



HAL
open science

Les données de la recherche à l'Université Bordeaux Montaigne

Julie Duprat

► **To cite this version:**

Julie Duprat. Les données de la recherche à l'Université Bordeaux Montaigne : Synthèse d'une enquête qualitative auprès des chercheurs. [Rapport de recherche] Université Bordeaux Montaigne. 2019. hal-02020141

HAL Id: hal-02020141

<https://hal.science/hal-02020141>

Submitted on 15 Feb 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.



Service Commun de Documentation

Université Bordeaux Montaigne – Service Commun de Documentation

Les données de la recherche à l'Université Bordeaux Montaigne

Synthèse d'une enquête qualitative auprès des chercheurs

Julie DUPRAT
21/12/2018

Table des matières

Table des matières	1
Introduction.....	3
Contexte de l'enquête	3
La mise en place d'une enquête qualitative.....	4
Partie I : Présentation des résultats de l'enquête	6
I. La typologie des données produites à l'Université Bordeaux Montaigne	6
a) Question préliminaire : définir le périmètre de l'enquête.....	6
b) Les données primaires collectées à l'Université Bordeaux Montaigne.....	7
c) Les données dérivées produites à l'UBM	9
d) Des données « légères » ?	10
II. La gestion actuelle des données de la recherche.....	12
a) Des pratiques très individuelles	12
b) La progressive mise en place d'une gestion collective dans les UMR.....	13
c) Archiver et partager ses données dans le cadre d'une recherche collective : une problématique récurrente.....	14
III. Partager ses données : une réflexion en cours	16
a) Partager ses données en SHS : une spécificité ?	16
b) Le partage des données de la recherche en SHS : des limites internes à la discipline ?...	17
c) Les arguments en faveur de l'ouverture des données de recherche.....	17
d) Les conditions posées par les chercheurs de l'UBM pour une ouverture des données ...	18
E) De nombreuses interrogations autour du droit des données de la recherche	20
IV. Les besoins des chercheurs de l'UBM	20
a) A la recherche d'outils adaptés	21
b) De nombreuses demandes pour la mise en place d'un accompagnement autour des données de la recherche	22
Des référents en matière de données de la recherche	22
Des nombreuses demandes de formations	23
Partie II : Des pistes de réflexion pour la création d'un service « données de la recherche » à l'UBM	25
I. Etablir des profils types pour améliorer la compréhension de la communauté des chercheurs de l'UBM.....	25
a) Les chercheurs menant un projet ou l'ayant récemment terminé	25
b) Les chercheurs bénéficiant d'un soutien technique	25
c) Les chercheurs préoccupés par la sauvegarde de la mémoire du laboratoire	26
d) Les chercheurs peu concernés par la question des données de la recherche	26
e) Un profil sous-représenté mais à exploiter : les jeunes chercheurs	26

II.	Accompagner les chercheurs dans la gestion de leurs données de la recherche	27
a)	Recommandations techniques	27
	Etat de l'art : plateforme d'archivage et de valorisation des données de la recherche	27
	Recommandations techniques par rapport aux pratiques et besoins des chercheurs de l'UBM	28
b)	Mobiliser les chercheurs de l'UBM autour des données de la recherche	28
	Stratégie de mise en place d'un service : quelle population cibler ?	28
	Accompagner et former les chercheurs	29
c)	Quel avenir pour le futur service « données de la recherche » ?	29
	Les principaux apports des entretiens avec les personnels d'appui à la recherche	29
	La mise en place d'un réseau des professionnels de la documentation	30
	Annexes	32
	Annexe 1 : Grille de questions	32

Introduction

Contexte de l'enquête

Alors que ces dernières années l'importance de l'ouverture des publications écrites par les chercheurs des universités françaises a été largement abordée, les regards se tournent désormais sur une autre de leurs productions avec les données de la recherche. Ces dernières, dont la définition ne cesse de varier en fonction des organismes¹, peuvent être définies simplement par toutes les productions, numériques ou non, collectées et réalisées par les chercheurs en amont de leur travail d'écriture proprement dit et à partir desquelles ils bâtissent leurs hypothèses. Les données de la recherche regroupent donc un ensemble hétéroclite de sources et matériaux de recherche, aussi appelés données primaires, ainsi que toutes les formes de traitement et d'analyses desdites sources, que l'on appelle alors des données dérivées.

L'importance de ces données dans le processus de recherche est de plus en plus criante. Ces trois dernières années, plusieurs évolutions importantes ont vu un passage obligé, sous l'impulsion décisive de l'Etat. Dès octobre 2016, la loi République Numérique (article 30) apporte des précisions sur le statut de « données publiques » des données de la recherche. Le 4 juillet 2018, la ministre Frédérique Vidal annonce la priorité à venir sur la Science Ouverte, en incluant un volet « données de la recherche ». A sa création à l'automne 2018, le Comité Science Ouverte, relevant du Ministère de l'Enseignement Supérieur et de la Recherche, a inclus un groupe de travail « données ». Enfin, les principaux organismes financeurs de projets de recherche (European Research Council et Agence Nationale de la Recherche) incitent, voire rendent obligatoire la présence de plan de gestion de données dans les projets déposés. Dès lors, la communauté scientifique, et par là les établissements d'enseignement supérieur et de recherche, sont tenus de se préoccuper de la diffusion des données produites par leurs travaux de recherche.

Plusieurs arguments viennent étayer l'idée d'un partage de ces corpus constitués ou produits par les chercheurs. Un argument scientifique tout d'abord : alors que ces données, une fois exploitées, sont souvent oubliées dans les ordinateurs ou les dossiers de leur producteur, il pourrait s'avérer utile de les mettre à disposition du reste de la communauté scientifique, afin de réinvestir ces corpus et les exploiter de manière différente. Le deuxième argument est éthique : dans un contexte où la recherche française est majoritairement financée par les fonds publics, et dans un souci de transparence commun au reste des données publiques, la question de l'ouverture des données de la recherche à la communauté civile est également prégnante. Enfin, le dernier argument est patrimonial : les corpus de données produits aujourd'hui sont en effet les archives scientifiques de demain et leur bonne préservation est donc essentielle pour la mémoire des projets scientifiques.

Dans cette optique, l'Université Bordeaux Montaigne, spécialisée en sciences humaines et sociales, est en train de mettre en place un service « données de la recherche » afin d'accompagner ses chercheurs dans la gestion et le partage de leurs données de recherche. Avant de mettre en place ce service, il a paru essentiel de se livrer à un état des lieux par le biais d'une enquête menée par une conservatrice-stagiaire de l'Ecole Nationale Supérieure des Sciences de l'Information et des Bibliothèques (ENSSIB) entre septembre et décembre 2018. Dans une logique *bottom up*, l'Université Bordeaux Montaigne

¹ Rémi Gaillard, De l'Open Data à l'Open Research Data : quelle(s) politique(s) pour les données de recherche ?, mémoire d'ENSSIB, 2014, pp. 15-17.

souhaite en effet partir des besoins exprimés par les chercheurs pour construire un service au plus près de leurs pratiques quotidiennes.

La mise en place d'une enquête qualitative

Le choix a été fait assez rapidement, en s'appuyant sur les retours des enquêtes effectuées précédemment sur le même sujet dans les universités de Rennes² et de Lille³, de se livrer à une **série d'entretiens qualitatifs**. Le sujet des données de la recherche, encore assez mal connu des chercheurs en SHS, se prête en effet peu à une enquête quantitative traditionnelle avec un questionnaire : les rapports de Rennes et Lille évoquent ainsi le fort taux d'abandon en cours de route du questionnaire face à des chercheurs non-familiarisés à ces questions et la nécessité de recourir systématiquement à des entretiens qualitatifs afin d'affiner leur réponse et d'avoir des réponses plus significatives. Ces entretiens qualitatifs, menés selon la méthode d'entretiens semi-directifs, permettent de plus d'avoir une vraie conversation avec les chercheurs et de prendre pleinement conscience de leurs problématiques et besoins, au cas par cas.

Le canevas des entretiens s'est inspiré en grande partie de celui de **l'enquête menée à Rennes** puisque les grandes problématiques, détaillées dans la suite de ce rapport, étaient les mêmes : typologie des données de la recherche, gestion actuelle de ces données, position sur leur partage, besoins et problématiques rencontrés. Il a également été fait le choix d'y rajouter une **question liminaire, visant à déterminer le degré de familiarité du chercheur** avec le terme « données de la recherche ». Le canevas utilisé, comprenant les questions principales mais également les relances utilisées par l'enquêtrice, peuvent être consultés en annexe de ce rapport.

Dans le cadre de cette enquête, **32 enseignants-chercheurs et 4 personnels d'appui à la recherche (personnels ITRF)** ont donc été rencontrés sur une période de six semaines, pour des entretiens d'une durée de 35 mn à 1h20 pour le plus long d'entre eux. L'objectif principal en établissant le panel était de rencontrer **tous les unités de recherche (UR) de l'Université Bordeaux Montaigne**, au nombre de seize⁴. L'objectif premier était de viser tous les directeurs d'unités de recherche, en complétant leurs entretiens avec différents chercheurs, et ce quel que soit leur âge : les jeunes chercheurs, en la personne des doctorants, ont ainsi été inclus dans l'enquête. Afin de pallier un défaut de représentativité inhérent à une enquête qualitative sur un panel restreint d'enquêtés, il a également été fait le choix d'équilibrer autant que faire se peut la proportion de chercheurs jeunes et plus âgés, ainsi que ceux participant ou ayant participé à un projet de recherche financé et ceux menant des travaux plus individuels.

Ces choix de constitution du panel ont donc permis de rencontrer 14 directeurs d'unité de recherche, ainsi que deux de leurs adjoints dans le cadre d'un entretien en binôme, sur les seize unités de recherche que compte l'Université Bordeaux Montaigne. Cette excellente participation des directeurs d'UR à l'enquête a du même coup conduit à une sur-représentativité des professeurs d'universités qui représentent presque les deux tiers des interrogés avec 20 répondants. Suivent ensuite les maîtres et

² Alexandre Serres et alii., Données de la recherche en SHS. Pratiques, représentations et attentes des chercheurs : une enquête à l'Université Rennes 2 : Rapport. Université Rennes 2, 2017.

³ Joachim Schöpfel. *Vers une culture de la donnée en SHS : Une étude à l'Université de Lille*. [Rapport de recherche] Université de Lille. 2018. [〈hal-01846849〉](#)

⁴ Pour la liste complète : https://www.u-bordeaux-montaigne.fr/fr/recherche/equipes_de_recherche.html

maîtresses de conférences au nombre de 7 et enfin les chargés de recherche CNRS et les doctorants, représentés chacun à hauteur de trois répondants.

Les UR de Bordeaux Montaigne ont par ailleurs tous pu être représentés, à l'exception de Géoresources et Environnement hébergé conjointement par l'Université Bordeaux Montaigne et Bordeaux INP. Le Centre d'Etudes des Mondes Modernes et Contemporaines (CEMMC) est le plus représenté au cours de cette enquête en raison d'un réseau déjà constitué par la chargée d'enquête : cette sur-représentativité, loin d'être à déplorer, s'avère être un atout pour le SCD qui pourra ainsi renforcer les liens déjà existants. Suit ensuite l'unité de recherche Culture, Littérature, Arts et Esthétique (CLARE) avec 4 répondants, les unités Médiation, Informations, Communication, Arts (MICA) et Textes, littérature : écriture et modèles (TELEM) avec trois chercheurs. La majorité des unités a été représentée par deux répondants, à l'instar des sept unités suivantes : Ameriber, IRAMAT-CRP2A, Ausonius, Cultures et littérature des mondes anglophones (CLIMAS), CLLE-ERSSàB, Centre François-Georges Pariset-Histoire de l'art, et enfin Passages. Enfin, quatre unités de recherche, parmi les plus petites en effectif de l'Université Bordeaux Montaigne, ont été représentées par un chercheur : Archeovision, IKER, L'Afrique dans le Monde (LAM) et enfin Sciences, Philosophie, Humanités (SPH).

Ces unités sont pour la très grande majorité multidisciplinaire et plusieurs chercheurs peuvent eux-mêmes osciller entre plusieurs thématiques de recherche. En gardant cette nuance en tête, il ressort cependant que la répartition par disciplines suit logiquement la participation des UR de l'UBM à cette enquête : l'Histoire et la Littérature sont surreprésentés avec 8 chercheurs chacun. Les disciplines les plus présentes sont ensuite les suivantes, par ordre décroissant : Sciences de l'Information et de la Communication (4 chercheurs), Archéologie, Arts et esthétique, Linguistique (3 chercheurs chacun) et enfin Géographie et Histoire de l'Art (2 chacun). Nous y reviendrons plus longuement au cours du rapport mais il s'est très vite avéré que la répartition par discipline était inopérante à l'Université Bordeaux Montaigne, le rattachement à une discipline ne préjugant qu'assez peu des pratiques et des positions des chercheurs sur les données de la recherche.

Enfin, parmi les répondants, 6 avaient conduit un projet financé les années précédentes et 11 étaient en train de mener un projet financé par l'ANR ou la région Nouvelle-Aquitaine au moment de l'enquête.

Partie I : Présentation des résultats de l'enquête

I. La typologie des données produites à l'Université Bordeaux Montaigne

a) Question préliminaire : définir le périmètre de l'enquête

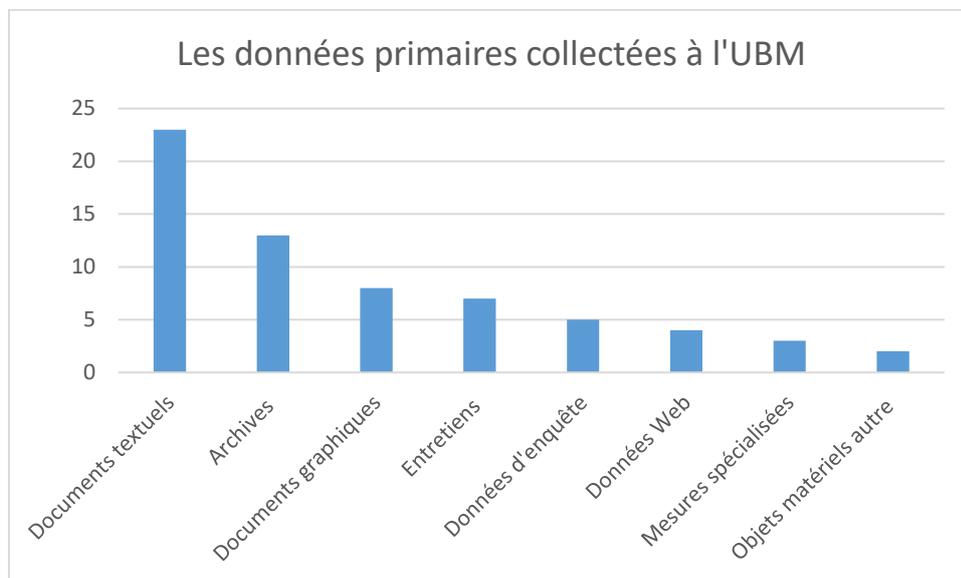
Avant d'aller plus en avant dans la description des résultats de l'enquête, il nous semblait important de revenir sur la définition des données de la recherche retenue pour cette enquête. Comme précisé ci-dessus, cette notion revêt en effet plusieurs définitions : ainsi, les données de la recherche sont souvent perçues avant tout comme un corpus de données numériques⁵. Nous avons fait le choix pour cette enquête de ne pas tenir compte, dans un premier temps, du format des données, considérant qu'exclure les données de recherche non-numériques de l'enquête aurait été susceptible d'écartier une grande partie des chercheurs de l'Université de l'enquête, comme l'ont d'ailleurs confirmé les entretiens. De même, nous nous sommes longuement interrogés sur la pertinence de garder une distinction entre données primaires et données dérivées, qui peut être parfois plus difficile à établir pour les SHS puisque la constitution du corpus engage déjà le chercheur dans un rapport subjectif à sa donnée. Nous l'avons cependant conservé, considérant à l'instar de l'enquête menée à Rennes que ces catégories permettaient de refléter les différentes étapes d'un travail de recherche. Pour être plus facilement compris des chercheurs, nous avons cependant choisi d'explicitier ces notions, préférant par exemple au terme « données primaires » celui de « corpus de sources » et aux termes « données dérivées » celui de « données obtenues après traitement ou analyse ».

L'entretien s'ouvrait avec une question préliminaire permettant de cerner la familiarité des chercheurs avec la question des données de la recherche : cette question s'est avérée essentielle puisqu'un tiers des répondants ne savait pas ce que recouvrait ce terme (soit 11 personnes). Si une minorité d'entre eux ont pu rapprocher les données de la recherche des archives scientifiques (« *Quelle est la différence avec l'archivage ?* »), la plupart ont été désarçonnés par ce terme. La confusion entre données de la recherche et données administratives du laboratoire a ainsi été fréquente (« *J'ai une idée de ce que sont les données de la recherche qu'on nous demande à nous : combien on est, combien on publie, dans quel type de revues ...* ») ainsi que celle entre données de la recherche et publications en ligne (« *Je m'interroge, qu'est-ce que vous voulez dire par là ? Les articles, les publications, les sites web ?* »).

L'autre partie des chercheurs interrogés pouvait définir tout ou en partie la notion de données de la recherche. Ainsi, au sein de ce groupe de 21 chercheurs informés, près de la moitié ne proposait qu'une définition partielle, ne retenant soit que les données primaires (« *Je vois à quoi ça peut renvoyer en sciences sociales [...] : les entretiens, les recueils statistiques, les questionnaires* »), d'autres au contraire que les données dérivées (« *Si je devais le résumer rapidement, je dirai tout ce qui peut être produit en seconde main après retraitement des données brutes* » ; « *tout le travail préparatoire, qui est individuel, dont une partie transparait dans les articles mais toutes les données, tableaux, statistiques que nous avons chez nous et qui finalement restent privés et obligent parfois quand on fait des recherches à recommencer les mêmes dépouillements, les mêmes investigations et qui sont rarement communiqués* »). Sur l'ensemble des répondants, seul un tiers pouvait donc fournir une définition « officielle » sur les données de la recherche (« *On distingue les données primaires et les données secondaires donc toutes les données dans un cadre très large* »). Quel que soit la réponse faite par le chercheur, nous nous sommes attachés à énoncer à la fin de ce premier échange la définition adoptée par l'Université Bordeaux Montaigne dans le cadre de l'enquête, afin d'être sûrs que les chercheurs n'excluent pas sciemment certains types de productions de leurs réponses.

⁵ Rémi Gaillard, *id.*, p.17

b) Les données primaires collectées à l'Université Bordeaux Montaigne



Cette première clarification effectuée, nous nous sommes ensuite penchés sur la typologie des données collectées à l'Université Bordeaux Montaigne, en commençant par les données primaires ou corpus de sources. La typologie que nous établissons ci-dessous, si elle a été inspirée par celle réalisée à Lille, s'en détache en partie pour être la plus fine possible.

Les résultats de ce premier sondage sont sans appel et rejoignent les conclusions établies à Lille : **70 % des chercheurs de l'UBM ont ainsi recours à un corpus textuel**⁶ contre 64 % à Lille. L'enregistrement même de ces corpus textuels est extrêmement varié : tandis qu'une minorité de chercheurs en fait une simple lecture sur support physique ou numérisé (« *Gallica par exemple ils sont en train de numériser, ils arrêtent pas de numériser, on trouve des choses en interrogeant* »), d'autres vont au contraire procéder à une sauvegarde en les recopiant ou en sélectionnant des extraits (« *j'ai la liste des extraits, des occurrences sur un document Word* ») ou encore en téléchargeant une version sur leurs ordinateurs. Ce type de corpus sources se retrouve très largement dans toute la communauté des chercheurs de l'UBM puisque toutes les disciplines sont représentées.

Le deuxième corpus de sources le plus fréquemment utilisé est celui des **archives puisque 40 % des chercheurs de l'UBM**, très majoritairement rattachés à une unité de recherche en histoire ou en littérature, s'appuient sur un corpus de ce type. Là encore, les résultats sont assez proches de ceux trouvés à Lille, où cette pratique concernait 34 % des répondants. Ces archives, si elles étaient initialement recopiées par les chercheurs, sont désormais photographiées par une très grande majorité (« *j'ai photographié des microfilms pour les retravailler ici [...] J'ai des bobines entières dupliquées et stockées sur mon ordinateur, c'est quasiment des archives à la maison* ») ou consultées directement en ligne sur des portails spécialisés (« *notre grand site à nous c'est un portail papyrologique [...] On utilise beaucoup ce portail qui nous renvoie aux collections qui existent, soit les textes, soit les collections du musée* »).

⁶ Par « corpus textuel », nous entendons ici des corpus d'œuvres éditées et publiées, par opposition aux documents d'archives non-publiés.

Le troisième corpus de sources le plus fréquemment utilisé est le corpus constitué par les entretiens, souvent conservés sous **format audiovisuel (captation sonore ou vidéo) : 27 % des chercheurs sont ainsi concernés par cette pratique**, en moindre proportion donc qu'à Lille (47 %). (« *Et puis le troisième type de support sur lequel on travaille, ce sont les entretiens* ».) Ces entretiens sont logiquement plébiscités comme sources par les linguistes, les géographes et les chercheurs en sciences de l'information et de la communication mais sont également de plus en plus saisis par les jeunes historiens comme source orale (« *En contemporaine, on a souvent des enquêtes, des entretiens* »).

Les entretiens sont suivis de très près par les documents graphiques auxquels 24 % des chercheurs ont recours. Cette dénomination de « documents graphiques » rassemble de fait des pratiques très variées, souvent liées aux profils disciplinaires : cartes et photographies en géographie et archéologie (« *une carte postale ou une vieille photographie est extrêmement important quand un bâtiment a pu disparaître* »), tableaux (« *les chercheurs en histoire de l'art, évidemment, réunissent des corpus importants de documentation iconographique* ») ou encore image satellitaire. Il est à noter que ces documents graphiques n'adoptent pas forcément un format numérique : ainsi, si certaines cartes ont été numérisées ou sont disponibles sur des portails spécialisés en format numériques (« *j'ai un collègue qui travaille sur les cartes du Web* »), d'autres sont toujours uniquement consultable sur un support matériel (« *Ça fait 25 ans que j'ai des cartes chez moi [...].* »).

Passés ces quatre ensembles, les pratiques sont par la suite plus éparses. On retrouve en cinquième position les **données d'enquêtes : 15 % des chercheurs de l'Université Bordeaux Montaigne**, tous en géographie ou en sciences de l'information et de la communication, utilisent ce type de recueil d'informations (« *des données qui proviennent d'enquêtes menées auprès des usagers de ces portails* »). Enfin, en queue de peloton, **trois corpus de sources sont à égalité, représentant chacun 9 % des sources utilisées : les données chiffrées et 3D issues de logiciels de mesures chez les archéologues, les objets matériels** (échantillons de terrain ou maquettes par exemple) **et enfin les données web**.

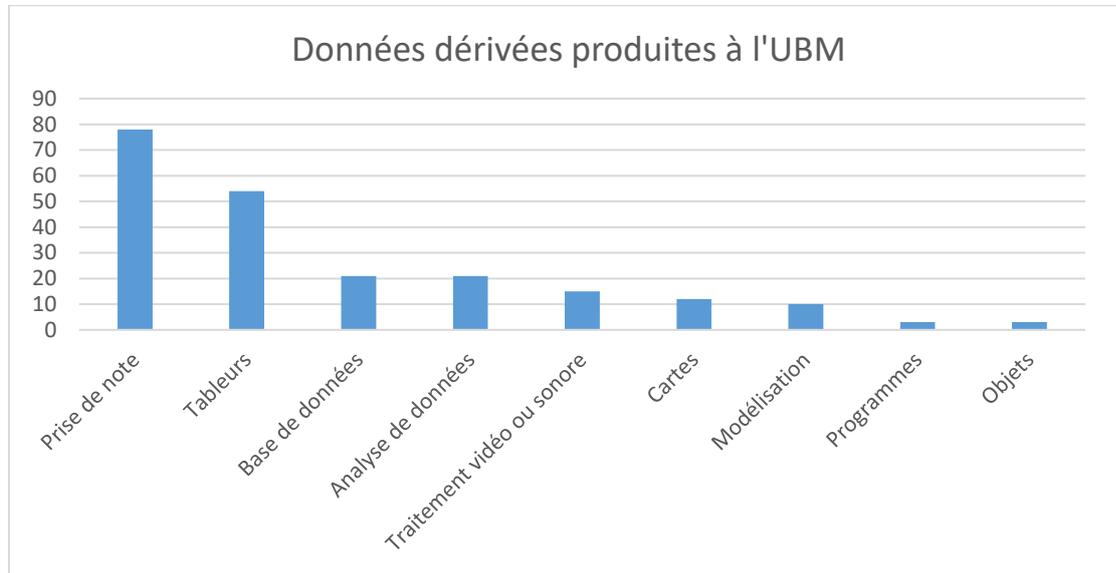
Ce dernier corpus de source, pour l'instant uniquement utilisé dans notre panel par des doctorants, a cependant été cité par d'autres chercheurs plus âgés ayant conscience de l'importance de cette source chez la jeune génération (« *Il y en a qui travaillent beaucoup sur Internet en ce moment* »). Cette nouvelle source pose d'ailleurs une vraie question sur la mise en place d'un processus d'archivage adapté par le biais des URL et/ou des captures d'écran pour retrouver facilement l'information (« *Et en terme de liens URL, répertorier les pages pour revenir dessus avec aussi le risque ... D'où l'importance de la numérisation, d'alimenter une base de données avec des captures d'écran, essentielles, car j'y ai déjà été confronté en fait avec des sites construits au début des années 2000 avec des sites qui ferment donc il n'y a plus de sources* »).

73 % des chercheurs manipulent plus d'un type de données mais contrairement aux résultats trouvés pour l'université de Lille où près de 43 % des répondants manipulaient au moins quatre types de corpus différents, cette diversité est plus restreinte pour l'Université Bordeaux Montaigne : **36 % des chercheurs n'ont ainsi recours qu'à deux types de sources, associant d'ailleurs bien souvent le corpus textuel et archivistique**. L'écart se creuse ensuite rapidement : 15 % des chercheurs ont recours à trois corpus de sources différentes et 12 % seulement à plus de trois sources différentes.

La typologie des données primaires collectées par les chercheurs de l'UBM diffère donc sensiblement des résultats des enquêtes de Lille et Rennes où, si le texte conservait la primauté, la majorité des sources utilisées provenait ensuite d'entretiens et d'observations. Les chercheurs de l'UBM ont donc un rapport encore plus étroit au texte comme source de recherche : après le duo de tête des

documents textuels et des archives, la répartition des réponses révèle des pratiques assez variées et diffuses autour de l'exploitation d'enquête, d'entretiens ou de documents graphiques.

c) *Les données dérivées produites à l'UBM*



A l'instar de ce qui a été observé dans les enquêtes de Rennes et de Lille où les données textes étaient majoritaires, les données dérivées les plus fréquemment produites à l'Université Bordeaux Montaigne sont du même ordre, **90 % des chercheurs procédant à une prise de note sous format numérique ou papier**.

La majorité des chercheurs, soit 78 % d'entre eux, a ainsi recours à une prise de note informatique : Word y tient le haut du pavé tandis que les solutions libres ne sont choisies que par une toute petite minorité (« *Moi je préfère utiliser Word, c'est l'outil le plus souple que j'ai à ma disposition* »).

Cette prise de note informatique n'empêche cependant pas la production de notes manuscrites : si pour une minorité d'entre eux, cette prise de notes n'est qu'une étape vers la prise de note informatique et n'est donc pas conservée (« *Vous avez des archives manuscrites ? R : Je les ai chez moi, c'est quand je prends des notes ... Mais ce n'est lisible que pour moi. Mais au bout d'un moment je jette car une fois que c'est propre ...* »), 36 % des répondants associent de manière complémentaire ces deux modes d'écriture (« *je prends des notes sur des cahiers et mon ordinateur* »). Enfin, il est à noter qu'une part non-négligeable des chercheurs de l'UBM, soit 12 % d'entre eux, n'ont recours qu'à des prises de notes papier pour analyser leur corpus (« *moi je prends des notes manuscrites et pas sur mon ordinateur* »).

Les tableurs, créés à partir d'outils de bureautique classiques, occupent ensuite la deuxième position des données dérivées les plus présentes à l'UBM. Ces **tableurs** sont produits majoritairement dans les communautés de chercheurs en histoire, linguistique et science de l'information et de la communication pour analyser des corpus quantitatifs et fédèrent ainsi **54 % des chercheurs de l'UBM**, un chiffre similaire aux communautés de chercheurs en SHS à Lille. Ces tableurs sont utilisés avec des objectifs très différents (« *L'avantage d'Excel c'est qu'on peut le modéliser comme on veut, c'est une structure vide sur laquelle on peut matérialiser plusieurs besoins* ») : centralisation d'informations (« *on les dépouille sur Excel* »), création de graphiques, exploitation statistique (« *Oui, après on fait des tableaux Excel, des écarts types, des statistiques* »).

Vient ensuite la production de base de données ou d'analyses statistiques à partir de logiciels spécialisés, concernant chacun 21 % des chercheurs de l'UBM. Les analyses statistiques sont plébiscitées par les communautés de chercheurs en linguistique ou aux disciplines empruntant aux sciences dures comme la géographie ou l'archéologie (« *on va réaliser un certain nombre d'analyses spatiales, statistiques ou réseaux* »). Les logiciels utilisés pour ces productions reflètent la très grande diversité des pratiques et la plupart, à l'exception notable d'Alceste, Sphinx et R, ne sont d'ailleurs cités qu'une fois comme Analec, Analore, Simi, Voyant Tools, Trope, Lexico, Tulip ou Iramuteq.

La production de base de données à partir de logiciels spécialisés est utilisée par un éventail plus varié de disciplines, au sein duquel on retrouve les chercheurs en histoire, géographie, archéologie, histoire et histoire de l'art qui s'appuient tous sur le logiciel propriétaire FileMaker (« *on utilise FileMaker parce que c'est très pratique et facile à utiliser* ») si l'on excepte une mention à la solution libre PostgreSQL.

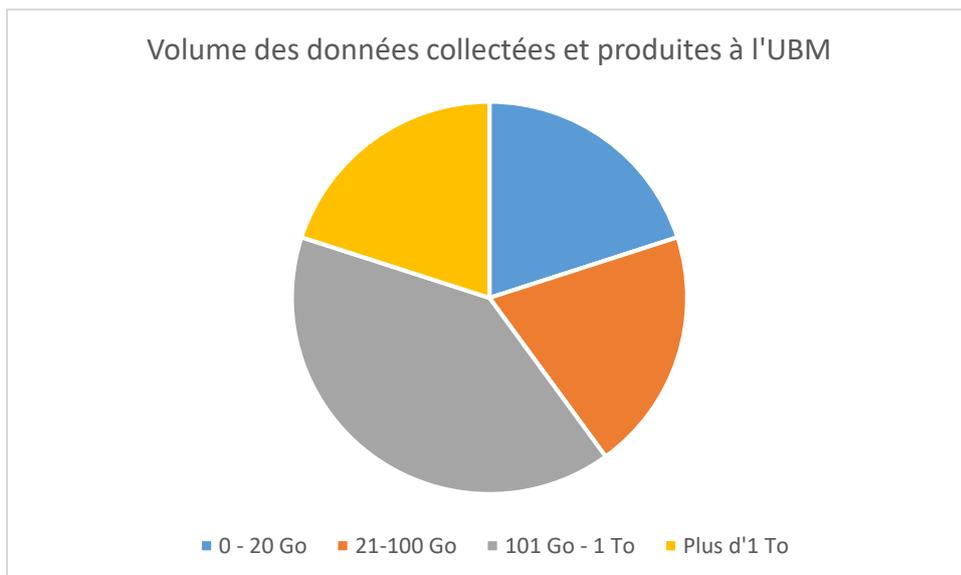
Les pratiques sont ensuite beaucoup plus éparses : **15 % des répondants produisent des données audiovisuelles issues du traitement de sources sonores et vidéos** à partir de logiciels comme Praat, Audacity ou les solutions fournies par Apple. Ils sont suivis de près par les chercheurs créant des **cartes (12 %), de la modélisation ou visualisation 3D (10 %), des programmes (3 %) ou encore des productions esthétiques pour 3 % d'entre eux** (« *Les documents qu'on archive, il y a disons des documents qui résultent d'expérience donc ça peut être des petites fabrications matérielles, petits objets [...]* »).

Les chercheurs de l'UBM sont donc marqués par un traitement qualitatif de leurs corpus de données. Leurs pratiques restent assez similaires aux observations faites à Lille et à Rennes puisqu'on retrouve le même podium de données dérivées, avec des légères différences de proportions : les documents textes, les tableurs et les bases de données se détachent ainsi, suivis d'une longue traîne de productions variées.

Nous aimerions conclure cette sous-partie sur la typologie des données de la recherche en mettant en lumière deux points qui nous ont frappé lors des entretiens semi-directifs. Ainsi, si certains chercheurs sont conscients de la diversité des pratiques et de la difficulté de généraliser à partir de sa propre expérience, certains au contraire font montre d'une relative méconnaissance des pratiques de leurs confrères (« *Même parmi les littéraires, il y en a peut-être qui prennent leur travail préparatoire sur Word, prendre des notes sur leur ordinateur ... Ça peut arriver, c'est pas impossible* »). Par ailleurs, la production de données dans un format majoritairement textuel et qualitatif a induit plusieurs chercheurs à penser qu'ils n'étaient pas concernés par l'enquête, associant en effet les données de la recherche à du quantitatif (« *Dans la recherche, tout n'est pas comptable, dénombrable, on ne produit pas de résultats [...] On a pas grand-chose à montrer en amont sauf à donner nos résumés de texte, ce qui n'intéressera personne* »).

d) Des données « légères » ?

Nous nous sommes ensuite intéressés au volume que pouvaient représenter les données collectées et produites dans leur format numérique. Cette question a été la plus difficile à cerner en entretien qualitatif par les chercheurs. Cet état de fait inciterait sans doute à revoir la méthodologie d'enquête sur ce point, en demandant peut-être en amont ou en aval des entretiens cette information assez technique. Malgré tout, nous avons pu avec la seconde moitié des chercheurs obtenir une assez bonne vision du volume des données des chercheurs de l'UBM qui se répartissent comme suit :



La prédominance des données pesant entre 101 Go et 1 To est confirmée par d'autres chercheurs qui, s'ils n'ont pas pu fournir une fourchette précise, ont pu évaluer la capacité de saturation de leurs espaces de stockage. A titre de comparaison, les volumes des données relevés dans les universités de Lille et Rennes semblent plus légers, la médiane des réponses se situant plutôt dans la fourchette 20-100 Go.

Cette prédominance des données pesant entre 101 Go et 1 To pourrait étonner pour des données essentiellement textuelles, nous l'avons vu, et donc théoriquement plus légères. De fait, deux formats influencent grandement sur ce poids :

- Les **téléchargements en PDF** de corpus textuels, tout d'abord, représentent une volumétrie conséquente, pointée par les chercheurs (« *J'ai une dizaine de clefs USB, de 32 Go. Ce sont des PDF. Un manuscrit d'une œuvre c'est 270 Mo* »).
- Par ailleurs, les **images fixes** utilisées comme support d'enregistrement sont très rapidement volumineuses : si l'imagerie satellitaire, pesant plusieurs Go, ne concerne qu'une extrême minorité des chercheurs de l'UBM, une grande partie d'entre eux ont malgré tout à gérer des photographies numériques d'archives pesant de plus en plus lourd (« *avec l'amélioration des appareils, la moindre photo c'est 2-3 Mo donc ça engorge très vite* »). Ces corpus d'archives photographiées représentent une masse de plus en plus volumineuse et en pleine croissance, certains chercheurs étant loin d'avoir fini leur carrière universitaire (« *ça fait 6 ans que je suis à l'université. 6 ans, 500 Go* »).

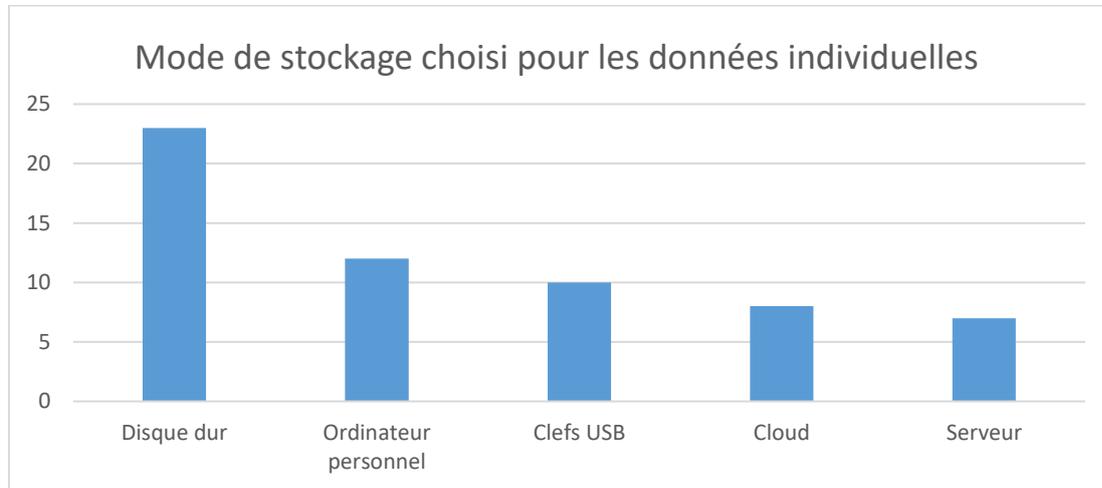
Si nous avons rassemblé la volumétrie comprise entre 101 et 1 To à l'imitation des précédentes enquêtes, il nous paraît cependant important de signaler que **le palier des 500 Go est souvent cité par les chercheurs pour évoquer leurs premières difficultés de stockage.**

Par ailleurs, et comme nous l'avons indiqué dans notre description typologique, plusieurs chercheurs conservent également de **nombreuses données papier** qui sont beaucoup plus difficilement quantifiables en terme de volumes : certains évoquent cependant « *plusieurs cahiers* » ou des « *milliers de fiches* ». Ces indications incitent donc à ne pas négliger ce type de données de la recherche qui constituent encore une partie des pratiques.

II. La gestion actuelle des données de la recherche

La deuxième problématique qui intéressait de près notre enquête était celle de la gestion de toutes ces données de recherche, qu'elles soient collectées ou produites.

a) Des pratiques très individuelles



Une première constatation s'est rapidement imposée : les chercheurs ont un stockage et un archivage de ces données organisées de **manière très individuelle** (« *C'est vraiment une démarche individuelle et ça peut même poser problème pour l'évaluation du laboratoire* »). En-dehors de leur ordinateur professionnel, **plusieurs espaces de stockage sont donc utilisés en fonction de leur préférence**. Celui qui remporte le plus de faveurs, à l'instar d'ailleurs des résultats trouvés lors des enquêtes menées à Lille et Rennes, est le **disque dur, utilisé par 66 % des chercheurs** comme premier espace de stockage. La capacité de stockage de ces disques durs externes et l'accessibilité de ces dispositifs expliquent ce succès, suffisant largement à la plupart des chercheurs pour stocker leurs archives de recherche (« *Et vous n'avez jamais été confronté à un problème de place sur vos disques durs ? / Non. J'utilise même pas le tiers, le quart de ce que j'ai...* »).

L'ordinateur personnel arrive ensuite en seconde position pour 36 % des répondants, un choix souvent corrélé à la volonté de travailler facilement depuis chez soi mais également afin de pallier un manque de confiance placé dans les ordinateurs professionnels (« *j'ai vidé sur l'espace de mon université et j'ai tout mis chez moi parce que j'ai plus peur, je ne sais pas si c'est rationnel* »).

Ce mode de stockage est suivi de près par le recours à des **clefs USB, utilisées par 30 % des chercheurs** (« plus je sature mes ordinateurs et j'externalise sur des serveurs externes. J'en viens à avoir des clefs USB énormes, spécifiques sur chaque projet »).

Les Clouds (stockage en ligne) arrivent en quatrième position, étant utilisés par seulement 21 % des répondants : de nombreux chercheurs leur opposent en effet le **manque de place** dans les versions gratuites ou craignent pour la **sécurité** de leurs données (« *Mais je n'ai pas eu le temps de me renseigner sur des Clouds fiables et puis les photos c'est très lourd, j'ai toujours peur de remplir le Cloud en un ou deux ans* »).

Enfin, **15 % d'entre eux, faisant partie d'une unité mixte de recherche, dispose d'un espace de stockage commun** à l'ensemble du laboratoire, une synchronisation étant effectuée de manière régulière vers un serveur à partir des ordinateurs professionnels.

La peur de perte des données étant encore extrêmement prégnante, la plupart des chercheurs multiplie les dispositifs d'archivage : si 21 % d'entre eux n'ont qu'un espace de stockage, la plupart font le choix d'associer plusieurs dispositifs. Ainsi, **27 % des chercheurs ont deux espaces de stockage, 30 % en ont trois et 12 % en ont au moins quatre** (« *j'ai un ordinateur maison très performant, avec des disques durs extractibles, des sauvegardes automatiques, je mets sur un Cloud, j'ai une clef USB, c'est sauvegardé plein de fois* »).

De la même façon, les pratiques de sauvegarde sont assez variées : tandis que certains pratiquent une sauvegarde régulière de manière quotidienne ou hebdomadaire (« *Je sauvegarde tous les jours* »), notamment grâce à la synchronisation de leurs espaces de stockage (« *En principe c'est automatique, si le disque dur est branché c'est enregistré* »), d'autres ne se livrent à cette opération que quelques fois dans l'année (« *... Dans mon irrégularité, je suis régulière, c'est moins de 6 mois* »).

Le stockage et l'archivage des données de la recherche est donc pour beaucoup une gestion individuelle, dont la plupart ne sont pas satisfaits, ayant conscience de leur méconnaissance des bonnes pratiques ou des outils appropriés (« *Comment conservez-vous vos données ?* "Alors ça c'est catastrophique" »). Cette gestion individuelle a également un réel coût financier qu'ils assument seuls.

Les archives papier, qui, nous l'avons vu, concernent encore une part non-négligeable des pratiques des chercheurs, sont conservées de manière encore plus individuelle : tous les chercheurs concernés mentionnent ainsi les conserver « *chez [eux]* » et seul Ausonius, nous le verrons, a commencé à se préoccuper de la gestion de cette masse.

b) La progressive mise en place d'une gestion collective dans les UMR

Les entretiens ont cependant permis de faire ressortir plusieurs exceptions puisque parmi quelques unités de recherche de l'UBM, les UMR associées au CNRS, ont entrepris de réfléchir à une gestion collective des données de la recherche, ce qui a amené à la mise en place de plusieurs dispositifs. Ces dernières bénéficient toutes de l'appui d'un centre de documentation et de personnels d'appui qui les ont aidés à mettre en place ce soutien technique. Nous rappellerons ces dispositifs dans les grandes lignes ci-dessous, sans inclure de citations pour préserver l'anonymat des répondants.

Le Centre de Documentation associé à l'UMR Passages mène un travail de sensibilisation auprès des chercheurs et a mis en place en partenariat avec la TGIR Huma-Num un système d'archivage des données de la recherche : en plus de la sauvegarde des données sur un serveur collectif, les chercheurs ont notamment été invité à déposer leurs données froides dans Nakala. Par ailleurs, le laboratoire propose désormais un outil de stockage et de partage en ligne ne dépendant pas d'initiatives privées en faisant appel à ShareDocs ainsi que d'une plateforme de valorisation des données de la recherche avec Navigae⁷ pour les photographies et les cartes. L'UMR **IRAMAT**, de la même façon, dispose également d'un serveur collectif où les sauvegardes s'effectuent sur une base régulière, ainsi que d'un informaticien chargé de faire l'intermédiaire avec le CINES pour l'archivage des données ; l'UMS **Archéovision** enfin, en raison de son statut d'unité mixte de soutien, est sans aucun doute le plus avancé sur la question de la gestion des données de la recherche avec un processus de stockage et d'archivage en plusieurs temps s'appuyant sur Huma-Num et le CINES. Cette unité a également développé son propre outil pour faciliter le dépôt des données chaudes.

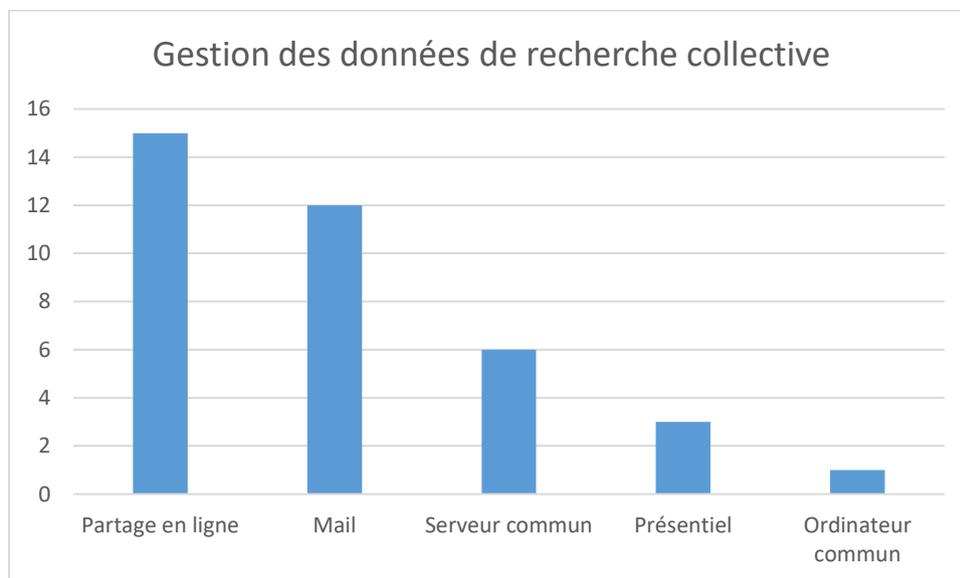
Au sein de **l'UMR Ausonius**, la venue d'une archiviste engagée comme contractuelle durant l'année 2017 a permis de sensibiliser les chercheurs à la question de l'archivage de leurs données de recherche

⁷ « Navigae » est un projet collectif, lié au consortium ImaGEO, qui regroupe les données de plusieurs établissements, dont l'UMR PASSAGES et le SCD de l'UBM (<https://imageo.hypotheses.org/>).

sous format papier. L'archiviste a effectué un travail de classement des archives papier et mis en place une série de recommandations affichées dans les bureaux des chercheurs. Cette initiative s'est posée en complément de l'action de plusieurs personnels d'appui à la recherche qui se chargent d'assurer l'archivage et la pérennité des bases de données produites dans le laboratoire ; l'outil ShareDocs est par ailleurs mis à la disposition des chercheurs.

Enfin, l'UMR IKER a mis en place une chaîne de stockage composée de plusieurs modules. Le premier est un serveur NAS synchronisant le stockage de toutes les ressources utilisées par les chercheurs de l'unité (données de la recherche, logiciel, e-books), uniquement accessible par l'ingénieur d'études en charge des données. Le second est un espace de stockage et de partage spécifiquement dédié aux données de la recherche, s'appuyant sur le logiciel Nuxeo, avec des dossiers à l'accès restreint aux chercheurs d'un même projet. Une fois que les publications ont été effectuées, les données de la recherche sont basculées sur le logiciel Omeka sur une copie compressée et en dernier lieu sur la bibliothèque numérique An-Persana.

c) *Archiver et partager ses données dans le cadre d'une recherche collective : une problématique récurrente*



Une des dernières questions qui nous a préoccupée au cours de ces entretiens était de savoir comment étaient partagées, stockées voire archivées les **données intéressant simultanément plusieurs chercheurs, notamment dans le cadre de projets de recherche**. Nous ne disposerons pas pour cet élément de point de comparaison avec les précédentes enquêtes, cet item étant spécifique à l'enquête menée à l'UBM. Cette question nous a permis de constater une diversité des pratiques plus grandes encore que pour le stockage des données « individuelles », une hétérogénéité provenant de la difficulté visible des chercheurs pour s'emparer de cette problématique.

Ainsi, la pratique la plus couramment utilisée par les chercheurs lorsqu'ils travaillent de manière collaborative est le recours à des **plateformes de stockage et de partage en ligne, pour 46 % d'entre eux**. Ces dernières peuvent prendre des formes variées : la très grande majorité des chercheurs se tournent ainsi vers des solutions privées, type **Dropbox ou GoogleDocs**, avec les limites de volume induits par les versions gratuites et les problèmes de sécurité liés à ces dispositifs comme nous l'avons vu plus haut. Pour contourner cette problématique, un répondant a même détourné les fonctions premières du réseau social **Facebook** pour l'utiliser comme plateforme d'échange de documents, l'espace y étant illimité, tout en soulignant sa volonté de s'en émanciper (« *Mais dans l'immédiat,*

comme on a travaillé de façon empirique, le mail était ... [...] Donc en fait ce qu'on a utilisé surtout comme système d'échange c'est Facebook, un groupe fermé. Parce que mine de rien, tel que c'est construit, c'est vraiment pratique : on fait des albums, on met des vidéos, des fichiers, des mises à jour de fichiers donc ça a été une plateforme d'échange. Après ça a été assez délicat comme organisation du travail [...] Mon objectif aussi c'est de sortir du réseau Facebook »). Cette utilisation de solutions privées est également liée aux limites des alternatives institutionnelles : le **Bureau Virtuel** mis à disposition des chercheurs n'est utilisé que par une toute petite minorité en raison de son faible espace de stockage (« Il y a déjà le BV qui est une plateforme de partage facile pour les documents textuels »), tandis que les **outils développés par le CNRS ou Huma-Num** ne sont connus que lorsque le chercheur dispose d'un accompagnement technique et ne concernent donc que 10 % d'entre eux.

Ce mode opératoire est suivi de près par **l'échange par mail, à l'aide de pièces jointes, plébiscité par 40 % d'entre eux**. Cette méthode, permettant notamment de maintenir des relations personnelles avec ses collègues (« Moi j'ai toujours trouvé ça plus sympa de recevoir des mails que d'aller sur une plateforme ») soulève cependant plusieurs difficultés dont la plus évidente est l'espace de stockage disponible au sein de la boîte mail institutionnelle qui n'est pas toujours suffisant (« La boîte mail de l'Université, on peut envoyer jusqu'à 8 Mo mais en général, quand on envoie un Word, ça va pas jusque-là. Si on envoie plusieurs PDF, il faut faire attention à ne pas dépasser, fractionner l'envoi. »). Cette méthode est d'autant moins efficiente que de très nombreux chercheurs ont signalé au cours de leurs entretiens la difficulté qu'ils avaient à gérer leurs courriers électroniques.

Passé ces deux groupes, les méthodes mises en places sont beaucoup moins systématiques. Ainsi, **15 % des chercheurs s'appuient sur un serveur ou une base de données en ligne** pour échanger et stocker leurs données (« C'est une fédération avec une quarantaine de chercheurs qui fournissent leurs données [...] Tout cela est mis sur une carte géolocalisée et collaborative »), parfois fourni par une autre université partenaire (« On a deux projets où c'est l'Université de Brest qui nous met à disposition un site qui sert de vitrine au projet de recherche et quand on se connecte en intranet on peut partager nos données etc. »), tandis que cet échange se fait en **présentiel pour 10 % d'entre eux**, sans passer par un outil de stockage et de partage à distance. Enfin, dans le cadre d'un projet collaboratif, les chercheurs travaillent sur un ordinateur fixe commun, dédié au projet, centralisant toutes les informations.

Ainsi, la question de l'archivage de ces données de recherche produites dans le cadre de projets communs pose encore de nombreuses problématiques : la question des données de la recherche produites à plusieurs est rarement pensée en amont du projet, amenant à une **mauvaise gestion courante** (« c'est un bazar pas possible et pour moi c'est un vrai souci et très honnêtement je ne sais pas m'y prendre, je ne sais pas gérer ça ») voire à une possible perte des données (« Il y a quelques années dans les années 90 on a fait un énorme travail de collecte. Je ne sais pas ce que chacun en a fait, s'ils ont gardé ou jeté comme moi. »).

La question du stockage et de l'archivage de ces données produites à plusieurs reste donc un **impensé** dans la plupart des cas (« "Je voulais voir avec vous le circuit de vos données, une fois que le projet est terminé. Avez-vous un circuit d'archivage ?" / "C'est une bonne question. Il n'y a pas de circuit d'archivage à part ce qui existe dans nos ordinateurs et sur notre serveur donc elles restent sur les bases de données chercheurs on va dire »), même si quelques membres de l'équipe projet peuvent être chargés officieusement de ce suivi (« Je me retrouve à être le référent pour ces trucs là car c'est moi qui maîtrise le mieux la bécane »).

III. Partager ses données : une réflexion en cours

Le troisième volet de l'entretien portait sur la position des chercheurs à propos de l'ouverture de leurs données de la recherche. Premier point d'étonnement : nous n'avons rencontré **aucun chercheur fermement opposé à cette pratique**, une réaction qui était peut-être cependant induite par le format même de l'entretien qualitatif et la présence, en la personne de l'enquêteur, d'un représentant du SCD qui a pu les pousser à nuancer leur position. Cette position a cependant été retrouvée de manière similaire dans l'enquête qualitative menée à Rennes et auprès des personnels d'appui interrogés dans le cadre de l'enquête. Les pratiques de partage restreintes en SHS sont donc moins issues d'une position de principe que de méthodes de travail et de diverses contraintes exprimées par les chercheurs au cours des entretiens.

a) Partager ses données en SHS : une spécificité ?

Si la question des pratiques de partage par les chercheurs n'a pas été abordée frontalement en tant que telle – à la différence des questionnaires en ligne produits par les universités des Rennes et Lille, plusieurs éléments sont ressortis au fil de la conversation qui ont permis de mettre en avant différentes conduites.

Avant d'aller plus en avant, il nous semblait tout d'abord important de rappeler que si la notion de données de la recherche – et a fortiori leur partage, reste un impensé pour beaucoup de chercheurs de l'UBM, **plusieurs d'entre eux ont déjà adopté, sans s'en rendre compte, une pratique de réutilisation de corpus produits par d'autres chercheurs, notamment via l'utilisation de corpus en ligne** : les index et thésaurus, les bases de données ou encore les portails iconographiques mis en place par d'autres universitaires sont ainsi régulièrement cités comme sources.

Malgré tout, le fait de partager ses propres données de la recherche n'est pas encore généralisé : la notion d'individualité des recherches, évoquées spontanément par dix-huit répondants, se retrouve logiquement dans **l'absence, a priori, de pratique de partage** (« *Déjà, il me semble que c'est assez spécifique au champ de domaine en études littéraires, c'est quand même un travail individuel. Sans parler du partage des données, le partage des réflexions, des méthodes est en fait assez peu présent* »).

Lorsque le chercheur échange ses données, c'est logiquement ensuite au sein de son groupe projet que cette pratique se déroule (« *Par exemple des données produites collaborativement, à plusieurs, de toute façon on va être bien contraint à partager* ») voire à l'échelle du laboratoire. Le principe de partage de manière plus générale à une communauté plus vaste est cependant loin d'être acquis même si une **minorité de chercheurs la revendique explicitement**, notamment pour une question d'éthique de la recherche (« *J'essaye d'inscrire au maximum mes travaux dans une logique de science ouverte, de partage des données parce que je suis aussi sur un terrain où il y a une méfiance très forte vis à vis des chercheurs* »).

Cette discussion sur le partage des données de la recherche a souvent pu être ouverte, dans un premier temps, en les interrogeant sur leurs pratiques d'Open Access, un échange qui a notamment permis de montrer que la rareté de ces pratiques est dû à un manque de temps du côté des chercheurs mais également par peur de manquer de visibilité sur les outils recommandés (« *Il n'y a pas de visibilité [sur HAL] sur combien il y a eu de lecteurs. En revanche j'ai de la visibilité tout de suite sur Academia et ça*

c'est important »). Ces contraintes seront sans doute applicables au partage des données de la recherche, ce qui peut déjà amener à réfléchir à un moyen de les dépasser.

b) Le partage des données de la recherche en SHS : des limites internes à la discipline ?

Plusieurs chercheurs ont souligné leur désarroi face à la question du partage des données de la recherche : s'ils ne sont pas opposés sur le principe, ils y voient **un « impensé » en SHS** (*« Déjà tout le monde n'en a pas, on est une minorité à avoir un produit intermédiaire entre le rien et la production finale donc le partage n'est jamais discuté »*) ou une pratique venue des sciences exactes qui ne trouverait pas son application dans le cas de données de la recherche plus qualitatives (*« peut-être qu'en sciences dures c'est pertinent de raisonner comme ça, il y a peut-être des données qui sont diffusables numériquement, pour nous c'est ... sans objet »*).

Ainsi, pour plusieurs chercheurs, le partage des données serait avant tout pertinent dans le cadre de **données quantitatives de recherche** : soulignant alors l'absence de telle production personnelle par manque de temps, ils sont plusieurs à pointer en revanche l'intérêt des gros corpus constitué par les doctorants au cours de leur thèse (*« mais ce que je trouve intéressant c'est pour nos doctorants, nos jeunes docteurs qui ont tous des données qu'ils n'ont pas publiées ou qu'ils ne publieront pas parce qu'ils ont fait des thèses de près de 2.000p et qui ont des données personnelles importantes »*).

L'autre limite soulignée par les répondants sur une pratique de partage provient du **rapport subjectif des chercheurs en SHS par rapport à leurs corpus** : la pertinence de la mise en ligne d'un corpus déjà biaisé, car sélectionné, a ainsi pu être questionné (*« C'est ça, il faut voir si ce sont les photographies, l'archive brute que l'on met directement ou si ça passe par notre filtre. Moi j'élimine, je ne prends que ce qui m'intéresse donc forcément il y a une sélection, un tri qui passe... »*).

Cette relation très étroite du chercheur à son corpus implique une deuxième limite au partage : la **réutilisation de ces corpus par le reste de la communauté scientifique serait loin d'être systématique**, le chercheur manquant d'informations pour comprendre comment le corpus a été constitué et pouvoir l'analyser avec un œil critique (*« Ce qui me gêne dans ce cas-là c'est que je ne sais pas comment ces bases de données ont été constituées et dans mon travail d'historienne ça me gêne tout de suite. Dès qu'on veut faire des statistiques, si on n'a pas une idée de comment ça été sélectionné, de la complétude ou non du corpus... »*). Autre corollaire, certains répondants ont également exprimé leurs doutes sur une possible réutilisation de leurs corpus mis en ligne par d'autres chercheurs, pointant cette fois-ci la nécessité dans plusieurs disciplines de construire son propre corpus pour être légitime en tant que scientifique (*« Le chercheur a besoin d'avoir des choses qu'il a réalisées lui-même »*).

Enfin, la contrainte sans doute la plus mise en avant est le **manque de temps** dévolu aux chercheurs qui soulignent bien souvent leurs difficultés à dégager des moments pour publier en ligne en premier lieu leurs articles, la question des données de la recherche leur apparaissant secondaire (*« La base de données créée à Paris 1, elle prend du temps pour mettre à disposition des autres un instrument mais c'est vrai que le temps qu'elle se retire dans son propre emploi du temps, elle produit rien elle-même. Parce qu'on est jugé sur ...A la fin se sont les publications qui comptent »*).

c) Les arguments en faveur de l'ouverture des données de recherche

Malgré ces limites, de nombreux chercheurs ont exprimé leur intérêt pour le partage des données, y voyant plusieurs avantages pour la recherche en SHS.

Les plus engagés d'entre eux dans le mouvement de science ouverte ont tout d'abord une **position éthique** vis-à-vis de cette question, rappelant que la recherche étant financée par les fonds publics, il serait naturel de diffuser les productions au plus grand nombre (« *Donc moi ça ne me pose aucun problème et par ailleurs je considère que tout ce que l'on produit dans le cadre de notre mission d'agent public, évidemment c'est ...* »).

La plupart des répondants ont surtout vu dans le partage des données un intérêt personnel, en percevant l'ouverture des données de la recherche comme un moyen **d'augmenter leur visibilité**, sur un plan scientifique (« *C'est un classique. Disons qu'il y a forcément une recherche de visibilité vis à vis de nos pairs scientifiques et aussi à un moment donné ce qui est important c'est aussi* »).

Cependant, le principal argument avancé en faveur de l'ouverture des données de la recherche a été celui de **l'émulation scientifique**. De nombreux chercheurs de l'UBM ont ainsi conscience de l'importance de leurs gisements de données, qui, une fois, dépouillés, restent inutilisés (« *J'ai collecté 30 versions manuscrites et là je me retrouve donc avec une masse de photos et je me dis que ça pourrait être utile de trouver un moyen de mettre à la disposition de la communauté scientifique l'ensemble des manuscrits disponibles. Mais comment faire ? Avons-nous des structures ?* »), certains trouvant cette accumulation de données froides particulièrement contre-productive (« *Plus on produit plus on oublie, j'ai envie de dire, c'est un peu bizarre [...] A la cave, il y a des masses de cartons de chercheurs partis à la retraite, on a une vingtaine de cartons à stocker avec toutes ces enquêtes, données, bases et dont on fait rien, zéro* » ou « *La production de données c'est aussi ne pas perdre les données produites* »).

Ce partage des corpus permettrait notamment de faciliter la **collaboration à l'international**, en favorisant par exemple l'accès à des collègues à des documents archivés à l'étranger (« *Il y a des sources en français conservées là-bas mais c'est compliqué, ça a un coût et c'est un peu intimidant pour les chercheurs qui n'ont pas l'habitude et il me semble que j'ai un rôle de passeur à jouer pour permettre à toutes les personnes qui s'intéressent à mon thème d'accéder à ces sources* »). Une telle émulation serait bien évidemment bénéfique et permettrait de faire avancer la recherche plus rapidement, notamment dans le cadre de chercheurs travaillant sur des sujets très spécifiques pour lesquels la participation d'autres scientifiques serait la bienvenue (« *Moi dans mon domaine, vu que cela n'intéresse personne, je n'ai aucune crainte, au contraire je les donne pour qu'on travaille à plusieurs ensembles* »).

d) Les conditions posées par les chercheurs de l'UBM pour une ouverture des données

Tirillés entre leur intérêt pour le partage des données de la recherche et leurs préoccupations sur la bonne réutilisation de ces données, les chercheurs ont donc pointé du doigt plusieurs pistes de réflexion pour mettre en place une politique d'ouverture des données adéquate.

Une première restriction posée à l'Open Access de manière quasi collective par les chercheurs a été de rappeler que le dépôt des données de la recherche en ligne ne pourrait s'effectuer qu'après publication ou à la fin du projet afin d'éviter toute tentative de plagiat (« *Donc ouvert au partage si cela intervient en bout de course, après une publication...* »), soulignant également que ces corpus en ligne devaient être bien attribués afin que d'autres ne se réapproprient pas leur travail (« *Mais il faudrait qu'il y ait une structure qui permette de tracer* »). Un seul répondant a évoqué la possibilité de partager ces données en cours de projet, ces dernières pouvant alors servir de levier auprès des institutions partenaires.

La question de la réutilisation de ces données interroge encore beaucoup et les répondants se sont donc préoccupés de la fiabilité des données de recherche mises en ligne, signalant la nécessaire contextualisation de ces données pour permettre une réutilisation adaptée (« *Si on faisait un espace*

de dépôt de données réutilisables, il faudrait vraiment le baliser pour qu'il y ait pas ce risque de se retrouver avec une dynamique contradictoire par rapport à ce qui a été produit initialement et les conditions de la production. On pourrait avoir des contresens phénoménaux sur l'analyse de ces données »). Ce travail descriptif du corpus mis en ligne représenterait alors une tâche supplémentaire pour les chercheurs, qui ont pu se montrer réservés sur cette question (« Si je donne le fichier brut sans donner les informations qui vont avec, la personne va se dire : "tiens c'est bizarre, je ne trouve pas le même résultat". Donc plein de petites choses qu'il faudrait pouvoir fournir avec ces fichiers et ça peut vite devenir très lourd. »).

Les chercheurs ont également identifié une contrainte externe pour le partage de leurs données, soulignant la nécessité de **reformaté certaines données** pour pouvoir les ouvrir de manière convenable (« *Pas vraiment, pas du tout par conviction, je crois vraiment qu'ouvrir les corpus c'est extrêmement important. Le seul qui s'y prêtait vraiment finalement, c'était plus par manque de formatage, il n'était pas conçu pour un usage grand public, ce n'était pas des données très lisibles en elles-mêmes et ça avait été des données codées pour un usage spécifique* » ou encore « *le support me pose problème. Comme tout est sur FileMaker, comment basculer ça sur Internet ? et on s'est retrouvé avec le même problème avec les projets de recherche ...* »).

Les chercheurs se sont également posé des questions sur les conditions concrètes de cette ouverture : que met-on en ligne ? Qui y a accès ?

Les postures sont en effet très variées sur le sujet : si les données papiers ne sont pas, dans leur grande majorité, prises en compte en raison notamment de leur illisibilité (« *Sur le principe oui mais le type de sources que j'ai, ce n'est pas possible donc soit c'est papier et écrit par moi ça n'a aucun intérêt surtout qu'avant j'écrivais à l'encre et l'encre s'efface et aujourd'hui j'ai plein d'archives qui sont illisibles de ma thèse* »), les données sous format numériques font plus débat.

Ainsi, si certains envisagent de partager leurs corpus de sources collectées, leurs archives brutes (« Il y a 10 ans on s'est posé la question de savoir comment on pouvait mettre tous les dépouillements que nous faisons, tous, en commun en rêvant d'un truc collaboratif dans les archives »), d'autres sont opposés à ce paradigme, arguant du temps qu'ils ont passé à rassembler ce corpus, mais restant ouverts à l'idée de partager leurs données dérivées (« *Ca me poserait pas plus de problème que ça mais je reconnais que sur les photos d'archives je serai plus ambivalente ... De façon très égoïste, je me suis tapée des après-midi entiers à tout photographier [...] mais je ne veux pas que ça serve de truc un peu facile, bon ... Je reconnais ma propre mesquinerie* »).

La ligne du partage entre ces deux positions réside entre deux conceptions différentes de l'activité de recherche, comme l'a souligné un répondant : « *Pendant très longtemps, la recherche c'était aussi trouver la donnée [...] Mais pour moi c'est pas ça la recherche, le travail du chercheur, c'est pas de trouver la donnée brute, il faut qu'elle soit immédiatement accessible, le travail du chercheur c'est de l'interpréter, de l'analyser ... il y a aussi une évolution du métier.* ».

Se pose enfin la question du **public concerné** par la publication de ces données et de la même façon, les réponses varient beaucoup d'un chercheur à l'autre : si la plupart des répondants n'envisagent cette publicité des données qu'à l'échelle de la communauté scientifique ou même à l'intérieur d'un projet (« *Je suis d'accord pour les partager à condition que la personne qui va les utiliser s'intègre au projet ... [...] Mais que ça soit 100 % public je pense pas que ça soit une bonne chose* »), plusieurs chercheurs étaient également ouverts à l'idée de partager les données à la communauté civile (« *Moi je préfère librement accessible, de toute manière c'est toujours compliqué quand c'est d'accès restreint. Quel type de public on va avoir, quel type de restriction ? Je ne trouve pas ça très satisfaisant, ça complique les choses ... Soit on restreint le groupe de recherche et dans ce cas-là c'est un outil de travail,*

ce qui est bien. Souvent la première étape c'est ça de toute façon. Soit on ouvre vraiment, je ne vois pas vraiment d'intermédiaire ») soulignant d'ailleurs pour plusieurs d'entre eux que ces derniers étaient les premiers à réutiliser leurs travaux (« Nous par exemple, il y a eu beaucoup de réutilisation d'analyses qui ont été faites mais jamais par la sphère scientifique. Toujours par des contributeurs d'Open Street Map, des guides, des statisticiens qui ont rebidouillé des données, ont fait des choses extrêmement intéressantes alors que les collègues scientifiques dans le même domaine vont préférer refaire une enquête, refaire un truc ... »).

E) De nombreuses interrogations autour du droit des données de la recherche

Les chercheurs se posent également beaucoup de questions autour du **droit des données** qu'ils pourraient diffuser, point abordé de manière spontanée par les chercheurs traitant avec des **données individuelles**, issues notamment d'entretiens. Conscients de la nécessaire anonymisation de leur corpus avant la mise en ligne, en lien avec la nouvelle réglementation européenne, ils se montrent donc prudents sur la question et avancent leurs besoins de formation en la matière (« *même pour ces données-là, les ouvrir davantage pourquoi pas mais en respectant la loi donc nécessairement anonymisation ce qui n'est pas forcément évident car on ne dispose pas forcément des compétences pour anonymiser par exemple ... La question de la voix par exemple : le débit, l'intonation, la tessiture ... C'est déjà une donnée personnelle donc on ne peut pas ouvrir ça parce que ça va renseigner de manière intrusive sur la personne »).*

Si toutes les données produites à l'UBM ne répondent pas à la même confidentialité, d'autres limites légales ont également pointées par les répondants : le **droit d'auteur** pour des textes inédits (« *Autre raison, des documents que je ne veux pas mettre sur le Cloud par principe par exemple des textes écrits par d'autre et dont on est le seul dépositaire. Un manuscrit qu'il ne publiera pas avant 10 ans ») ou le **droit d'image** pour les documents graphiques (« *En revanche, pour l'illustration on a encore des ayants droits très attachés à l'image de la marque, de la propriété intellectuelle »).**

Les chercheurs ont également mentionné le **droit des archives**, à plus ou moins bien escient : si certains évoquaient le caractère restrictif des archives récentes pouvant contenir des données personnelles (« *Au début je me disais : les gens sont morts et puis après j'ai pensé aux descendants et à la question de l'anonymisation [...] on ne peut pas ouvrir les choses comme ça il y a des enjeux d'anonymats ») ou le statut des archives privées (« *Et un autre problème se pose : je vais dans les fonds privés donc souvent les photos que je prends c'est reproduction interdite donc ça c'est impossible de le mettre sur un site, c'est pour mes propres recherches, on ne peut pas céder le droit à quelqu'un »), le statut des archives publiques est en revanche parfois mal connu (« *C'est à dire que si c'est donner des photos mais si je donne l'ensemble de mes photos à la BU, est-ce qu'il y aurait une personne chargée de contacter chaque institution dans lesquelles j'ai pris des photos pour leur dire : on crée un fonds, on le met en ligne »).***

Enfin, la nécessité de préserver certains **sites archéologiques** du pillage a également été pointée par des spécialistes pour justifier de ne pas divulguer toutes les coordonnées géographiques (« *les services archéologiques sont très sensibles sur la question de la divulgation de certains sites archéologiques qui ne sont pas protégés et qu'il ne vaut mieux pas faire connaître, par rapport aux détecteurs de métaux par exemple »).*

IV. Les besoins des chercheurs de l'UBM

Les entretiens menés avec les répondants ont également eu l'avantage de faire remonter plus facilement leurs besoins, soit qu'ils les aient exprimés de manière claire ou suggérés à travers les difficultés et problématiques rencontrées au quotidien. Les chercheurs ont par ailleurs régulièrement

abordé les problèmes liés à la gestion de leur bibliographie ou de leur boîte mail mais ces points ne rentrant pas dans la problématique de la gestion des données de la recherche, nous nous contenterons de les signaler sans plus nous y appesantir.

Les chercheurs ont de fait exprimé leurs besoins sur deux principaux sujets : la mise en place d'un appui technique et l'importance d'un accompagnement.

a) *A la recherche d'outils adaptés*

Les chercheurs ont ainsi signalé la nécessité de disposer de plateformes ou de serveurs adaptés pour répondre à une double problématique : la **saturation des espaces de stockage personnels d'une part et la nécessité de centraliser sur un même espace de stockage les données collectées et produites par un même chercheur d'autre part** (« *Ça manque, ça manque énormément effectivement, un lieu de stockage à la fois fiable et qui fasse pas partie des propriétaires mais c'est vrai qu'avec tous les chercheurs et toutes les données produites il faudrait un gros espace de stockage, mais aussi la nécessité de centraliser des données disséminées dans un espace sécurisé* »).

Ce souhait s'inscrit de manière plus générale dans la volonté de gérer plus facilement cette masse de données de recherche à l'échelle du laboratoire, dans les cas où l'importance des données de la recherche et leurs potentialités ont bien été comprises (« *Plusieurs chercheurs dans d'autres disciplines rencontrent des difficultés et j'aimerais bien qu'on puisse être une force de demande, de propositions pour dire : est-ce qu'on peut nous fournir des outils pour gérer nos données ?* »).

Dans les faits, ce souhait générique s'accompagne parfois de **préconisations plus précises** :

- La modularité des **modes d'accès aux datasets** est ainsi à prendre en compte (« *oui je serai favorable à une politique à l'échelle de l'Université, surtout si on trouve, suffisamment en tout cas pour être exprimé dans un univers partagé intelligemment avec des passerelles, passwords* »),
- la possibilité de pouvoir accéder aux données de **manière déconnectée**, pour s'adapter aux chercheurs travaillant sur des terrains éloignés (« *Quand on fait du terrain, [...] dans des zones où il n'y a pas forcément de connexion internet et où on a besoin parfois d'avoir avec nous des matériaux de recherche pour pouvoir sur le terrain retrouver des éléments qu'on a pu voir* ») ou ne souhaitant pas, et ils sont plusieurs, déposer leurs données sur un Cloud par crainte d'être piraté⁸.
- De manière plus générale, la volonté d'avoir un **outil souple et facile d'utilisation** (« *Avant on avait un serveur qui n'était sans doute pas très bien normé mais qui nous permettait très facilement de ... Il y a aussi cette question, est-ce qu'on met en place des usines à gaz, qui répondent à toutes les normes mais sont absolument inutilisables ou est-ce qu'on fait des choses pratiques* »)
- La question des **métadonnées et de l'indexation** des corpus qui doivent pouvoir être adaptés aux pratiques de recherche (« *C'est assez difficile de se mettre d'accord avec l'éditeur sur les mots clés qui vont sortir dans une base de données. Ça c'est un problème. La manière de concevoir les métadonnées, parce que c'est un système information et qu'on fonctionne par mots et non par phrases, est-ce que ce n'est pas déjà une forme d'appauvrissement ? C'est compliqué de trouver le bon système de référencement. [...] C'est finalement des systèmes qui sont moins souples.* »).

⁸ Cf. III.a

Les chercheurs imaginent par ailleurs plusieurs formes à un possible espace de stockage des données : plateforme accessible à l'échelle du laboratoire ou de la communauté scientifique, utilisation des ressources d'Huma-Num ou encore utilisation du récent Data Center installé à Bouliac pour archiver les données de manière fiable.

La question de la visibilité de ces corpus, souvent disséminés entre plusieurs entrepôts, se pose toutefois ce qui a amené plusieurs chercheurs à proposer la mise en place d'un **portail pour favoriser la visibilité des productions au sein de l'UBM** (« Sans doute un portail ou un point d'accès en ligne clairement identifié avec des interlocuteurs, une chaîne de dépôt classique qui permettrait le dépôt avec des métadonnées sérieuses (qui dépose ? quels sont les autres auteurs ? types de données ?) , un jeu de métadonnées pour un dépôt jusqu'à une certaine charge et puis un référent, une référente pour des données plus complexes, ne serait-ce que du multi-fichier »).

Plusieurs historiens ont également formulé le souhait de voir émerger la mise en place d'une **plateforme de partage permettant de diffuser les nombreuses photographies faites en centre d'archives** (« Il y a 10 ans on s'est posé la question de savoir comment on pouvait mettre tous les dépouillements que nous faisons, tous, en commun en rêvant d'un truc collaboratif dans les archives... Dans les archives on fait tout un travail qui ne reste que dans nos ordinateurs et on s'était interrogé à plusieurs sur comment le mettre en commun »).

Dans le cadre de projets financés, l'autre besoin qui a émergé est celui de disposer d'un outil adapté pour **partager les données en mode collaboratif** (« La question de la place qui est toujours problématique et le fait surtout, de travailler à distance, de pouvoir échanger les données, sans envoyer tout de suite un mail très lourd et risque d'encombrer une boîte »). De fait, les limites des outils actuellement disponibles ont été pointées par plusieurs répondants : opérateurs privés, échange restreints aux chercheurs d'une même institution et compliquant de fait les collaborations à l'échelle nationale ou internationale... Toutes ces limites ont ainsi été très bien résumées par le répondant suivant :

« J'essaie et puis c'est bien que nos institutions essaient de nous offrir des sites pratiques mais l'équivalent de DropBox j'ai pas trouvé. Il y a My CoRe qui marche mais avec des gens du CNRS, qui ont envie de s'en servir [...] Je sais que j'utilise aussi sur un projet un équivalent DropBox de l'Université, NextCould, ça ça marche très bien mais je ne peux échanger qu'avec des gens qui ont une adresse professionnelle ici. ».

b) De nombreuses demandes pour la mise en place d'un accompagnement autour des données de la recherche

[Des référents en matière de données de la recherche](#)

Le deuxième point crucial au-delà de l'aspect technique, qui était parfois difficile à imaginer pour les répondants, est surtout le **besoin d'accompagnement et de formations de manière courante autour de la gestion des données de la recherche**. Plusieurs gradations peuvent ainsi être distinguées autour de la forme que prendrait cet accompagnement, selon le degré d'autonomie plus ou moins poussé des chercheurs.

Ainsi, tandis que certains sont surtout intéressés par la mise en place d'une cellule de **veille** autour de ces questions (« Ça serait très très intéressant d'avoir cette veille non pas réglementaire mais d'outils mis à disposition. Ça existe sûrement, Huma-Num le fait peut-être mais il y en a tellement dans tous les sens qu'on est pas tout le temps vigilant et ça ça serait intéressant de pouvoir repérer des outils et signaler des outils qui pourraient répondre à nos besoins »), d'autres, plus nombreux, ont explicitement signalé leur besoin de **référent** en matière d'archivage ou de gestion des données de la recherche («

Après je pense que c'est des problématiques qui vont se poser au SCD à un moment ou un autre et il y aura forcément besoin à un moment qu'il y ait un archiviste au niveau de l'établissement qui fera ce travail de manière transversale »).

La question de l'accompagnement des chercheurs pour le dépôt de leurs données de recherche se pose également au vu des difficultés que peuvent déjà rencontrer les chercheurs dans le dépôt d'articles en *open access* (« *C'est plus que ça prend du temps. Mais j'ai des collègues qui sont pas des manches mais ont des problèmes d'identifiants, d'accès que moi je n'ai pas mais je sais que certains en ont et ce ne sont pas des handicapés de HAL* »).

Il paraît également essentiel de proposer un accompagnement plus poussé pour les chercheurs en cours de **montage de projet, ou venant d'obtenir des financements**, afin de résoudre tout à la fois les questions de partage en mode collaboratif évoqué ci-dessus mais aussi pour prodiguer des conseils pour les chercheurs ne disposant pas à demeure d'un appui technique (« *A l'avenir je formaliserai tout par écrit, je ferai une convention de recherche au préalable en définissant les tâches, le statut de la propriété intellectuelle. Ça je l'ai pas fait* »).

Des nombreuses demandes de formations

Enfin, au-delà du rôle de référent que peut jouer le service « données de la recherche » de l'UBM, plusieurs besoins en terme de formations sont également ressortis.

La question qui taraude sans doute le plus les répondants est celle des **conditions légales** autour de la diffusion de leurs données de recherche : ainsi, suite logique des obstacles identifiés par plusieurs d'entre eux pour diffuser librement leur données de la recherche⁹, les chercheurs ont souvent répété leur ignorance des obligations légales et en sous-texte, leur besoin d'éclaircissement et de réponses claires à ce sujet (« *en fait on n'est pas du tout sensibilisé à ces questions juridiques* »), notamment autour de la RGPD, du droit des archives et du droit des images.

Les chercheurs ont par ailleurs conscience de leur manque de formation autour des outils informatiques (« *On a toujours besoin d'une formation logiciel* »), réalisant que leur manque de familiarité à certains logiciels freine leur recherche :

- Utilisation des **outils bureautiques** (« Sinon on manque de personnel technique, ça peut aller du traitement de texte, des potentialités du texte du traitement de texte que souvent on ne maîtrise pas complètement. Je pense qu'on pourrait bénéficier de plus d'expertise dans ces domaines »)
- Utilisation d'un **logiciel de base de données** (« *une formation à la fois pour la montée ou pour le maniement de la base de données serait la bienvenue* » ou « *ici on n'arrive toujours pas à trouver un espace pour dispenser une formation informatique adaptée à l'histoire, autre chose que du traitement de texte ... On n'arrive pas à dégager d'espace et des compétences. [...] On bidouille sous FileMaker de manière artisanale* »)
- Utilisation d'un **logiciel d'analyse de données** (« *J'ai vu des chercheurs travailler aussi avec des logiciels de lexicographie, d'analyse de discours, je trouve ça passionnant même si j'ai pas eu moi l'occasion de le mobiliser* »).

Ce chercheur résume ainsi très bien les différentes attentes sur le sujet :

« *Des formations sur Wordpress, pour gérer le site Internet. Des formations sur la gestion des bibliographies, j'ai déjà fait une formation sur Zotero mais comme je n'ai pas pratiqué depuis, j'ai tout*

⁹ Cf. IV.3

perdu et j'aimerais refaire. J'aimerais faire une formation sur Alceste par exemple donc tout ce qui est exploitation des données ça serait intéressant ». Ils ont par ailleurs été plusieurs à préciser que ces formations étaient à privilégier à l'échelle du laboratoire, sous la forme de petits groupes, afin de mettre en place des réponses personnalisées (« *Ce qui n'était pas mal c'était la séance sur HAL avec la bibliothèque mais aussi parce que toute l'équipe en avait besoin. S'il y a formation, il faut qu'elle soit pensée de manière très très ciblée comme ça, quasi individuelle en fait »*).

Enfin, de manière implicite, plusieurs chercheurs ont également signalé leur besoin d'être sensibilisés aux **bonnes pratiques d'archivage** afin d'améliorer leur gestion quotidienne :

- Sur une question de **choix** tout d'abord (« *En même temps c'est toujours le questionnement des archives : est-ce qu'il faut vraiment tout garder ? Est qu'il n'y a pas une sélection à opérer ? Sur quels critères ? »*)
- À travers plusieurs récits de pertes de données, sur la manière de **conserver de manière sécurisée ses productions** (« *J'ai eu un accident une fois, j'ai mis tous mes fichiers sur mon Bureau puis à la corbeille [... Je les ai récupérés mais il a fallu que je trie 3.000 enregistrements [...] Le lendemain, mon ordinateur m'a lâché et j'avais un disque dur externe qui m'a lâché en même temps »*).
- Prise de conscience de la migration nécessaire de certaines données de la recherche vers un **format plus lisible afin de garantir leur pérennité** (« *Mon premier souci c'est que je dois voir avec l'UBM, la cellule audiovisuelle, pour qu'il me fasse des copies de mes entretiens vidéos. C'est un vieux format en HI8 et ça doit être transcrit mais je ne l'ai pas fait, ça fait 3-4 ans qu'il faut que je le fasse »* ou « *Moi j'utilise aussi au compte-goutte une base de données réalisée par un de mes anciens collègues. Avant de partir, il a téléchargé sa base de donnée sur un vieil ordinateur, que je surveille comme la prune de mes yeux. Je l'utilise de temps en temps et ponctuellement pour les collègues »*).

Partie II : Des pistes de réflexion pour la création d'un service « données de la recherche » à l'UBM

A la suite de ce compte-rendu détaillé des entretiens et de leurs principaux renseignements sur les pratiques de la recherche à l'UBM, il nous paraissait pertinent de terminer ce rapport en évoquant des perspectives à envisager pour la mise en place d'un service « données de la recherche ».

I. Etablir des profils types pour améliorer la compréhension de la communauté des chercheurs de l'UBM

Les entretiens qualitatifs fournissent une matière extrêmement riche à exploiter mais parfois difficile à synthétiser dans la perspective d'une mise en place de service puisqu'il y a presque autant de réponses possibles que d'individus. Pour faciliter le repérage des acteurs clefs et de leurs besoins, nous avons donc fait le choix d'établir à la suite de cette enquête plusieurs profils- types de chercheurs, indépendamment de leur discipline, de leur unité de recherche ou de leur statut, ces critères s'avérant peu opérants à l'UBM. Cette classification par profil a donc été faite selon les besoins des répondants en matière de stockage et de partage de leurs données de recherche.

Il est bien entendu évident que tous les chercheurs ne se répartissent pas de manière stricte dans ces groupes et ces typologies sont donc plus à regarder comme des grandes lignes de synthèse.

a) *Les chercheurs menant un projet ou l'ayant récemment terminé*

La préoccupation principale de ce groupe est de pouvoir collaborer en mode projet : indépendamment de leurs disciplines, ces chercheurs sont caractérisés par leur implication dans des projets financés, qu'ils soient en fin de projet ou au contraire lors de sa genèse. Manquant d'infrastructures techniques et de référents en la matière, ils ont de multiples besoins :

- Nécessité d'une plateforme pour stocker et partager les données à l'intérieur de l'équipe projet
- Formalisation d'un circuit d'archivage des données de recherche produites dans le cadre du projet
- Souhait de rendre visible leurs travaux, à la fois à l'internationale mais aussi auprès de la communauté civile

Plusieurs actions peuvent donc être montées pour les aider au mieux dans la gestion de leurs données : **accompagnement** aux projets de recherche, autour des **plans de gestion de données** notamment ; aide au **choix de plateformes** adaptées pour le **stockage et le partage des données** ; **aide à la valorisation** une fois le projet terminé, par **l'hébergement des bases de données** et corpus créés ainsi que leur signalement.

b) *Les chercheurs bénéficiant d'un soutien technique*

Disposant bien souvent de personnel spécialisé et d'un circuit de stockage et d'archivage de données de la recherche en cours de création ou déjà opérant, les priorités de ces chercheurs diffèrent sensiblement et se concentrent donc plutôt autour du partage des données de la recherche, afin d'aller plus loin dans le principe d'Open Data. Plusieurs besoins ont ainsi pu être identifiés :

- Améliorer la visibilité du laboratoire à travers le partage des données de la recherche
- Préserver les travaux des chercheurs en fin de carrière et notamment leur documentation papier

Auprès de ce groupe, le service « données de la recherche » peut donc travailler dans deux directions : accompagner les chercheurs dans le choix d'une plateforme de partage pertinente et dans l'utilisation de formats interopérables et réfléchir à la prise en charge des archives papier qui restent un impensé.

c) Les chercheurs préoccupés par la sauvegarde de la mémoire du laboratoire

Ce groupe, caractérisé par la prépondérance des historiens et civilisationnistes, se préoccupe plus particulièrement de la pérennité des productions effectuées à l'intérieur de l'unité, ayant notamment conscience des potentialités des corpus quantitatifs et de leur réutilisation. Leurs besoins se font sentir sur plusieurs points :

- Volonté de faire migrer les bases de données déjà produites pour assurer leur pérennité
- Volonté de mettre en valeur les corpus quantitatifs, notamment ceux issus des travaux des étudiants et doctorants
- Pour les plus avancés sur la question de l'OA, une volonté enfin de centraliser les corpus de sources primaires

Les efforts du service « données de la recherche » peuvent alors se porter sur plusieurs points : aide à la recension des bases de données déjà produites et susceptibles d'être illisibles dans les années à venir ; accompagnement des UR dans la migration de ces productions ; et enfin réflexion à un moyen de gérer et valoriser les corpus photographiques.

d) Les chercheurs peu concernés par la question des données de la recherche

Ce sous-groupe est constitué par des chercheurs aux pratiques souvent très individuelles : peu concernés par la question du partage des données, ils ne rencontrent en outre pas de difficultés à stocker leurs données de la recherche, ces dernières étant peu volumineuses ou majoritairement sous format papier. Par ailleurs, ce sous-groupe a tendance à dévaloriser sa production, se posant notamment la question de l'intérêt que ces données pourraient avoir pour d'autres personnes.

Le travail du service « données de la recherche » auprès de ce type de chercheurs serait donc dans un premier temps de les sensibiliser à l'intérêt des données de la recherche, leur bonne gestion et diffusion pouvant être un point clef pour l'élaboration de nouvelles pistes de recherche. Cette enquête aura cependant sans doute été une première porte d'entrée autour de ces questions.

e) Un profil sous-représenté mais à exploiter : les jeunes chercheurs

Enfin, et même s'ils ont été peu représentés au sein de l'enquête, un groupe se distingue par l'homogénéité de ces pratiques, en la personne des jeunes chercheurs. Ayant naturellement recours aux outils informatiques, ces derniers sont en premières ligne dans l'exploitation de corpus quantitatifs et ont souvent été sensibilisés à l'importance de l'Open Access. De plus, ces derniers se montrent demandeurs de formations autour de plusieurs questions : logiciels d'analyse de données, *datavisualisation* ou encore archivage des données du web. Ces différentes caractéristiques font donc des corpus créés ou exploités par les jeunes chercheurs autant de données susceptibles d'intéresser le reste de la communauté scientifique ou civile.

Le service « données de la recherche » pourrait ainsi mener plusieurs actions à destination des jeunes chercheurs afin de valoriser leur travail : initiation aux bonnes pratiques de stockage des données dès le début de leur recherche ; réflexion sur une diffusion adaptée de leurs données de recherche garantissant leur propriété intellectuelle : mise en place de différentes formations autour des logiciels d'exploitation de données.

II. Accompagner les chercheurs dans la gestion de leurs données de la recherche

a) *Recommandations techniques*

Etat de l'art : plateforme d'archivage et de valorisation des données de la recherche

Avant de procéder à des premières recommandations, il paraissait essentiel de faire une veille approfondie sur les outils déjà disponibles autour des données de la recherche en SHS. En effet, il a déjà été démontré que monter sa propre plateforme autour de cette question n'apportait pas vraiment de plus-value, cette solution n'étant d'ailleurs pas plébiscitée par les chercheurs (« *tout le monde parle d'Open Science et que tout le monde veut faire sa petite plateforme, on démultiplie les dispositifs et finalement les données ne sont plus accessibles puisque personne ne réussit à les retrouver* »).

Nous avons donc mené une recension des plateformes déjà disponibles, en nous aidant du moteur spécialisé re3data et en sélectionnant des entrepôts de données de la recherche selon plusieurs critères : tout d'abord, ces espaces de stockage devaient être adaptés aux SHS en n'étant donc pas spécialisés dans la diffusion de données de recherche issus des sciences dures ; par ailleurs, ces entrepôts se devaient de respecter les principes FAIR ou FACILE, garantissant la visibilité, l'ouverture et la pérennité des données de la recherche ainsi que l'interopérabilité de leurs données de recherche. Enfin, ces entrepôts devaient être en français ou au moins en anglais pour faciliter le dépôt par les chercheurs de l'UBM. Cette série de critères nous a permis de dégager une petite vingtaine d'entrepôts. Les entrepôts ainsi sélectionnés se partagent entre les entrepôts disciplinaires, concernant aussi bien le champ archéologique, statistique ou sociologique, au nombre de douze, et les entrepôts multidisciplinaires au nombre de neuf.

Plusieurs critères sont à prendre à compte pour la sélection d'un entrepôt :

- Le **volume des données** doit tout d'abord être suffisant pour pouvoir héberger les données produites à l'Université Bordeaux Montaigne, plusieurs téraoctets étant donc nécessaire puisqu'une partie non-négligeable des chercheurs de l'UBM produisent une centaine de gigaoctets au cours de leur recherche comme nous l'ont appris les entretiens.
- Par ailleurs, et même si les données produites par les chercheurs de l'Université Bordeaux Montaigne sont principalement des documents textuels ou des images fixes, deux **supports** faciles à archiver et parmi les plus présents dans les entrepôts de données, ces derniers doivent être capable de prendre en charge d'autres types de données de la recherche régulièrement utilisées par les chercheurs de l'UBM, en l'occurrence les données sur support audiovisuel, les bases de données produites par les chercheurs et enfin des données 3D.
- La question du **coût** est également primordiale : si les conditions financières sont rarement détaillées, il faut veiller à privilégier une solution tout à la fois économique et pérenne et prendre en considération ses évolutions dans le temps, le coût de dépôt pouvant augmenter avec le volume des données déposées.
- Enfin, les chercheurs ayant manifesté des positions très variées sur l'ouverture de leurs données, il semble pertinent dans un premier temps de proposer un entrepôt dont l'accès à chaque jeu de données peut être **modulé selon les besoins**, pour prendre notamment en compte les contraintes législatives.

Plusieurs points, moins fondamentaux, peuvent également être pris en considération. Les chercheurs ont ainsi régulièrement exprimé leur volonté d'avoir des métadonnées souples, pouvant s'adapter aux pratiques de leurs disciplines et notamment au dépôt multi-auteurs. La modification des dépôts déjà effectués par l'ajout de nouvelles versions divise plus la communauté : tandis que certains n'y voient guère d'intérêt, soulignant même les problématiques que cela pourrait engendrer, cette option serait

au contraire intéressante notamment dans la communauté des historiens qui travaillent sur un temps plus long et peuvent être amenés à revoir leurs relevés.

Recommandations techniques par rapport aux pratiques et besoins des chercheurs de l'UBM

En gardant ces critères en tête, nous avons pu tirer plusieurs conclusions pour le stockage et le partage des données de la recherche produite à l'UBM.

Les outils mis en place par Huma-Num nous semblent particulièrement intéressants, ayant par ailleurs déjà été adoptés par plusieurs centres de documentation rattachés à des UMR de l'université. En effet, cette structure propose des outils pour chaque moment du cycle de la donnée :

- ShareDocs comme Cloud en ligne
- Huma-Num Box pour les données tièdes en grande masse
- Nakala comme espace de stockage avec un accès restreint avant une éventuelle éditorialisation sur Internet...

La TGIR Huma-Num assure de plus une pérennité des données de recherche dont elle a la charge puisqu'elle est associée au CINES pour l'archivage des données froides et propose également un catalogue, ISIDORE, s'imposant progressivement comme un outil de référence pour la recherche de données en libre accès en SHS (publications et données de la recherche), ce qui garantit une visibilité de l'UBM.

Huma-Num, s'il dispose d'un méta-catalogue avec ISIDORE, ne propose pas de plateforme de partage des données de la recherche à proprement parler et il est donc nécessaire que la valorisation des données de la recherche se fasse en utilisant d'autres portails. Deux options s'offrent alors : recourir à des entrepôts disciplinaire ou multidisciplinaire. Ces derniers, s'ils sont souvent plus connus des chercheurs, sont cependant très majoritairement utilisés par des communautés de chercheurs en sciences dures et/ou anglophones, ce qui ne favorise pas la visibilité de la recherche française en SHS.

Les entrepôts disciplinaires, déjà connus pour certains des chercheurs interrogés, semblent donc plus à même de répondre à leurs besoins. Certains en particulier peuvent être privilégiés, tel HAL pour le dépôt de logiciels, MediHAL pour le dépôt d'archives graphiques ou audiovisuelle ou encore Cocoon, conçus pour les enregistrements oraux et leurs transcriptions, tous trois moissonnés dans ISIDORE, ce qui garantit leur visibilité. Ces entrepôts disciplinaires ne résolvent cependant pas la valorisation de certaines données, produites principalement en histoire ou en esthétique, pour lesquelles aucun entrepôt répondant aux principes FAIR n'est en place ; cependant, la valorisation de ces corpus de données pourrait s'effectuer à partir d'autres plateformes comme les blogs Hypothèses ou diverses bibliothèques numériques (1886, Anpersana) déjà moissonnés dans ISIDORE.

Enfin, dans le cadre de projets européens, Zenodo apparaît également comme un partenaire de choix : cette infrastructure permet le stockage des données de la recherche à hauteur de 500 Go par jeux de données, leur accès aux membres d'un même projet puis leur diffusion une fois le projet fini, cette clause étant désormais obligatoire pour les projets H2020.

b) Mobiliser les chercheurs de l'UBM autour des données de la recherche

Stratégie de mise en place d'un service : quelle population cibler ?

Afin de mettre en place ce service « données de la recherche », nous proposons de nous servir des profils-types établis plus haut.

Dans l'immédiat, les **chercheurs en cours de projet** sont ainsi ceux ayant manifesté le plus de besoins autour des données de la recherche et leur accompagnement doit donc être une priorité. Une première recension des chercheurs engagés dans de tels projets est donc nécessaire afin de faciliter la

prise de contact et de constituer un premier panel de chercheurs potentiellement intéressés. L'accent pourra par ailleurs être mis sur les projets tout récemment lancés afin de mettre en place des bonnes pratiques (plan de gestion de données de la recherche, réflexion sur les formats des données ...) dès la genèse du projet, ce qui permettra d'avoir des projets précurseurs, vitrine de la politique de gestion des données de la recherche au sein de l'Université Bordeaux Montaigne.

L'autre priorité est d'intégrer les **jeunes chercheurs** dans cette politique, ce groupe ayant l'avantage d'être curieux envers les outils d'analyse et de produire des corpus quantitatifs intéressants déjà sous format numérique. En faisant le lien avec la direction de la recherche et l'Ecole Doctorale, le service « données de la recherche » pourrait ainsi être amenée à familiariser les doctorants à la bonne gestion de leurs données et à leur partage lors des formations de rentrée.

Dans un dernier temps, le service « données de la recherche » pourra se pencher sur la question de la **préservation de la mémoire des unités de recherche**, qu'elle soit sous format papier ou numérique : recenser les bases de données sous format numérique déjà produites et veiller à leur pérennité, en opérant notamment des mutations de format ; assurer le bon archivage des données sous format papier, voire rétroconvertir les corpus quantitatifs.

Accompagner et former les chercheurs

Alors que les outils pour gérer les données de la recherche se multiplient mais sont encore mal connus des chercheurs en SHS pour qui ce sujet est relativement neuf, le rôle d'un service « données de la recherche » se situerait moins sur un plan technique que sur un plan humain, en accompagnant les chercheurs dans la prise en charge de leurs données de recherche.

En nous appuyant sur les besoins en terme d'accompagnement détaillés plus tôt, les efforts d'un service « données de la recherche » doivent se porter sur les points suivants :

- Informer les chercheurs de l'existant et assurer une veille autour des données de la recherche
- Accompagner les chercheurs dans le montage de projets internationaux, notamment autour de la rédaction des plans de gestion des données de la recherche
- Aider les chercheurs volontaires au stockage de leurs données de recherche puis à leur partage en proposant des solutions adaptées à leurs pratiques
- Offrir un soutien technique autour des formats libres et interopérables en veillant notamment à la pérennité des bases produites
- Former les chercheurs à la législation autour des données de recherche
- Former les chercheurs à des outils correspondant à leurs besoins : bureautique, base de données et logiciel d'analyse de données
- Familiariser les chercheurs aux bonnes pratiques d'archivage et être référent pour la bonne conservation des données sous format non-numérique

c) Quel avenir pour le futur service « données de la recherche » ?

Les principaux apports des entretiens avec les personnels d'appui à la recherche

Les entretiens avec les personnels autres que chercheurs (ingénieurs d'étude ou de recherche, documentalistes...) ont fait émerger plusieurs problématiques qu'il nous semble intéressant de rappeler afin de penser les évolutions du futur service « données de la recherche ».

Le questionnement principal porte sur la valorisation de ces jeux de données et leur réutilisation, pour lesquels la question des métadonnées et de l'indexation est centrale.

- Construire ces métadonnées en collaboration avec les chercheurs paraît essentielle mais leur mobilisation sur ces questions n'est pas toujours évidente ; certains jeux de données posent

par ailleurs plus de problèmes que d'autres, l'indexation des corpus iconographiques étant ainsi particulièrement délicate.

- Au-delà du choix de vocabulaire, la forme de ces métadonnées est également à explorer en gardant à l'esprit les possibilités du web sémantique. Certains centres de documentation se penchent par ailleurs de plus en plus sur l'aide à l'indexation en TEI de documents textuels¹⁰, une nouvelle mission qui nécessite cependant une formation adaptée des professionnels de la documentation.
- Toujours autour de la valorisation des données de la recherche, les centres de documentation s'interrogent également sur les personnes susceptibles de réutiliser ces jeux de données et quels indicateurs pourraient permettre d'évaluer ces pratiques.

Une autre difficulté à laquelle sont confrontés ces personnels est celle de la **législation** autour des données de la recherche : certains types de données sont ainsi particulièrement délicats à partager comme les corpus photographiques dès lors qu'une personne y figure. La question se pose aussi pour les carnets de recherche, données sensibles pour de nombreux chercheurs mais qui tombent tout de même dans la catégorie des données publiques ce qui amène à réfléchir à de nouvelles méthodes d'enquête.

Enfin, la question du **plan de gestion des données** est également saisie par certains répondants en appui à la recherche qui ne parviennent pas à trouver pour l'instant de solutions concrètes adaptées aux besoins et pratiques de leurs chercheurs.

La mise en place d'un réseau des professionnels de la documentation

Le chantier des données de la recherche à l'UBM, qui reste encore à monter complètement, ne peut pas se passer à notre sens de la mise en place d'un réseau de professionnels de la documentation à l'échelle locale, afin d'harmoniser les pratiques mais également de mutualiser outils et expériences.

Au sein de l'UBM, ce service devra ainsi associer plusieurs acteurs de l'Université afin de mettre en commun leurs atouts : la direction de la Recherche pour son poids politique et son rôle de coordination, le Service Commun de la Documentation pour son expertise technique dans le domaine de la science ouverte, l'École Doctorale également afin d'impliquer les jeunes chercheurs dans la problématique des données de la recherche. Les compétences des collègues spécialistes de l'Information Scientifique et Technique dans les UMR, parfois rattachés au CNRS (IKER, Ausonius et PASSAGES), a tout intérêt à être mis à profit pour développer ce réseau vers des spécialisations plus disciplinaires (géomatique, cartographie, fouille de texte...). Enfin, la Direction du système d'information et du numérique (DSIN) sera un appui technique incontournable.

D'autres acteurs, peut-être moins évidents, pourraient également être sollicités. Ainsi, les Maisons des Sciences de l'Homme s'engagent de plus en plus à l'échelle nationale pour aider à la diffusion des données de la recherche : tout récemment la MSH Sud a lancé un projet, intitulé Numerev, afin de mettre à disposition et de favoriser la réutilisation des matériaux de recherche ; par ailleurs, ce sont ainsi les MSH qui assurent la gestion des Huma-Num Box mises à disposition pour le stockage de données tièdes par la TGIR, un rôle qui pourrait donc être confié à la MSHA.

¹⁰ Cette réflexion peut être approfondie en visionnant l'intervention d'Emmanuelle Morlock à l'INHA le 12 novembre 2018, « Réaliser des éditions de sources textuelles durables, est-ce possible ? » ainsi que Tiphaine Van de Weghe et alii., « Des données particulières : les données de la recherche en SHS » dans *BDA 2018*, Bucarest, Roumanie, 2018.

Les autres établissements partenaires du site universitaire bordelais que sont l'Université de Bordeaux, Sciences Po Bordeaux, Bordeaux INP et Bordeaux Sciences Agro pourront également être des alliés dans la construction d'une offre de services cohérente sur la place bordelaise.

Enfin, il pourrait être intéressant de se mettre en relation avec les services d'archives locaux pour valoriser les corpus photographiques réalisés de manière individuelle par les chercheurs lors de leurs déplacements en centre d'archives.

Annexes

Annexe 1 : Grille de questions

- Avez-vous déjà entendu le terme « données de la recherche » ?
 - Comment le définiriez-vous ?
 - Que mettriez-vous dans ce périmètre ?
 - Faites-vous une différence entre les données collectées et les données produites ?
 - Qu'est-ce qui vous pose problème pour le définir ?

- Pouvez-vous préciser quel(s) type(s) de données vous collectez et produisez ?
 - Pouvez-vous me donner des exemples concrets ?
 - Sur quelles sources travaillez-vous en ce moment ?
 - Comment les collectez-vous ?
 - Sous quel format sont ces données sources (énumérer si nécessaire : papier, photos, relevés, AV...) ?
 - Avez-vous perçu une évolution des sources collectées dans votre discipline ?

- Que produisez-vous à partir de vos données sources ?
 - En ce moment, quelles données produisez-vous en dehors de vos publications et de leur rédaction ?
 - Pouvez-vous me donner un exemple concret ?
 - Sous quel format sont produites ces données (énumérer si nécessaire : tableur, traitement de texte, logiciel, bdd, TEI...) ?
 - Quel est l'outil que vous utilisez le plus au quotidien ?
 - Au quotidien, utilisez-vous des logiciels payants ou gratuits ? Libres ou propriétaire ?
 - Connaissez-vous les formats que vous utilisez le plus dans votre production numérique ? (Excel, word, jpeg, pdf ...)

- Avez-vous des données sensibles ou confidentielles ?
 - Pourquoi sont-elles confidentielles ?

- Combien pèsent toutes ces données ?
 - Pouvez-vous donner une estimation ?
 - Ces données peuvent-elles être stockées par votre ordinateur ou avez-vous besoin d'un stockage plus puissant ?

- Produisez-vous des données en liaison avec des chercheurs d'autres laboratoires ?
 - Comment faites-vous pour travailler ensemble ?
 - Comment faites-vous pour pouvoir y accéder ou modifier ?

- Pourriez-vous dire quelle est la part de données numériques et non-numériques dans votre quotidien ?
 - Comment gérez-vous ces données au quotidien ?

- Comment sont sauvegardées vos données ?
 - Avez-vous plusieurs endroits de stockage ?
 - Faites-vous des sauvegardes régulières ?
 - Avez-vous déjà connu des pertes de données ?

- Comment décrivez-vous ces données pour les retrouver ?
 - Comment nommez-vous vos fichiers ?
 - Avez-vous un classement personnel ? officiel ?
 - Quelles sont vos règles pour décrire les fichiers ?

- Y a une organisation dans l'unité de recherche (UR) pour la conservation des données ?
 - Votre UR produit-elle des guides pouvant servir de supports ?
 - Avez-vous déjà évoqué cette question avec votre directeur d'UR ?
 - Y a-t-il une coordination entre collègues ?
 - Comment se coordonnent les membres d'un même projet ?

- Rencontrez-vous des difficultés particulières pour gérer vos données ?
 - Avez-vous un problème qui revient souvent ?
 - Une méconnaissance sur la gestion des données a-t-elle déjà impacté votre projet de recherche ?
 - Pouvez-vous me décrire la dernière fois où vous avez eu un problème dans la gestion de vos données ?

- Quelle est votre opinion personnelle sur le partage des données ?

- Connaissez-vous les obligations légales liées aux données de la recherche ?
 - Avez-vous déjà entendu parler des obligations liées aux programmes H2020 ou ANR ?
 - Cette question a-t-elle déjà été évoquée pour d'autres sources de financement ?
 - Comment ressentez-vous l'évolution de cette question ? (est-elle plus ou moins prégnante dans votre travail quotidien ?)

- Avez-vous déjà rendu accessibles vos données ?
 - **Si oui** : Dans quel contexte ?
 - Était-ce une volonté personnelle ou étiez-vous contraint ?
 - Regrettez-vous cette ouverture ?
 - A qui ces données sont-elles accessibles ?

- Savez-vous si votre corpus de donnée est toujours accessible ?
 - **Si non** : pourquoi ?
 - Quelles sont vos craintes concernant le partage des données ?
- Avez-vous déjà utilisé les données de la recherche publiées par un autre chercheur ?
 - Dans quel contexte ?
 - Qu'avez-vous ensuite fait de ces données ?
- Avez-vous déjà visité un entrepôt de données ?
 - Si oui, lesquels ?
 - Qu'est-ce qui vous plait ? Vous plait moins ?
 - Avez-vous une préférence ?
 - Au contraire, avez-vous un entrepôt que vous n'appréciez pas ? Pourquoi ?
 - Si non, savez-vous que cela existe ?
- Quels sont vos besoins concernant le stockage des données ?
- Quel serait pour vous le critère principal pour choisir un entrepôt où stocker ses données ?
- Cet entrepôt doit-il être ouvert, en libre accès, ou en accès restreint ? Pourquoi ?
- Aimerez-vous pouvoir faire vous-même le dépôt ou préféreriez-vous déléguer cette tâche ?
 - Quelle place accordez-vous à l'ergonomie du site ?
 - Ressentez-vous le besoin d'être formé sur ces questions ?
 - Comment envisageriez l'organisation d'une chaîne de dépôt de données ?
- Avez-vous besoin que vos données puissent être citées ?
- Quelle importance accordez-vous à la visibilité de l'entrepôt sur le plan national et international ?
- Voulez-vous lier vos données à des publications ?
- Avez-vous un corpus de données composé de plusieurs supports différents ?
 - Si oui, quelle forme prennent ces corpus ?