



Maison <sup>DES</sup>  
Sciences  
de l'Homme  
EN Bretagne

**Données de la recherche en SHS**  
**Pratiques, représentations et attentes des**  
**chercheurs : une enquête à l'Université Rennes 2**

**Synthèse des résultats**

**Alexandre Serres (sous la dir. de), URFIST de Rennes**  
**Marie-Laure Malingre, URFIST de Rennes**  
**Morgane Mignon, MSHB**  
**Cécile Pierre, SCD Rennes 2**  
**Didier Collet, SCD Rennes 2**



Novembre 2017

## Table des matières

Préambule.....	3
1. Les données socio-démographiques de l'enquête statistique.....	4
2. Quelles données de recherche ?.....	5
2.1 Les résultats de l'enquête statistique.....	5
2.2 Les résultats des entretiens.....	7
2.3 Synthèse sur les données collectées et produites.....	8
3. Quelles pratiques de stockage et d'archivage ?.....	9
3.1 Les résultats de l'enquête statistique.....	9
3.2 Les résultats des entretiens.....	11
3.3 Synthèse sur les pratiques de stockage et d'archivage.....	11
4. Pratiques de partage et de diffusion.....	13
4.1 Résultats de l'enquête statistique.....	13
4.2 Quelles variables sur les pratiques de partage des données ?.....	14
4.2.1 La variable générationnelle.....	14
4.2.2 Les variables du statut et des responsabilités.....	15
4.2.3 Les variables scientifiques.....	15
4.2.4 Les variables liées à la nature des données.....	16
4.3 Résultats des entretiens sur le partage des données.....	17
5. Quels besoins, quelles attentes ?.....	19
5.1 Les résultats de l'enquête statistique.....	19
5.2 Les résultats des entretiens.....	21
5.3 Synthèse sur les besoins et attentes.....	22
Liste des six propositions sur les données de recherche.....	23
pour l'Université Rennes 2.....	23

## Préambule

Comme son titre l'indique, ce document vise à présenter une synthèse des principaux résultats d'une enquête menée à l'Université Rennes 2 au début 2017 auprès des enseignants-chercheurs de cette université. Les personnes intéressées par les résultats complets pourront lire le rapport intégral et consulter les trois volumes des annexes, l'ensemble étant disponible sur HAL. Compte tenu de la richesse du matériau recueilli, de la taille du rapport et des annexes, nous avons décidé de réaliser cette synthèse d'une vingtaine de pages, pour les lecteurs pressés.

Cette enquête, lancée par l'URFIST de Rennes, en partenariat avec la MSHB et le SCD de Rennes 2, a été soutenue par les instances de Rennes 2, notamment les Vice-Présidents Recherche et Documentation et la Commission Recherche. S'inspirant en bonne partie de l'enquête menée à l'Université Lille 3<sup>1</sup>, elle a été élaborée pendant l'automne 2016 et a été ouverte de février à avril 2017, uniquement auprès des chercheurs et enseignants-chercheurs de Rennes 2. Les doctorants n'étaient pas concernés.

L'enquête visait trois objectifs principaux :

- **établir un état des lieux** : avoir un panorama aussi complet que possible de la situation des données de recherche à l'Université Rennes 2 : quelles données, quelles pratiques des chercheurs, quelles perceptions notamment de la question de l'ouverture et du partage des données, quelles réticences éventuelles, quels besoins, etc. ;
- **pouvoir élaborer une offre de services** : pour l'URFIST, le SCD et la MSHB, l'enquête visait à recueillir les besoins des enseignants-chercheurs, afin de pouvoir mieux y répondre, en termes d'accompagnement, de formation, de services, de ressources ;
- **alimenter la recherche sur les pratiques informationnelles des chercheurs** : cet objectif intéressait plus particulièrement l'URFIST de Rennes, le nouveau GIS « Réseau des URFIST » et la MSHB. Il s'agissait de mieux connaître les représentations des enseignants-chercheurs, de mieux comprendre les réticences, les points de blocage, afin d'alimenter le débat scientifique autour des données de recherche en SHS.

L'enquête s'est faite selon deux modalités :

- **une enquête quantitative, statistique**, en ligne ; un questionnaire de 32 questions était proposé aux enseignants-chercheurs de Rennes 2, découpé en quatre parties : outre des informations socio-démographiques (sur l'âge, le statut, la discipline, etc.), une partie des questions portait sur la typologie des données collectées et produites, une autre sur les pratiques de stockage, d'indexation et d'archivage, une troisième sur les pratiques et les représentations relatives au partage des données (les réticences ou les appétences), et une quatrième partie sur les besoins, les services à proposer ;
- **une enquête qualitative**, avec des entretiens d'une demi-heure, menés auprès de collègues volontaires ou sollicités par l'équipe. Plus d'une vingtaine d'entretiens souvent passionnants, très riches, qui nous ont permis d'aller plus loin dans la compréhension des pratiques.

Outre les annexes qui présentent l'ensemble des résultats statistiques, des croisements selon différentes variables et des extraits des entretiens, annexes disponibles sur HAL, nous avons également mis en consultation libre, sur l'application TableauPublic<sup>2</sup>, les résultats statistiques. En tant qu'adeptes et défenseurs du Libre Accès, nous ne pouvions pas faire moins !

---

<sup>1</sup> Prost Hélène, Schöpfel Joachim. *Les données de la recherche en SHS. Une enquête à l'Université de Lille 3. : Rapport final*. [Rapport de recherche] Lille 3. 2015. <http://hal.univ-lille3.fr/hal-01198379>

<sup>2</sup> <https://public.tableau.com/profile/urfist.rennes#!/>

# 1. Les données socio-démographiques de l'enquête statistique

Sur un échantillon de 495 enseignants-chercheurs à Rennes 2, nous avons eu **143 réponses** (dont 109 complètes), soit un **taux de 28,8 %** de répondants, ce qui est assez significatif.

Les données factuelles recueillies dans la première partie du questionnaire concernaient : la tranche d'âge, le statut, les responsabilités actuelles, le laboratoire d'appartenance, la section CNU concernée. Ces éléments (obligatoires) permettent de dresser des profils et de faire l'objet de nombreux croisements.

- **L'âge des répondants :**

La tranche d'âge des 30-50 ans représente près des 2/3 des répondants et celle des plus de 50 ans le 1/3 restant. Cette proportion est similaire à celle observée sur la pyramide des âges des enseignants-chercheurs de Rennes 2 : considéré isolément, le critère de l'âge n'influe pas sur le taux de réponse.

- **Les statuts :**

Toute proportion gardée, la distribution des répondants par statut est analogue à celle que l'on observe dans l'effectif global de Rennes 2 : 80 maîtres de conférence, 43 professeurs, 12 HDR, 3 chercheurs, 2 ingénieurs. Mis à part les chercheurs, le taux de participation s'accroît à mesure que l'on progresse dans les statuts et en sens inverse des effectifs : il est de 23 % pour les maîtres de conférences, pourtant les plus nombreux (345) et monte à 50 % pour les HDR (soit 12 sur un total de 24). La taille de l'effectif de départ n'affecte pas en elle-même le taux de participation, en revanche, ce dernier est sensible au statut et s'élève avec lui.

- **Les responsabilités :**

Le recueil des responsabilités exercées par les répondants peut contribuer à définir la logique avec laquelle chacun s'empare de la problématique des données. Les catégories prédéfinies du questionnaire semblent couvrir une large partie des situations mais il faut tenir compte du fait que les responsabilités sont susceptibles d'évoluer et de s'accumuler au cours d'une carrière. La répartition des répondants par types de responsabilités exercées donne le classement suivant par ordre décroissant : encadrement de thèse (77), responsabilités éducatives (64), scientifiques (58), éditoriales (51), politiques (36), administratives (14). Le croisement avec les tranches d'âge ne bouscule pas fondamentalement ce classement, il confirme la tendance à l'accumulation des responsabilités des plus de 50 ans avec toutefois, pour cette tranche, une baisse sensible des responsabilités éducatives (près de 50 %) et plus modérée pour les responsabilités scientifiques (10 %). Dans la tranche d'âge n'excédant pas 50 ans (celle qui comprend le début de carrière), une majorité de répondants ne possède qu'une seule responsabilité. Il faut souligner également que 40 % des répondants assurent de 3 à 6 responsabilités, des profils susceptibles de multiplier les points de vue sur cette problématique des données.

- **Les laboratoires d'appartenance :**

Les laboratoires sont tous représentés mais de manière très inégale : dans le classement des répondants par unité de recherche, le premier tiers (sur un total de 19<sup>3</sup>) concentre près de la moitié des participants (ESO, LP3C, Tempora, APP, CELLAM). Ce classement, rapporté aux

---

<sup>3</sup> L'enquête a été réalisée avant le nouveau contrat, instituant le nouveau laboratoire Arts plastiques.

effectifs, n'est pas bouleversé et la sous-représentation de certains laboratoires, dont rien ne permet de supposer un intérêt moindre pour cette thématique, demeure inexpliquée à l'issue de cette enquête. Les laboratoires les mieux représentés dans ce classement sont parmi ceux qui possèdent certes les plus gros effectifs, mais de manière tout aussi importante, la plus grande diversité des statuts. Si l'on prend l'exemple du laboratoire ESO, il se place au 7<sup>ème</sup> rang en termes d'effectifs à Rennes 2 (28) mais il occupe la première place du classement du nombre de répondants au questionnaire par unité de recherche (15), c'est également celui qui possède la plus grande variété de statuts (MCF, professeur, HDR, chercheur, ingénieur) et le seul à compter dans ses rangs 2 ingénieurs.

- **Les sections CNU :**

Une large majorité des répondants n'est rattachée qu'à une seule section CNU (87 %), seuls 10 % possèdent 2 qualifications. Le critère de la section CNU a tendance à disperser parfois ce qu'un laboratoire pluridisciplinaire rassemble autour de ses axes de recherche mais nous retrouvons la prépondérance déjà observée de certaines disciplines (psychologie, histoire, STAPS, par exemple) et la faible représentation de certaines d'entre elles (les langues et la littérature, les sciences de l'éducation).

- **Données sur les entretiens :**

Dans le prolongement du questionnaire, les répondants étaient invités à participer à des entretiens qui constituaient la seconde phase de l'enquête. Cette démarche individuelle excluait la possibilité d'un échantillon représentatif. Cependant, la répartition par tranches d'âge des participants correspond à celle du questionnaire et une majorité de laboratoires existants à Rennes 2 (13 sur 19) sont représentés dans ce recueil qualitatif. Contrairement au questionnaire, les maîtres de conférences et les professeurs y figurent à parts égales et l'on compte également 2 HDR et un ingénieur. Les responsabilités éducatives et scientifiques sont de loin les mieux représentées parmi les participants. Les déséquilibres coïncident logiquement avec ceux constatés dans le questionnaire : les unités absentes (6) sont essentiellement celles qui enregistrent le plus faible taux de réponse au questionnaire et les mieux représentées possèdent les effectifs les plus nombreux de répondants et la plus grande diversité de statuts (4 membres de Tempora, 3 d'ESO, par exemple).

## 2. Quelles données de recherche ?

### 2.1 Les résultats de l'enquête statistique

Les chercheurs collectent ou créent une très grande diversité de données « sources » (i.e. les données collectées, à la base d'un travail de recherche). Toutes les catégories, à part la 3D, obtiennent plus de 20 % de réponses. Il s'agit principalement de données textuelles (72 %), mais les données d'enquête (53 %) et les données d'observation (43 %) sont citées par presque la moitié des répondants. Cette hiérarchie recoupe celle observée dans l'enquête de Lille. Le multimédia est également assez présent (1/3 travaillent sur des données image ou multimédia). La question permettant des réponses multiples, il est intéressant de constater que 83 % des répondants manipulent plus d'un type de données, dont 42 % 4 types ou plus.

On constate un lien étroit entre certains types de données, citées ensemble dans les réponses. Au sein de chaque laboratoire, *a minima*, les chercheurs manipulent au moins une catégorie de données identique (textes pour APP, données d'enquête pour ESO...).

La hiérarchie des réponses pour un laboratoire ne respecte pas forcément le classement des réponses globales. On constate une différence entre les laboratoires en langues ou arts, lettres et communication, qui classent en premier les données textuelles, et les unités de sciences sociales et sciences humaines qui privilégient, suivant leur discipline, d'autres types de données (données d'enquête / sondage, données d'observation, données expérimentales...).

Toutefois, cette orientation disciplinaire recouvre un paysage assez nuancé et diversifié au sein d'un même laboratoire, puisque les répondants de 13 laboratoires sur 19 ont cité au moins une fois 7 types de données sur les 9 que comportait l'enquête.

L'analyse du croisement des données sources et des catégories CNU recoupe les observations sur les laboratoires : prédominance du texte pour les langues, sciences du langage, arts, histoire et SIC ; des données expérimentales pour la psychologie ; des données d'observation pour les sections ethnologie, géographie, urbanisme et STAPS ; des données d'enquête et de sondage pour la sociologie et les sciences de l'éducation, avec là aussi de multiples types de données utilisés pour chaque discipline.

Si l'on s'intéresse au support matériel des données (question 7), il s'avère que les données nativement numériques sont encore très peu importantes en quantité (10 % seulement). 45 % des données ne sont initialement ni nativement numériques, ni numérisées. Cependant, 44 % des répondants numérisent *a posteriori* les données non numériques récoltées. On aboutit donc à plus de 75 % de données numériques au final (soit nativement numériques, soit numérisées avant la collecte, soit numérisées par le chercheur après collecte).

On peut remarquer aussi le nombre assez important de « Sans réponse » (20 %) sur une question qui semble simple.

Une analyse croisée des réponses par laboratoire confirme la primauté des données non-numériques mais montre aussi des disparités entre les unités, et au sein de la même unité.

Il y a une différence de pratiques entre SHS, où les données numérisées sont majoritaires pour 8 laboratoires, et Arts-Lettres-Langues, où les données principalement manipulées sont non numériques. Cependant, il y a des usages diversifiés au sein du même laboratoire, la quasi-totalité des laboratoires citant au moins deux supports.

L'analyse des croisements avec les disciplines CNU confirme cette hétérogénéité des pratiques. Sur les 23 disciplines représentées à Rennes 2, 19 mentionnent deux supports au moins (pour un total de 122 réponses, soit 89 % des réponses). On retrouve la distinction entre SHS et Arts-Lettres. Par contre, au niveau plus fin par discipline, il est difficile de tirer des conclusions.

Contrairement à ce qui est observé pour le type de données sources, le laboratoire et la discipline CNU ne semblent donc pas discriminants pour le support matériel des données collectées, même si une logique globale se dégage au niveau des grands champs disciplinaires.

Comme pour les données sources, le texte est majoritaire pour les données résultat ; la production des chercheurs semble de prime abord plus homogène (49 % des répondants ayant répondu à la question n'ont coché qu'une catégorie).

L'analyse des données produites ventilée par laboratoire confirme d'une part la prédominance du texte pour une majorité d'entre eux (pour 13 unités sur 19), et d'autre part trace, comme pour les données sources, une différence entre champs disciplinaires.

Langues, arts et littératures produisent surtout du texte (autour de 90 % des répondants l'ont sélectionné dans ces laboratoires) et des données graphiques (notamment au PREFics) et audiovisuelles (APP...). Les sciences sociales respectent la hiérarchie globale (données textuelles, puis graphiques, puis statistiques). Les sciences humaines produisent plus de données statistiques.

Les données audiovisuelles, qui représentent 13 % des réponses, sont réparties assez largement et de façon transversale aux champs disciplinaires, entre 11 unités. Par contre, les données 3D restent l'exception.

On constate là encore une grande diversité : 14 unités produisent au moins trois types de résultats différents, et 7 en produisent quatre.

On peut noter aussi des particularités pour certaines unités : la forte production de données statistiques et graphiques au LP3C (psychologie), l'importance de la 3D pour M2S (STAPS).

La ventilation par discipline CNU confirme ce constat de singularités identifiables : prédominance des données statistiques en psychologie et en mathématiques, et des données graphiques en géographie et en SIC. Pour la plupart des disciplines, on voit la prédominance d'une catégorie (plus de 50 % des réponses), avec éventuellement une deuxième mineure. Trois disciplines font preuve d'une répartition moins tranchée (trois catégories se partageant entre 20 et 30 % des réponses) : les sciences de l'éducation, les SIC et les STAPS.

La question 10 permet de préciser la nature des données produites.

Les données textuelles sont là encore privilégiées pour plus de  $\frac{3}{4}$  des répondants. On peut remarquer également que 42 répondants produisent des bases de données et/ou des programmes et applications (soit un pourcentage cumulé de 42,5 % des réponses).

Tous ces résultats recourent ceux de l'enquête de Lille 3, à la notable exception des données de visualisation et 3D, beaucoup moins présentes à Rennes 2 qu'à Lille, consolidant une certaine homogénéité des pratiques en SHS à l'échelle macro, même si elles varient beaucoup au niveau micro d'une discipline et d'un laboratoire, voire d'un individu à l'autre.

## 2.2 Les résultats des entretiens

Les échanges confirment l'extrême variété des données, tant collectées que produites, qu'il s'agisse de leur nature ou des formats et de la manière d'en parler. Les échanges révèlent aussi l'utilisation de nombreux logiciels et formats de données propriétaires, problématiques pour un archivage à long terme. Parmi les explications avancées par les chercheurs figurent la manipulation de données différentes selon les projets de recherche, et la variable disciplinaire. A l'intérieur d'une même discipline, la langue, la période étudiée et les écoles de pensée induisent aussi le choix des sources, des méthodes et des données manipulées. Inversement certaines communautés de pratiques peuvent transcender la discipline.

Qu'est-ce qu'une donnée ? La difficulté des spécialistes à donner une définition de cette notion protéiforme fait écho à l'interrogation des producteurs eux-mêmes sur l'ambiguïté du terme et ce qu'il recouvre. A partir de quand une information ou un matériau collectés deviennent-ils une donnée ? Certains identifient information collectée, veille et données. Inversement, un chercheur souligne le caractère exceptionnel et délibéré de la collecte.

La distinction entre donnée source brute et donnée produite n'est pas non plus tranchée. La donnée n'est pas figée dans le temps ou dans des états successifs bien identifiés pour les chercheurs.

De la même façon, la frontière entre donnée de travail, hypothèse et résultat est poreuse.

La donnée, pour un chercheur en prosopographie, peut être le résultat final d'un travail, une hypothèse évolutive, ce qui est le cas aussi de l'édition de textes pour un historien.

La publication, résultat des données, est-elle elle-même une donnée ? Cette problématique, soulevée notamment par des historiens et des philosophes dont la principale production concerne les textes, se retrouve aussi dans la bouche d'un géographe, qu'on pourrait penser *a priori* plus enclin à distinguer données et publications.

Deux chercheurs en SIC et linguistique vont plus loin en remettant en cause le terme utilisé. La donnée est un construit, choix délibéré et fruit du travail du chercheur. Le travail et le regard du chercheur donnent de la valeur à la donnée, parfois insignifiante autrement.

Cette valeur est aussi mise en exergue par la conscience du long et pénible travail de construction que représentent la collecte et le traitement des données, soulignée dans de nombreux entretiens

Quelle que soit la discipline, la question de la reconnaissance du travail accompli par la propriété intellectuelle revient de façon récurrente dans les échanges, de façon transversale aux disciplines.

Cette revendication s'exprime au niveau des chercheurs individuels, mais aussi au niveau de l'équipe. La recherche collective, qu'il s'agisse de la vie des équipes de recherche internes au laboratoire ou de projets collaboratifs entre équipes, complique d'ailleurs cette question des droits d'auteur, allant parfois jusqu'à bloquer l'exploitation des données.

Le paradoxe souligné par certains, c'est que le travail de construction et de partage des données est peu reconnu institutionnellement. La publication vient comme une « *légalisation* », une « *forme officielle* » de travaux qui ont demandé sacrifices et investissements.

Paradoxalement, l'attachement aux données s'exprime davantage en creux, dans la perte, leitmotiv récurrent des entretiens, qu'elle soit réelle avec « *des expériences douloureuses de perte des données* » (incidents matériels, obsolescence des supports et formats de sauvegarde) ou fantasmée (vol).

Mais autant la perte de la donnée « vive » et active est ressentie vigoureusement et brutalement (un « *incendie* »), autant la perte de la donnée « dormante » de projets de recherche antérieurs est constatée mais vécue sans drame.

La masse et la diversité des données suscitent aussi quelques commentaires. Ce foisonnement est diversement perçu, suivant le point de vue où l'enquête se place : producteur, auteur... Les chercheurs se satisfont généralement de leur méthode personnelle, et ne semblent pas en demande.

Le format des données et leur disponibilité en ligne ont changé les pratiques.

Plusieurs chercheurs issus de disciplines différentes saluent la facilité d'accès et la simplification de leur travail apportées par la mise en ligne, mais aussi les avancées liées à l'open access et l'open data (disponibilité, gratuité). Cette disponibilité des données en ligne a un impact sur les pratiques et les lieux de travail. Certains font cependant écho d'un besoin d'orientation face à la profusion des sources.

L'émergence des humanités numériques bouleverse aussi la façon même dont se fait la science en SHS, et plusieurs chercheurs de Rennes 2 témoignent de cet impact.

Mais les données sont-elles pour autant une problématique largement partagée et débattue ? Si certains individus expriment leur conviction de l'importance du sujet, la conscience individuelle ne semble pas encore relayée par un débat plus large au sein des unités. Les données constituent encore largement un « *impensé* ».

## 2.3 Synthèse sur les données collectées et produites

Le paysage des données qui se dessine à travers l'enquête et les entretiens est celui d'une grande diversité. Si à un niveau macro, les pratiques des champs disciplinaires et des chercheurs présentent des similitudes, un chercheur individuel manipule des données variées.

La masse des données, la difficulté à distinguer les informations à archiver dans un projet et les questions relatives à la propriété intellectuelle ne doivent pas être sous-estimées.



Les échanges et résultats du questionnaire permettent de cerner les préoccupations des chercheurs et le peu de temps qu'ils ont à consacrer à ces nouvelles problématiques. Ils permettent aussi d'identifier les thématiques qui ont du sens pour eux, et sur lesquelles on peut s'appuyer :

- **Tenir compte de la valeur affective attachée aux données**
- **Proposer un recours et des informations pratiques et non une injonction théorique notamment sur les problématiques qui sont ressorties comme sensibles**
- **Cibler les interventions prioritaires**
- **Démontrer par l'exemple (perte) la nécessité de garder les données vivantes**
- **Faire des données un objet de débat dans les instances universitaires**

Comme en ont témoigné certains chercheurs, les données transforment la pratique même de la science. Les pressions des financeurs font évoluer le paysage rapidement. Il est donc indispensable de faire et de partager une veille sur ce sujet.

### 3. Quelles pratiques de stockage et d'archivage ?

Après un examen des types de données collectées et produites, l'enquête a donné l'occasion d'interroger les pratiques des enseignants-chercheurs de Rennes 2 en matière de stockage et d'archivage. Ces dernières sont-elles normalisées ou relèvent-elles davantage de méthodologies personnelles ? Des politiques d'archivage sont-elles déjà mises en place ? Et si oui, quelles sont leur portée et leur efficacité ? En traitant la question de l'archivage, l'enquête visait par ailleurs à mieux évaluer la manière dont les enseignants-chercheurs se sont emparés des enjeux de sécurisation et de documentation des données de recherche.

#### 3.1 Les résultats de l'enquête statistique

L'enquête statistique a permis d'aborder les thématiques suivantes : le stockage des données (volume, lieu et fréquence des copies de sauvegarde), la politique d'archivage et la responsabilité de cette politique, ainsi que les pratiques de description des données dans une visée d'archivage à long terme.

Les enseignants-chercheurs ont exprimé des besoins très hétérogènes en matière de volume de stockage. Ces derniers oscillent entre un espace inférieur à 1 Go (8 % des réponses) et un espace supérieur à 1 To (11 %). Environ un tiers des répondants (34 %) estime avoir besoin d'un espace entre 20 et 100 Go. Cette diversité montre qu'il est difficile de présumer des besoins réels d'espace de stockage des enseignants-chercheurs en SHS tant leurs pratiques sont variables, y compris au sein d'un même champ disciplinaire ou d'une unité de recherche. Ainsi, il serait cohérent, bien que plus complexe en termes de gestion, de mettre en place un service personnalisé de stockage. L'espace non utilisé par certains pourrait être d'une grande utilité à d'autres dont les données sont plus volumineuses. Compte tenu de l'émergence et de la multiplication de nouvelles méthodes de recherche intrinsèquement liées au numérique, la

proportion d'enseignants-chercheurs ayant besoin d'espaces de stockage conséquents est vraisemblablement amenée à croître au fil des années à venir.

Interrogés sur le lieu de stockage de leurs données, les répondants ont déclaré utiliser un stockage local et personnel pour une grande majorité d'entre eux : le disque dur externe arrive en tête (77 %), puis l'ordinateur professionnel (62 %) et l'ordinateur personnel (57 %). En revanche, les services de stockage institutionnels, proposés par l'université ou par les Très Grandes Infrastructures de Recherche, ne remportent qu'une faible adhésion : seulement un répondant sur dix a recours au serveur de l'université et 6 % des personnes interrogées ont déjà entreposé des données sur les serveurs d'une TGIR. Quant au stockage sur le *cloud*, il est adopté par 27 % des répondants et semble progressivement se développer. Enfin, 87 % des répondants cumulent plusieurs types ou supports de stockage.

Les enseignants-chercheurs témoignent de pratiques disparates en matière de copies de sauvegarde. Ces dernières sont faites le plus souvent de manière irrégulière – c'est le cas pour un tiers des répondants – ce qui est révélateur d'une absence globale de politique, collective ou personnelle, de gestion des données.

Une écrasante majorité des répondants (93 %) a notifié l'absence de politique d'archivage à l'échelle de leur laboratoire. Seulement deux laboratoires de Rennes 2 sont identifiés comme ayant développé une politique commune : l'UMR Costel<sup>4</sup> et l'équipe d'accueil M2S<sup>5</sup>. Il s'agit de deux unités qui sont confrontées au traitement et au stockage de données lourdes (données d'observation, données expérimentales et données 3D principalement). Ainsi, il semble que plus les données à stocker sont nombreuses et volumineuses, plus cela confère un caractère d'urgence à la mise en œuvre effective d'une politique adaptée.

Lorsqu'une politique d'archivage est établie dans le laboratoire, c'est toujours à un membre de l'équipe de recherche qu'en revient la responsabilité (100 %). En revanche, lorsqu'aucune politique d'archivage n'a été décidée à l'échelle de l'unité de recherche, il n'y a pas non plus de mise au point d'une politique personnelle équivalente pour les deux tiers des répondants. L'archivage étant majoritairement individuel, peu de dispositions sont alors prises pour garantir la pérennité des données produites par des chercheurs non-permanents du laboratoire (doctorants, chercheurs associés, etc.).

L'archivage à long terme nécessite la réalisation d'une description des données par les enseignants-chercheurs. Environ la moitié des répondants attribue généralement des informations supplémentaires, ou métadonnées, à leurs données. Elles sont de type majoritairement scientifique (48 %), puis administratif (25 %) et enfin technique (12 %). Lorsque les répondants associent des métadonnées à leurs données, ils élaborent le plus souvent une documentation personnelle (95 %). Ils sont peu nombreux en revanche à utiliser un cahier de laboratoire (16 %) ou à s'appuyer sur un standard de métadonnées (11 %). Cela montre d'une part que ces standards sont très peu utilisés par les enseignants-chercheurs eux-mêmes, mais très certainement aussi qu'ils sont de fait largement méconnus.

---

<sup>4</sup> Climat et occupation du sol par télédétection, composante Rennes 2 de l'UMR LETG (Littoral, Environnement, Télédétection, Géomatique), unité rattachée à l'UFR Sciences sociales

<sup>5</sup> Mouvement, Sport, Santé, unité rattachée à l'UFR STAPS

## 3.2 Les résultats des entretiens

Les entretiens ont confirmé deux éléments mis en évidence par l'enquête statistique concernant le stockage : d'une part, les volumes de données produits et gérés par les enseignants-chercheurs sont très inégaux et, d'autre part, l'emploi d'un stockage local et personnel reste prééminent pour la majorité des personnes interrogées. Bien que ce type de stockage soit largement utilisé, il présente de nombreuses limites dont les enseignants-chercheurs ne sont pas dupes : les risques de pertes de données sont nombreux (perte et vol du support, incompatibilité des formats, etc.), mais aussi il ne permet qu'une collaboration restreinte et la synchronisation des données sur différents supports est fastidieuse. Or, ces deux dernières fonctionnalités – l'aspect collaboratif et la synchronisation – constituent l'atout principal d'un autre type de stockage, le *cloud*. Néanmoins, plusieurs enseignants-chercheurs se déclarent méfiants face à ces services de stockage en ligne. Par défaut ou par méconnaissance de l'offre existante, ils se rabattent pour la plupart sur des services de type « grand public », comme Google Drive ou Dropbox, alors que les données traitées peuvent être sensibles. L'absence de transparence dont font preuve les firmes privées qui proposent ces services est problématique et n'est pas sans alerter les chercheurs sur le manque de sécurisation de leurs données.

Comme l'enquête statistique l'a montré, la mise en place d'un archivage raisonné semble encore inexistante dans les pratiques des enseignants-chercheurs. Aucune stratégie personnelle et systématique d'archivage des données numériques n'a été présentée et détaillée lors des entretiens. Néanmoins, bien que les enseignants-chercheurs n'aient pas développé ces pratiques, cela ne signifie pas pour autant qu'ils n'ont pas conscience d'un manque criant en la matière. Plusieurs freins ont été soulevés : une insuffisance de services adéquats, des lacunes dans l'ergonomie de certains outils proposés, l'aspect chronophage de l'archivage, tant pour les enseignants-chercheurs que pour les ingénieurs, et un manque de compétences spécifiques. Au moins un tiers des personnes interrogées sollicitent expressément l'appui d'ingénieurs pour les accompagner sur le plan technique dans le stockage et l'archivage de leurs données.

## 3.3 Synthèse sur les pratiques de stockage et d'archivage

Le stockage et l'archivage sont deux problématiques complémentaires qui traversent tout le cycle de vie des données. Ainsi, la conservation des données ne doit pas être pensée comme une étape finale, mais plutôt comme un processus dont la logique d'ensemble est à prévoir dès la production de ces mêmes données. Alors que les enseignants-chercheurs font part de l'importance qu'ils accordent à cette question du stockage et de l'archivage, que ce soit par le fait même d'avoir répondu à cette enquête ou explicitement lors des entretiens individuels, leurs pratiques réelles laissent sous-entendre le contraire. Comment interpréter dès lors l'écart entre la théorie et la pratique ?

Les enseignants-chercheurs ont majoritairement exprimé la demande d'être, au minimum, guidés dans la conservation de leurs données de recherche. Cela implique de prendre la mesure de la diversité de leurs pratiques effectives, de leurs besoins et de leurs préoccupations. **Quelques priorités ont pu être identifiées : la mise à disposition d'espaces de stockage importants** (quantifiés en To) **et collaboratifs**, afin de simplifier le travail en équipe, ainsi qu'un souci accru pour la sécurisation des données.

Il ressort de l'enquête que les enseignants-chercheurs de Rennes 2 assurent globalement un stockage artisanal de leurs données sans application d'une stratégie au long cours. C'était déjà

le constat qui émanait de l'enquête de Lille 3<sup>6</sup> : les pratiques de stockage et d'archivage sont globalement personnelles et réalisées avec les seuls moyens directement à disposition (stockage local sur un support privé, avec une documentation minimale et non standardisée) bien qu'essentiellement faillibles et soumis à de nombreux risques (perte ou vol de matériel, illisibilité des formats). Certaines précautions de base sont tout de même prises, comme le cumul des supports de stockage.

Les lacunes en matière d'archivage se révèlent liées à deux principales contradictions inhérentes au paysage actuel de la recherche en France. La première contradiction se trouve entre les ressources, aussi bien humaines que financières, des équipes de recherche et leurs besoins réels. Les laboratoires n'ont que trop rarement la possibilité de se doter chacun d'un matériel de stockage conséquent et sécurisé. Et quand bien même c'est le cas, encore faut-il du personnel qualifié et suffisamment nombreux pour assurer le bon fonctionnement de ce matériel et la conservation des données qui y sont stockées. Une seconde contradiction subsiste entre les différentes temporalités de la recherche. Les chercheurs sont encouragés à assurer un archivage pérenne de leurs données. Or, les jeux de données qu'ils produisent sont souvent liés à différents projets de recherche dont les durées sont limitées et dont l'enchaînement n'est pas propice à un bon archivage. Les données d'un projet précédent se trouvent bien souvent abandonnées, là encore parce que l'archivage ne constitue pas une priorité pour les chercheurs au regard de leurs autres activités.

Les enseignants-chercheurs sont incités à faire preuve de plus en plus d'autonomie dans les tâches les plus élémentaires du stockage et de l'archivage. **Cela recouvre un ensemble de bonnes pratiques, auxquelles ils doivent être sensibilisés et formés, relatives au choix des formats de données** (en privilégiant les formats ouverts), **au nommage de leurs fichiers, à la diversification des supports de stockage, à la sélection des données à conserver** (le tri faisant partie intégrante de l'archivage) **et à la normalisation des métadonnées**. Cette autonomie est importante et doit être préservée le plus possible car c'est ce qui garantit aux enseignants-chercheurs de garder le contrôle sur leurs données.

Toutefois, il ne s'agit ici que d'une première étape et cette question de l'archivage doit être impérativement traitée par la suite à une échelle supérieure. Sur le plan national, les Très Grandes Infrastructures de Recherche en SHS se sont déjà emparées de cette préoccupation avec la mise en œuvre d'une offre de service adaptée en matière de stockage et d'archivage. L'utilisation de ces dispositifs par les chercheurs reste malgré tout minoritaire d'après l'enquête, laissant suggérer que des actions de communication, de sensibilisation et de formation sont là encore nécessaires pour étendre les pratiques. Mais ces actions sont-elles suffisantes ? En effet, la méconnaissance de l'offre de service et le manque de compétences des enseignants-chercheurs en matière de gestion de données ne constituent pas les seuls facteurs faisant obstacle à la mise en place d'un archivage systématique des données de recherche. Deux autres éléments explicatifs se sont détachés au cours de l'enquête. D'une part, dans certains cas, les entretiens ont révélé qu'il devient pressant de renforcer les activités de formation par **l'accompagnement à long terme d'un personnel spécialisé et dédié**, en nombre suffisant, qui soit sur le terrain – à l'échelle de l'établissement ou des unités de recherche – et en prise avec le quotidien des enseignants-chercheurs afin de nouer de réelles relations de collaboration. L'archivage numérique est un travail à part entière, souvent invisible mais essentiel pour assurer la pérennité et la valorisation des travaux de recherche. D'autre part, les outils et services de soutien à la recherche déjà mis en place pourraient encore être améliorés afin de dépasser les contraintes qu'ils imposent – non à dessein – aux chercheurs. Parmi elles, on peut citer par exemple les diverses démarches impliquées par le stockage et l'archivage des données vécues

---

<sup>6</sup> Hélène Prost, Joachim Schöpfel, *Les données de la recherche en SHS. Une enquête à l'Université de Lille 3 : Rapport final*, op. cit., p. 15-18.

comme une charge supplémentaire pour les enseignants-chercheurs, une ergonomie des outils qui n'est pas toujours transparente ou adéquate pour les utilisateurs ou encore l'impression d'une perte de contrôle sur les données. Autant d'éléments qui peuvent paraître parfois minimes, parfois irréductibles, mais qu'il s'agirait néanmoins de prendre en compte et d'identifier plus précisément pour tenter de les dissiper.

## 4. Pratiques de partage et de diffusion

### 4.1 Résultats de l'enquête statistique

Sur le partage des données, l'enquête statistique a cherché à cerner trois éléments : les pratiques réelles, les opinions et représentations, les souhaits et besoins exprimés. Ce qui frappe le plus est sans doute l'écart, parfois les contradictions, entre ces trois aspects.

On peut résumer les pratiques de partage des données de recherche selon deux perspectives :  
- si l'on voit le verre à moitié vide, on relèvera que les  $\frac{3}{4}$  des répondants n'ont jamais déposé leurs données de recherche en libre accès, que plus de la moitié n'ont jamais téléchargé les données d'autres chercheurs (56 %), ou donné l'accès à quiconque (54 %) ;  
- à l'inverse, si l'on considère le verre à moitié plein, on observera que plus de 40 % ont déjà partagé leurs données avec d'autres chercheurs (46 %), ou téléchargé les données des autres (44 %), et qu'environ un quart des répondants (24 %) ont déjà déposé leurs données de recherche en libre accès.

Concernant les facteurs d'incitation au partage des données, on voit la prédominance très nette de deux facteurs : d'une part, la recherche de visibilité de ses travaux de recherche (65 %), d'autre part la recherche de contacts et de coopérations scientifiques (54 %). A ces deux premières réponses s'ajoute la reconnaissance de la communauté scientifique (32 %). Autrement dit, la dimension sociale de la science l'emporte largement, dans les facteurs d'incitation au partage. Mais la dimension plus scientifique occupe également une place importante : le souhait de voir se développer d'autres recherches à partir de ses jeux de données (40 %), le souci de validation des résultats (11 %), ou l'adhésion aux valeurs du Libre Accès (40 %), si l'on classe ce motif dans les raisons scientifiques. Le facteur institutionnel (l'obligation faite par le financeur de la recherche) est également non négligeable, en recueillant un tiers des réponses (33 %). On relèvera la faiblesse du facteur d'incitation financière (7 %).

A l'inverse, quels sont les principaux obstacles à la diffusion des données de recherche en libre accès, selon les chercheurs ? Sans surprise, ce sont les freins juridiques liés à la nature des données qui arrivent en tête avec 58 réponses, soit 54 % des répondants, et 22,8 % des réponses. Les chercheurs en SHS travaillent souvent sur des données personnelles, protégées. Le manque de temps arrive immédiatement en deuxième position avec 55 réponses, soit 51 % des répondants. La troisième réponse est la crainte du plagiat, qui recueille 40 réponses, soit 37 % des répondants. Le quatrième frein invoqué, cité par 29 personnes, soit 27 % des répondants, est sans doute le plus intéressant à relever : les obstacles scientifiques à la diffusion des données constituent une raison de fond, une véritable objection au partage des données.

Une répartition différente des réponses donne à voir une autre interprétation intéressante de ces résultats. En effet, les neuf facteurs de frein peuvent se répartir en deux ensembles : d'un côté, les raisons « externes », liées aux aspects juridiques, scientifiques, économiques, institutionnels... des données de recherche ; de l'autre côté, les facteurs que nous qualifierons

d'« internes », dans la mesure où ils sont davantage liés aux perceptions, aux représentations personnelles des enseignants-chercheurs. Et le tableau des deux ensembles montre clairement que la majeure partie des obstacles à la diffusion des données de recherche provient des représentations, des perceptions des enseignants-chercheurs, et non des contraintes extérieures (plus de 61 % des réponses contre 38 %). En clair, l'enquête confirme une intuition assez courante, visible également dans les résultats d'autres enquêtes (comme celle de Lille 3) : les principaux freins à la diffusion des données de recherche sont à chercher chez les producteurs des données eux-mêmes (pour diverses raisons), plus que dans les contraintes objectives liées aux données.

Un écart intéressant entre les pratiques, les représentations et les opinions peut être décelé dans la question 24, qui portait sur la position de principe sur la diffusion des données, puisque 66 % des répondants se déclarent en faveur du partage des données, contre 27 % qui s'y opposent.

Trois items d'une autre question confirment cette position de principe nettement favorable au partage des données :

- l'item « *Je suis prêt à partager mes données de recherche avec les autres* » recueille une nette majorité (67 réponses, soit plus de 65 %) en faveur du partage ;
- une très nette majorité se déclare également d'accord avec l'énoncé portant sur l'importance d'une déontologie des données dans le partage ;
- enfin, un peu moins de la moitié des répondants (48, soit 49,4 %) se déclarent « familiers avec les enjeux de l'Open access » contre 37 (soit 38 %) en désaccord.

Concernant le choix des plateformes de dépôt des données de recherche, on peut relever un autre écart entre les souhaits et les pratiques réelles, puisque si 41 % des enseignants-chercheurs souhaiteraient déposer leurs données sur des entrepôts spécialisés, seulement 6 % d'entre eux les stockent réellement sur le serveur d'une TGIR.

Au final, l'enquête statistique montre une articulation complexe entre pratiques, représentations et opinions, que l'on pourrait résumer ainsi : des pratiques de libre accès minoritaires, qui s'expliquent par différentes perceptions négatives de la diffusion des données, mais qui sont contredites par une position de principe favorable au partage des données.

Le recueil des besoins confirme que le partage des données ne constitue pas une priorité pour les enseignants-chercheurs, puisque plus de la moitié (55 %) souhaitent *éventuellement* des conseils relatifs à la publication des données, 21 % les refusent, et seulement 23 % les considèrent *prioritaires*. Et dans le classement des items par ordre décroissant de priorité<sup>7</sup> les conseils relatifs à la publication des données de recherche arrivent en dernier.

## 4.2 Quelles variables sur les pratiques de partage des données ?

Trois ensembles de variables ont été déterminées : les variables « socio-démographiques » (l'âge, le statut et les responsabilités), scientifiques (le laboratoire, la discipline) et celles liées à la nature des données.

### 4.2.1 La variable générationnelle

D'après les croisements effectués sur la question 1 (la tranche d'âge des répondants), il semble bien que la variable générationnelle intervienne, certes de manière tendancielle, dans le positionnement des enseignants-chercheurs face au partage des données. Le résultat des croisements montre en effet que les dispositions les plus favorables au partage et à la diffusion

---

<sup>7</sup> Voir l'Annexe 1 la figure 34, p. 39.

des données, notamment la sensibilité aux valeurs du Libre Accès, se trouvent plutôt chez les plus de 50 ans. En effet, dans la quasi-totalité des items, on relève des valeurs plus élevées chez les plus de 50 ans, avec des différences notables, de + 5 points (pour la prédisposition au partage des données) à + 20,3 points, écart le plus important, portant sur l'adhésion aux valeurs du Libre Accès comme facteur d'incitation au partage. Ainsi, aussi bien sur les pratiques de diffusion que sur les opinions, les représentations ou les besoins exprimés, les enseignants-chercheurs de plus de 50 ans sont proportionnellement plus nombreux que leurs collègues plus jeunes à exprimer des positions favorables au partage des données.

Dans les facteurs d'incitation au partage des données, on relève quelques différences sur les quatre items suivants : une meilleure visibilité, la reconnaissance de la communauté scientifique, l'adhésion aux valeurs du Libre Accès, avec une part plus grande chez les plus de 50 ans, tandis que le souci de pouvoir faire valider ses résultats à partir des jeux de données est plus fort chez les 30-50 ans.

Sur les facteurs de frein au libre accès, on relève également quelques différences : sur la question des compétences (informationnelles, numériques, etc.), il semble que la variable de l'âge intervienne, les enseignants-chercheurs entre 30 et 50 ans se déclarant plus compétents que les « seniors ».

#### 4.2.2 Les variables du statut et des responsabilités

Le statut ne semble pas avoir d'influence particulière sur les réponses concernant le partage et la diffusion des données. Les quelques variations observées paraissent analogues à celles observées pour la variable de l'âge, et comme le nombre de professeurs est sans doute plus élevé chez les plus de 50 ans que chez les 30-50 ans, on peut probablement mettre ces variations sur le compte de l'âge et non sur celui du statut. Ainsi, la différence la plus notable porte sur la demande de conseils relatifs à la publication des données de recherche, plus forte chez les professeurs que chez les maîtres de conférence, comme chez les plus 50 ans.

On observe également des différences sensibles dans le choix du support de diffusion des données sur le web : les professeurs choisissent plutôt les archives ouvertes (33,3 % contre 18,8 % des MCF) et les plateformes spécialisées (22,8 % contre 19,8 % chez les MCF), tandis que les maîtres de conférence semblent préférer plutôt un réseau social académique (à 15,8 % contre 10,5 % chez les professeurs), ainsi que les supports de l'université (le blog du laboratoire, à 11,8 % et surtout un serveur de l'université : 14,8 % contre 5,2 % seulement chez les professeurs). Mais les valeurs de référence restent peu importantes et il faut donc rester prudent dans l'interprétation.

Nous avons cherché à savoir si l'exercice de responsabilités avait une incidence quelconque sur les pratiques et les représentations des enseignants-chercheurs sur le partage de données. Mais au final, sur l'ensemble des questions, les écarts entre les différents types de responsabilités ne sont pas vraiment significatifs, et il est plus prudent, en l'état, de renoncer à voir un effet de la responsabilité exercée sur les pratiques de partage de données.

#### 4.2.3 Les variables scientifiques

Variable paraissant essentielle pour déterminer les positions et les pratiques sur le partage des données, la variable scientifique n'a pourtant pas fourni de réponses très probantes et claires. Il est en effet très difficile de tirer des conclusions des croisements effectués sur deux questions : la question n° 4 sur le laboratoire d'appartenance et la n° 5 sur la section du CNU. La raison vient probablement de la grande dispersion des réponses : 19 items pour le laboratoire, 29 pour la section CNU. La majorité des résultats pour plusieurs réponses sont donc très faibles et ne permettent aucune leçon généralisable.

#### 4.2.4 Les variables liées à la nature des données

Est-ce que le type de données collectées (textuelles, expérimentales, d'observation, etc.), le support (nativement numérique ou non), le type de données produites (statistiques, textuelles, etc.) ont une incidence quelconque sur les pratiques de partage ? Il semble bien que le support matériel et le type de données collectées peuvent avoir une incidence sur la pratique de partage :

- les données web et les données 3D sont les plus accessibles, de très loin ; elles favorisent également davantage l'accord pour une diffusion en libre accès (à 83,3 %) ; mais les valeurs sont assez faibles et il est difficile de généraliser ;
- à l'inverse, les images fixes (50 %), les données audiovisuelles (51,9 %) et les données textuelles (55 %) sont les moins accessibles en dehors du chercheur ;
- les données audiovisuelles (16,6 %) et expérimentales (20 %) sont le moins stockées en libre accès sur le web ;
- les données expérimentales semblent susciter le moins de prédisposition à être diffusées en libre accès (56 %).

Il convient toutefois de rester prudent dans l'interprétation, car hormis le cas des données web et données 3D, les écarts entre les autres types de données ne sont pas très probants.

Le support matériel semble être une variable plus forte que la nature même des données, car les écarts sont plus nets : les données nativement numériques sont les plus accessibles (un peu plus de 70 %), les plus stockées en libre accès sur le web (58,3 %) et les plus prédisposées à être partagées (83,3 %). Mais la faiblesse des valeurs de référence incite, là encore, à une certaine prudence.

Concernant le rôle de la nature des données produites, dans l'inclination au partage, on peut faire les observations suivantes :

- la spécificité des données 3D, de simulation : comme pour les données collectées, les données de simulation produites apparaissent comme les plus propices à la diffusion et au partage ;
- pour les données audiovisuelles, une différence sensible entre données collectées et produites : peu accessibles (51,9 %) et très peu stockées en libre accès (16,6 %), lorsqu'il s'agit de données collectées (ce qui peut s'expliquer sans doute par des raisons juridiques), les données audiovisuelles produites par les chercheurs sont largement accessibles (80 %) et davantage stockées en libre accès sur le web (33,3 %) ; de même, elles suscitent un taux d'accord très important (80 %, contre 61,9 %) pour une diffusion en libre accès ;
- pour les données textuelles, l'accessibilité des données produites semble un peu plus faible (51,9 %) que celle des données collectées (55 %) ; mais curieusement, la part des données produites, stockées en libre accès sur le web, est nettement plus forte : 48,7 % contre 25 % pour les données collectées. Mais l'interprétation de cette contradiction, qui n'est peut-être qu'apparente, n'est guère aisée ;
- Quant aux données statistiques et graphiques, elles sont les moins stockées en libre accès.

Au final, la nature des données collectées et produites, et plus encore le support matériel, semblent bien avoir un effet sur les pratiques de partage et de diffusion des données, mais qui reste difficile à interpréter. Les effets les plus probants paraissent liés d'abord à la nature nativement numérique des données, qui semble favoriser la diffusion et le partage. On discerne également la place plus importante des facteurs juridiques attachés à certains types de données (comme les images, les photographies, les films, les textes), qui peuvent freiner la mise en libre accès des données. De même, les données chiffrées, statistiques, paraissent moins favoriser la diffusion en libre accès, aussi bien en tant que données collectées que données produites. Mais la complexité des croisements, ainsi que la faiblesse de certaines valeurs, incitent à une grande prudence dans l'interprétation de ces résultats.



### 4.3 Résultats des entretiens sur le partage des données

Dans l'ensemble, la quasi-totalité des collègues interviewés déclarent leur accord de principe avec l'idée générale d'un partage, d'une diffusion des données.

S'ajoutant à des positionnements plus personnels, la variable disciplinaire peut incliner fortement au partage des données, notamment en sciences expérimentales. Par ailleurs, les évolutions récentes de certaines disciplines, notamment sous l'effet du numérique, jouent un rôle essentiel dans le positionnement face à la diffusion des données, et ce en lien étroit avec la question de l'intégrité scientifique, comme l'affirme un enseignant-chercheur en psychologie. Mais ce positionnement majoritairement favorable au partage est souvent nuancé par différentes interrogations, des doutes, sinon de sérieuses réserves.

Nous avons cherché à mieux cerner les raisons qui inciteraient à rendre les données de recherche accessibles en libre accès en interrogeant les chercheurs sur les différents items proposés dans la question 25.

La comparaison avec l'enquête statistique montre une assez grande distorsion, sur l'item de l'obligation du financeur, qui a recueilli 33 % des 107 répondants, alors que dans les entretiens, elle n'est pas considérée comme un facteur important d'incitation au partage des données.

Sur la reconnaissance de la communauté scientifique, on relève également un certain écart, mais pas de véritable contradiction. Pour la majorité des chercheurs interviewés, elle ne semble pas être un facteur suffisant ou pertinent pour partager les données, notamment parce que la reconnaissance de la communauté scientifique passe avant tout par les publications.

Sur l'item de la visibilité, qui arrive en tête des réponses de l'enquête statistique (65 % des 107 répondants), l'enquête qualitative semble contredire nettement ces résultats et montre des positions beaucoup plus nuancées.

Les deux enquêtes paraissent plus concordantes sur l'occasion de nouveaux contacts, comme facteur incitatif au partage. L'item arrive en deuxième position dans l'enquête quantitative (avec 54 % de répondants sur 107), et il recueille dans les entretiens une majorité de réponses positives, tout en suscitant également quelques réserves.

Avant-dernier item dans l'enquête statistique, le souci de pouvoir faire valider ses résultats à partir de ses jeux de données obtient encore moins de succès dans les entretiens, puisqu'aucun enseignant-chercheur interviewé ne reconnaît ce facteur comme une incitation au partage.

En revanche, les deux enquêtes paraissent cohérentes dans leurs résultats sur « le souhait de voir se développer d'autres recherches à partir de ses jeux de données », qui recueille une majorité de réponses positives dans les entretiens.

Placée également en troisième ou quatrième position dans l'enquête statistique, l'adhésion aux valeurs du Libre Accès a suscité en revanche moins de réactions dans les entretiens, même si une attitude favorable au libre accès des résultats de la science est assez largement partagée. Enfin, dans les deux enquêtes, les possibles retombées financières n'incitent guère les enseignants-chercheurs à ouvrir leurs données : dans l'enquête statistique, il s'agit du dernier item, cité par seulement 7 % des répondants ; et dans les entretiens, personne ne l'a mentionné comme facteur d'incitation.

D'autres facteurs, ne figurant pas dans le questionnaire et pouvant constituer aussi des incitations au partage des données, ont été cités par quelques collègues interviewés : l'obligation faite aux chercheurs par les revues, l'enjeu de conservation, de sécurisation des données, le partage des connaissances théoriques permis par le partage des données.

Concernant les facteurs de frein au partage, l'enquête statistique proposait quatre types de facteurs « externes » : les freins juridiques, les raisons économiques, scientifiques et

institutionnelles. Les entretiens en ont mis en lumière deux autres : l'écosystème de publication et les pratiques scientifiques disciplinaires.

Item partagé avec l'enquête statistique, le frein juridique a été abondamment évoqué lors des entretiens par de nombreux chercheurs, avec la question des données personnelles, constituée au moins de deux sortes de données juridiquement sensibles : les données personnelles de nature judiciaire et les données liées à la protection de la vie privée.

Les entretiens ont confirmé la très faible part des aspects économiques, dans les obstacles à la diffusion des données.

Placé en quatrième position dans l'enquête statistique, l'item des raisons scientifiques comme frein au partage des données a été souvent cité dans les entretiens et parfois abondamment commenté par quelques enseignants-chercheurs, pour qui il constitue le principal, voire le seul véritable frein au partage. Deux aspects, étroitement imbriqués, participent de cette dimension scientifique des données de recherche en SHS, et lui donnent toute sa complexité :

- d'une part, la question épistémologique de la définition même des « données de recherche » collectées, *i.e.* leur caractère construit ;

- d'autre part, la question scientifique, méthodologique, de leur indispensable contextualisation.

Un autre frein au partage, non prévu dans le questionnaire, a été évoqué par quelques enseignants-chercheurs : le souci de l'économie des revues, des éditeurs en SHS, la question de l'embargo dans le libre accès, etc. Le poids des revues, de la publication d'articles, d'ouvrages, de communications, etc., transparaît très nettement dans la quasi-totalité des entretiens. La diffusion des données de recherche apparaît soit comme secondaire par rapport aux publications, soit comme une possibilité, plus ou moins difficile à réaliser, parfois impraticable. En aucun cas, sauf sans doute dans les disciplines plus proches des sciences expérimentales (psychologie, géographie physique, étude du mouvement), le partage des données de recherche n'est une priorité pour les enseignants-chercheurs de SHS, d'abord soucieux de publier articles et ouvrages.

Enfin, un témoignage nous a paru emblématique d'un sixième facteur de frein à la diffusion des données de recherche : le poids des pratiques scientifiques disciplinaires, notamment le caractère artisanal, très individuel, sinon individualiste, de la recherche en SHS. Ce n'est certes pas une nouveauté et cette solitude du chercheur, notamment en histoire, a évidemment des incidences sur son positionnement face au partage des données : d'abord en termes de « culture disciplinaire », puisque la diffusion des données n'est pas du tout inscrite dans la culture collective de la discipline, contrairement à d'autres ; mais aussi en termes d'incitation, de motivation au partage.

On le voit : tous ces facteurs de frein au partage des données, que nous venons d'évoquer et que nous avons qualifiés d'externes, se conjuguent plus ou moins étroitement, pour susciter des réticences, sinon des oppositions parfois très fortes à l'idée de partage des données, réticences qu'il importe de comprendre et de prendre en compte.

Pour avoir un aperçu complet des obstacles au libre accès des données de recherche, il faut ajouter les facteurs que nous avons nommés « internes », car ils nous semblent plutôt liés aux individus eux-mêmes, à leurs représentations, leurs craintes, leurs ressentis, leurs compétences, leurs valeurs...

Sur la crainte du plagiat, l'enquête qualitative confirme ici la quantitative, car quasiment toutes les personnes interrogées ont mentionné, à un moment ou un autre, la peur, voire la hantise du plagiat, du vol de données... D'autres ont évoqué le régime de concurrence auquel sont soumis *de facto* les chercheurs, et qui justifie de prendre des précautions par rapport au pillage.

Confirmant l'enquête statistique, très peu de chercheurs ont reconnu le facteur de « l'arrière-cuisine » que l'on ne saurait montrer, comme un frein important au partage.

De même, le manque de temps est beaucoup moins cité dans les entretiens et ne semble pas vraiment constituer un frein au partage des données.

Si 26 % des 108 répondants ont cité le manque de connaissances comme facteur de frein au partage des données, presque aucun enseignant-chercheur interrogé ne l'a mentionné comme un obstacle. Cela ne signifie pas, *a contrario*, que tous les chercheurs maîtrisent parfaitement la description des jeux de données ou les modalités de leur dépôt ! Ce relatif silence ne témoignerait-il pas plutôt d'une non-prise en compte, ou d'une sous-estimation de l'importance de la description des données ?

Quant au manque de compétences (numériques, techniques, informationnelles...), cité par 19 % des 108 répondants, il n'a pas été évoqué dans les entretiens, sauf pour démentir qu'il puisse constituer un frein.

Enfin, parmi les facteurs « internes », propres aux représentations personnelles des enseignants-chercheurs, deux autres réponses faites par deux collègues témoignent de la diversité des représentations du partage des données : la réticence à toute visibilité personnelle sur internet, sentiment beaucoup plus partagé qu'on ne le croit chez beaucoup d'enseignants-chercheurs, et une sorte d'excès de modestie scientifique, avec le doute sur l'intérêt possible de ses données.

## 5. Quels besoins, quelles attentes ?

### 5.1 Les résultats de l'enquête statistique

L'enquête statistique a consacré quatre questions à l'investigation des besoins et attentes liés à la gestion des données de recherche, par une triple approche des pratiques, des positionnements et des souhaits. Il s'agissait de cerner le niveau et les perspectives d'appropriation et de sonder les chercheurs sur les aspects à privilégier dans le cadre d'une démarche d'appui.

Un premier constat, c'est que la notion de plan de gestion de données est assez largement méconnue, puisque 89 % des chercheurs interrogés ont indiqué ne pas savoir de quoi il s'agissait. Seuls 7 laboratoires sur 19 sont représentés dans les 12 réponses positives ; parmi celles-ci, 6 sont le fait de trois laboratoires de Sciences sociales : ESO-Rennes, Tempora, LiRIS ; 3 réponses positives viennent de 2 laboratoires d'Arts Lettres Communication, APP et PREFics, 2 viennent d'un laboratoire de Langues, le LIDILE, enfin 1 réponse vient d'un laboratoire de Sciences humaines, le CREAD.

La question suivante, optionnelle, interrogeait le positionnement personnel des chercheurs sur ce que pourrait être une stratégie institutionnelle en matière de gestion de données de recherche : outre des items relatifs au partage des données, elle évoquait la mise à disposition d'un espace de stockage des données sur le long terme, la recommandation, au niveau de l'université, d'un plan de gestion de données, la préconisation de métadonnées ainsi que d'un système de nommage standard de fichiers. La question proposait une gradation dans les réponses, depuis « tout à fait d'accord » jusqu'à « pas du tout d'accord ».

Sur l'analyse des réponses, le constat le plus évident est celui d'un large consensus dans les attentes en matière de stockage à long terme des données : près de 90% des chercheurs, avec un positionnement très fort des professeurs, s'accordent sur l'intérêt de pouvoir stocker les données au-delà de la durée de vie d'un projet de recherche (réponses « tout à fait d'accord » et « plutôt d'accord »). On voit que cinq laboratoires sont particulièrement intéressés par le stockage long des données de recherche : ESO Rennes, LP3C, Tempora, APP, LIDILE. La

constitution d'un corpus personnel de données accessible sur le long terme, notamment en l'absence d'une politique d'archivage dans les laboratoires, apparaît comme un enjeu important. La question qui se pose dès lors est de savoir si l'université est l'échelon adéquat par rapport à ce besoin de la recherche.

Les trois éléments portant justement sur les attentes à l'égard d'une stratégie d'université pour la gestion de données de recherche sont révélateurs. Si le Plan de Gestion de Données est ignoré d'une très grande partie des répondants, il n'y a pas d'opposition au développement d'un PGD au niveau de l'université, puisque seuls 2 répondants le refusent et que 65 % des répondants sont tout à fait d'accord ou plutôt d'accord. 17 laboratoires sur 19 se sont exprimés sur ce point et 4 sont particulièrement demandeurs d'un PGD au niveau de l'université (ESO, LIDILE, APP et LiRIS). Reste que la part des « sans opinion » (environ 1/3 des répondants) est prégnante et montre la nécessité d'un travail d'information et de sensibilisation avant toute autre démarche. Les chercheurs sont aussi peu nombreux, on l'a vu plus haut, à utiliser et sans doute à connaître les standards de métadonnées. Les réponses à la question 29 sur la préconisation par l'université d'un standard de métadonnées le confirment : en effet, les réponses « sans opinion » dominent, avec 42 réponses sur 82, traduisant probablement chez beaucoup une vision peu claire de ce que sont les métadonnées et de leur intérêt. Toutefois, 41 % des répondants souhaiteraient un standard de métadonnées établi par l'université. Rares au final sont ceux qui en refusent le principe (6 seulement). Reste que beaucoup n'expriment pas d'avis sur le sujet. La répartition des réponses par laboratoire reflète également ces incertitudes autour des standards de métadonnées.

Enfin, sur le souhait de voir l'université préconiser un système de nommage standard de fichiers, on retrouve une forte proportion de « sans opinion », soit près de la moitié des répondants. Cependant, parmi ceux qui expriment un avis, une majorité (37 % des répondants) se déclarent en faveur d'un standard de nommage.

On s'est intéressé ensuite, avec la question 30, aux besoins concrets que les chercheurs pouvaient faire remonter en termes de services, infrastructures ou outils propres à faciliter la gestion de leurs données. Cette question s'inspire très largement de l'enquête de Lille, dont elle reprend textuellement sept éléments sur huit : espace d'archivage sécurisé et fiable, conseils généraux, conseils sur les questions techniques, conseils juridiques, conseils relatifs à la publication et à la citation des données, assistance technique, aide sur le PGD. Une offre de formation sur la question des données de recherche a été ajoutée. Les réponses sont hiérarchisées selon des degrés de priorité.

Si l'on regarde l'expression des besoins prioritaires, le taux de réponse montre que, comme dans l'enquête de Lille, s'expriment **trois besoins forts** : d'abord la **mise à disposition d'un espace d'archivage sécurisé et fiable** (plus de 57 % des réponses), ce qui semble cohérent avec le souhait exprimé auparavant d'un stockage long des données et fait écho aux problèmes de perte de données et aux réticences vis à vis du stockage sur le *cloud*, évoqués dans les entretiens ; puis **les demandes de conseils techniques** (41,23 %) et **juridiques** (38,04 %) ; viennent ensuite le besoin **d'assistance technique** (37,77 %) et de **conseils généraux sur les données de recherche** (30,61 %), avec une préférence pour un accompagnement assez diversifié, plutôt informel, ciblé et personnalisé (le besoin d'une formation à proprement parler recueille 27,47 % de réponses « en priorité », l'accompagnement sur la publication et la citation des données, 23,40 %, enfin l'assistance sur le PGD, 27,16 %).

Si l'on considère les réponses favorables (besoins prioritaires et éventuels), les résultats indiquent une attitude globalement ouverte face à l'ensemble des propositions : peu de chercheurs estiment inutile une aide, quelle qu'elle soit, et beaucoup de chercheurs se sont

sentis concernés par les différentes possibilités d'accompagnement. La dernière question, qui proposait d'indiquer en texte libre d'éventuelles suggestions complémentaires, a mis en avant deux éléments originaux : d'une part le besoin d'un archivage collaboratif institutionnel, d'autre part l'intérêt et la possibilité d'une politique multi-sites et multi-établissements.

## 5.2 Les résultats des entretiens

Les besoins et attentes qui se sont exprimés lors des entretiens reflètent la grande diversité des situations de recherche ; il est toutefois possible de dégager des thématiques qui touchent de manière transversale une assez large partie des chercheurs.

Un des besoins les plus couramment abordés durant ces échanges est la **nécessité d'un archivage institutionnel** permettant de sortir des solutions de fortune ou « artisanales ».

Beaucoup observent un manque de rigueur dans le traitement des données au sein de leur laboratoire et, en raison d'un manque symptomatique de temps, souhaitent la mise à disposition de procédures très simplifiées ou l'appui de professionnels de l'information dédiés auxquels il serait possible de confier cette tâche. Simultanément, plusieurs expriment une frustration que les données, une fois la recherche effectuée, demeurent difficilement accessibles, voire inaccessibles pour une réutilisation ultérieure, empêchant ainsi leur partage et a fortiori leur valorisation. La solution résiderait pour eux à la fois dans un **lieu de stockage pérenne et sécurisé** et dans un **travail sur les métadonnées** en vue d'améliorer **l'accessibilité des données**. Une solution institutionnelle serait privilégiée si elle offrait les mêmes services que certains prestataires payants (souplesse, ergonomie, autonomie de l'utilisateur) et le recours durable à un professionnel de l'informatique dédié à ces questions recueille une large adhésion. Un cadre méthodologique et une politique d'archivage contribueraient de manière décisive à installer des pratiques normées et collectives. Le plan de gestion de données (PGD), une fois sa finalité précisée, pourrait en constituer une pièce centrale.

Une autre demande forte est celle d'un **espace de travail collaboratif**. En l'absence d'outil institutionnel, les chercheurs se tournent vers des outils propriétaires (Google Drive, Dropbox). Ils se montrent cependant réservés, d'une part sur le fait que l'offre devient payante pour des capacités de stockage étendues, d'autre part sur la politique d'utilisation de ces outils en termes de sécurité, enfin sur les garanties de conservation des données. **Un service institutionnel** aux fonctionnalités équivalentes serait favorablement accueilli, à condition de bien mesurer son champ d'action, potentiellement plus large que l'université pour des projets inter-établissements.

Circonscrire la notion de données et ses implications s'avère difficile, de là le besoin exprimé d'une **clarification sur les différents aspects de la gestion de données**. Pour répondre à ce besoin, certains directeurs d'unité préconisent un important travail de pédagogie, ou au moins « *une sensibilisation auprès des collègues* ». Le sentiment largement partagé est celui de manquer d'information face à un paysage numérique en pleine mutation, y compris en ce qui concerne le cadre réglementaire.

La problématique des données s'inscrivant dans un environnement en régulière évolution, certains enseignants-chercheurs souhaitent « *une formation continue en termes juridiques et techniques* », avec des formations sur la méthodologie ou les outils. Il faut souligner l'importance de la **demande de conseil juridique** pour l'exploitation des données de recherche, que ce soit au niveau de la collecte ou de la diffusion de données. Enfin le soutien pourrait aussi

être sollicité pour établir un plan de gestion de données dont la pratique est largement méconnue, mais dont l'enjeu est souvent perçu.

L'une des principales attentes des chercheurs concerne la possibilité de solliciter un personnel qualifié, un « référent », un « expert », qui soit en mesure de suggérer les « bonnes pratiques » liées à la gestion des données.

Le soutien passe aussi par l'établissement d'un lien plus étroit entre le technicien et le chercheur. L'expertise technique d'un ingénieur (notamment en humanités numériques) demeure fondamentale, mais l'efficacité de son action sera fonction de son implication personnelle. Il faut donc un profil alliant de solides compétences techniques à une bonne connaissance des objets de recherche, qui peut s'acquérir au contact prolongé des chercheurs et de leurs problématiques.

Au final, deux grands types de besoin se sont exprimés : un **soutien technique et informationnel**, à renforcer et pérenniser, une plus grande **coopération entre enseignants-chercheurs et personnels IST**.

### 5.3 Synthèse sur les besoins et attentes

**L'enjeu d'un stockage sécurisé des données** est parfaitement bien saisi et fait écho, aussi bien dans les réponses au questionnaire que dans les entretiens, à l'inquiétude devant le risque de perte de données. De fait, la demande la plus forte et la plus consensuelle va à la mise à disposition d'un espace d'archivage sécurisé. **L'accessibilité des données sur le long terme** et leur organisation au sein d'une structure institutionnelle adaptée est une nécessité qui a également été mise en avant. Seraient ainsi à envisager une sensibilisation à ces questions, un dispositif d'appui méthodologique, enfin la mise en place d'une infrastructure universitaire permettant le dépôt de données dans un espace sécurisé et pérenne.

**Le stockage partagé des données** constitue aussi un souhait récurrent, dans la mesure où il permettrait aux membres d'un même projet de recherche ou d'un même laboratoire de travailler sur ces données au sein d'un espace collaboratif avec des outils partagés. Reste à déterminer quel est l'échelon le plus approprié pour proposer ce type de dispositif.

Les questions posées sur les besoins et attentes ont mis en évidence une méconnaissance de certains concepts, outils, méthodes liés à la gestion des données et ont montré **qu'un travail d'explicitation et de clarification** est à engager, en préalable, en parallèle ou dans le cadre d'actions de sensibilisation, selon des modalités qui peuvent être diverses.

**L'accompagnement à la gestion de données, en tant que stratégie institutionnelle**, est un processus auquel les chercheurs se montrent largement ouverts, même si beaucoup, peu familiers de cet univers, restent sans opinion à ce sujet. De la même manière, les services proposés recueillent un assentiment assez général. Même si **l'archivage sécurisé, les conseils sur des questions techniques et les conseils juridiques** semblent trois besoins prioritaires, il semble y avoir **une attente globale d'assistance à la gestion des données de recherche**, impression que viennent conforter les entretiens. Parallèlement à des **modes classiques de formation**, s'exprime aussi **le besoin d'un « référent »** qui accompagne le chercheur de manière ciblée et personnalisée, en interne, pour diffuser les bonnes pratiques.

Il s'agira donc de concevoir un dispositif, à la fois souple et pensé en amont, à géométrie variable et doté de moyens adaptés pour répondre aux attentes exprimées et développer des conditions satisfaisantes de gestion des données de recherche.

## **Liste des six propositions sur les données de recherche pour l'Université Rennes 2**

- 1/ Créer un groupe de travail chargé de la réflexion sur une politique des données de recherche**
- 2/ Développer des services mutualisés pour la sensibilisation et la formation**
- 3/ Offrir un service juridique spécialisé**
- 4/ Mieux adapter les espaces internes de stockage des données de recherche aux besoins individuels et collectifs des chercheurs**
- 5/ Sensibiliser à la description des jeux de données et à la rédaction des Plans de Gestion de Données**
- 6/ Sensibiliser aux solutions institutionnelles de partage et d'archivage des jeux de données**

(pour le détail des propositions, voir le Rapport complet de l'enquête).