# Les données de la recherche en SHS.

# Une enquête à l'Université de Lille 3

## Rapport final

Hélène Prost (CNRS)

Joachim Schöpfel (Lille 3)

Villeneuve d'Ascq, juillet 2015

Les auteurs :

Hélène Prost est ingénieur d'études à l'Institut de l'Information Scientifique et Technique (INIST-CNRS) et membre associé au laboratoire GERiiCO de l'Université de Lille 3. Elle participe à différents projets de recherche relatifs à l'évaluation des collections, la fourniture de documents et l'analyse des usages, ainsi qu'à la littérature grise et le libre accès. Elle est l'auteur de plusieurs publications.

helene.prost@inist.fr

Joachim Schöpfel est maître de conférences en sciences de l'information et de la communication à l'université de Lille 3, directeur de l'Atelier National de Reproduction des Thèses et membre du laboratoire GERiiCO. Il s'intéresse à la communication scientifique, en particulier au libre accès et à la littérature grise, et à l'évolution des fonctions et métiers de l'information scientifique. Il est membre de la Bibliothèque Numérique Scientifique BSN (référentiel de la recherche, open access et numérisation).

joachim.schopfel@univ-lille3.fr

*L'enquête a été préparée dans le cadre d'un projet d'immersion CNRS, en partenariat avec l'Université Humboldt de Berlin et avec le soutien financier du service des Relations internationales de l'Université de Lille 3. Nous remercions toutes les personnes ayant contribué de près ou de loin à la réalisation et la réussite de l'enquête, en particulier Peter Schirmbacher, Maxi Kindling et Elena Simukovic (Berlin) et Isabelle Westeel, Cécile Malleret et Stéphane Chaudiron (Lille).*

## Résumé

Les données de la recherche deviennent l'un des nouveaux défis de la gestion scientifique. Produites dans le cadre des projets de recherche, ces données posent des questions inédites aux laboratoires, bibliothèques et services informatiques des universités : comment conserver ces données, comment les signaler et mettre à disposition d'autres chercheurs, comment faire le lien avec les publications, comment les intégrer dans une politique du libre accès à l'information scientifique ? Avant de se lancer dans un projet de données de la recherche, un établissement doit faire un état des lieux sur le terrain pour mieux connaître les producteurs de ces données, leurs pratiques et besoins dans la gestion des données et leurs outils, mais aussi la nature concrète de ces données. L'Université de Lille 3 a réalisé une étude sur les pratiques, besoins et attentes en matière de gestion des données de la recherche auprès de son personnel scientifique. L'étude est pilotée par le laboratoire GERiiCO et le SCD de Lille 3. Il fait partie d'une démarche concertée en faveur de la gestion et du partage des données de la recherche mise en œuvre à partir de 2013, avec plusieurs analyses, séminaires et publications.

L'enquête a été préparée avec l'Université Humboldt de Berlin. Le questionnaire contient 22 questions et a été mis en ligne en avril et mai 2015. Il a reçu 270 réponses (taux de réponse 15%). Toutes les disciplines sont représentées, ainsi que toutes les catégories des personnels scientifiques.

Les personnes interrogées décrivent un large éventail de données sources. Les corpus (documents textuels) sont de loin la source la plus importante (64%), suivi par les enquêtes et entretiens (47%), observations (41%), expériences (36%) et archives (34%). Quant à la typologie des données produites par les chercheurs comme résultat de leur travail, les données textuelles occupent de loin le premier rang (75%), suivies des tableaux (49%), bases de données (37%), visualisations ou modèles multidimensionnels (32%) et données AV (photo, son, vidéo), avec 20-25%.

Le stockage en local est de loin le mode de sauvegarde privilégié, choisi par 207 participants, que ce soit sur leur ordinateur privé pour 83% des répondants, ou sur leur ordinateur professionnel pour 49% d'entre eux. 40 répondants (19%) stockent « dans le cloud », alors que 8% ont des données sur le serveur d'une autre institution. En réseau, 12% des répondants se tournent vers le serveur de l'université.

La majorité ne partage pas ses données avec d'autres. Et ceux qui le font, partagent d'abord et surtout avec les collègues du groupe de travail (équipe scientifique) (34%). Très peu (<5%) ouvrent leurs données davantage et partagent avec l'institution, d'autres chercheurs ou « tout le monde », dans une démarche *open data* au sens strict du terme.

36% ont déjà pratiqué, d'une manière ou d'une autre, le partage des données – que ce soit avec les collègues d'une équipe scientifique ou d'un projet, via une archive ouverte ou lors de la publication d'un article, ou en téléchargeant les résultats d'autres chercheurs. Apparemment, 20-25% ont davantage d'expérience que les autres (« précurseurs »). 18% autres répondants se disent prêts et ont l'intention d'adopter ces pratiques à l'avenir (« motivés »). 5-10% indiquent qu'ils n'ont pas l'intention de partager leurs données à l'avenir (« réticents »). 30% manquent d'information et de connaissance concernant le partage des données de la recherche (« ignorants »).

En tête de la liste des besoins figure l'espace d'archivage sécurisé et fiable pour ses propres données, suivi des conseils pour la gestion, des conseils techniques (normes, métadonnées etc.) et des conseils juridiques (autour de 50%). Moins nombreux sont les chercheurs qui souhaitent des conseils ou de l'assistance pour la publication ou diffusion des données (autour de 40%), encore moins ceux qui

demandent de l'aide pour préparer un plan de gestion (29%) ou un conseil dans le domaine de l'éthique (24%).

Les résultats ont été analysés par rapport aux disciplines et statuts des personnes interrogées. Ils ont été également comparés avec les résultats de trois autres enquêtes (Berlin, Strasbourg, LIBER Europe). A l'issue de l'enquête, nous allons mener des interviews avec un échantillon restreint d'enseignants-chercheurs et doctorants volontaires, afin de mieux connaître certains pratiques et besoins et aussi, pour valider les résultats de l'enquête. En même temps, nous allons contribuer à développer une offre de service sur le campus de l'Université de Lille 3 pour accompagner et faciliter la gestion des données de la recherche des équipes scientifiques. Cette offre pourrait s'articuler autour de trois axes : la formation des étudiants et chercheurs à la gestion des données, une proposition de conseil et assistance, et une contribution à la création d'une nouvelle infrastructure pour la curation, le stockage, l'archivage et la diffusion des données.

## Table des matières


**1. Introduction** ..... 7

1.1. Les universités ..... 7

1.2. Les bibliothèques universitaires ..... 7

1.3. Connaissance des pratiques ..... 8

1.4. Notre projet de recherche ..... 9

1.5. A propos du terme « données » ..... 9

**2. Méthodologie** ..... 11

**3. Résultats** ..... 12

3.1. Statuts des répondants ..... 12

3.2. Disciplines ..... 13

3.3. Typologie des données ..... 14

3.4. Stockage des données ..... 15

3.5. Diffusion des données ..... 18

3.6. Utilisation d'autres données ..... 18

3.7. Motivation et attitude à l'égard de la diffusion en *open access* ..... 19

3.8. Préférences par rapport au dispositif ..... 21

3.9. Besoins ..... 22

**4. Discussion** ..... 24

4.1. Taux de réponses et groupes ..... 24

4.2. Différences par statut des répondants ..... 24

4.3. Différences disciplinaires ..... 25

4.4. Comparaison des résultats de Lille avec d'autres enquêtes ..... 26

**5. Perspectives** ..... 28

**6. Références** ..... 29

Annexe – Le questionnaire ..... 31

# 1. Introduction

Les données de la recherche deviennent l'un des nouveaux défis de la gestion scientifique. « La science du XXI[e] siècle est la science de l'exploration des données » (André 2015, p.79). Produites dans le cadre des projets de recherche, ces données posent des questions inédites aux laboratoires, bibliothèques et services informatiques des universités : comment conserver ces données, comment les signaler et mettre à disposition d'autres chercheurs, comment faire le lien avec les publications, comment les intégrer dans une politique du libre accès à l'information scientifique ?

## 1.1. Les universités

La CPU a inscrit les données de la recherche tout en haut de son agenda, lors du colloque annuel 2015 à Strasbourg. « Assurer une maitrise publique des données de la recherche », ainsi commencent les 10 propositions de la CPU pour les *Universités 3.0*[1]. A l'heure où l'Etat français mène une politique exemplaire d'ouverture des données publiques et où les entreprises font face à la révolution *Big Data*, les universités relèvent le défi des données de la recherche. L'enjeu est double : mettre en place des infrastructures sécurisées pour la gestion et la préservation des données scientifiques, et assurer leur diffusion pour stimuler la recherche et l'innovation. L'irruption du numérique bouscule les modes d'apprentissage, l'organisation des universités, la définition du lien social mais avant tout aussi, la façon de faire de la recherche. Impossible aujourd'hui d'imaginer la science autrement que numérique. Elle est devenue ouverte[2], grâce à l'outil numérique, et elle est devenue participative. La diffusion libre des résultats de la recherche sur Internet augmente l'impact du chercheur et de l'institution. En même temps, elle correspond à un formidable retour sur investissement pour la société civile.

Cependant, pour la circulation des données de la recherche, les universités ont pris du retard par rapport aux organismes de recherche, comme le fait remarquer Catherine Rivière, PDG de GENCI : « Les communautés (astrophysique, climat…) affichent des données. Les universités, pas vraiment. »[3] Il est temps, selon les mots de la CPU, d'organiser et de systématiser la mise à disposition des résultats scientifiques et des données brutes de la recherche ; on s'oriente ainsi vers un statut de la donnée ouverte, pour la recherche et pour l'innovation.

## 1.2. Les bibliothèques universitaires

Déjà engagées en faveur du libre accès à l'information, les bibliothèques universitaires se trouvent en première ligne pour la gestion des données de la recherche. Or, quel peut-être le rôle d'une BU ? Pierre Naegelen (2015) en cite trois :

- « Analyser les statistiques de consultation de la documentation en vue d'évaluer les besoins des chercheurs ;
- Etre un acteur de l'indexation des données (…) ;
- Constituer une cellule d'appui pour la gestion des données de la recherche : l'élaboration d'un (…) Plan de Gestion des Données (PGD) ».

Dans un rapport de 2011, LIBER, la Ligue des Bibliothèques Européennes de Recherche, détaille les opportunités pour les bibliothèques et autres « centres de données ». L'inventaire proposé détaille

[1] http://www.cpu.fr/actualite/colloque-annuel-de-la-cpu-les-10-propositions/

[2] « Along with "the exponential growth of data" and "the availability of digital technologies", (Jean-Claude Brugelmann) argued, key drivers included "the demand for accountable, responsive and transparent science" and "the need to address faster societal challenges". » *Times Higher Education* 28 juin 2015 https://www.timeshighereducation.co.uk/news/remove-barriers-open-science-says-head-eu-policy-unit

[3] Intervention lors de la table ronde « Science ouverte et méga-données : le numérique, un facteur d'accélération de la création et de la valorisation des savoirs ». Colloque annuel 2015 de la CPU à Strasbourg.

17 actions ; il se lit comme une « wish list » dont l'objectif est assez ambitieux (cf. la synthèse dans Reilly et al. 2011, p.9). Cependant, il s'agit d'opportunités, de pistes de développement potentiel, pas d'un programme à mettre en place. Certaines suggestions sont assez vagues, comme par exemple :

- *Réduire les obstacles pour faciliter la diffusion des données.*
- *Promouvoir de bonnes pratiques de la gestion des données.*
- *Coopérer avec des producteurs de données.*

D'autres idées par contre sont précises et concernent les compétences cœur des BU[4]. Des exemples :

- *Encourager l'utilisation d'identifiants uniques.*
- *Contribuer au développement de métadonnées et de pratiques de citation communes.*
- *Contribuer au développement de normes uniformes pour la citation de données.*
- *Encourager des liens croisés entre publications et données.*
- *Conserver ("curer") et archiver les données.*

Tandis que d'autres propositions mobilisent les capacités des BU à former, conseiller et assister les chercheurs, comme par exemple :

- *Promouvoir l'usage de normes et outils partagés par les chercheurs.*
- *Fournir et aider à comprendre les métadonnées des données.*
- *Former les chercheurs aux bonnes pratiques disciplinaires en matière de production de données (format de conservation, documentation d'expériences,…).*

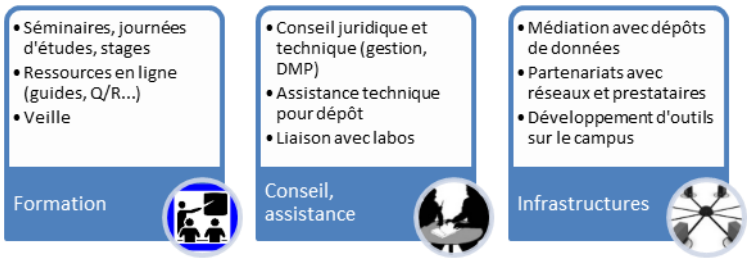



*Figure 1 : Le cadre pour un projet « données de la recherche »*

D'une manière plus synthétique et à partir d'un projet de l'University of Amherst, Massachussetts, Reznik-Zellen et al. (2012) évoquent trois axes pour fixer le cadre d'une nouvelle offre de service au sein de la BU : la formation des étudiants et chercheurs à la gestion des données, une proposition de conseil et assistance, et une contribution à la création d'une nouvelle infrastructure pour la curation, le stockage, l'archivage et la diffusion des données. Figure 1 présente un schéma pour illustrer ce cadre qui pourrait être utile pour le développement d'une offre de service sur le campus.

### 1.3. Connaissance des pratiques

Or, il faut connaître ces pratiques et plus encore, les besoins réels des chercheurs avant de mettre en place une politique en faveur des données de la recherche. Développer de nouveaux services et outils à destination des chercheurs est une chose. Les faire adhérer, c'en est une toute autre et nécessite « (…) en pointer l'intérêt. Il faut privilégier des approches *bottom-up*, seules à même de mobiliser le plus grand nombre » (André 2015, p.94).

Avant de se lancer dans un projet de données de la recherche, un établissement doit faire un état des lieux sur le terrain pour mieux connaître les producteurs de ces données (équipes, unités,

[4] Cf. aussi la synthèse dans LIBER (2012)

disciplines), leurs pratiques et besoins dans la gestion des données et leurs outils, mais aussi la nature concrète de ces données (formats, tailles etc.). Comment font-ils, par exemple, pour sauvegarder les résultats de leur travail ? Qui a accès à ces données ? Seraient-ils disposés à partager leurs données en ligne, et si oui, sur quel genre de plateforme ? Certains chercheurs appliquent déjà des procédures systématiques s'apparentant à un plan de gestion des données (*data management plan*) tandis que d'autres poursuivent des stratégies et techniques relevant davantage de la boite à outils du *personal knowledge management* (PKM) avec les moyens du bord, peut-être efficaces, mais souvent limités.

L'université Humboldt de Berlin a mené en 2013 une enquête pour préparer sa propre politique institutionnelle en matière de données de la recherche et aussi, pour développer un pôle de recherche dans le domaine de l'eScience. Cette enquête s'appuyait notamment sur des études menées en Grande Bretagne par trois universités (Cambridge, Glasgow, Imperial College London), sous l'égide du JISC[5]. L'idée était d'évaluer les pratiques courantes des chercheurs du campus et d'explorer l'intérêt pour certains services à mettre en place. Le résultat fut contrasté : une certaine motivation à « jouer le jeu de l'ouverture des données » mais aussi quelques objections et réticences sérieuses, notamment par rapport à la protection des données personnelles et confidentielles (Simukovic et al. 2014). Toujours est-il que cette enquête a été le déclencheur d'une démarche institutionnelle avec des projets de recherche, des formations et des réalisations au niveau de l'infrastructure documentaire et informatique.

### 1.4. Notre projet de recherche

Notre démarche va dans le même sens – faire un état des lieux, analyser la situation et préparer le terrain pour mettre en place une offre de service adaptée et utile aux doctorants et chercheurs du campus. La démarche a été mise en œuvre à partir de 2013, avec plusieurs sources de financement (MESHS, GERiiCO, Ecole doctorale SHS, Relations internationales, ANRT). Les principales étapes :

1. Une synthèse de la question des données de la recherche dans les thèses de doctorat, pièce-maîtresse de la documentation universitaire (Jacquemin et al. 2013 ; Schöpfel et al. 2014).
2. Trois séminaires sur la valorisation des données de la recherche en sciences humaines et sociales début 2015[6].
3. Une étude sur les données de la recherche dans les thèses de doctorat de l'Université de Lille 3 (Prost et al. 2015, Schöpfel et al. 2015).
4. Un livre blanc sur les données de la recherche dans les thèses de doctorat (Chaudiron et al. 2015).

Une enquête à l'instar de celle de Berlin a complété cette concertation. Mais avant de présenter la méthodologie et les résultats, il reste à savoir ce que recouvre ce terme de « données de la recherche ». De quoi parlons-nous au juste ?

### 1.5. A propos du terme « données »

LIBER définit les données de la recherche comme un « first class research object » à la base du travail des chercheurs, donc plus comme une source (en amont) que comme un résultat (en aval) (Reilly et al. 2011, p.4). Le gouvernement américain considère comme *research data* « *the recorded factual material commonly accepted in the scientific community as necessary to validate research findings* »[7], autrement dit, tout ce qui est accepté, enregistré ou publié et utile pour valider la production

[5] Cf. Data Asset Framework http://www.data-audit.eu/
[6] Cf. DRTD-SHS http://drtdshs2015.sciencesconf.org/
[7] Cf. OMB Circular 110 https://www.whitehouse.gov/omb/circulars_a110#36

scientifique – une définition clairement en aval du processus de la recherche et sans préciser de quoi il s'agit exactement.

D'une manière pragmatique, l'équipe de l'Université Humboldt de Berlin a séparé « données sources » (= les données utilisées par les chercheurs pour leurs recherches) et « données résultats » (= les données produites comme résultats des recherches), en proposant deux listes (Simukovic et al., 2014, pp.744-745) :

- Données sources : « *observations, expériences, simulations, images, enquêtes et interviews, statistiques et données de références, fichiers logs et données d'utilisation, documents textuels* » ;
- Données résultats : « *images, visualisations et modèles multidimensionnels, enregistrements sonores, enregistrements vidéo, textes, tableur, bases de données, logiciels et applications, des données spécifiques par domaine ou instruments/équipements* ».

Au lieu de tenter une définition, le répertoire des archives de données *re3data*[8] indexe le contenu de plus de 1200 sites à partir d'une typologie de 14 données différentes : « données archivées*, données audiovisuelles, données de configuration, bases de données, images, données de réseau, plein texte, données brutes, formats de données scientifiques et statistiques, logiciels, code source, documents administratifs, documents administratifs, et texte structuré* ». Et comme cela ne suffit pas, *re3data* contient une 15e catégorie « autres » pour tout le reste.

D'une manière plus conceptuelle, Francis André propose une classification avec cinq categories de données (2015, pp.81-82) :

- Les données d'observation
- Les données d'expérimentation
- Les données de simulation
- Les données dérivées, résultat d'un traitement, d'une sélection, d'une compilation ou d'une agrégation de données
- Les données de références, servant d'éléments canoniques

Et il poursuit : « Cette approche classificatoire introduit indirectement l'idée que la donnée scientifique est un objet dont les caractéristiques évoluent selon l'étape du processus de recherche (…) La donnée scientifique est (…) un objet complexe, dynamique et vivant qui prend toute sa place dans la science d'aujourd'hui » (ibid.). C'est sans doute l'approche la plus réaliste, du moins d'après nos propres résultats qui montrent une grande variété de données, souvent sans pouvoir clairement distinguer entre données « sources » et données « résultats » (Prost et al. 2015).

Un dernier mot à propos des données en question. Contrairement aux grands équipements et structures scientifiques (CERN, *Human Genome Project* etc.), on trouvera sur un campus de Lettres avec des disciplines relevant des SHS sans doute pas de *big data* mais surtout ce qu'on appelle par opposition des *small data*, voire des *dark data*, c'est-à-dire «*data that is not easily found by potential users (…) unpublished data (and) research findings and raw data that lie behind published works which are also difficult or impossible to access as time progresses* » (Heidorn 2008, p.281 et p.285). Donc plutôt de petits volumes, produits ponctuellement (enquêtes, tests, observations, fouilles etc.) plutôt qu'en continu mais, seul point en commun avec le *big data*, d'une grande variété et hétérogénéité. Une raison de plus pour une enquête exploratoire, ouverte.

[8] http://www.re3data.org/

## 2. Méthodologie

Nous avons adapté l'enquête de Berlin (Simukovic et al. 2013) au contexte de l'Université de Lille 3. La structure générale est la même. Les questions sont regroupées en six sections :

1. Informations sur le répondant (statut, domaine)
2. Typologie des données (sources et résultats)
3. Pratiques de stockage et sauvegarde
4. Pratiques de partage
5. Avis par rapport au partage et aux archives de données
6. Besoins

Pour la section 2 (typologie), nous avons également exploité les résultats de notre étude sur les données dans les thèses de Lille 3 (Prost et al. 2015). Comme à Berlin, le questionnaire était anonyme, avec très peu d'information sur la personne. Mais nous avons demandé à la fin du questionnaire si le répondant était intéressé pour participer à une interview sur la gestion des données ; dans ce cas, il pouvait laisser ses coordonnées. Aucune question n'était obligatoire pour valider le questionnaire.

Le questionnaire – 22 questions (cf. annexe) – a été chargé sur la plateforme de la MESHS[9]. Après plusieurs semaines de test, l'enquête a été lancée en avril 2015, du 8 avril au 10 mai. Elle a été accompagnée par une communication à l'ensemble du personnel de l'Université de Lille 3 (avec rappels) et une information plus ciblée via l'Ecole doctorale SHS (doctorants), la Direction de la Recherche (directeurs de laboratoires) et GERiiCO. Cette communication par messagerie institutionnelle, outre que d'expliquer la démarche, comportait aussi le lien vers le questionnaire sur le serveur de la MESHS.

L'enquête a été pilotée par le laboratoire GERiiCO (Stéphane Chaudiron, Joachim Schöpfel) et le SCD (Isabelle Westeel). Ont participé au montage du questionnaire et à l'analyse des résultats Hélène Prost (CNRS, membre associée GERiiCO) et Cécile Malleret (SCD). Le service Relations Internationales de l'Université de Lille 3 a facilité la préparation de l'enquête, en particulier par le financement d'un court séjour à l'Université Humboldt de Berlin en juillet 2014.

[9] https://formulaires.meshs.fr/ avec le logiciel *Limesurvey*

# 3. Résultats

L'enquête s'adressa en premier lieu aux enseignants-chercheurs et doctorants sur le campus de l'Université de Lille 3 (1800 personnes). Elle a reçu 270 réponses, ce qui correspond à un taux de réponse d'environ 15%.

Les participants ont répondu aux 22 questions de l'enquête dans des proportions inégales. Aucune des questions n'était obligatoire. Résultat : selon les questions, le taux de réponse varie entre 12 et 82%. Ont reçu le plus grand nombre de réponses les deux premières questions, concernant le statut (82%) et la discipline (80%) du répondant.

## 3.1. Statuts des répondants

Les doctorants et maîtres de conférences ont été les plus nombreux à répondre à l'enquête (73 et 69 répondants), suivies par les professeurs (40), les chercheurs (16) et les ITA/IRTF/BIATOSS (13) (Figure 2).

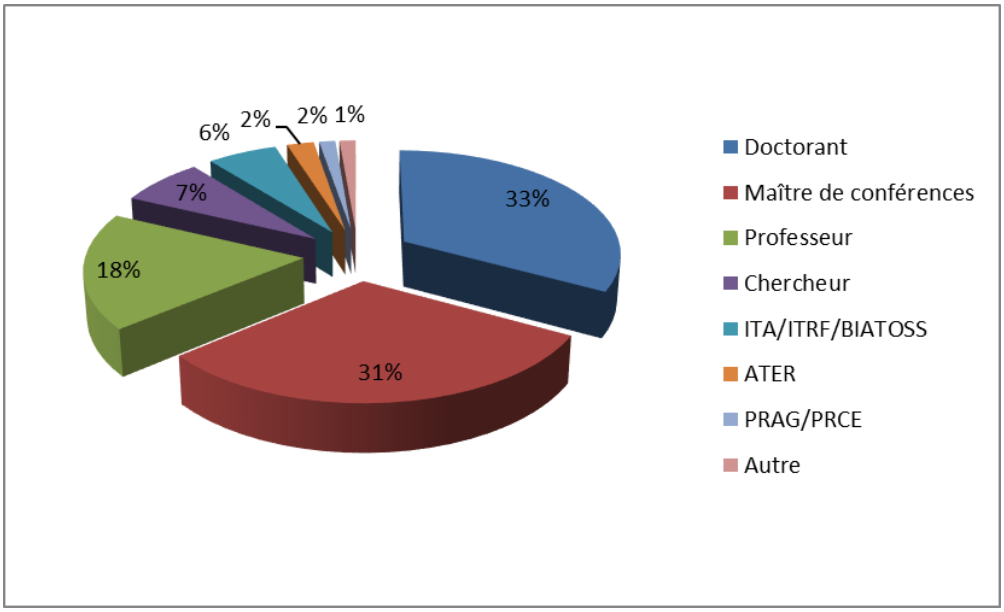



*Figure 2 : Statut des personnes ayant répondu à l'enquête (N=222)*

En mettant le nombre de répondants en rapport avec la répartition du personnel scientifique de l'Université de Lille 3[10], on obtient un résultat contrasté de la représentativité des différentes catégories :

- Professeur d'université : 26%
- Maîtres de conférences : 17%
- Doctorants : 13%
- BIATOSS (seulement catégorie A) : 9%

Ces chiffres indiquent qu'un quart des professeurs ont répondu à cette enquête. Quant aux maîtres de conférences et doctorants, environ un sixième des deux catégories ont répondu.

[10] Source : Direction des Ressources Humaines : *Bilan social 2013 de l'Université Charles de Gaulle-Lille 3*. Villeneuve d'Ascq, 2014.

## 3.2. Disciplines

80% des répondants ont renseigné leur domaine scientifique, en indiquant, soit la section scientifique dont ils dépendent, soit leur discipline scientifique. Nous avons rassemblé les différentes réponses dans 21 domaines principaux (Figure 3).

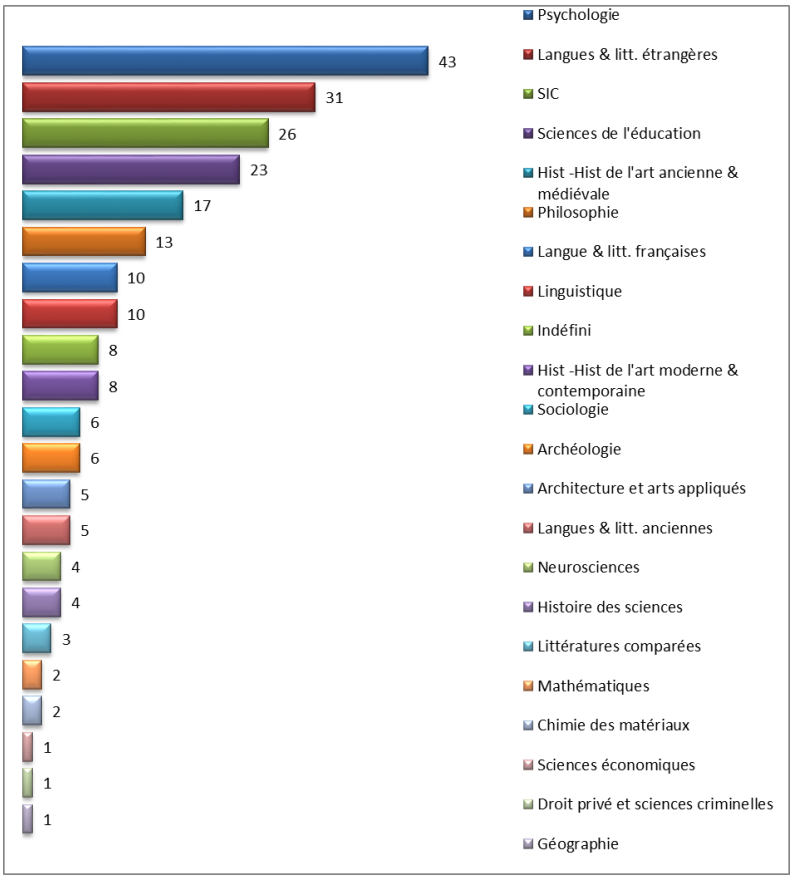



*Figure 3 : Les disciplines des répondants (N=217)*

La psychologie regroupe le plus grand nombre de répondants, soit 43 ou 19% des participants ; viennent ensuite les langues et littératures étrangères, les sciences de l'information et de la communication, et les sciences de l'éducation.

Si on met ces chiffres en rapport avec la répartition du personnel scientifique par discipline (Bilan social 2013, hors doctorants), on obtient une autre image de la représentativité des réponses ; dans l'ordre décroissant :

- Histoire : 43%
- SIC : 38%
- Sciences de l'éducation : 36%
- Psychologie : 30%
- Langue & littérature françaises : 27%
- Linguistique : 27%
- Philosophie : 25%
- Langues & littératures étrangères : 14%

Cela donne une image intéressante des disciplines concernées et intéressées. On pourrait distinguer deux groupes : le premier groupe avec une bonne représentativité, c'est-à-dire avec plus d'un tiers

de répondants (Histoire, SIC, Sciences de l'éducation) ; un deuxième groupe un peu moins représentatif, avec 25-30% de répondants (Psychologie, Langue & littérature françaises, Linguistique, Philosophie). De l'autre côté, on trouve une sorte de « longue traîne » composée de disciplines avec moins de 10% de répondants, avec notamment l'archéologie et la sociologie, l'architecture & arts appliquées, les langues & littératures anciennes, l'histoire des sciences, les neurosciences, les mathématiques et l'économie.

### 3.3. Typologie des données

De quelles données s'agit-il ? Quels sont les données de la recherche que les chercheurs produisent comme résultat de leur travail scientifique ? Nous avons évoqué plus haut les difficultés à définir ces données. Notre enquête adopte une approche pragmatique, à l'instar de l'enquête de Berlin (Simukovic et al. 2014), et pose deux questions : l'une sur les données sources qui sont utilisées par les chercheurs pour faire leur travail ; et l'autre sur les données résultats, produites par les chercheurs pendant et à l'issue de leur travail scientifique. Quant aux catégories (typologie), nous avons adapté les catégories de l'enquête de Berlin, en mettant l'accent sur la recherche en SHS et en intégrant les catégories de l'étude sur les données dans les thèses de doctorat (Prost et al. 2015).

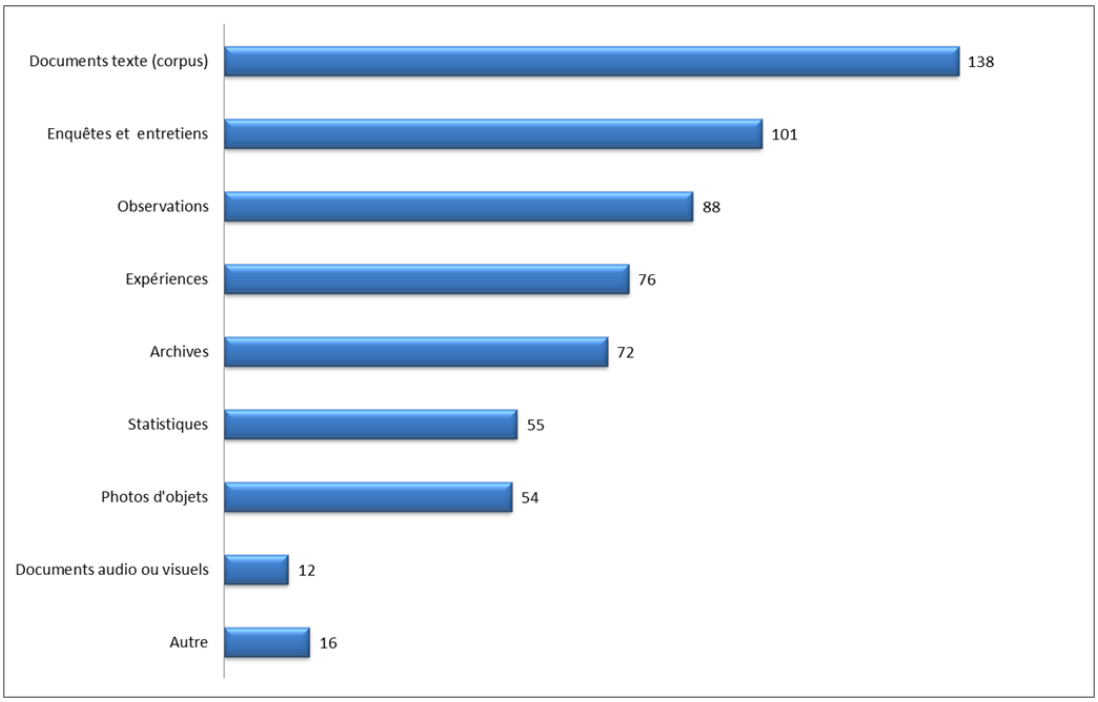



*Figure 4 : Données sources (N=214)*

79% des participants ont répondu à la question sur les données sources (Figure 4), en indiquant un large éventail de données exploitées. Les corpus (documents textuels) sont de loin la source la plus importante (64%), suivi par les enquêtes et entretiens (47%), observations (41%), expériences (36%) et archives (34%). Moins importants : statistiques et matériel audiovisuel, des dessins, gravures ou fiches de cours. Pas de simulations ou fichiers logs, du moins dans notre échantillon.

La figure 5 montre les réponses à la 2e question concernant la typologie des données produites par les chercheurs comme résultat de leur travail. De nouveau, les données textuelles occupent de loin le premier rang, parfois avec des annotations (75%), suivies des tableaux (49%), bases de données (37%), visualisations ou modèles multidimensionnels (32%) et données AV (photo, son, vidéo), avec 20-25%.

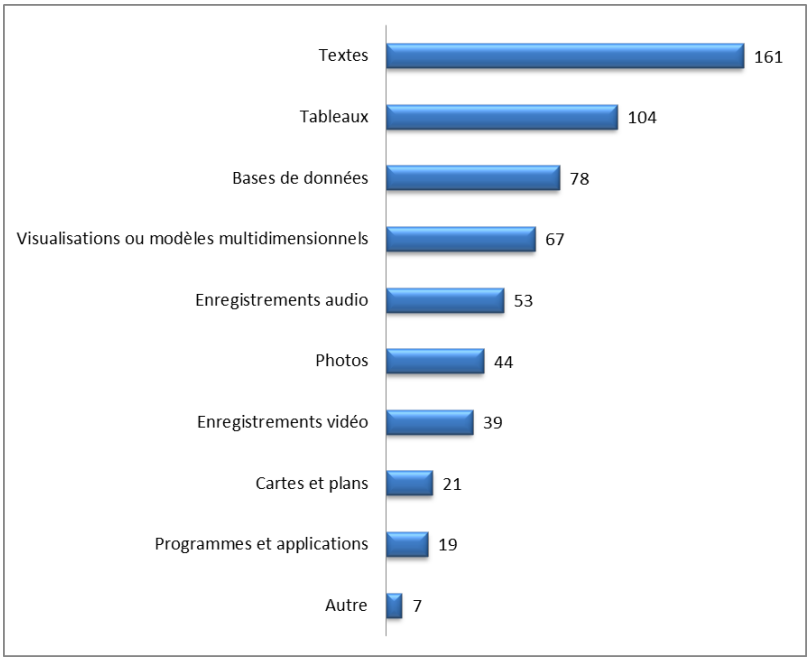



*Figure 5 : Données résultats (N=211)*

78% des participants ont répondu à cette question. La diversité des réponses illustre la richesse des données produites, mais reflète parfois aussi la difficulté de délimiter la frontière entre les sources utilisées et les données produites. Par exemple, les photos peuvent être des données sources à partir desquelles le doctorant construit l'argumentaire de sa thèse. Dans le cadre des thèses en Histoire de l'Art, elles sont à la fois sources et données produites, car le doctorant peut récolter des photos classées dans un inventaire, mais il peut également prendre lui-même des photos d'œuvres ou monuments.

## 3.4. Stockage des données

212 personnes, soit 79% des participants ont indiqué un ou plusieurs espaces principaux de stockage pour leurs données. Le stockage en local est de loin le mode de sauvegarde privilégié, choisi par 207 participants, que ce soit sur leur ordinateur privé pour 83% des répondants, ou sur leur ordinateur professionnel pour 49% d'entre eux. 40 répondants (19%) stockent « dans le cloud », alors que 8% ont des données sur le serveur d'une autre institution. En réseau, 12% des répondants se tournent vers le serveur de l'université (Figure 6).

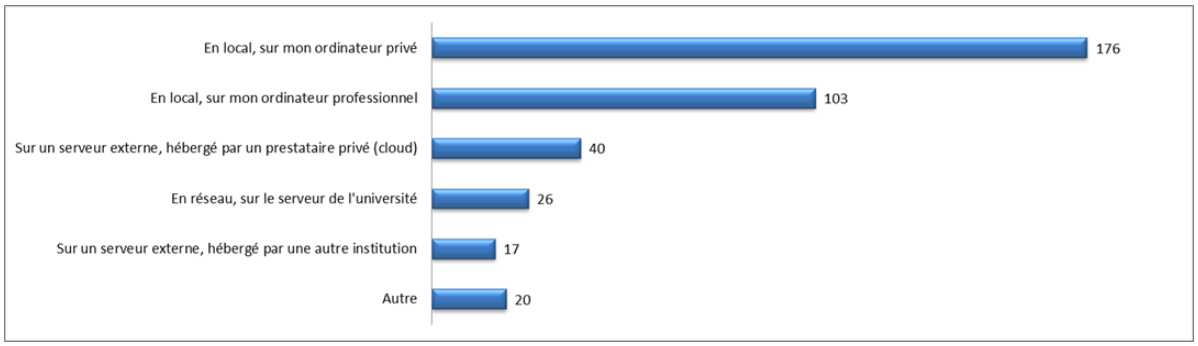



*Figure 6 : Endroits de stockage (N=212)*

117 personnes indiquent utiliser deux ou plusieurs lieux de stockage. Dans la plupart des cas, cela veut dire : stockage sur l'ordinateur privé plus quelque part ailleurs, notamment l'ordinateur professionnel.

204 personnes, soit 76% des participants ont mentionné un autre support de stockage. 81% d'entre eux stockent leurs données sur un disque dur externe, 60% sur une clé USB, et 13% sur CD ou DVD (Figure 7).

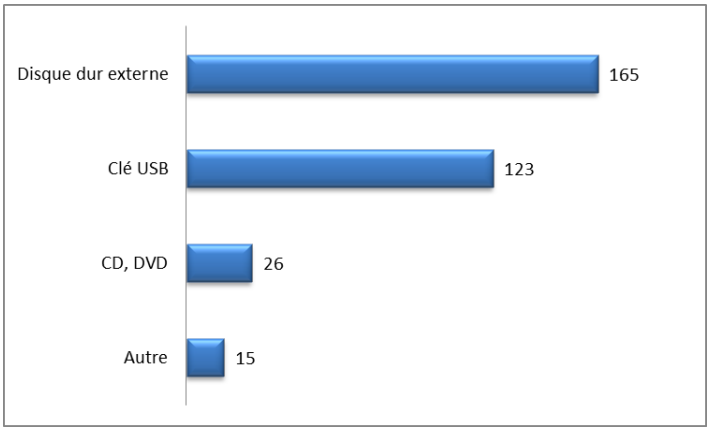



*Figure 7 : D'autres supports de stockage (N=204)*

Parmi les autres supports, 8 participants mentionnent un stockage sur le web, que ce soit « le service Dropbox ou le « cloud ». Ici aussi, le stockage « multi support » est la règle, avec la combinaison « clé USB plus disque dur externe ».

L'enquête révèle également que la sauvegarde sur papier n'a pas totalement disparu : il reste l'un des modes principaux de stockage pour 5 participants, sous forme de « notes manuscrites » ou « document papier » par exemple.

210 personnes, soit 78% des participants ont répondu à la question relative à l'espace nécessaire pour le stockage de leurs données. 38 répondants ont des difficultés pour donner une estimation. 51% ont besoin de 100 Go au plus. 13 d'entre eux réclament un espace supérieur à 1 To, parmi lesquels 6 appartiennent au domaine de la Psychologie (Figure 8).

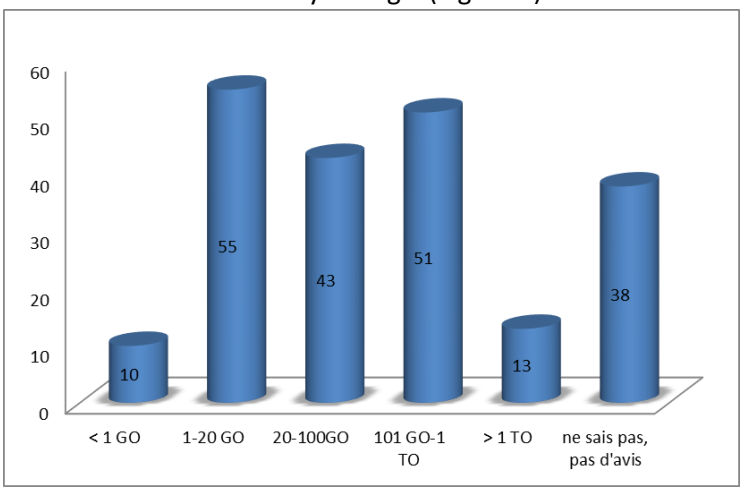



*Figure 8 : Taille de l'espace de stockage (N=210)*

A quelle fréquence sauvegardent-ils leurs données ? Alors que 28% déclarent sauvegarder quotidiennement leurs données, presqu'autant (26%) le font d'une manière irrégulière « en fonction

des besoins ». D'autres sauvegardent moins souvent, une fois par semaine (22%), par mois (15%) ou par trimestre (6%) (Figure 9).

Ceci étant, la question relative à la fréquence de sauvegarde des données révèle les différentes interprétations attribuées au terme « sauvegarde ». Les réponses laissent penser soit à un enregistrement des mises à jour, dont la sauvegarde est quotidienne, soit à un archivage des dossiers, dont la fréquence est plus espacée.

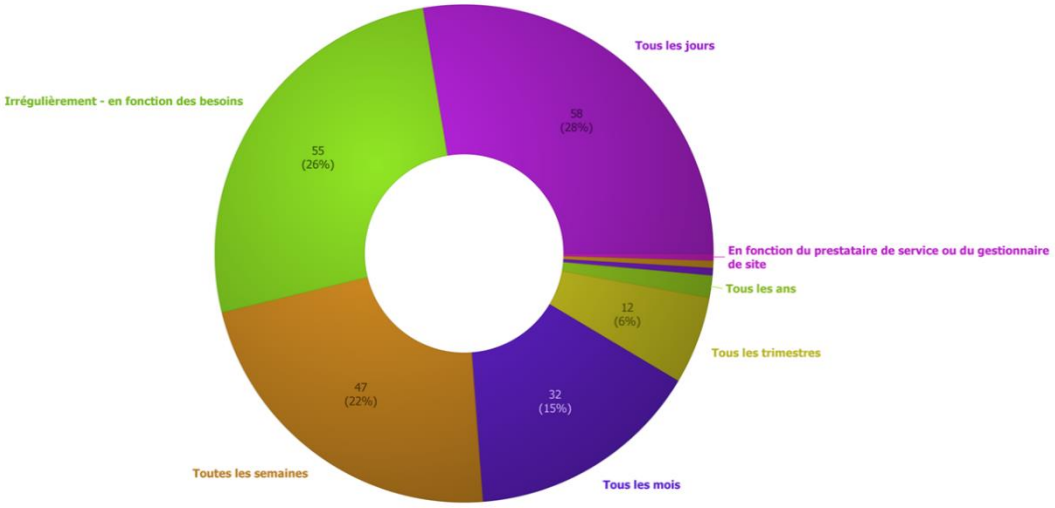



*Figure 9 : Fréquence de sauvegarde (N=210)*

97% des répondants se déclarent eux-mêmes responsables de la sauvegarde des données. Le chef de projet en a la responsabilité dans 8 % des cas. En revanche, à ce jour le personnel de la bibliothèque n'est pas sollicité pour mener à bien la sauvegarde des données. Cette absence peut s'expliquer par le fait que la majorité des données sont sauvegardées en local mais sans doute aussi tout simplement par l'absence d'une offre de service (Figure 10).

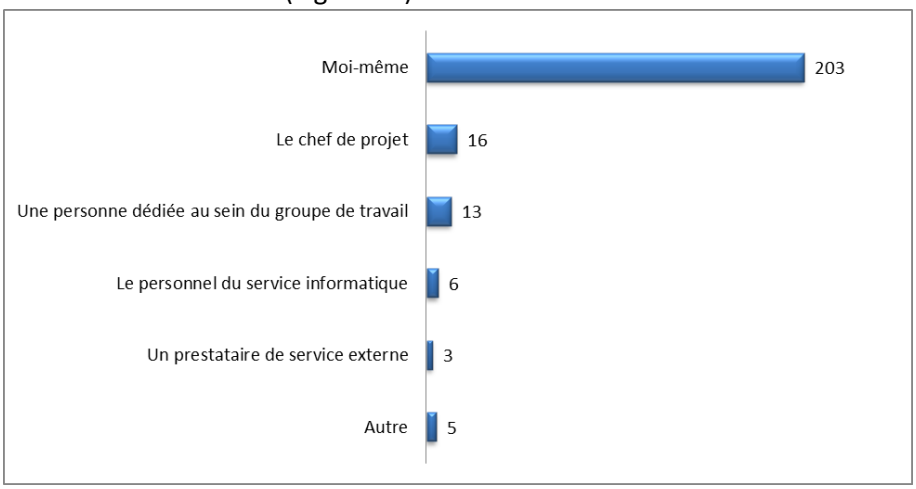



*Figure 10 : Responsabilité de la sauvegarde (N=210)*

Dans l'ensemble, cela donne l'impression d'une pratique personnelle, privée plus souvent que professionnelle, avec les moyens du bord (disque dur, clé USB) mais avec un certain souci de sécurité (stockage à plusieurs endroits et sur plusieurs supports, sauvegardes plus ou moins régulières). Par contre, l'enquête ne révèle aucun élément qui permettrait de dire que cette façon de faire ait posé

un problème majeur à l'un des collègues interrogés – qu'il y ait eu une grosse perte accidentelle de données, par exemple. Un fait dont il faudra tenir compte pour la suite.

## 3.5. Diffusion des données

A la question « qui, en dehors de vous, peut accéder à vos données ? », 77% des participants ont répondu. Pour deux tiers (64%), la réponse est simple et sans équivoque : « personne d'autre » (Figure 11).

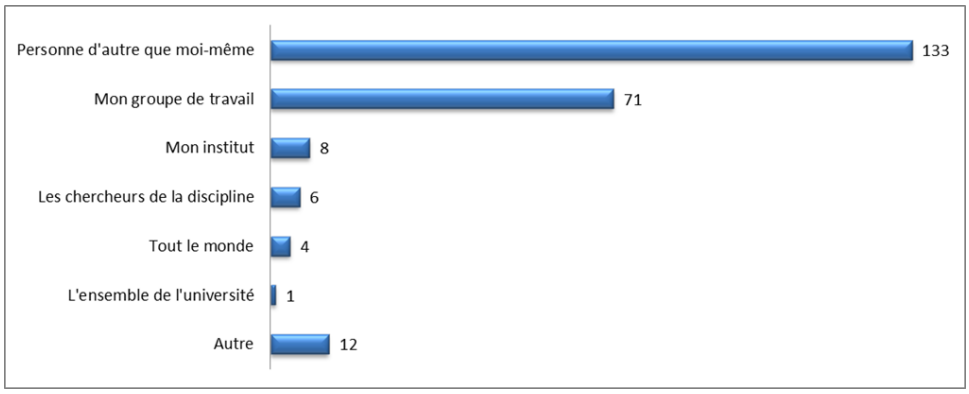



*Figure 11 : Accès aux données (N=209)*

Autrement dit, la majorité ne partage pas ses données avec d'autres. Et ceux qui le font, partagent d'abord et surtout avec les collègues du groupe de travail (équipe scientifique) (34%). Très peu (<5%) ouvrent leurs données davantage et partagent avec l'institution, d'autres chercheurs ou « tout le monde », dans une démarche *open data* au sens strict du terme.

## 3.6. Utilisation d'autres données

« Avez-vous déjà téléchargé des données à partir d'une archive ouverte ? » Seulement 58% répondent à cette question. Et seulement une minorité de 38% confirme l'avoir déjà fait, correspondant à 22% de l'ensemble de l'échantillon (Figure 12).

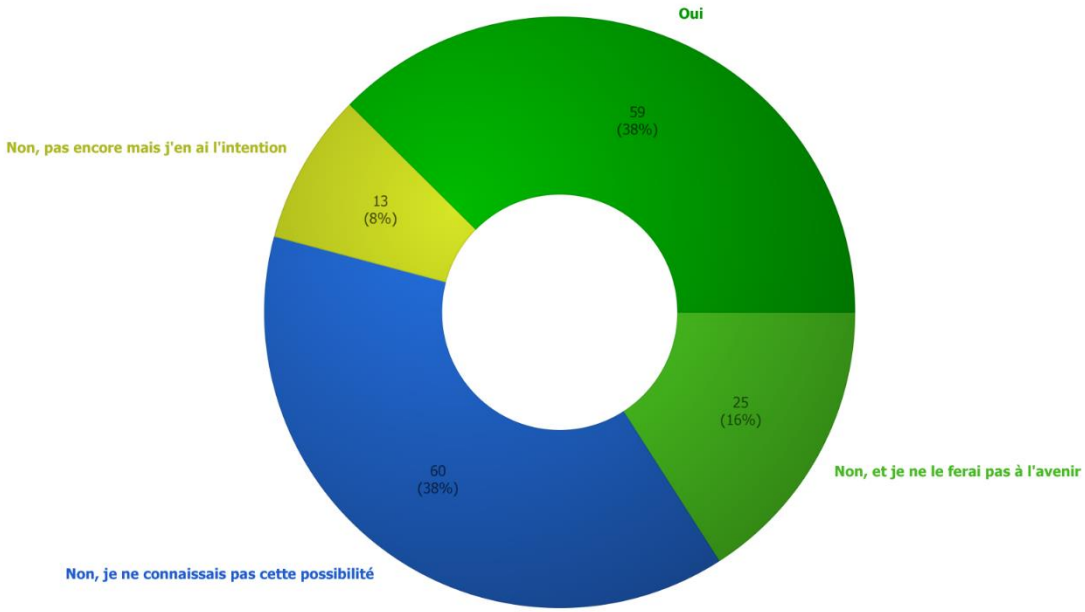



*Figure 12 : Téléchargement de données sur d'autres sites (N=157)*

Les autres – trois sur cinq - n'ont jamais téléchargé d'autres résultats pour leur recherche (62%). Certains disent qu'ils ne le feront pas non plus à l'avenir (16%) ; d'autres au contraire ont l'intention de le faire (8%). Mais la plupart disent tout simplement qu'ils ne connaissaient pas cette possibilité. Ignorance plutôt que bonne ou mauvaise volonté - ce n'est pas vraiment surprenant en sciences humaines et sociales, avec, à ce jour une offre plutôt limitée.

## 3.7. Motivation et attitude à l'égard de la diffusion en *open access*

Nous avons posé plusieurs questions pour mieux connaître l'intérêt et l'opinion des collègues par rapport au partage des données de la recherche, aussi pour avoir une idée s'ils sont oui ou non motivés de contribuer à un tel dispositif.

La première question portait sur un éventuel dépôt de données dans une archive ouverte. Seulement une petite minorité – un sur six – a déjà déposé ses données de cette façon (16%). Les autres 84% n'ont pas encore pratiqué un tel dépôt (Figure 13).

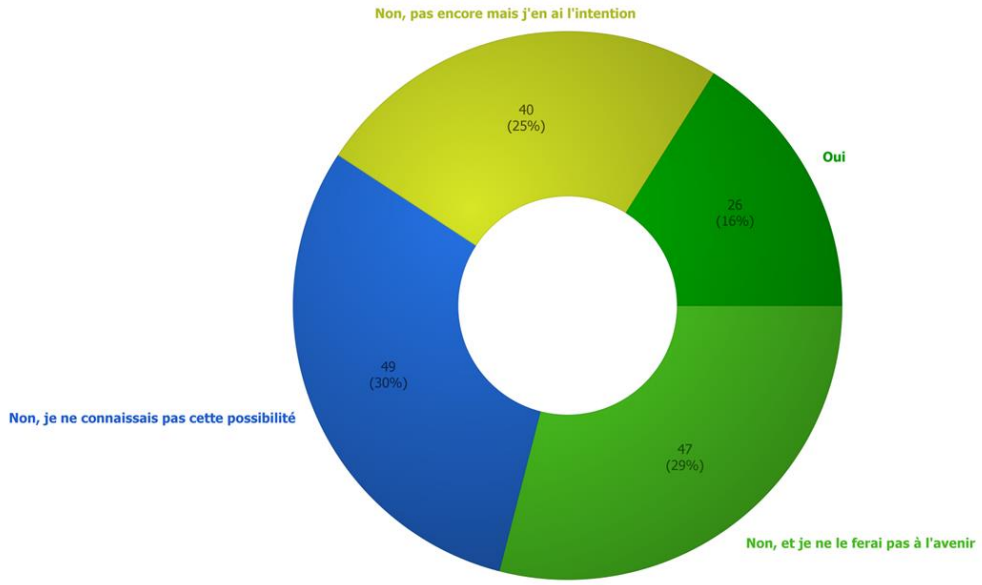



*Figure 13 : Dépôt de données dans une archive ouverte (N=162)*

Parmi ces 84%, on peut grossièrement distinguer trois groupes de taille égale : les uns ignoraient cette possibilité, d'autres ont l'intention de le faire à l'avenir, puis un troisième groupe dit clairement qu'ils n'ont pas l'intention de partager leurs données de cette manière.

Le résultat est un peu différent quand il s'agit non pas de la diffusion via une archive ouverte mais de la publication des données en lien avec la publication d'un article de revue. Presque la moitié dit d'avoir déjà publié des données de cette manière. Et très peu de répondants (7%) refusent le principe d'une telle publication (Figure 14).

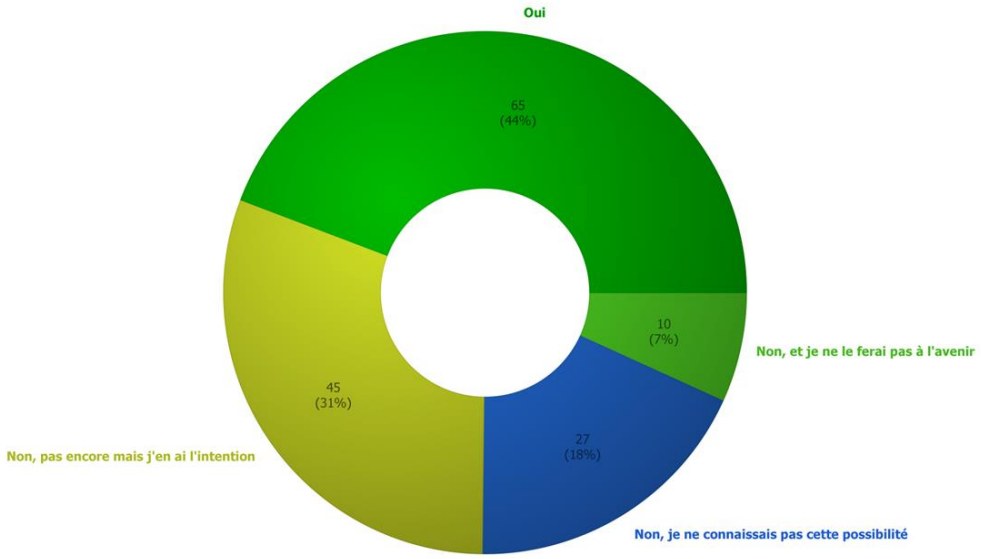



*Figure 14 : Publication de données avec des articles (N=147)*

Seulement 24% des répondants disent connaître des recommandations en matière de données. Ces recommandations ne sont pas nécessairement d'ordre technique, il s'agit davantage de règles éthiques[11] voire politiques sur « ce qu'il convient de faire avec les données de la recherche ».

Seraient-ils d'accord pour déposer leurs données dans une archive ? Peu (31% seulement) répondent à cette question, ce qui est déjà un élément de réponse en soi. Parmi ceux qui répondent, une courte majorité de 54% se prononcent en faveur d'un dépôt, 25% sont plutôt opposés tandis que les autres (21%) sont indécis. Vu le taux de réponse, cela ne veut sans doute pas dire grand-chose. Plus intéressant peut-être le grand nombre de commentaire. Parmi les raisons pour ne pas déposer, plusieurs arguments reviennent : en tête le caractère sensible ou confidentiel des données, le risque de plagiat et la surcharge de travail. D'autres raisons évoquées sont l'illisibilité des données brutes (autrement dit, la nécessité de passer du temps pour les rendre lisibles pour d'autres) ou tout simplement la protection de la propriété intellectuelle : « ces données m'appartiennent »…

La dernière question portait sur la nature des données à déposer. 187 personnes ont répondu à cette question à choix multiple. Seulement 23 (12%) disent très clairement de ne pas vouloir déposer leurs données (Figure 15).

[11] Voir par exemple l'avis récent du COMETS (2015) du CNRS.

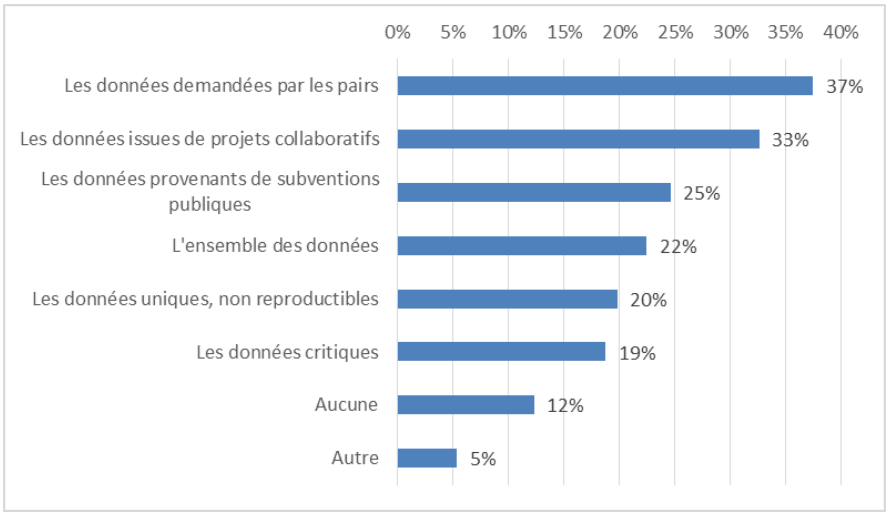



*Figure 15 : Type de données à déposer (N=187)*

Mais seulement 22% se disent prêts à déposer l'ensemble de leurs résultats. Les autres réponses se divisent en deux groupes :

- Communauté scientifique : Environ un tiers des personnes déposeraient des données à la demande de collègues (sans doute un peu comme l'envoi d'un manuscrit à d'autres collègues) ou bien dans le cadre d'un projet collaboratif. Autrement dit, le dépôt et partage est dominé par les relations collégiales au sein d'une communauté de recherche.
- Données uniques et/ou critiques : Environ 20% des personnes répondent qu'elles déposeraient volontiers des données uniques, non reproductibles et/ou des données particulièrement importantes (critiques). Ici, l'argument n'est pas la communauté mais la nature même des données en question.

Un quart seulement des personnes interrogées évoquent la raison politique du *open data* : dépôt des données issues de projets financés avec des subventions publiques.

## 3.8. Préférences par rapport au dispositif

« Quelle archive de données utiliseriez-vous le plus probablement ? » 173 personnes ont répondu à cette question, parfois avec deux ou trois options ; et on ne peut pas dire qu'il y ait une véritable préférence pour l'un ou l'autre type d'archive, même si l'archive internationale est le plus souvent mentionnée (cf. Figure 16).

Mais comment choisir telle ou telle archive ? Quels seraient les critères de choix ? Ici, les avis sont moins partagés et deux critères se détachent très clairement : 54% disent se fier aux conseils de leurs collègues tandis que 51% indiquent la réputation de l'organisme hébergeur comme critère de choix.

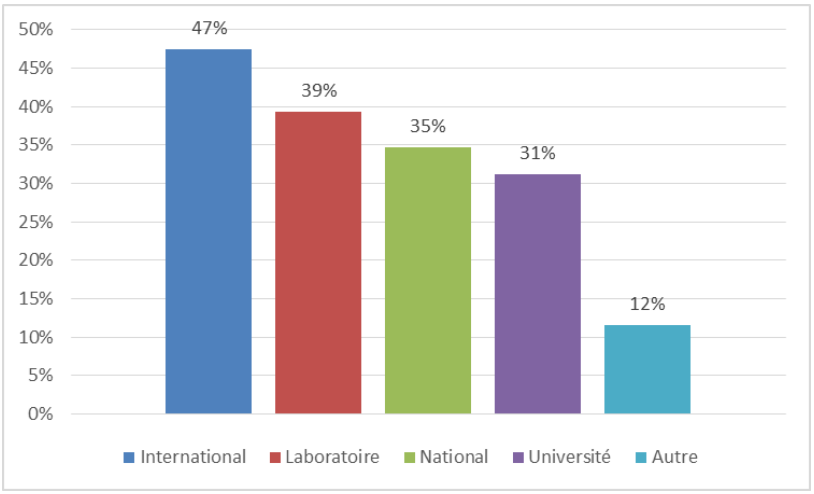



*Figure 16 : Préférence pour type d'archive (N=173)*

Ceci étant, apparemment très peu (17) ont réellement déjà utilisé d'autres dispositifs que ceux de l'université de Lille 3, notamment dans le cadre de projets collaboratifs où quand la personne interrogée travaillent partiellement dans ou avec un autre organisme. Mais cela reste une exception.

## 3.9. Besoins

Pour finir, nous avons posé la question des besoins, en proposant une liste ouverte de huit services potentiellement intéressants pour les doctorants, enseignants-chercheurs etc. Plusieurs réponses ont été possibles (Figure 17).

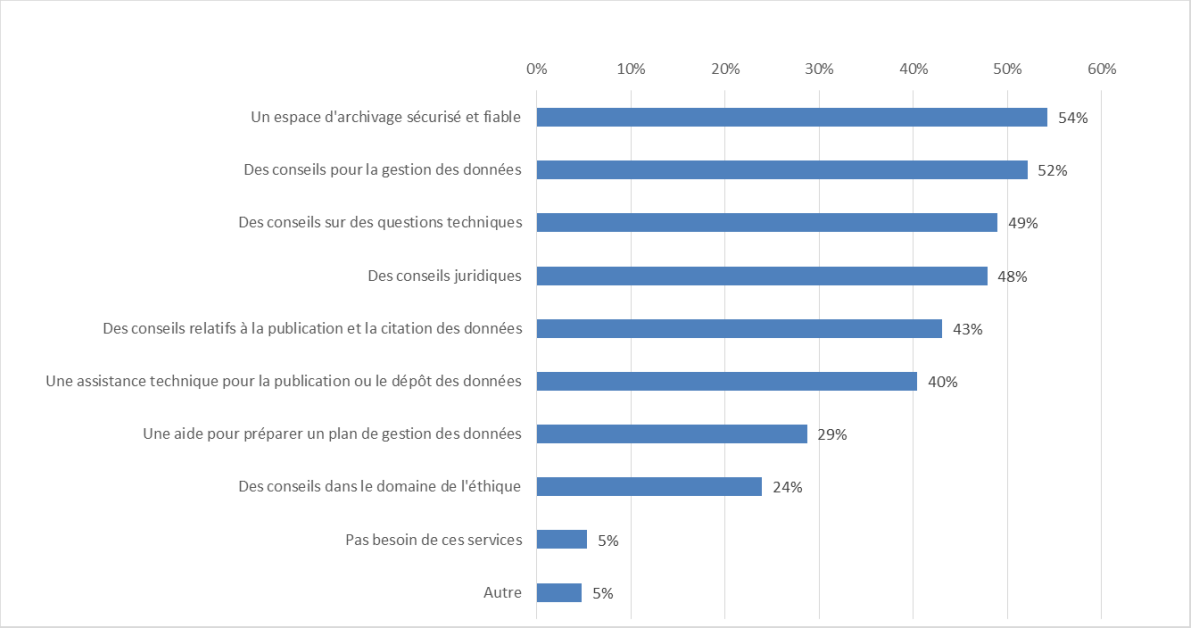



*Figure 17 : Les besoins par rapport aux services (N=188*

Selon les réponses, la plupart des services proposés est demandée. En tête de la liste, l'espace d'archivage sécurisé et fiable pour ses propres données, suivi des conseils pour la gestion, des conseils techniques (normes, métadonnées etc.) et des conseils juridiques (autour de 50%).

Moins nombreux les chercheurs qui souhaitent des conseils ou de l'assistance pour la publication ou diffusion des données (autour de 40%), encore moins ceux qui demandent de l'aide pour préparer un

plan de gestion (29%) ou un conseil dans le domaine de l'éthique (24%). Souvent, les répondants ont coché plusieurs besoins. Et seulement 5% disent de ne pas avoir besoin de tels services.

# 4. Discussion

## 4.1. Taux de réponses et groupes

Les personnes interrogées avaient le choix de répondre oui ou non aux questions. Aucune des questions n'était obligatoire. Par conséquent, aucune question n'a reçu 100% de réponses. A regarder de près, nous pouvons distinguer deux groupes de questions :

Gestion des données : Toutes les questions concernant les pratiques habituelles de la gestion des données ont reçu un taux de réponse moyen élevé de 78%. Il s'agit en particulier des questions concernant la source et la typologie des données, le stockage et la sauvegarde au quotidien.

Partage des données : Cependant, quand il s'agit de questions sur les pratiques de partage – déposer ou diffuser ses propres données, utiliser les données d'autres chercheurs – le taux de réponse moyen tombe à 49%.

Il y a sûrement plusieurs raisons à cela. L'absence de réponse peut aussi bien signifier « ne sais pas », ou « rien à signaler » que « ne souhaite pas répondre ». Toujours est-il, pour notre enquête ce profil de réponse est porteur d'un double message : d'une part, il souligne une relative manque d'assurance par rapport à ces nouvelles pratiques. On pourrait interpréter cette incertitude d'une manière positive comme un besoin d'apprendre, de mieux connaître, de pratiquer. D'autre part, cette différence de taux de réponse indique aussi qu'il faut être plus prudent dans l'interprétation des questions de cette 2e catégorie, ne pas trop généraliser, ne pas tirer des conclusions hâtives là où la moitié ou davantage des répondants n'a pas souhaité s'exprimer. -

Par rapport aux pratiques des chercheurs, Reilly et al. (2011) ont parlé d'une règle « 60-40 » : 60% des chercheurs veulent bien utiliser les données des autres tandis que seulement 40% sont prêts à partager leurs propres données. Les résultats de notre enquête confirment cette règle seulement partiellement. Effectivement, 41% des répondants de Lille 3 ont déjà diffusé leurs données via une archive ouverte ou ont l'intention de le faire. Mais seulement 46% (et non pas 60%) sont favorables à la réutilisation des résultats d'autres chercheurs. De nouveau, cela incite à une certaine prudence par rapport à une démarche sur le campus.

Pour mieux comprendre les différences entre les répondants, nous avons effectué une analyse croisée des réponses aux questions sur les pratiques de partage et sur la motivation à déposer les données dans une archive (questions 10, 11 et 14). Voici quelques éléments :

- 36% ont déjà pratiqué, d'une manière ou d'une autre, le partage des données – que ce soit avec les collègues d'une équipe scientifique ou d'un projet, via une archive ouverte ou lors de la publication d'un article, ou en téléchargeant les résultats d'autres chercheurs. Apparemment, 20-25% ont davantage d'expérience que les autres (« précurseurs »).
- 18% autres répondants se disent prêts et ont l'intention d'adopter ces pratiques à l'avenir (« motivés »).
- 5-10% indiquent qu'ils n'ont pas l'intention de partager leurs données à l'avenir (« réticents »).
- 30% manquent d'information et de connaissance concernant le partage des données de la recherche (« ignorants »).

## 4.2. Différences par statut des répondants

Nous avons comparé les réponses des professeurs d'université, maîtres de conférences et doctorants par rapport à plusieurs variables. L'échantillon n'est pas grand, et vu la taille des trois groupes (40 PR,

69 MCF, 73 doctorants), il faut se méfier du biais du petit nombre et éviter les généralisations. Voici quelques différences entre ces trois groupes, tout de même.

Pratiques de partages : D'après les réponses, les enseignants-chercheurs ont bien plus d'expériences que les doctorants en matière de partage des données (Tableau 1). Les doctorants sont aussi les plus nombreux à dire qu'ils ne connaissaient pas ces possibilités.

| | **Professeurs** | **MCF** | **Doctorants** |
|---|---|---|---|
| ***Dépôt dans archive*** | 27% | 24% | 5% |
| ***Publication avec article*** | 61% | 47% | 25% |
| ***Téléchargement*** | 56% | 34% | 27% |

*Tableau 1 : Pratiques de partage par statut*

La part de ceux qui refusent ce partage varie moins d'un groupe à l'autre (autour de 30% pour le dépôt, 5-10% pour la publication avec un article, 10-20% pour le téléchargement), à une exception près : les professeurs se déclarent bien plus réticents à télécharger les données d'autres chercheurs (28%) que les MCF (17%) ou les doctorants (13%).

Motivation : La part de ceux qui refusent le partage à l'avenir est limitée et assez comparable (PR 27%, MCF 31%, doctorants 28%) et la différence se trouve dans la motivation à partager : 63% des doctorants se disent favorable à un futur dépôt, ce qui est davantage que les enseignants-chercheurs (PR 45% et MCF 44%) qui sont plus nombreux à dire qu'ils devaient encore réfléchir avant de se décider. Quel type de données ? Peu de différences ici, sauf pour les données issues de projets « publiques » (mises en avant par les professeurs) et celles issues de projets collaboratifs (mentionnées surtout par les maîtres de conférences). Quel type d'archive ? Tout le monde est d'accord pour dire qu'ils favorisent un site international. Les doctorants sont les plus nombreux à parler d'une archive locale (= laboratoire, institut…).

Besoins : Les enseignants-chercheurs mettent en avant l'aide et le conseil technique, avec l'assistance pour les plans de gestion des données. Les doctorants sont davantage intéressés par les questions éthiques et juridiques, et demandent conseil pour la publication des données.

### 4.3. Différences disciplinaires

Comparer les pratiques et besoins par disciplines révèle quelques différences et particularités. Cependant, à cause du petit nombre des répondants il faut rester prudent et éviter une interprétation hâtive. Voici quand-même quelques éléments, par rapport aux domaines de l'histoire, de la psychologie, des lettres (langues et littératures française et étrangères), des sciences de l'éducation et des sciences de l'information et de la communication (SIC).

Pratiques de partage : Nous avons comparé les réponses aux trois questions sur le dépôt dans une archive de données, la publication avec un article et le téléchargement de données d'autres chercheurs. Le tableau 2 montre une expérience contrastée.

| | **Histoire** | **Lettres** | **Psychologie** | **Education** | **SIC** |
|---|---|---|---|---|---|
| ***Dépôt dans archive*** | 17% | 22% | 4% | 12% | 14% |
| ***Publication avec article*** | 52% | 39% | 24% | 40% | 59% |
| ***Téléchargement*** | 44% | 50% | 16% | 25% | 41% |

*Tableau 2 : Pratiques de partage dans cinq disciplines*

D'après les réponses, les chercheurs en psychologie et, dans une moindre mesure, en sciences de l'éducation ont à ce jour moins d'expérience avec le partage des données que ceux des autres domaines. Ils sont également plus nombreux à admettre qu'ils ne connaissaient pas ces possibilités.

Motivation et besoin : On peut observer la même tendance quand on compare les réponses à la question sur la motivation. Tandis qu'une large majorité des répondants en histoire se déclare favorable à rendre leurs résultats plus accessibles par un dépôt dans une archive de données (70%), l'enthousiasme de partage est plus mitigé en sciences de l'éducation (60%), SIC (57%) et lettres (50%) et plutôt limité en psychologie avec autant d'avis favorables que défavorables (40)%. Quant au type de données à déposer, voire le genre d'archive préférée, il n'y a que peu de différences entre les disciplines, sauf pour les lettres qui préfèrent très largement une solution internationale à tout archivage au niveau du laboratoire – sans doute le reflet de l'orientation internationale « intrinsèque » de la recherche en langues et lettres étrangères. Pas de différences significatives non plus dans l'expression des besoins. Les réponses indiquent une certaine homogénéité « SHS » avec un peu de variance autour d'une moyenne plutôt que des différences fondamentales entre disciplines. Mais de nouveau, il ne faut pas sur-interpréter ces résultats, vu la taille limitée de l'échantillon et du nombre de répondants par discipline.

### 4.4. Comparaison des résultats de Lille avec d'autres enquêtes

Afin de mieux comprendre les résultats de notre enquête, nous les avons comparés avec trois autres enquêtes sur les pratiques et besoins par rapport aux données de la recherche :

- Berlin : Université Humboldt de Berlin en 2013 ; N=499 (Simukovic et al. 2013, 2014)
- Strasbourg : Université de Strasbourg 2014-2015 ; N=644 (Rege 2015)
- LIBER : enquête européenne 2008-2009 ; N=1840 (Reilly et al. 2011, Kuipers & van der Hoeven 2009)

Voici quelques éléments d'analyse.

Taux de réponse : avec environ 15%, l'enquête de Lille se situe dans la moyenne. Strasbourg a récolté 7%, Berlin 26% mais ajoute que le taux normal pour une telle enquête sur un campus était de 13-17%.

Echantillon : l'enquête de Lille est caractérisée par l'importance du groupe des doctorants (33%) qui ne représentent que 23% à Berlin et 10% à Strasbourg. De l'autre côté, la part des enseignants-chercheurs s'élève à 49%, comparée à 60% à Strasbourg et environ 70% à Berlin.

Pratiques de gestion de données : un taux relativement bas (16%), comparé à Strasbourg (31%) ou Berlin (25-35%). Le stockage des données se fait d'abord sur l'ordinateur personnel (privé), tandis qu'à Berlin le matériel professionnel a une plus grande importance, tout comme dans l'enquête LIBER qui mentionne l'ordinateur personnel seulement en 4$^{e}$ position, après les supports mobiles et le serveur institutionnel.

Partage : idem, un taux relativement bas de collègues qui disent partager leurs données avec les collègues (34%), par rapport à Berlin (50%) et l'enquête LIBER (58%).

Besoins : l'enquête met en avant les besoins en matière de stockage et conseil pour la gestion, suivi du conseil et de l'assistance technique et juridique. C'est assez comparable avec l'enquête à Berlin (stockage, conseils techniques et juridiques) mais différent du résultat à Strasbourg qui révèle des besoins en matière financière et pour un soutien technique.

Attitude : les collègues de Lille paraissent plus prudents, avec 40% d'avis favorables pour le partage des données, où à Berlin ils étaient 60%. De l'autre côté, la part de ceux qui disent ne pas vouloir déposer leurs données dans une archives de données est assez comparable (12% à Lille, 15% à Berlin).

# 5. Perspectives

L'enquête sur les données de la recherche est la première du genre à l'Université de Lille 3. Elle fait partie d'un ensemble d'études et de projets pour développer les pratiques et compétences pour la gestion, la conservation et le partage des résultats de la recherche. Une telle démarche prendra du temps, dans la mesure où il faudra coordonner plusieurs structures, développer une offre de services à plusieurs facettes et prioriser les actions, notamment par rapport aux doctorants et jeunes chercheurs et dans le cadre du montage des projets européens. De même, il faudra tenir compte des autres projets dans le domaine des données de la recherche dans le périmètre de l'Université de Lille, sur les campus sciences et technologies, droit et santé.

Les résultats de l'enquête montrent une image contrastée des pratiques, motivations et attentes. Tous sont obligés à gérer leurs données d'une manière ou d'une autre, parfois avec les moyens du bord et une approche individuelle voire privée, parfois d'une manière plus professionnelle et collective. Une partie significative des personnes interrogées – environ un tiers – a déjà une certaine expérience avec le partage des données, d'autres sont motivées ou attendent d'en savoir davantage. Très peu de collègues semblent réellement opposés à de telles pratiques.

Les réponses laissent penser qu'une partie des différences sont liées aux disciplines scientifiques tandis que d'autres s'expliquent davantage par l'expérience des personnes interrogées. Ce sont des arguments pour une démarche différentielle, par discipline (laboratoire) et catégorie de personnel. Mais malgré la représentativité de l'ensemble de l'échantillon (15%), il faut rester prudent par rapport à une interprétation (trop) détaillée des résultats.

Quel sera la suite ? L'enquête est aussi un outil de communication et de sensibilisation, pour susciter l'attention des personnels scientifiques et préparer la mise en place d'une démarche institutionnelle. Pour une telle démarche, le groupe des 20-25% « précurseurs », transversal aux différentes catégories professionnelles et disciplines, peut servir de « fer de lance » en matière de bonnes pratiques et de modèle, avec un effet viral sur les autres personnels « pratiquants », « motivés » et/ou en attente de plus amples informations.

Par ailleurs et malgré l'anonymat du questionnaire, 35 répondants (13%) ont laissé leurs coordonnées pour poursuivre l'enquête sous forme d'une interview et aller plus loin dans l'analyse des pratiques et besoins. On pourrait donc approfondir certains aspects de l'étude par une enquête qualitative, c'est-à-dire des interviews avec ces collègues prêts à partager leurs expériences et attentes. Cette 2$^{e}$ partie de l'étude pourrait fournir les éléments d'information – bonnes pratiques, interrogations, solutions etc. – pour accompagner la mise en place progressive d'un dispositif à même de répondre aux besoins des chercheurs et laboratoires et aux contraintes imposées par les projets scientifiques et organismes de recherche.

Les trois axes de cette démarche ont été décrites plus haut – formation, assistance et conseil et infrastructure (cf. Figure 1). Un tel projet sur le campus de l'Université de Lille 3 ne parait pas seulement faisable, à condition de fédérer l'effort de plusieurs services et de prioriser les actions, mais il est tout simplement indispensable.

# 6. Références


F. André (2015). `Déluge des données de la recherche ?'. In L. Calderan, P. Laurent, H. Lowinger, & J. Millet (eds.), *Big data : nouvelles partitions de l'information. Actes du Séminaire IST Inria, octobre 2014*, pp. 77-95. De Boeck; ADBS, Louvain-la-Neuve.

S. Chaudiron, et al. (2015). *Livre blanc sur les données de la recherche dans les thèses de doctorat*. Université de Lille 3, Villeneuve d'Ascq.

COMETS (2015). *Les enjeux éthiques du partage des données scientifiques*. Comité éthique du CNRS, Paris. http://www.cnrs.fr/comets/spip.php?article123

P. B. Heidorn (2008). `Shedding Light on the Dark Data in the Long Tail of Science'. *Library Trends* 57(2):280-299. https://www.ideals.illinois.edu/bitstream/handle/2142/10672/heidorn.pdf?sequence=2

B. Jacquemin, et al. (2013). `Ouvrir les données de la recherche pour la veille scientifique. Le cas des thèses électroniques'. In *VSST'2013, Nancy, 23-25 octobre 2013*. http://archivesic.ccsd.cnrs.fr/sic_01070495/fr/

T. Kuipers & J. van der Hoeven (2009). `*Insight into digital preservation of research output in Europe. Survey report*'. PARSE insight, n/a. http://www.parse-insight.eu/downloads/PARSE-Insight_D3-4_SurveyReport_final_hq.pdf

LIBER working group on E-Science / Research Data Management (2012). `*Ten recommendations for libraries to get started with research data management*'. LIBER Association of European Research Libraries, The Hague. http://libereurope.eu/wp-content/uploads/The%20research%20data%20group%202012%20v7%20final.pdf

P. Naegelen (2015). `Données de la recherche : quel positionnement et quels rôles pour les bibliothèques ?'. In *Données en partage : enjeux et acteurs des données de la recherche. URFIST Toulouse, 15 juin 2015*. http://fr.slideshare.net/pierrenaegelen/donnes-de-la-recherche-quel-positionnement-et-quels-rles-pour-les-bibliothques

H. Prost, et al. (2015 à venir). `Hidden Treasures. Opening Data in PhD Dissertations in Social Sciences and Humanities'. *Journal of Librarianship and Scholarly Communication* 3(1).

A. Rege (2015). `Retour sur l'enquête sur les pratiques de publication scientifique et de production de données de la recherche'. In: *Université de Strasbourg, Projet AOC, COPIL* 27 mars 2015.

S. Reilly, et al. (2011). `*Report on Integration of Data and Publications*'. ODE Opportunities for Data Exchange, The Hague. http://www.stm-assoc.org/2011_12_5_ODE_Report_On_Integration_of_Data_and_Publications.pdf

R. Reznik-Zellen, et al. (2012). `Tiers of Research Data Support Services'. *Journal of eScience Librarianship* 1(1):27-35. http://escholarship.umassmed.edu/jeslib/vol1/iss1/5/

J. Schöpfel, et al. (2014). `Open Access to Research Data in Electronic Theses and Dissertations: An Overview'. *Library Hi Tech* 32(4):612-627. http://dx.doi.org/10.1108/LHT-06-2014-0058

J. Schöpfel, et al. (2015 à venir). `Dissertations and Data (Keynote Address)'. In *GL17 International Conference on Grey Literature, 1-2 December 2015, Amsterdam*.

E. Simukovic, et al. (2013). `Umfrage zum Umgang mit digitalen Forschungsdaten an der Humboldt-Universität zu Berlin'. *Humboldt-Universität zu Berlin, Zentraleinrichtung Computer- und*

*Medienservice (Rechenzentrum)*, Berlin. http://edoc.hu-berlin.de/docviews/abstract.php?id=40341

E. Simukovic, et al. (2014). `Unveiling Research Data Stocks: A Case of Humboldt-Universität zu Berlin'. In *iConference, 4-7 March 2014, Berlin*, pp. 742-748. http://hdl.handle.net/2142/47259

## Annexe – Le questionnaire

Les données de la recherche (« données brutes ») deviennent un enjeu majeur pour la gestion et la diffusion des connaissances scientifiques. Sous forme d'enregistrements factuels (chiffres, textes, images et sons), elles sont utilisées comme sources principales pour la recherche scientifique, et elles sont généralement reconnues par la communauté scientifique comme nécessaires pour valider des résultats de recherche. De plus en plus souvent, surtout pour les projets européens, il faut fournir un plan de gestion de ces données. Pour mieux aider et orienter notamment les jeunes chercheurs et doctorants dans cette démarche, l'Université de Lille 3 (SCD, GERiiCO) en partenariat avec l'Université Humboldt de Berlin et avec le soutien logistique de la MESHS lance aujourd'hui une enquête pour mieux connaître les pratiques et besoins en matière des données de la recherche. Il s'agit d'un état des lieux anonyme, à destination des enseignants, chercheurs, doctorants et autres personnels scientifiques de Lille 3. L'enquête sera en ligne jusqu'au 30 avril 2015. Les résultats seront publiés et diffusés sur la plateforme HAL-Lille3. L'enquête contient 22 questions. Répondre à ces questions prendra environ dix minutes. La participation est anonyme.

Contact : joachim.schopfel@univ-lille3.fr

### Champ disciplinaire et statut

1. Quelle est votre section scientifique ?

   Si vous n'appartenez pas à une section, indiquez votre discipline :

2. Vous êtes :

   - professeur
   - maître de conférences
   - chercheur
   - doctorant
   - ITA/ITRF/BIATOSS
   - autre

### Typologie des données

3. Quelles sont vos sources pour les données ? (plusieurs réponses possibles)

   - observations
   - expériences
   - simulations
   - photos d'objets
   - enquêtes et  entretiens
   - statistiques
   - fichiers  logs et données d'usage (web)
   - documents texte (corpus)
   - autre

4. De quel type de données s'agit-t-il ? (plusieurs réponses possibles)

- photos
- visualisations ou modèles multidimensionnels
- enregistrements audio
- enregistrements vidéo
- textes
- tableaux
- bases de données
- programmes et applications
- données spécialisées ou spécifiques

Si oui, veuillez détailler (plusieurs réponses possibles)

- analyse statistique
- corpus de texte / annotations
- données de géolocalisation (SIG)
- cartes et plans
- images satellite
- données de télédétection
- spectres
- série de mesure
- modélisation du climat
- données de santé
- questionnaires
- Autre: _____

## Pratiques de stockage et d'archivage

5. Où stockez-vous vos données ? (plusieurs réponses possibles)

    - en local, sur mon ordinateur privé
    - en local, sur mon ordinateur professionnel
    - en réseau, sur le serveur du laboratoire, du département, de l'institut
    - en réseau, sur le serveur de l'université
    - sur un serveur externe, hébergé par une autre institution
    - sur un serveur externe, hébergé par un prestataire privé (cloud)

6. Stockez-vous vos données sur un autre support ? (plusieurs réponses possibles)

    - disque dur externe
    - clé USB
    - CD, DVD
    - serveur institutionnel
    - aucun support
    - autre

7. Estimez l'espace nécessaire pour le stockage de vos données

    - < 1 GO
    - 1-20 GO
    - 21-100 GO  au milieu 20-100Go

- 101 GO-1 TO
  - 1 TO
- ne sais pas, pas d'avis

8. A quelle fréquence sauvegardez-vous vos données

   - tous les jours
   - toutes les semaines
   - tous les mois
   - tous les trimestres
   - tous les ans
   - irrégulièrement, en fonction des besoins
   - en fonction du prestataire de service ou du gestionnaire de site

9. Qui est la personne responsable du stockage, de la sauvegarde et de l'archivage des données ?

   - moi-même
   - mon assistant
   - un doctorant
   - le chef de projet
   - une personne dédiée au sein du groupe de travail
   - le personnel du service informatique
   - le personnel de la bibliothèque
   - un prestataire de service externe
   - autre

## Pratiques de partage

10. Qui, en dehors de vous, peut accéder à vos données ?

    - mon groupe de travail
    - mon institut
    - l'ensemble de l'université
    - les chercheurs de la discipline
    - les données de recherche proviennent d'une enquête publique
    - personne d'autre que moi-même
    - autre

11. Avez-vous au moins une fois :

| | Oui<br>1 | Non, pas encore mais j'en ai l'intention<br>2 | Non, je ne connaissais pas cette possibilité<br>3 | Non, et je ne le ferai pas à l'avenir<br>4 |
|---|---|---|---|---|
| stocké vos données dans une archive ouverte | | | | |
| téléchargé à partir d'une archive des données d'autres chercheurs | | | | |

publié vos données avec un article

12. Si, à la question 11, vous avez répondu «1» ou «2», veuillez indiquer le site (nom ou url)

13. Connaissez-vous des directives ou recommandations pour la gestion des données de recherche ? (plusieurs réponses possibles)

    - recommandations de mon groupe de travail / laboratoire / établissement
    - recommandations de mon organisation professionnelle
    - recommandations de l'organisme financeur externe (ANR, EU, autre)
    - je ne connais pas très bien la situation
    - autres commentaires

## Motivations et besoins

14. Seriez-vous prêt(e), sur le principe, à déposer vos données de recherche pour les rendre accessibles ?

    - très probable/oui
    - plutôt oui
    - je dois encore réfléchir
    - plutôt non
    - vraisemblablement non/ non

    Pourquoi ?

15. Quelles données de recherche déposeriez-vous ? (plusieurs réponses possibles)

    - les données de la recherche critique dont la conservation est particulièrement importante
    - les données de recherche, sur lesquelles sont basées mes publications
    - les données de recherche demandées par les pairs
    - toutes les données de recherche provenant des fonds publics
    - les données de recherche uniques, non reproductibles
    - les données de recherche issues de projets collaboratifs
    - je voudrais stocker l'ensemble de mes données de recherche
    - je n'utiliserais pas ce service
    - autre

17. Quelle archive de données utiliseriez-vous le plus probablement ?

    - une archive de données de mon laboratoire / département / institut
    - une archive de données disciplinaire et internationale
    - une archive de données disciplinaire et nationale
    - une archive de données centralisée à l'université
    - autre

18. Quels critères vous aideraient à sélectionner une archive de données appropriée ?

| | Assez probable / Oui | Plutôt oui | Incertain / ou pas décidé | Plutôt non | Peu probable / Non |
|---|---|---|---|---|---|
| Recommandations de mes collègues | | | | | |
| Recommandations de mon organisation professionnelle | | | | | |
| Evaluation ou classement indépendant | | | | | |
| Réputation de l'organisme hébergeur | | | | | |
| Je préfère prendre seul ma décision | | | | | |
| Autre (veuillez préciser) | | | | | |

Donnez-nous des exemples d'autres critères

19. Utilisez-vous l'infrastructure technique d'autres institutions (dans le cadre de projets collaboratifs)

    - oui, régulièrement
    - oui, dans le cadre de tels projets collaboratifs, de temps en temps
    - oui, mais plutôt dans des cas exceptionnels
    - non, je ne participe pas à ces projets collaboratifs

Pourquoi ?

20. De quelles offres de services aimeriez-vous bénéficier à la BU ? (plusieurs réponses possibles)

    - des conseils relatifs à la publication et la citation des données de la recherche
    - des conseils d'ordre général pour la gestion des données de la recherche
    - des conseils sur des questions techniques (par exemple, métadonnées, normes, archivage à long terme)
    - des conseils juridiques (par exemple, les restrictions d'accès, le traitement des données sensibles, l'utilisation de licences)
    - une assistance technique (par exemple, pour la publication des données dans une revue)
    - une aide pour préparer un plan de gestion des données
    - un espace d'archivage sécurisé et fiable pour mes données de recherche
    - je n'ai pas besoin de services
    - autre

## Intérêt

Intérêt pour une interview et les résultats du sondage

21. Seriez-vous prêt(e) à participer à une interview pour témoigner de votre expérience et de vos attentes dans le domaine des données de la recherche ?

    - oui
    - non

    Si oui, laissez-nous vos coordonnées (adresse électronique)

22. Êtes-vous intéressé(e) par les résultats de l'enquête ?

    - oui
    - non

    Si oui, laissez-nous  vos coordonnées