu'ils soient documentalistes ou personnels des bibliothèques, est de collecter l'information scientifique et de la diffuser, par le biais de médiations documentaires, à des communautés de recherche plus ou moins diversifiées. Les professionnels de l'IST interviennent donc traditionnellement en amont du processus de recherche, fournissant aux chercheurs une information qui servira de matière première à leurs recherches. Or, en raison du développement des technologies, ces professionnels se trouvent soumis à une remise en cause régulière de leur fonction (Fabre et Gardiès, 2008). C'est pourquoi ils ont tenté au cours de ces vingt dernières années de se positionner également sur la valorisation de l'information scientifique, en aval du processus de recherche. Directement concernés par l'augmentation du prix des abonnements aux revues scientifiques (Chartron, 2010), ils ont notamment rallié le mouvement du libre accès aux résultats de la recherche et développé des archives ouvertes pour la diffusion des publications. Leur positionnement sur la question de la gestion et du partage des données de recherche s'est alors imposé comme une suite logique. Se proposant d'accompagner les chercheurs dans la documentation et la diffusion de ces données, les professionnels de l'IST ont développé des services de différents ordres :

- Des supports d'information, comme le site DataPartage [1] (développé par la délégation Information Scientifique et Technique de l'INRA [2]) ;

- Des formations à destination des doctorants et des chercheurs, comme en proposent les URFIST [3] (13 formations sur le thème des données de la recherche étaient organisées dans 6 des 7 URFIST au cours du premier semestre 2019) ;
- Des dispositifs d'aide personnalisée, comme à l'Université de Strasbourg, où le service des bibliothèques propose aux chercheurs l'assistance d'un *data librarian* [4];
- Des supports pour la rédaction de plans gestion de données (*Data Management Plan* en anglais), comme l'outil DMP OPIDoR [5] développé par l'Inist-CNRS [6] ;
- Des entrepôts de données, comme celui du CDSP [7], BeQuali [8], qui permet de diffuser les résultats d'enquêtes qualitatives en sciences sociales.

Au vu de ces initiatives, la présente communication vise à questionner l'adéquation des services de données avec les pratiques de recherche. Par « service », on entend la fourniture de ressources humaines et/ou techniques pour gérer les données à une ou plusieurs étapes d'un projet de recherche. Dans cette étude, seuls les services proposés par des structures publiques, c'est-à-dire affiliés à une institution ou une infrastructure de recherche publique, seront étudiés. L'objectif est d'apporter des éléments de réponse à la question : les services développés et mis en avant par les professionnels de l'IST sont-ils utilisés par la communauté scientifique ?

Pour tenter d'y répondre, une enquête a été réalisée auprès de 46 chercheurs issus de différentes disciplines. L'approche choisie est une approche qualitative, visant à comprendre pourquoi et comment les chercheurs ont recours à des services de données. L'objectif n'est pas de quantifier l'usage qu'ils en font, à partir de statistiques d'utilisation par exemple. Le catalogue Cat OPIDoR, commandité par la Bibliothèque Scientifique Numérique (renommée aujourd'hui Comité pour la Science Ouverte [9]), a servi de base d'étude pour cette enquête. Une première partie de la communication sera consacrée à l'exposition du contexte de recherche ; une seconde partie à la présentation des premiers résultats de l'enquête.

1 – CONTEXTE DE RECHERCHE

1.1 Cat OPIDoR : les services de données vus par des professionnels de l'IST

Cat OPIDoR [10] est un catalogue en ligne, dont le but est de répertorier les services français dédiés aux données scientifiques (de leur acquisition à leur archivage, en passant par leur diffusion). En 2015, le groupe *Données de la recherche* de la Bibliothèque Scientifique Numérique (BSN10) a en effet commandité un travail de cartographie des services français dédiés aux données de la recherche (Rebouillat, 2017). Deux méthodes ont été employées pour identifier ces services :

- Une veille a été mise en place pour repérer les colloques ayant lieu sur le thème des données de recherche. Leur programme a alors été consulté, afin d'identifier d'éventuels services présentés par les intervenants
- Les membres du groupe BSN10, ceux de l'association EPRIST [11] et le groupe de travail pour l'accès ouvert du consortium COUPERIN [12] ont été interrogés sur l'existence de services de données dans leurs établissements d'appartenance.

A l'issue de ce recensement, il a été décidé de développer un catalogue en ligne, Cat OPIDoR, permettant de consulter les services recensés et d'en répertorier de nouveaux de manière collaborative. La conception du catalogue a été menée en partenariat avec l'Inist-CNRS (O. Aberkane, A. Ciolek, M. C. Jacquemot, V. Rebouillat), qui héberge et modère aujourd'hui l'outil. Cat OPIDoR a été conçu pour un public multiple, comprenant la communauté scientifique, les

personnels d'appui à la recherche, les fournisseurs de services et les financeurs de la recherche.

Au cours du recensement initial, l'existence de deux ordres de services a été constatée :

- Un premier ordre de services, à la fois très divers et très spécifique, consacré à la collecte, l'échange et l'analyse de données dans les différentes communautés disciplinaires ;
- Un second ordre de services, plus généraliste et plutôt tourné vers l'ouverture des données.

Le premier ensemble de services, que nous appellerons « services disciplinaires », rassemble des dispositifs nés au sein des communautés scientifiques pour répondre à un besoin particulier. Ils sont gérés par des chercheurs et des ingénieurs. Il peut s'agir de plateformes technologiques, donnant accès par le biais de prestations de services à des instruments de collecte de données (en biologie notamment).

Le second ensemble de services, que nous nommerons « services de gestion et d'ouverture des données », est apparu avec le mouvement d'ouverture des données de la recherche. Ces services sont donc relativement nouveaux (postérieurs à 2010) et ont principalement été développés par la sphère des professionnels de l'IST. Il s'agit des services que nous avons évoqués en introduction.

Pour Cat OPIDoR, l'équipe projet a choisi de prendre en compte ces deux ordres de services. Il a été plus facile d'identifier les services de gestion et d'ouverture car le projet était commandité par des professionnels de l'information scientifique. En revanche, il a été plus difficile de recenser les services disciplinaires, car ceux-ci font partie de communautés scientifiques très spécifiques, dont il faut connaître le langage et les réseaux. A la date du 12 mai 2019, Cat OPIDoR répertoriait 231 services, se différenciant :

- Par leur nature (neuf types de services ont été identifiés, allant de l'acquisition à la diffusion de données, en passant par l'archivage et la formation à la gestion et à l'ouverture des données) ;
- Par le périmètre de leur public cible (les services peuvent avoir une vocation institutionnelle ou bien cibler un public plus large, national voire international – c'est le cas de quelques services, comme VizieR [13], SIMBAD [14], SEANOE [15] et ArkeoGIS [16], qui s'adressent à la communauté de recherche dans son ensemble) ;
- Par le profil professionnel des personnels fournissant le service (professionnels de l'information scientifique et technique, informaticiens, personnels de recherche...).

A la suite de ce travail, plusieurs interrogations sont nées concernant l'utilisation de Cat OPIDoR par la communauté scientifique :

- Quels services de données utilisent les chercheurs ?
- Connaissent-ils Cat OPIDoR et les services qui y sont répertoriés ?
- Ce catalogue et ces services ont-ils une utilité pour eux ?

Il a donc été décidé de soumettre Cat OPIDoR et les services qu'il répertorie à l'avis d'un échantillon de chercheurs, issus de différentes disciplines.

1.2 Méthodologie de l'enquête

L'enquête a pour terrain d'étude l'Université de Strasbourg. Nous avons conservé le même échantillon de chercheurs que celui sondé dans le cadre de notre précédente enquête (Rebouillat, à paraître), à savoir seize chercheurs issus de différentes disciplines. A l'origine, ceux-ci faisaient partie d'une population sélectionnée sur la base de l'excellence scientifique. Nous avons en effet choisi des projets de recherche bénéficiant d'un financement public (type H2020 [17] ou ANR [18]) et faisant partie des laboratoires de l'Université de Strasbourg les mieux notés par l'AERES [19]. N'ayant obtenu sur les seize chercheurs qu'un nombre très limité de réponses (5), nous avons choisi d'élargir le périmètre de la population sondée. Nous nous sommes adressés à l'ensemble des chercheurs [20] de dix des seize laboratoires initiaux, afin d'obtenir un panel varié de disciplines. Sur 228 chercheurs contactés, nous avons obtenu 41 réponses favorables.

La première population a été sondée en février 2019 ; la seconde en mars et avril 2019. Le taux de participation total est de 27% (tableau 1).

Domaine scientifique	Laboratoire	Discipline du laboratoire	Nombre de chercheurs contactés	Nombre d'entretiens menés	Taux de participation
Vie & Santé	Laboratoire 1	Biologie	1	1	100 %
	Laboratoire 2	Chimie et biologie	30	3	10 %
	Laboratoire 3	Chimie et biologie	22	5	23 %
	Laboratoire 4	Écologie	19	3	16 %
	Laboratoire 5	Neurosciences	20	4	20 %
	Laboratoire 6	Neurosciences	10	1	10 %
Sciences & Technologies	Laboratoire 7	Astronomie	16	7	44 %
	Laboratoire 8	Sciences de l'ingénieur	33	5	15 %
	Laboratoire 9	Sciences de la Terre	23	3	13 %
Sciences humaines et sociales	Laboratoire 10	Géographie	22	4	18 %
	Laboratoire 11	Sciences sociales	20	8	40 %
	Laboratoire 12	Théologie	17	2	12 %
Total			233	46	27 %

Tableau 1 : Taux de participation à l'enquête par laboratoire

L'objectif de l'enquête était de recueillir le point de vue des chercheurs sur les services répertoriés dans Cat OPIDoR. C'était en tout cas l'objectif qui avait été fixé pour la première population interrogée. Nous avons conçu un ensemble de questions, que nous leur envoyions par mail (annexe 1). Ces questions portaient sur l'ergonomie, la pertinence et l'utilisation éventuelle des services répertoriés dans Cat OPIDoR. Nous demandions aux seize chercheurs de parcourir le catalogue et de porter leur attention sur un ou deux services, avant de répondre au questionnaire. Nous recueillions leurs réponses par écrit (par retour de mail) ou de vive voix (dans le cadre d'un entretien téléphonique).

Pour la seconde population, étant donné que nous ne connaissions pas les répondants, leurs pratiques de recherche et la place des données dans ces pratiques, nous avons dans un premier temps cherché à savoir : avec quels outils les données étaient traitées et diffusées ; d'où provenaient ces outils ; quel était le degré de satisfaction des chercheurs par rapport à ces outils. En deuxième partie venait la présentation de Cat OPIDoR et le recueil de l'opinion des chercheurs sur le catalogue et les services répertoriés. Pour cette seconde population, nous avons choisi exclusivement la forme de l'entretien, présentant un meilleur taux de réponses que par questionnaire (guide d'entretien en annexe). Nous avons donc contacté les 228 chercheurs par mail, leur demandant s'ils acceptaient de nous accorder un entretien téléphonique ou de visu (selon leur préférence). D'une durée d'environ trente à quarante-cinq minutes, les entretiens étaient conduits de manière semi-directive.

Le propos de cette communication repose donc sur les réponses couplées de la première et de la seconde population. Il s'agit de présenter ici le fruit d'un premier travail d'analyse, encore partiel, qui sera complété par la suite par une étude plus détaillée des résultats de l'enquête.

2 – RÉSULTATS ET ANALYSE

Les résultats présentés ici ne se veulent pas représentatifs d'une discipline. Ils tentent seulement de dégager des tendances à partir de pratiques individuelles décrites par les chercheurs interrogés. Par ailleurs, notre panel présente les deux biais suivants. Un premier biais est de ne pas être parfaitement équilibré en termes de domaines scientifiques. Le domaine Vie & Santé est surreprésenté (6 des 12 laboratoires contactés) par rapport aux domaines des Sciences & Technologies (3 laboratoires) et des Sciences humaines et sociales (3 laboratoires). Le second biais du panel est qu'il se limite très probablement à des chercheurs auxquels les termes de « données » et de « services » sont éloquents (bien que plusieurs synonymes aient été listés dans le mail de demande d'entretien). Plusieurs réponses négatives penchent dans ce sens. Onze chercheurs ont en effet décliné l'entretien, au motif qu'ils ne se sentaient pas assez compétents pour parler des services et des données utilisés dans leur discipline. En termes de participation, on trouve le taux de réponse le plus élevé en astronomie (laboratoire 7). Le laboratoire en question est en effet adossé au Centre de Données astronomiques de Strasbourg (CDS) [21], qui depuis 1972 collecte et distribue des données astronomiques au niveau international. La question des données et de leur partage est donc inscrite dans la culture du laboratoire. Ses membres en sont familiers. Un des taux de participation les plus bas concerne la théologie (laboratoire 12), où le terme de « données » semble ne pas être utilisé (cela sera évoqué plus loin).

2.1 Les services utilisés par les répondants

Aucun des chercheurs interrogés ne connaissait l'existence de Cat OPIDoR. Le catalogue a été perçu comme intéressant et potentiellement utile. Trois répondants (en biologie, chimie/biologie et écologie) ont pointé comme inconvénient le périmètre national du catalogue, précisant qu'un recensement international serait plus utile dans le cadre de leurs recherches.

Par ailleurs, rares étaient les chercheurs connaissant ou utilisant un des services répertoriés dans Cat OPIDoR. Ils en connaissaient ou utilisaient au mieux un ou deux : le synchrotron

SOLEIL [22] (pour un des chercheurs en chimie et biologie), PROGEDO [23] (en sciences sociales), Reclnat [24] (en écologie), SIMBAD [25] et Vizier [26] (en astronomie), le pôle HPC Unistra [27] (en sciences de l'ingénieur).

De manière générale, les chercheurs rencontrés ont peu recours à des services de données. Ils collectent, analysent, gèrent et partagent leurs données avec les moyens à disposition dans leur laboratoire.

- Parfois leurs recherches ne requièrent ni outil particulier ni expertise externe. C'est le cas de 3 des 46 chercheurs rencontrés, en droit (laboratoire 11), en histoire (laboratoire 11) et en théologie (laboratoire 12). Ces chercheurs manipulent essentiellement des textes (juridiques, historiques...). Ils n'utilisent d'ailleurs pas le terme de « données » mais celui de « sources ». Leur démarche consiste à lire, analyser et interpréter ces textes selon une approche herméneutique, ne requérant donc aucun service de données.
- Un autre facteur explicatif est la méconnaissance des services existants. Plusieurs chercheurs ont déclaré se sentir dépassés par le flot d'informations. Deux d'entre eux (en sciences de la Terre et en géographie) ont explicitement rendu compte de difficultés à repérer quels services existaient. Il est probable que la diversité et le manque de visibilité des services rendent ce repérage malaisé. A cette difficulté s'ajoute un manque de temps global, non favorable à une recherche d'informations approfondie. Cat OPIDoR a été jugé utile sur ce point. Les chercheurs ont supposé qu'il pourrait leur permettre de rechercher un service dans le cadre d'un besoin précis.
- Parfois encore, les chercheurs font sciemment le choix de ne pas utiliser de services, considérant que le travail d'analyse requiert une nécessaire proximité avec l'origine des données. Par « origine » on entend le terrain d'étude. Plusieurs chercheurs considèrent en effet qu'ils doivent collecter les données par eux-mêmes. En sciences sociales, les anthropologues et les sociologues disent collecter et traiter les données de manière solitaire. La nature qualitative de leur analyse nécessite, selon eux, de se confronter personnellement au terrain. Ils n'utilisent donc ni service ni logiciel.

Lorsque les chercheurs ont effectivement recours à un service de données, il s'agit le plus souvent d'un service de collecte. Ce type de service est en effet davantage utilisé que les services de diffusion de données. C'est notamment le cas des dispositifs coûteux, qu'un laboratoire n'est pas en mesure d'acquérir à lui seul. On pense aux dispositifs d'enquête, tels les panels de population en démographie (le *European Social Survey* [28], par exemple). On pense également aux télescopes en astronomie, aux séquenceurs ADN en biologie, chimie et écologie. Dans ce dernier cas, les chercheurs ont recours à des prestations de service : ils sous-traitent l'acquisition des données à des plateformes dédiées, moyennant paiement.

Les services d'ouverture de données sont peu utilisés des chercheurs, car ces derniers privilégient la publication d'articles ou d'ouvrages. Le recours à des services de type entrepôts de données ou agence d'attribution de DOI reste donc limité.

La question de l'ouverture des données est certes un sujet qui parle aux chercheurs mais qui, à quelques exceptions près, ne fait pas partie de leurs préoccupations quotidiennes. Le fait que l'évaluation de la recherche et l'attribution de financements reposent essentiellement sur le nombre de publications ne leur permet pas de consacrer plus de temps et d'intérêt aux

données. Fecher et al. (2015) qualifient le système d'échange de l'information scientifique d' « économie de la réputation » (*reputation economy*). La production scientifique n'est partagée que si elle permet au chercheur qui la transmet d'accroître sa réputation auprès des pairs. Dans ce système de communication de l'information scientifique, donnant la priorité à la publication, les données de recherche ne possèdent donc qu'une faible valeur d'échange (Fecher et al., 2017). D'où un impact sur leur gestion et leur diffusion.

Le fait que les chercheurs interrogés utilisent peu les services existants ne signifie pas cependant qu'ils n'ont pas de besoins en termes de données. Les résultats de l'enquête révèlent en effet des besoins pour le stockage et l'organisation des données. Ces besoins s'expriment toutefois en arrière-plan. Les préoccupations des chercheurs étant principalement tournées vers la publication, les problèmes liés aux données restent en effet au second plan et les besoins sont rarement clairement exprimés. Les chercheurs tentent de trouver des solutions par eux-mêmes, avec les moyens dont ils disposent dans leur laboratoire. Derrière cette question de priorité se dessine donc une autre réalité : celle d'un manque de ressources humaines et financières (Schöpfel, 2018).

2.2 Un rapport privilégié aux éditeurs

En 2017, l'éditeur Springer Nature (Stuart et al., 2018) a mené une des plus vastes enquêtes internationales sur les pratiques de partage des données de recherche (7719 répondants) [29]. L'objectif était de comprendre quelles étaient les pratiques des chercheurs en matière de partage de données, au moment où ceux-ci publient un article scientifique. L'étude a montré que 63% des répondants partageaient les données sous-jacentes à une publication dans des entrepôts de données et/ou comme matériel supplémentaire à l'article. Les chercheurs en biologie étaient les plus nombreux à partager ces données (75 % d'entre eux), suivis par les sciences de la Terre (68%), les sciences médicales (61 %), les sciences physiques (59 %) et les « autres sciences » (46%), rassemblant sciences humaines, sciences sociales, informatique et mathématiques.

Dans notre enquête, la proportion de chercheurs partageant les données sous-jacentes aux publications est légèrement supérieure : 70% (le tableau 2 détaille la répartition par laboratoire). Les données sont majoritairement partagées sous forme de fichiers liés à l'article, dans la section appelée « matériel supplémentaire ». Dans notre panel, seuls les biologistes et les astronomes ont recours à des banques de données (ou entrepôts de données). Pour les astronomes, ce sont les observatoires (là où sont situés les télescopes par exemple), qui se chargent de rendre les données disponibles (chaque observatoire dispose de sa propre archive de données). Les chercheurs ne sont donc pas directement à l'initiative du partage. En biologie, ce sont essentiellement les données omiques qui sont concernées par le dépôt dans des banques de données. Lorsque les biologistes soumettent un article dans une revue prestigieuse, ils sont contraints de mettre à disposition les séquences ADN. L'article ne sera pas publié, tant que celles-ci n'auront pas été déposées dans une banque de données dédiée (GenBank [30], en l'occurrence).

Domaine scientifique	Laboratoire	Discipline du laboratoire	Nombre de chercheurs rencontrés	Nombre de chercheurs ayant déjà partagé des données sous-jacentes à une publication	Proportion
Vie & Santé	Laboratoire 1	Biologie	1	1	100 %
	Laboratoire 2	Chimie et biologie	3	3	100 %
	Laboratoire 3	Chimie et biologie	5	5	100 %
	Laboratoire 4	Écologie	3	3	100 %
	Laboratoire 5	Neurosciences	4	4	100 %
	Laboratoire 6	Neurosciences	1	1	100 %
Sciences & Technologies	Laboratoire 7	Astronomie	7	2	29 %
	Laboratoire 8	Sciences de l'ingénieur	5	2	40 %
	Laboratoire 9	Sciences de la Terre	3	2	67 %
Sciences humaines et sociales	Laboratoire 10	Géographie	4	3	75 %
	Laboratoire 11	Sciences sociales	8	2	25 %
	Laboratoire 12	Théologie	2	0	0 %
Total			46	28	70 %

Tableau 2 : Proportion par laboratoire de chercheurs ayant déjà partagé des données sous-jacentes à une publication

Il convient néanmoins de nuancer ces résultats. Car, si 70% des chercheurs interrogés disent mettre à disposition les données sous-jacentes aux publications, il ne s'agit pas de pratiques systématiques. A l'exception des biologistes, cela est même relativement rare. Les chercheurs rendent leurs données publiques, lorsque cela leur est demandé – en particulier par les éditeurs et les relecteurs. Dans ce cas, ils suivent un protocole bien défini par la revue (dépôt dans une banque de données particulière ou dépôt comme matériel supplémentaire à l'article).

La demande des éditeurs s'inscrit dans une logique d'intégrité scientifique. La fourniture des données joue un rôle d'administration de la preuve. Elle serait gage de qualité et permettrait de limiter les cas de fraude. Selon les chercheurs, les revues sont de plus en plus nombreuses à exiger la mise à disposition des données sous-jacentes.

Or, même s'ils émettent des réserves quant à l'efficacité de ce système, les chercheurs respectent cette clause d'ouverture des données. L'éditeur PLOS faisait ce même constat, ayant instauré une politique de mise à disposition des données pour ses propres revues (Byrne, 2017). Un des chercheurs rencontrés explique qu'il n'est pas dans son intérêt de refuser la requête de l'éditeur. Pour lui, refuser c'est peut-être perdre l'opportunité d'être publié. Or la publication est primordiale pour la suite de ses recherches (pour l'obtention de nouveaux financements, pour une éventuelle promotion...). En raison de la prédominance des

publications dans le système d'évaluation, les éditeurs semblent donc avoir une forte influence sur la communauté scientifique.

CONCLUSION

L'exemple de Cat OPIDoR montre que les services développés et mis en avant par les professionnels de l'IST correspondent pour une faible partie à ceux qu'utilisent les chercheurs, quelle que soit leur discipline d'appartenance. Trois explications peuvent être avancées à cela. (1) Notre échantillon n'est pas représentatif des usages de la communauté de recherche. Les services de gestion et d'ouverture des données sont peut-être davantage utilisés dans des établissements comme l'Irstea ou l'Ifremer, qui ont été précurseurs en termes de libre accès (via l'instauration de mandats d'ouverture, le développement d'infrastructures dédiées...). (2) L'offre de services proposée par les professionnels de l'IST arrive trop tôt pour rencontrer les besoins des chercheurs. L'offre a anticipé la demande. (3) Il y a un hiatus entre la communauté scientifique et les professionnels de l'IST. Ces derniers n'ont pas de pouvoir d'influence sur les communautés de recherche en matière de données. A l'heure actuelle, il semble que les chercheurs soient davantage influencés par la sphère de l'édition scientifique. Si un éditeur leur demande de rendre leurs données accessibles en échange de la publication d'un article ou d'un ouvrage, ils le feront. Car la publication est primordiale pour eux. Elle conditionne leur réussite professionnelle et la poursuite de leurs recherches. Le partage des données n'est pas une priorité, car il ne permet pas d'en tirer des avantages. C'est une des raisons pour lesquelles les chercheurs utilisent si peu les services de gestion et de partage de données, conçus dans une perspective de science ouverte par les professionnels de l'IST.

Depuis quelques années, les politiques nationales et européennes rendent la science ouverte de plus en plus tangible (Chartron, 2018). Néanmoins, si les pratiques de partage des données tendent à se généraliser, se tourneront-elles vers les services développés par la sphère de l'IST ? Ceux-ci perdureront-ils ou bien seront-ils supplantés par l'offre de services d'autres acteurs ? Si l'objectif poursuivi est bien d'œuvrer pour une amplification de la mise à disposition des données de la recherche, alors une coopération discutée entre les acteurs ne peut qu'être encouragée. Au vu des résultats de l'enquête, plusieurs initiatives pourraient se renforcer :

- D'une part, l'évolution du système d'évaluation scientifique vers une prise en compte plus grande des données de recherche.
- D'autre part, l'adossement de certains services de données au rôle éditorial, car l'enquête a montré que la publication des articles demeurait la priorité principale de la dynamique des chercheurs. En ce sens, une offre de services provenant des éditeurs devrait rencontrer un avis favorable des chercheurs. Il lui faudrait néanmoins être ajustable en fonction des contextes, avec des services variés entre codage, métadonnées, citations, liens, voire stockage des données.
- Enfin, la montée en puissance d'entrepôts de données, alignés selon des standards partagés, que pourraient prendre en charge les institutions de recherche, afin de veiller à l'indépendance de leurs données. Il s'agirait alors de proposer des services efficaces et durables pour les chercheurs.

Une coordination entre acteurs publics de la recherche et éditeurs scientifiques serait positive pour le partage des tâches liées à la gestion et à la valorisation des données, ainsi que pour l'optimisation des coûts qu'engendrera inéluctablement le pilotage de ces données.

REMERCIEMENTS

Nous tenons à remercier les 46 chercheurs qui ont accepté de participer à l'enquête. Nous remercions également les ex-pilotes du groupe *Données de la Recherche* de la Bibliothèque Scientifique Numérique, Francis André et Paul-Antoine Hervieux, pour nous avoir permis d'utiliser les résultats de la cartographie des services de données à des fins de recherche. Enfin, nous remercions Ourida Aberkane, Anne Ciolek et Marie-Christine Jacquemot de l'Inist-CNRS, qui ont conçu Cat OPIDoR et qui contribuent encore aujourd'hui à son amélioration.

NOTES

- [1] <https://www6.inra.fr/datapartage>
- [2] Institut National de la Recherche Agronomique
- [3] Unités régionales de formation à l'information scientifique et technique
- [4] https://bu.unistra.fr/opac/article/la-gestion-des-donnees-de-la-recherche-a-luniversite-de-strasbourg/services_donnees
- [5] <https://dmp.opidor.fr/>
- [6] Institut de l'Information Scientifique et Technique
- [7] Centre de Données Socio-Politiques (unité mixte de services de Sciences Po et du CNRS)
- [8] <https://bequali.fr>
- [9] <https://www.ouvrirlascience.fr/presentation-du-comite/>
- [10] Catalogue pour une Optimisation du Partage et de l'Interopérabilité des Données de Recherche (<https://cat.opidor.fr/>)
- [11] Association des responsables de l'information scientifique et technique des organismes de recherche français publics ou d'utilité publique (<https://www.eprist.fr/>)
- [12] COUPERIN est le consortium unifié des établissements universitaires et de recherche pour l'accès aux publications numériques (<https://www.couperin.org>). Son groupe de travail pour l'accès ouvert (GTAO) a pour mission de promouvoir et d'implémenter le libre accès dans les établissements membres du consortium.
- [13] <http://vizier.u-strasbg.fr/viz-bin/VizieR>
- [14] <http://simbad.u-strasbg.fr/simbad/>
- [15] <https://www.seanoe.org/>
- [16] <http://arkeogis.org/>
- [17] Horizon 2020, programme de financement 2014-2020 de la Commission européenne (<http://www.horizon2020.gouv.fr/>)
- [18] Agence Nationale de la Recherche (<http://www.agence-nationale-recherche.fr/>)
- [19] Agence d'Evaluation de la Recherche et de l'Enseignement Supérieur. Nous nous sommes basés sur le rapport d'évaluation des unités de recherche 2011-2012 (<https://www.hceres.fr/LISTE-ALPHABETIQUE-DES-ETABLISSEMENTS-ET-ORGANISMES-EVALUES/UNIVERSITE-DE-STRASBOURG-UNISTRA>), dernier rapport rendu public dans sa totalité, au moment où nous avons constitué notre échantillon en octobre 2017.

- [20] Maîtres de conférences, professeurs d'université, chargés de recherches et directeurs de recherche.
- [21] <http://cdsweb.u-strasbg.fr/index-fr.gml>
- [22] <https://www.synchrotron-soleil.fr/>
- [23] <http://www.progedo.fr/>
- [24] <https://www.recolnat.org/fr/>
- [25] <http://simbad.u-strasbg.fr/simbad/>
- [26] <http://vizier.u-strasbg.fr/viz-bin/VizieR>
- [27] <https://services-numeriques.unistra.fr/les-services-aux-usagers/hpc.html>
- [28] <https://www.europeansocialsurvey.org/>
- [29] Un questionnaire a été envoyé à chacune des 249 000 personnes inscrites à nature.com, biomedcentral.com et springer.com.
- [30] <https://www.ncbi.nlm.nih.gov/genbank/>

RÉFÉRENCES BIBLIOGRAPHIQUES

- BORGMAN Christine L. (2015). *Big data, little data, no data: scholarship in the networked world*. Cambridge MA: The MIT Press, 2015, 416 p.
- BYRNE Meg (2017). Making Progress Toward Open Data: Reflections on Data Sharing at PLOS. Billet du blog *PLOS Blogs* [en ligne]. Disponible sur : <https://blogs.plos.org/everyone/2017/05/08/making-progress-toward-open-data/> (consulté le 17 mai 2019)
- CHARTRON Ghislaine (2010). Scénarios prospectifs pour l'édition scientifique. *Hermès, La Revue* [en ligne], 2010, vol. 5n n°2, p.123-129. Disponible sur : <https://www.cairn.info/revue-hermes-la-revue-2010-2-page-123.htm> (consulté le 17 mai 2019)
- CHARTRON Ghislaine (2018). L'Open science au prisme de la Commission européenne. *Education et sociétés* [en ligne], 2018, vol. 41, n°1, p.177-193. Disponible sur : <https://www.cairn.info/revue-education-et-societes-2018-1-page-177.htm> (consulté le 17 mai 2019)
- CUKIER Kenneth (2010). Data, Data Everywhere: A special Report on Managing Information. *The Economist* [en ligne]. Disponible sur : <http://www.economist.com/node/15557443> (consulté le 17 mai 2019)
- FABRE Isabelle, GARDIES Cécile (2008). L'accès à l'information scientifique numérique : organisation des savoirs et enjeux de pouvoir dans une communauté scientifique. *Sciences de la Société* [en ligne], 2008, p.84-99. Disponible sur : <https://hal.archives-ouvertes.fr/hal-00802763> (consulté le 17 mai 2019)
- FECHER Benedikt, FRIESIKE Sascha, HEBING Marcel, LINEK Stephanie (2017). A reputation economy: how individual reward considerations trump systemic arguments for open access to data. *Palgrave Communications* [en ligne], 2017, vol. 3, n° article 17051. Disponible sur : <https://www.nature.com/articles/palcomms201751> (consulté le 17 mai 2019)
- FECHER Benedikt, FRIESIKE Sascha, HEBING Marcel, LINEK Stephanie, SAUERMAN Armin (2015). *A Reputation Economy: Results from an Empirical Survey on Academic Data Sharing*.

- DIW Berlin Discussion Paper n°1454 [en ligne]. Disponible sur : <https://ssrn.com/abstract=2568693> (consulté le 17 mai 2019)
- LEONELLI Sabina (2013). Why the Current Insistence on Open Access to Scientific Data? Big Data, Knowledge Production, and the Political Economy of Contemporary Biology. *Bulletin of Science, Technology & Society* [en ligne], 2013, vol. 33, n° 1-2, p.6-11. Disponible sur : <https://doi.org/10.1177/0270467613496768> (consulté le 17 mai 2019)
- LEONELLI Sabina (2015). What Counts as Scientific Data? A Relational Framework. *Philosophy of Science* [en ligne], 2015, vol. 82, n°5, p.810-821. Disponible sur : <https://www.jstor.org/stable/10.1086/684083> (consulté le 17 mai 2019)
- MAX PLANCK GESELLSCHAFT (2003). *Berlin Declaration on Open Access to Knowledge in the Sciences and Humanities* [en ligne]. Disponible sur : <http://openaccess.mpg.de/Berlin-Declaration> (consulté le 17 mai 2019)
- REBOUILLAT Violaine (2017). Inventory of Research Data Management Services in France. In : CHAN Leslie, LOIZIDES Fernando (dir.). *Expanding Perspectives on Open Science: Communities, Cultures and Diversity in Concepts and Practices* [en ligne]. Amsterdam : IOS Press, 2017, p.174-181. Disponible sur : <http://ebooks.iospress.nl/publication/46651> (consulté le 17 mai 2019)
- REBOUILLAT Violaine (à paraître). Les données scientifiques face aux enjeux de la recherche en Sciences, Technologie et Médecine : enquête exploratoire à l'Université de Strasbourg. *Etudes de communication* [en ligne], à paraître, 52.
- SCHÖPFEL Joachim (2018). *Vers une culture de la donnée en SHS : Une étude à l'Université de Lille*. Rapport de recherche [en ligne]. Villeneuve d'Ascq : Université de Lille, 2018, 47 p. Disponible sur <https://hal.archives-ouvertes.fr/hal-01846849>
- SCHÖPFEL Joachim, KERGOSIEN Eric, PROST Hélène (2017). Pour commencer, pourriez-vous définir « données de la recherche » ? Une tentative de réponse. Colloque *INFORSID 2017*, Toulouse. Disponible sur : <https://hal.archives-ouvertes.fr/hal-01530937> (consulté le 17 mai 2019)
- STUART David, BAYNES Grace, HRYNASZKIEWICZ Ian, ALLIN Katie, PENNY Dan, LUCRAFT Mithu, ASTELL Mathias (2018). *Whitepaper: Practical Challenges for Researchers in Data Sharing*. Figshare [en ligne], 30 p. Disponible sur : <https://doi.org/10.6084/m9.figshare.5975011.v1> (consulté le 17 mai 2019)
- TENOPIR Carol, DALTON Elizabeth D., ALLARD Suzie, FRAME Mike, PJESIVAC Ivanka, BIRCH Ben, POLLOCK Danielle, DORSETT Kristina (2015). Changes in Data Sharing and Data Reuse Practices and Perceptions among Scientists Worldwide. *PLoS ONE* [en ligne], 2015, vol. 10, n°8: e0134826. Disponible sur : <https://doi.org/10.1371/journal.pone.0134826> (consulté le 17 mai 2019)

ANNEXES

Annexe 1 : Mail envoyé à la première population d'enquêtés

Bonjour,

Nous nous sommes rencontrées l'an dernier. J'étais venue vous interroger sur les données scientifiques du projet xxx.

Je vous remercie chaleureusement pour l'entretien que vous m'avez consacré. Il m'a été d'une grande utilité dans mes recherches.

En 2017 j'ai contribué à la conception d'un répertoire appelé [Cat OPIDoR](#). Cet outil recense les services qui sont proposés aux chercheurs pour collecter, analyser, stocker et diffuser les données scientifiques.

Je m'interroge sur l'utilisation de ces services par la communauté de recherche. J'aimerais beaucoup avoir votre avis sur ce point.

Pourriez-vous prendre le temps de parcourir [Cat OPIDoR](#) et me faire part de vos observations (voir questions ci-après) ?

Je vous remercie par avance.

Bien cordialement,

Violaine Rebouillat.

Ma thèse porte sur trois aspects des données scientifiques :

- La place que celles-ci occupent dans les pratiques de recherche (c'était l'objet de notre entretien),
- Les services qui sont proposés aux chercheurs pour collecter, analyser, stocker et diffuser ces données,
- L'adéquation entre ces services et les attentes des chercheurs.

Je travaille actuellement sur le troisième aspect, à partir du catalogue [Cat OPIDoR](#), dont je vous parlais plus haut. Cat OPIDoR répertorie au niveau national les services dédiés aux données scientifiques.

Pourriez-vous me donner votre avis sur cet outil et les services qu'il recense ?

Voici quelques questions qui vous guideront peut-être.

- Le contenu de Cat OPIDoR vous est-il plutôt opaque ou limpide ?
- Connaissez-vous l'un des services répertoriés ? L'avez-vous déjà utilisé ?
- Parmi les différents types de services présentés, duquel (ou desquels) auriez-vous l'utilité dans vos recherches ?

Pour information, la liste des services répertoriés est accessible depuis la barre menu de gauche : [Service](#). On peut également naviguer par [domaine scientifique](#) ou par [type de service](#).

Si vous préférez me faire part de vos observations par téléphone, nous pouvons tout à fait convenir d'un rendez-vous.

Annexe 2 : Guide d'entretien utilisé pour la seconde population d'enquêtés

Introduction

- Présentation de l'enquête
- Objectif de l'entretien
- Signature du formulaire de consentement éclairé

Contexte

- Thématique de recherche
- Approche et méthodologie scientifiques
- Utilisation du terme de « données » (oui/non, exemples, définition)

Utilisation de services de données (utiliser aussi des synonymes : outils, plateformes...)

- Oui/non
- Description du/des service(s) utilisé(s)
- Mode d'accès à ce(s) service(s)
- Utilité de ce(s) service(s)

Cat OPIDoR

- Avis sur le catalogue (ergonomie, contenu, usage)
- Connaissance des services répertoriés

Conclusion

- Besoins en termes de services
- Opinion sur le mouvement d'ouverture des données scientifiques