



Maison ^{DES}
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

ANNEXE 3 :
Extraits des entretiens

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

1. Quelles données ?	3
1.1 Caractérisation des données.....	3
1.2 Types de données : une distinction difficile	3
1.3 La donnée comme construction	4
1.4 Collecter des données, un lourd travail... utile à l'appropriation	5
1.5 Valeur de la donnée.....	5
1.6 La perte, révélateur de la valeur.....	6
1.7 Nombre et support des données	7
1.8 Evolutions des pratiques liées aux données.....	8
1.9 Regards sur la thématique des données	9
2. Stockage et archivage des données de recherche	10
2.1 Les pratiques de stockage	10
2.2 Les pratiques d'archivage	11
2.3 Les facteurs de freins à la gestion des données	12
2.4 Les besoins des enseignants-chercheurs en matière de stockage et d'archivage.....	13
2.4.1 Outils et infrastructures de stockage.....	13
2.4.2 Besoins	14
3. Pratiques et représentations en matière de partage des données	15
3.1 Quelles positions générales sur le principe du partage ?	15
3.2 Les facteurs d'incitation au partage des données.....	16
3.3 Les facteurs de frein au partage des données	19
3.4 Témoignages, interrogations, souhaits.....	24
4. Besoins, attentes	25

En préambule :

Nous reproduisons dans cette annexe de nombreux extraits des 21 entretiens passés avec les enseignants-chercheurs et ingénieurs de recherche de l'Université Rennes 2. Ils ont été classés selon nos thématiques, soigneusement anonymisés, comme nous en étions convenus avec les collègues ; chaque tiret correspond à un locuteur différent, et plusieurs extraits de la même personne peuvent être regroupés derrière un même tiret.

Nous remercions une fois de plus ces collègues qui ont accepté ces entretiens, d'une grande richesse.

1. Quelles données ?

1.1 Caractérisation des données

- **Diversité :**

- « *La donnée que je manipule est diverse* » ;

- « *C'est très variable, je vais vous parler de plusieurs projets si vous voulez.* » ;

- « *Je ne manipule pas le même type de données* ».

- **Différences disciplinaires :**

- « *Après, il y a des pratiques qui sont différentes selon les disciplines ; principalement en géographie et en sociologie ; et selon les disciplines, on mobilise des matériaux qui sont pas forcément les mêmes, et avec des niveaux d'appropriation des outils numériques variables d'un collègue à l'autre* » ;

- « *Je pense que c'est parce que j'appartiens à un labo d'Histoire dont les besoins ne sont peut-être pas les mêmes que ceux d'un laboratoire de sciences appliquées, où là les données sont beaucoup plus massives, beaucoup plus sensibles.* » ;

- « *La recherche, du coup, ça ne tient pas que à ce qui est juste consigné, comme ça pourrait l'être probablement en biologie ou climat ; en sciences sociales, de type [...] sociologie, c'est vrai que c'est un peu compliqué, la nature de la donnée* ».

- **Autres facteurs de diversité :**

- « *Q : Chacun travaille sur des données très spécifiques, des enquêtes, des textes... R : oui c'est ça, et puis dans nos langues, et puis selon les périodes sur lesquelles tu travailles* » ;

- « *Parce que nous, notre discipline, ce sont les Sciences du langage, mais en Sciences du langage, il y a des choses totalement différentes, et incompatibles. Mais à l'intérieur, il y a des groupes de gens qui font des choses et là ça pourrait être un échelon intéressant* » ;

- « *Ce n'est pas tant une question de discipline, mais de paradigme de recherche avec des méthodes qui ne sont pas les mêmes ; nous ici, on partage des méthodes avec les gens d'ESO, qui sont en géographie sociale, qui travaillent comme nous, ils ont dû vous tenir le même discours, ou certains sociologues ou ethnologues, etc. bien qu'on soit pas dans la même discipline ; mais il y a des grandes catégories de méthodes, c'est quelque chose de structurant ; et puis il y a aussi les cultures disciplinaires. Je pense qu'un des niveaux, si tant est que ça existe, ce serait la communauté des chercheurs et enseignants-chercheurs, de tel domaine, qui reste à définir.* ».

1.2 Types de données : une distinction difficile

- **Données de gestion et données de recherche :**

- « *Je distingue deux types de données : on a les données de recherche et les données de gestion et de pilotage [du laboratoire]. Donc on a nos données de gestion et ensuite il y a la question des données de recherche, je crois que ce n'est pas le même statut.* ».

- **Veille, informations et données de recherche :**

- « Moi j'utilise Zotero, mais est-ce que ça fait partie des informations de la recherche, sachant que c'est plutôt des PDF, des textes, enfin si ce n'est pas des corpus que je constitue ? » ;
- « On suit régulièrement ce qui se fait dans le domaine de la recherche, par le biais des publications, des éditions, des articles, et [...] de bouche à oreille, on sait qui fait quoi, qui a dit quoi, [...] La collecte des informations, en partie c'est ça et puis c'est des ouvrages » ;
- « J'ai récolté mes choses, je les trie, je travaille, [...] j'ai deux-trois journaux que je dépouille tous les jours rapidement, je me fais ma propre revue de presse, je regarde ce qu'il y a, etc. Quand j'ai des documents intéressants, je les enregistre sur PDF et je les stocke. Je pense qu'il y a plein de choses que je pourrai faire avec, mais je n'ai pas trop le temps d'y penser ».
- « Les données de recherche, on ne va pas en collecter tous les jours non plus, je ne sais pas ».

- **Données source et données résultat :**

- « C'est difficile de savoir où commence le traitement. On a le fichier brut, on a le fichier traité, et après en sortie, on a des positions au cours du temps, des gros tableaux, c'est comme ça que ça se lit, et ensuite on fait nos calculs avec MATLAB » ;
- « Il y a toute la donnée qu'on produit en continu dans le cadre de projets pour arriver à un truc qu'on peut finaliser, et toute la donnée dérivée, c'est-à-dire la donnée de travail qu'on produit entre le début et la fin, est-ce qu'on doit l'effacer, la supprimer, l'archiver, c'est très compliqué. » ;
- « Q : Il y a les données sources et les données résultats ? R : Oui, ce n'est pas notre système de catégorisation, en tout cas, c'est une différence de cet ordre-là. » ;
- « Comment situer les données, par rapport à ce processus d'élaboration dont je parlais un peu ? Est-ce que les données, c'est les données brutes, ou est-ce que les résultats, on les appelle aussi données ? » ;
- « Il y a toute cette, disons, ambiguïté du statut de données dans ce type de travail, où la vraie donnée pour moi c'est le résultat » ;
- « Moi je serais même assez partisan de mettre en ligne des choses raisonnées qui sont plus proches des résultats de recherche que des données elles-mêmes ».

- **Données et publications**

- « Et pour nous, où est la limite entre un article et les données ? » ;
- « Q : Donc s'il y avait un partage ou une diffusion en libre accès des données produites, ça porterait sur quoi alors ? R : Sur l'article » ;
- « Je travaille sur les dépôts des textes sur les archives ouvertes et en préparant plusieurs textes plutôt anciens, je me rends compte qu'il y a quand même des problématiques complexes, donc la version, est-ce que c'est la version qui a été déposée ou est-ce qu'au bout de 15 ans, il faut plutôt revoir la version parce qu'on n'a pas toujours la même... Ce n'est pas tout à fait les données mais en tout cas je découvre une série de problématiques qui sont liées ».

- **Des données évolutives :**

- « Si tout d'un coup je retrouve un fragment et je complète le nom... les dates de mes sources peuvent changer tout d'un coup vingt ans après, on décale tout de cinquante ans, on reprend, donc il faut inventer un système où on puisse gérer cette mobilité [...], c'est toutes ces étapes de déconstruction en fait des données que j'essaye de rendre apparentes... » ;
- « Q : Parce que vos données résultats, c'est l'article en fait, vous travaillez à partir de corpus, de parchemins, de photos. R : Oui et non, parce que les textes que je transcris avec les critères de l'édition scientifique, je pourrais aussi les mettre en ligne du jour en lendemain. Ca, c'est une question qui se pose d'ailleurs, est-ce que l'édition papier va perdurer, moi j'en ai fait, je continue à en faire, je ne sais pas si c'est une bonne idée, le papier ça reste inerte et une fois que c'est imprimé, c'est terminé ».

1.3 La donnée comme construction

- « Le problème, on dit des données, mais en fait pour moi, ce n'est pas des données, c'est des construits. [...] le mot « données », concrètement ça m'embête [...] ça veut dire que les choses sont données. Si

vous avez construit tout un appareil méthodologique qui permet de prélever des éléments, et donc en aucun cas ce n'est donné, mais c'est construit, prélevé, et donc du coup ça pose problème déjà, parce que les données, on a l'impression que c'est quelque chose qui est donné, physique, factuel, que les choses sont comme ça, en fait, non » ;

- « Ma collecte de données, c'est justement ce dont je parlais [...] c'est à dire tous les échanges avec les artistes, mais qui ne sont pas méthodiques, ce n'est pas des formes d'enquêtes, des formes d'interviews, c'est divers types d'échanges à diverses occasions qui peuvent avoir l'air de rien du tout mais qui sont des données sur lesquelles je m'appuie réellement ».

(Voir aussi ci-dessous Partie 3 sur les pratiques de partage).

1.4 Collecter des données, un lourd travail... utile à l'appropriation

- « Là il y a 600 pages de texte copié à partir de sources de l'époque. Là j'ai passé des semaines et des semaines à faire ça. » ;

- « J'ai passé deux ans de ma vie à traiter, exploiter une base d'entretiens, c'est du boulot aussi, mais une heure d'entretien, je sais le temps que ça représente après si on veut transcrire, exploiter, coder » ;

- « Voilà, la vraie donnée c'est ça, ça c'est des heures de boulot, d'analyse, etc. » ;

- « Cela peut me prendre quinze jours pour traiter cinq personnes d'une même famille, parce que c'est très très long en fait, il y a une recherche sur les textes qui est très très longue » ;

- À propos d'un projet ANR : « Généralement quand c'est de la collecte, de toute manière on recrute un docteur sans poste qui va faire tout le boulot de première main, qui va passer un an à saisir des données » ;

- « Comment reconnaître le travail de celui qui s'est coltiné des heures et des heures pour aller chercher tel manuscrit au fin fond des dépôts d'archives ? » ;

- « Rentrer des données prosopographiques c'est juste l'enfer » « Donc c'est un travail de bénédictin en fait » ;

- « Je travaille en fait assez "en moinesse", en moniale » ;

- « Nos données qu'on s'est embêtés à produire » ;

- « Je garde les photos pour des documents exceptionnels ou alors pour des documents, comment dirais-je, de nature sérielle : des listes, des tableaux... Mais j'aime bien travailler quand même intellectuellement la source. » ;

- « Moi j'écris dans des carnets, des cahiers, des bouts de papier, etc., ce qui fait que le numérique, ou alors c'est plutôt l'informatique, c'est un moment de reprise et de ressaisie et c'est pour ça que mes premières données, ce sont mes premiers documents de traitement de texte, ce sont des ressaisies de mes textes. ».

1.5 Valeur de la donnée

• Importance scientifique :

- « Quand on se constitue un terrain comme ça, c'est quelque chose de très précieux car c'est aussi la condition du maintien dans un niveau d'échange scientifique assez élevé » ;

- « On ne va pas filer notre stock de données pour que d'autres les travaillent. ».

• Importance affective :

- « Mes écrits, chacun dit que ça vaut quand même, c'est mon enfant en quelque sorte » ;

- « J'avais jusqu'à maintenant une sorte de vision instrumentale et utilitaire de mes propres données, ce qui m'intéressait, c'était le résultat de mes données. Mais la donnée elle-même était une sorte de structure un peu neutre que j'interrogeais. Une sorte d'impensé, un peu [...] Une fois que j'ai soutenu la thèse et publié mon livre, je me suis désintéressé complètement de ces disquettes, c'est révélateur de quelque chose » [...alors que...], j'ai quand même toujours veillé à conserver tous mes articles depuis le début, c'est-à-dire tout ce que j'ai produit [...] ça dit quelque chose aussi du fait que les données en tant que telles étaient considérées comme purement utilitaires, alors qu'aujourd'hui je pourrais continuer à les exploiter. ».

- **Importance économique :**

- « On sait qu'elles sont plébiscitées par des entreprises [...], ça me dérangerait par exemple, que ce soit réexploité sans le dire ou sans bénéfice pour l'université Rennes 2 » ;
- « C'est les photos sur lesquelles on va travailler, celles qui sont payantes, il y a une priorité pour la personne qui les a achetées, sur un contrat de recherche et ainsi de suite » ;
- « Tant qu'on n'a pas fini d'exploiter ces données-là, les collègues ont un peu de mal à les ouvrir à d'autres ; au fond, ça a coûté des sous, de l'énergie à collecter ».

- **L'enjeu de la propriété intellectuelle :**

- « Et puis il a quand même un élément de confidentialité, c'est-à-dire des heures et des heures qui ont été passées à monter une donnée, je ne sais pas où commence et où s'arrête le droit d'auteur mais il y a toute cette question quand même » ;
- « Il faut qu'il y ait une forme de, non pas de droit d'auteur, mais au moins de propriété intellectuelle, ne serait-ce que morale, c'est-à-dire établir très précisément quelle est la personne qui est allée, enfin qui est à l'origine de l'accessibilité » ;
- « Ensuite, il y a l'image brute et puis l'image traitée. Et l'image traitée, elle a une propriété, elle a un propriétaire au sens de droit d'auteur » ;
- « Je pense que les informations ou bien les productions personnelles, ça reste des droits d'auteur, entre guillemets, même si ce n'est pas formellement inscrit de cette manière-là, puisqu'il n'y a rien qui le légalise, tel un ouvrage ou autre » ;
- « Les gens sont inscrits dans une équipe, et pas seulement des agents de Rennes 2, ils sont membres d'une équipe et leurs données, leurs recherches font partie de l'équipe » ;
- « Quand vous travaillez à plusieurs sur un projet, ces données appartiennent à qui ? [...] Pour l'instant on manque d'un cadre pour savoir si systématiquement les gens qui ont récolté les données ont le même statut que ceux qui vont ensuite les analyser, les traiter et finalement produire une analyse scientifique » [...] « on a été confronté au problème il y a quelque temps. En l'occurrence, j'ai écrit un article que je n'ai jamais pu soumettre parce que je risquais de m'exposer à des problèmes, j'ai préféré laisser mes résultats de côté, c'est un peu dommage parce que c'était un projet chronophage, intéressant et scientifiquement porteur ».

- **Une absence de reconnaissance institutionnelle :**

- « Notre métier, c'est pas de créer des bases de données pour créer des bases de données, c'est de publier donc on produit les résultats, pas la boîte noire ; donc rendre publique la boîte noire, ce n'est pas une habitude, on n'est pas valorisé scientifiquement pour avoir passé du temps à publier nos boîtes noires. » ;
- « Je suis en ... section CNU, et au fond je ne peux pas habiller sur l'outil. C'est un peu paradoxal parce que cela me prend énormément de temps [...], j'essaye de trouver ce qui va me permettre de rédiger le plus vite possible un mémoire, là où je pourrai travailler en rentrant le moins de données possible ».

1.6 La perte, révélateur de la valeur

- **Perte d'années de travail :**

- « Dans les manipulations, une sauvegarde de l'ordinateur a écrasé ce qui était déposé sur le disque, donc c'était une perte vraiment considérable pour moi. C'est comme après un incendie, je repartais à zéro » ;
- « Moi, j'ai une collègue de Tarbes, son ordinateur s'est scratché, il ya trois mois, elle a perdu quatre ans de travail, parce que les DSI ne veulent pas fournir d'espaces de stockage » ;
- « Ce n'est pas externalisé quelque part. Donc si on me vole tous mes ordinateurs, tout est parti. » ;
- « Comme mon ordinateur portable est l'ordinateur sur lequel je travaille [...] si demain je me le fais piquer, je perds un mois de travail » ;
- « Et si mes ordinateurs tombent en panne ou me sont dérobés, c'est disparu. ».

(Voir aussi ci-dessous : **Pratiques de stockage et d'archivage**)

- **Données dormantes / inaccessibles :**

- « Après j'ai eu des déboires d'enseignant-chercheur, [...] qui légitiment une réflexion sur les pratiques. [...] j'ai perdu toutes les données de ma thèse [...] J'avais fait une base de données de près de 1000 [enregistrements...], des données croisées. Sauf qu'à l'époque [...] j'avais enregistré et sauvegardé ça sur disquette. Et aujourd'hui, j'ai plus les moyens de convertir ça. Données pas perdues physiquement, mais inaccessibles. Jamais pensé à les convertir » ;
- « Avant même que les équipes de recherche n'existent, on avait déjà construit des bases de données multilingues [...], c'est un gros dossier que l'on a sur un vieux serveur qui ne fonctionne plus, on sait plus très bien qu'en faire alors que c'était quand même de la recherche de qualité avec beaucoup de données en plusieurs langues, [...], cet outil là n'est plus utilisable. » ;
- « J'ai deux types de sources vidéo, j'en ai qui datent de la fin des années 90, début 2000 et je crois que je n'arrive même plus à les lire maintenant parce qu'il y a les codecs qui ne fonctionnent plus » ;
- « Cet archivage à long terme des données, si on ne le fait pas, on n'a plus aucune trace des travaux qui ont été faits par nos doctorants par exemple mais aussi de nos vieilles manips à nous qu'on a dans les cartons bien sûr. Et c'est des choses qui ont un impact quand même, parce qu'on se base quand même sur les thèses de nos doctorants pour dire "ça n'a pas été publié mais j'ai un doctorant qui a fait ça il y a quelques années", ce sont des choses qui restent en fait, qui ont un impact à long terme et pour lesquelles les données ont vraiment une durée de vie extrêmement courte [...] Je pense qu'il y a un gâchis à ce niveau là quand même » ;
- « Je dirai que globalement avec le temps, on perd le matériau issu du recueil initial » ;
- « On a toute une série de données, d'enregistrements, [...] qui sont sous des formes de bandes, qu'il faudrait numériser ; tu te rends compte de ce que ça représente comme travail ! Comment tu fais ça ? C'est un problème d'exploitation des données [...] Sur le nombre de choses qu'on a faites, il y a des tas de choses qui se perdent, qui sont du matériel. Mais en même temps, on se dit pourquoi faire ? [...] il faut avoir un projet par rapport à ça, il ne faut pas exploiter pour exploiter, mais il y a des mémoires qui vont se perdre, si tu n'as pas le personnage, l'intérêt ... » ;
- « J'ai des tas de stocks de photos sous forme de diapos et je voulais toutes les scanner en fait mais il y en a plusieurs milliers ».

1.7 Nombre et support des données

- **Le foisonnement des données :**

- « Parce que nos données explosent, on a de plus en plus de vidéos, [...] les fichiers deviennent de plus en plus lourds, en plus avec les années qui passent, moi je stocke des choses maintenant depuis 20 ans » ;
- « Pas de données massives pour l'instant ; plutôt du little data, mais très varié » ;
- « Cela ne veut pas dire pour autant qu'on n'a pas énormément de données - pas en poids mais plutôt en terme de foisonnement. [...]il s'accumule sur mon disque dur, ça s'accumule, ça s'accumule, et puis... En fait c'est compliqué aussi après d'y revenir parce qu'on ne fait pas le tri [...] par peur d'effacer quelque chose d'important ou qui pourrait nous servir, on garde, mais derrière, ça implique un foisonnement de données qui nous empêche de revenir » ;
- « Et quand je discute avec mes collègues de psycho, dans toutes les disciplines, on voit bien que l'on nous demande de plus en plus de données [...], les enquêtes sont de plus en plus encombrantes. Il faut qu'on apprenne à travailler ensemble et à traiter ensemble cette masse de données ».

- **Format des données :**

- « Je pense que les données, il y en a toujours eu une multitude, une diversité, je ne pense pas qu'il y en ait plus aujourd'hui. Par contre, elles sont pour beaucoup numériques, et là, ça change énormément la donne. [...] Mais après pour moi, ça n'augmente pas forcément les données sur lesquelles je vais pouvoir travailler, moi, dans la mesure où une donnée numérique, elle n'a pas de sens si elle ne peut pas être mise en lien. » ;
- « La plupart est numérique, après ce qu'on va avoir en format papier, ça va être tout ce qui est consentement de participation ».

1.8 Evolutions des pratiques liées aux données

- **Open access et open data :**

- « On est très contents que les collectivités territoriales rendent publiques leurs données. J'ai souvenir d'une époque où l'on essayait de monter des bases de stockage d'informations, de mutualiser les données qu'on arrachait à l'INSEE, à l'époque où l'information était rare et chère. Maintenant comme tout est plus ou moins public, ça a beaucoup facilité notre travail » ;

- « Il y a certaines choses comme Gallica, ou d'autres... RetroNews pour la numérisation des périodiques, qui sont des outils absolument essentiels. Qui sont mis gratuitement en plus au service du public. Et ça, ça fait avancer la recherche. » ;

- « Quand j'ai fait ma thèse il y a longtemps à Paris, je travaillais dans une bibliothèque où toutes les sources étaient là. Donc je passais mon temps à les sortir des gros corpus, etc. Et donc heureusement que maintenant tout est accessible en ligne » ;

- « Je sais que j'aimerais bien travailler sur de grandes enquêtes nationales, je sais que maintenant ils ouvrent beaucoup plus, on peut plus facilement récupérer des données pour faire des analyses secondaires de statistiques effectuées à l'échelle nationale ou autre, je trouve que c'est une avancée. » ;

- « Alors pour l'instant en France, les photos aériennes sont en accès quasiment toujours gratuit sur Géoportail, sauf les plus récentes, mais ça, ça va. Par contre dans d'autres pays, il faut les acheter, les scanner et tout de suite, ça occupe de la place ».

- **Changement de pratiques :**

- « J'ai toujours plaisir à aller dans des Archives et à travailler sur des documents papiers. Après, peut-être de moins en moins envie de faire des retranscriptions. C'est-à-dire que je prends beaucoup de photos maintenant et je fais ça chez moi. » ;

- « La SNCF, on ne cherche plus à stocker leurs données puisqu'elles sont accessibles. [...] Rennes, l'agglomération, la Métropole, on ne cherche plus à développer des espaces de stockage puisqu'on a accès directement à leurs données. Pour nous, c'est une logique différente : on va se servir quand on a besoin. » ;

- « Pendant un moment, j'ai archivé les données de houle, maintenant elles sont en accès public donc je n'ai plus besoin de le faire mais... ».

- **Un impact générationnel :**

- « Je ne sais plus quand est né Gallica [...] mais ça a quand même été décisif... Et je m'aperçois qu'en fait les gens, ou des gens plus âgés, ne l'utilisent pas. C'est-à-dire que moi, pour mon directeur de thèse ou pour d'autres personnes, je passais pour quelqu'un qui trouvait des choses, plein de détails, alors qu'en fait c'était ultra facile d'aller sur Google Books ou Gallica, de faire de la recherche en plein texte » ;

- « Toutes les générations de chercheurs n'évoluent pas à la même vitesse. Il y a des collègues plus âgés qui travaillent encore à l'ancienne » ;

- « Il y a peut-être une question aussi générationnelle, c'est à dire que là on est en train de travailler pour des chercheurs qui ont 50- 40 ans, 35 ans aujourd'hui ils ne sont pas très... et puis ceux qui arrivent n'auront même plus besoin de ces services ».

- **Un besoin d'orientation :**

- « Le problème aussi de l'informatique, moi je le vois bien, c'est qu'on a accès à tellement de choses, ça devient tellement énorme, que en fait tu te noies ; j'ai un peu l'habitude du truc, mais il y a des moments, je me dis mais ce n'est pas possible, et comme on n'arrive pas ; pour moi, c'est la forêt, et en plus ça va très très vite. » ;

- « Le premier besoin, c'est celui dont je vous parlais, c'est la clarification, il faudrait mieux pouvoir s'orienter dans toutes les plateformes qui existent, dans tous les types de programmes, parce que ne serait-ce qu'en histoire médiévale [...], il y a une telle profusion de plateformes qu'on finit par ne plus savoir par quel bout les prendre, et là je parle d'un point de vue d'utilisateur, et pas d'un point de vue

de producteur de données, parce que quand on est producteur, on maîtrise, on sait dans quelle structure on s'inscrit. » ;

- « Dans le domaine du monde arabe, il y a les sources occidentales, entre guillemets, en caractères latins, il y a les pays arabes, les universités arabes, les mécènes arabes, parce qu'il y a des sites que les étudiants ne connaissent pas, il y a des sites dans lesquels on trouve des manuscrits arabes, du dixième, douzième siècle que les mécènes qui ont les moyens mettent à disposition des étudiants ou des chercheurs et ça s'est développé sur les dix dernières années ».

- **Le renouvellement de la science par les données :**

- « En tout cas, pour moi effectivement les données permettent de faire des thèses différentes. » ;

- « Et puis après effectivement, tous ces logiciels comme Gephi, Paraview, c'est là où on va vers les humanités numériques, c'est que ça m'a sorti de ce qu'on a appelé la lecture de près pour aller vers d'autres pratiques de géographe, vers la spatialisation des données. Donc ça, ça a bouleversé ma façon de travailler ; ça a radicalisé ma façon de travailler » ;

- « J'aime bien faire de la recherche en plein texte, ça c'est important pour moi. Ca, ce sont des choses qu'on ne pouvait pas faire avant, des recherches en plein texte ou des recherches en plein texte de mots-clés dans les oeuvres. Avant on était quand même tributaires de la mémoire physique et de la lecture » ;

- « On passe des revues payantes papier, qui coûtent cher à diffuser, aux revues gratuites sur internet à large spectre de diffusion où on peut stocker beaucoup de données, ce qui fait que de plus en plus, on peut diffuser des ensembles de données plus importants et de plus en plus publier des choses qui ont moins d'impact scientifiquement mais qui sont plus intéressantes méthodologiquement ».

1.9 Regards sur la thématique des données

- **L'importance du sujet :**

- « Quand j'ai vu l'enquête, je me suis dis : il faut vite répondre, c'est quand même dans l'air du temps. » ;

- « Les données vont être essentielles » ;

- « C'est une question que l'on se posait moins il y a 10 ans : la mise à disposition des données de la recherche. Aujourd'hui, c'est aussi une demande à la fois institutionnelle et de nos étudiants ».

- **Une problématique encore peu présente dans les unités :**

- « Q : C'est une problématique évoquée entre vous et vos collègues ? R : On commence à en parler, oui. Un petit peu. [...] J'essaye de faire un peu de sensibilisation pour qu'on engage des bonnes pratiques, et qu'on aie une réflexion, ça commence. C'est hétérogène en fonction des collègues et des projets, mais ça prend quand même, il y a des choses qui sont intéressantes ou encourageantes. » ;

- « C'est des questions sur lesquelles on ne s'est pas forcément interrogé. Q : on met le doigt sur des problèmes. R : oui certains points, nous on commence à se poser la question » ;

- « Entre chercheurs, je n'ai jamais abordé cette question là, même récemment. On n'a pas du tout de politique de labo là dessus. Entre collègues, on n'a jamais vraiment discuté de cela. » ;

- « Maintenant je ne sais pas si nous, nos données qui sont produites dans des petits laboratoires ont vocation à être ouvertes. Ce n'est pas quelque chose dont on discute beaucoup et sur lequel on réfléchit beaucoup ; c'est quand même assez récent. » ;

- « Je pense que l'on a pas mal de besoins, en même temps on n'en a jamais discuté en labo » ;

- « Là il se trouve que je suis dans un moment particulier de ma vie à l'université, parce qu'on a une équipe qui s'est scindée ; donc on reconstitue une équipe et la question du numérique, de la gestion des données, de ce qu'on doit faire ne se pose même pas. [...] Enfin c'est assez surprenant, ce n'est pas un élément structurant, une question structurante de l'équipe, alors qu'en fait il y a des chercheurs qui s'intéressent au numérique. Donc ce n'est pas si évident que ça. ».

2. Stockage et archivage des données de recherche

2.1 Les pratiques de stockage

- **Le volume de données à stocker :**

- « Ça n'est pas négligeable, de temps en temps, il faut que je nettoie, je suis souvent en limite de capacité. Là j'atteins une limite que je n'ai pas connue auparavant. » ;

- « Quand je discute avec mes collègues, dans toutes les disciplines, on voit bien que l'on nous demande de plus en plus de données, ça devient de plus en plus volumineux à stocker, les enquêtes sont de plus en plus encombrantes. Il faut qu'on apprenne à travailler ensemble et à traiter ensemble cette masse de données. » ;

- « J'ai des fichiers Word qui ne prennent pas beaucoup de place » ; « Donc pour ce qui est du stockage, je n'ai jamais eu vraiment de difficultés. Encore une fois malgré tout, je limite mes investigations en matière de photo. Ce qui là pèse beaucoup plus lourd. » ;

- « Tous les enregistrements vidéos que j'avais fait, je ne sais pas...ils sont sur un disque dur [...] effectivement ça pèse un peu, mais ce n'est pas non plus... je vis très bien avec mon disque dur. ».

- **Le choix du lieu de stockage :**

- « J'ai un TimeMachine sur mon Mac » ; « J'ai une Dropbox, mais sur lequel je ne mets que des fichiers que je dépose temporairement [...]. Et dès que j'en ai plus besoin, je le retire exprès, pour pas que quelqu'un d'autre puisse y avoir accès hors de ma vigilance. » ;

- « Je sauvegarde sur disque dur. Je n'utilise pas les réseaux de la fac. » ;

- « Je fonctionne avec un ordinateur fixe qui a une grosse capacité de stockage, un portable qui me suit partout et 3 disques durs externes. Voilà, je bouge beaucoup en plus. J'ai toujours un disque dur externe dans le sac. » ;

- « Toutes mes données sont sur trois disques : le disque dur de mon portable, le disque dur externe qui est chez moi et le disque S. » ;

- « J'utilise des espaces sur le disque S de la DSI et sur disque dur externe. » ;

- « Pour le labo, on a un NAS, un disque dur en ligne, en fait c'est deux disques, il y a une copie de sauvegarde, si jamais un disque venait à griller, moi j'utilise en plus les outils de l'INRIA. » ;

- « Je stocke sur mon ordinateur, sur mon disque dur et après, il y a des documents qui sont partagés via le Drive. ».

- **Les copies de sauvegarde :**

- « Moi ce que je fais par exemple, au fur et à mesure que je passe d'un ordinateur à l'autre, je transfère évidemment de façon automatisée tout ce que j'avais sur l'autre et je vérifie que les fichiers que je veux garder sont restés lisibles ; et s'ils ne sont pas restés lisibles, je cherche une solution pour les rendre lisibles. » ;

- « Je ne fais pas les sauvegardes ni tous les jours ni toutes les semaines, ça c'est pas bien. » ;

- « Oui je sauvegarde. En fait, je les ai en trois endroits différents, et même en deux endroits spatialement différents. Et même maintenant en quatre, parce que là il se trouve que comme j'arrive au bout d'une recherche très lourde, j'ai tout mis sur une clé et j'ai demandé à une collègue de laisser la clé dans le grenier de sa maison de vacances. ».

- **La synchronisation des données :**

- « Tous les jours il y a au moins une ligne, mais en général c'est plusieurs pages, plusieurs paragraphes qui sont transformés, qui sont rajoutés, etc. Et donc la question des sauvegardes effectivement pose problème. [...] Je ne dispose pas d'un système global capable de prendre l'ensemble de mes données et de les gérer, même si je dispose du Time Machine. ».

- **Les pertes de données :**

- « J'ai perdu par exemple – je sais où elles sont mais elles ne sont pas convertibles – j'ai perdu toutes les données de ma thèse, des données relativement lourdes, j'avais fait des bases de données [...]. Sauf qu'à l'époque, j'avais un PC, pas portable, j'avais enregistré et sauvegardé ça sur disquette. Et aujourd'hui, je n'ai plus les moyens de convertir ces données. Elles ne sont pas perdues physiquement, mais inaccessibles. Je n'ai jamais pensé à les convertir. » ;

- « La collègue avec laquelle je travaille et qui est chargée des mises à jour, etc. mais aussi des mises en page des divers documents, on lui a volé son ordinateur. Donc toutes les données sont parties » ; « Là c'est une perte aussi irrémédiable, parce qu'il y avait des choses qu'elle avait préservées, mais pas tout. Donc voilà vu du côté des histoires dramatiques pour lesquelles il faut sans doute penser aux réponses. » ;

- « On est confrontés, sur le plan pratique, sur l'archivage des questionnaires papier, dans le cadre de certains programmes, je pense au programme XXX, quand les chercheurs ont obtenu le financement pour ce programme, il y avait une clause selon laquelle il fallait garder les formulaires papier 5 ans après la fin du projet » ; « Les réponses sont saisies sous format numérique, mais on a encore ces formulaires papier qui encombrant nos armoires » ; « Je dirais que globalement avec le temps, on perd le matériau issu du recueil initial. » ;

- « J'ai un archivage de mes données sur un disque dur. Si vous me volez mon ordi portable et mon disque dur externe qui est là et celui qui est à la maison, j'aurai perdu 6 ans de recherches avec aucun moyen de récupérer ces données là. Donc c'est plus de ça qu'on a peur. » ;

- « Q : Vous n'avez jamais eu de crash ? » « R : Si, si et grave et il y a longtemps, il y a vingt ans maintenant à peu près, quand j'étais en DEA. J'avais mon premier portable et il a lâché huit jours avant la soutenance, à l'époque la récupération de données c'était une histoire. Donc j'avais soutenu fin juillet. Donc j'ai connu... Quand j'étais en thèse, j'étais vraiment obsédé, j'avais acheté un graveur, je gravais tous les jours. Maintenant, s'il y a un pépin purement électronique, les données sont sauvegardées, on espère ne pas perdre grand chose. ».

2.2 Les pratiques d'archivage

- **Sur la question d'une politique institutionnelle :**

- « Pas de politique de laboratoire pour l'instant sur la façon dont on procède, ni pour la collecte, ni pour l'archivage, ni pour la valorisation » ; « Chacun fait comme il peut ou comme il sait faire ou pas faire. » ;

- « La question de l'archivage... On ne la traite pas pour l'instant. ».

- **Les politiques personnelles d'archivage :**

- « Sur l'archivage, je me rends compte que je suis nulle en archivage, il me faudrait un cours de rangement. Je mets des choses de côté, et je mets "À trier". » ;

- « Je me rends compte que j'archive mes projets de façon pas très ordonnée. C'est à dire qu'un projet se termine, on passe à un autre, je l'archive tant bien que mal. En fait l'archivage, c'est comme un enterrement, c'est à dire que si on doit y revenir, c'est presque une exhumation, c'est vraiment très dur. Et donc ça, ça me peine un peu parce qu'en fait on n'est pas dans une logique de valorisation des travaux antérieurs. » ;

- « Justement, quand on stocke ses données sur Dropbox, dès que la Dropbox est fermée, il n'y a plus rien. » ; « Donc c'est vrai aussi qu'il y a une bonne partie des données qu'on collecte qui ne sont pas stockées de façon viable à long terme. ».

2.3 Les facteurs de freins à la gestion des données

- **Le manque d'appropriation des outils existants :**

- « Il y a le disque S¹, mais en tout cas, quand j'essaye d'y accéder par l'ENT, avec mon Mac portable, je n'arrive jamais à transférer un fichier dessus » ; « Il doit y avoir des problèmes de compatibilité des formats ou il n'a pas été pensé pour ça, c'est tout. » ;

- À propos du disque S : « J'estime que ce n'est pas assez ergonomique, que ça suppose des bonnes connexions partout, mais je me trompe peut-être. » ;

- « Il faut aussi qu'il y ait aussi une forme de souplesse, d'ergonomie » ;

- « Encore une fois, les outils existent déjà, mais ils ne sont pas forcément très efficaces, ou en tout cas tout le monde n'y adhère pas d'emblée. ».

- **Le manque de temps :**

Pour archiver :

- « Je ne stocke pas tout, parce que je ne suis jamais à jour, c'est comme la corde sur une poulie, ça continue à tourner, à peine on sort la tête d'un article ou d'une intervention, qu'on a à subvenir à une autre demande en quelque sorte. Bien sûr, sur le temps, etc. » ;

- « L'archivage, l'entretien d'une année sur l'autre, la réactualisation des outils informatiques, là ça dépasse un peu les compétences et le temps que les chercheurs ont à consacrer à ça. ».

Pour se former :

- « Mais je n'ai pas forcément le temps d'aller à la formation, et des fois en bidouillant, tout seul, ça prend un peu de temps mais je vais m'en sortir. » ;

- « Alors malheureusement, et bien là il y a une formation à l'URFIST vendredi... donc il y a plein de choses qui sont proposées, mais par manque de temps... Quand on n'est pas juste chercheur, quand on est enseignant-chercheur avec des tâches administratives, on n'arrive pas à tout faire, et donc la dimension formation, voilà... » ; « Je ne sais pas s'il y a une question de communication d'un côté, et puis aussi un peu de manque de temps de notre part, et du coup comme on va au plus pressé, il y a plein de choses qu'on voit passer mais on n'ouvre même pas le mail, parce que ce n'est pas possible. Donc je pense qu'il y a plein de choses... et puis ça se développe de plus en plus, donc c'est vrai que Rennes 2 offre énormément de choses, plein d'outils, plein de choses... donc ça c'est bien. Nous, on manque de temps... ».

- **Les habitudes de travail :**

- « Il existe peut-être, sans doute, des systèmes meilleurs et des possibilités de les gérer autrement, mais moi dans ma petite pratique quotidienne, je suis un peu tiraillé. » ;

- « Par exemple, puisque je suis sur Mac, j'utilise ce qui s'appelait Sherlock autrefois, Spotlight maintenant, on fait "pomme-F" et puis on trouve tous les fichiers ou il y a tel mot par exemple. Est-ce que je pourrai faire ça en ligne à distance sur un site de stockage extrêmement massif ? Si je peux pas le faire pour moi, il n'y a pas de gain. » ;

- « Il est certain que plus de formations créent aussi des besoins. Là je tourne, je suis un peu stagnant. Mais je n'ai pas l'impression de limiter mes investigations intellectuelles par défaut de compétences techniques. Après il est très possible que si je prenais le temps pendant des mois de me former, j'aurais des idées supplémentaires qui viendraient enrichir mon travail intellectuel. Mais aujourd'hui je suis intuitivement convaincu que ce ne serait pas considérable. Du coup, vu le peu de temps qu'on a, c'est un effort que je ne fais pas. ».

- **Le refus de prendre en charge la gestion de ses données :**

- « Mes données sont stockées sur un disque dur et il est dans mon bureau [rires]. Il est mis à jour une fois par an. Et franchement je ne me vois pas prendre en mains un logiciel quelconque qui permettrait

¹ Le disque S est un espace personnel attribué aux personnels de Rennes 2 sur l'intranet.

que je fasse la mise en partage en direct, je n'ai pas le temps, franchement je n'ai pas le temps ! J'ai appris beaucoup de choses, ça je ne veux pas l'apprendre. ».

2.4 Les besoins des enseignants-chercheurs en matière de stockage et d'archivage

- « Ce qui me paraît clair, c'est qu'on a tous le même constat : on manque d'outils, on manque de soutien pour bien gérer, administrer nos données. ».

2.4.1 Outils et infrastructures de stockage

- **Stocker de grands volumes de données :**

- « Il faut plus de place pour stocker les photos ». ; « Voilà, mon disque S, il est plein. Il est complètement plein. ».

- **Travailler en collaboration :**

- « Cette demande d'espace collaboratif est très forte et régulière » ;

- « On a des besoins d'outils collaboratifs pour arriver à travailler en projet avec des données multiples qui viennent s'agréger. On est obligé de passer par des outils type Google ou autres et cela pose de gros problèmes [...] Le vrai problème que l'on a en ce moment, c'est sur Dropbox, si je veux augmenter ma taille de stockage, je suis obligé de payer de ma poche puisque l'université ne veut pas prendre en charge, ce que je peux comprendre. Mais en même temps, on n'est pas capable de donner des outils pour travailler... Parce que je n'ai pas le choix, je suis obligé de collaborer avec des collègues en Suisse ou ailleurs. » ;

- « Oui, quand je bosse avec XXX, c'est tout de suite des gros fichiers, c'est des données assez lourdes... Et bien soit je me déplace et ça me donne l'occasion d'aller lui dire bonjour, soit je lui mets sur WeTransfer et hop c'est bon. Si tu veux, on n'a pas besoin de travailler en temps réel sur le même jeu de données. ».

- **Sécuriser ses données :**

- « J'ai des collègues qui ont ouvert une Dropbox, on a commencé à l'utiliser et puis on s'est ravisé, on a changé d'avis. On l'utilise pour la coordination des TD entre collègues de première année mais je n'irai pas plus loin. » ;

- « Je me suis questionnée justement pour stocker aussi une partie des données « sur un nuage », et finalement je n'ai pas franchi le pas. Je pense que je ne sentais pas suffisamment de garantie en termes de sécurité. Je ne suis pas sûre que moi, en les stockant par moi-même sur mon petit appareil, je garantis plus de sécurité... Mais en tout cas pour le moment, j'ai l'impression de mieux sécuriser mes données en les gardant par devers moi. ».

- **Organiser et pérenniser ses données :**

- « Je vois, quand on fait l'HCERES, on stocke... Je stocke tout ce qui sort, j'essaye de stresser, non de stocker, c'est un lapsus révélateur ! De voir un petit peu mais je n'ai pas que ça à gérer. Quand on a fait le bilan HCERES, [...] on demande à chaque collègue de faire un bilan, au bout de 5 ans » ; « Chaque année, le responsable de l'axe récupère les choses, mais c'est super compliqué, ça prend du temps » ; « Je trouve que c'est important qu'on ait un lieu unique avec un rangement, dans lequel on sait qu'on peut trouver les choses. Peut-être centralisé. Que ce soit physique ou virtuel. » ;

- « Effectivement, peut-être passer d'un espace de partage projet, qui a une durée de vie de cinq à six ans, à un espace de partage pérenne... L'espace de travail projet, même s'il y a des consignes sur la façon de déposer les fichiers, ça devient vite le bazar et passer à un espace plus pérenne, ça demande aussi un temps... ».

- **Synchroniser ses données :**

- « Un truc qui m'arrangerait bien, c'est si on pouvait, enfin ça existe peut-être, disposer d'une procédure de sauvegarde... [...] Non, mon problème c'est un problème pratique, comme mon ordinateur portable est l'ordinateur sur lequel je travaille... C'est-à-dire que si demain je me le fais piquer, je perds un mois de travail. Et donc le problème c'est que, en fait, comme c'est mon ordinateur perso [...], je suis obligé de faire tout ce travail de sauvegarde régulière. » ; « Il faudrait une sorte de sauvegarde automatique des fichiers, mais en ligne, qui me dispenserait de faire cette gymnastique régulière. ».

2.4.2 Besoins

- **Soutien technique :**

- « Et puis après un soutien technique manque quand même cruellement, enfin, dans le labo ou à l'université [...]. Et après ce sont des questions aussi personnelles, c'est à dire qu'il faut trouver la personne qui soit et compétente et disponible. » ; « Donc je pense que c'est vraiment très important qu'il y ait des informaticiens formés en sciences dures, vraiment en informatique, mais qui aient envie de travailler là. » ;

- « Penser par exemple que les techniciens et les ingénieurs qui sont dans le traitement peuvent aussi gérer à eux seuls l'administration de la donnée, je pense que c'est une erreur. On ne peut plus. Ici au niveau du labo, le labo a grossi, grossi, grossi sans vergogne... Par contre en terme de personnels on est restés au même stade. Donc l'étendue des missions a explosé, en terme de volumes de données aussi, donc c'est très compliqué. On a besoin de soutien du point de vue logistique, matériel et humain, tout simplement. » ;

- « Les ingénieurs dans le labo, ce serait l'idéal. » ; « C'est quand même un gros plus quand on commence à faire des choses très pointues. On a besoin de ce genre de personne. » ; « On a besoin de gens qui ont des compétences méthodologiques très particulières qui vont nous aider et nous pousser sur de nouvelles méthodologies. Après, je ne sais pas si l'université aura les moyens, le budget... Ça suppose faire des choix. On est dans un labo qui est pluridisciplinaire, on a des sociologues, des historiens, des collègues en psycho, des juristes, on n'a pas tous les mêmes besoins. Si on recrute quelqu'un d'hyperspécialiste qui ne peut servir qu'à 2 ou 3 personnes du labo ce n'est pas très utile. » ;

- « Je crois vraiment que l'on est maintenant dans une période où on a besoin de professionnels de l'informatique qui soient dédiés à ça [l'archivage] mais qui soient dédiés aussi, pas aux équipes, je sais bien que chaque équipe ne peut pas avoir un ingénieur informatique, ça serait très exagéré certainement, mais qui soit dédié de façon pérenne, qu'on ne change pas d'ingénieur tous les ans. Qu'on n'ait pas quelqu'un pour trois semaines et puis le temps de lui expliquer... Quelqu'un qui soit aussi un peu partie prenante, qui pourrait nous permettre d'améliorer pourquoi pas les procédures qu'on utilise. ».

3. Pratiques et représentations en matière de partage des données

3.1 Quelles positions générales sur le principe du partage ?

- **L'adhésion personnelle au principe du partage des données :**

- « *Moi je suis entièrement pour les archives libres* », « *Concernant les résultats, moi je suis entièrement pour le partage* » ;
- « *Oui, plutôt favorable sur l'idée générale ; pas au nom de principes libéraux, mais plutôt en termes de valorisation, j'y vois aussi un outil de valorisation de la recherche, sous d'autres formes que nos écrits* » ;
- « *Quand on travaille dans le public, on sait qu'on est financé par l'Etat, c'est un peu logique de rendre accessible à tous le fruit de nos recherches.* » ;
- « *De mon point de vue, moi je suis (pour) à 100 %* » ;
- « *Mes articles, dès que c'est possible ils sont en libre accès* » ;
- « *Par contre je suis pour la mise à disposition de mes données ; (mais) je veux d'abord présenter les résultats, ce que j'en ai déduit. Après je les mets à disposition, il peut y avoir d'autres analyses, des analyses qui vont contre les miennes etc. C'est l'objectif de la science, il n'y a aucun souci* » ;
- « *Elles (les données) ne sont pas en libre accès. Mais elles le seraient que ça ne serait pas un problème* » ;
- « *Mais on ne peut être que favorable sur le fond en tout cas. Et nous on en bénéficie aussi. C'est-à-dire que quand nous on travaille sur le même sujet que d'autres chercheurs qui rendent leurs données accessibles, c'est toujours bénéfique pour nous. Donc c'est difficile d'y être vraiment opposé* ».

- **La variable disciplinaire :**

- « *Je pense que philosophiquement, on est tous favorables au partage de données dans les sciences expérimentales* » ;
- « *Je travaille sur la marche, les interactions entre marcheurs (...), et typiquement, pour valider les modèles qui sont créés, on a besoin de se comparer avec des données réelles ; et vous imaginez bien que recueillir des données sur une foule, on ne fait pas ça tous les jours, donc, il y a quelques labos qui font ça, et ils mettent à disposition leurs jeux de données.* » ; « *c'est vraiment des données finalement brutes, c'est des trajectoires de marche (...), on a créé un modèle, qui est censé reproduire des trajectoires de vrais humains, et pour les confronter, j'ai besoin de beaucoup de données.* » ;
- « *On a une crise de répliquabilité en ce moment en psychologie. Pas mal de champs de recherche en psychologie ont de gros problèmes à être répliqués, des résultats classiques qu'on ne retrouve pas quand on essaie de refaire les mêmes expériences. De plus en plus de questionnements donc, autour de la bonne façon d'analyser les données, de la bonne façon de répliquer les données, de refaire les expériences à l'identique. (...) La psycho est en train de rentrer doucement dans l'ère du "on enregistre nos expériences avant de les faire, en déclarant à l'avance combien de participants on va collecter, quelle méthode on va utiliser, quelles analyses on va faire et comment on va traiter nos données", dans l'idée que si on déclare ça avant, ça nous empêchera de trafiquer notre analyse après pour trouver des résultats qui nous intéressent. Et c'est à cause de ça aussi que la demande des revues de stocker les données se développe de plus en plus. C'est quelque chose qui n'existait pas il y a 5 ans par exemple. Quand j'ai commencé, il n'y avait aucune revue qui demandait qu'on rende nos données accessibles, ça n'existait pas du tout, et là c'est en train de devenir majoritaire, au contraire...* » ;
- « *Les équipes internationales avec lesquelles on collabore, on voit la différence, parce que quand on bosse avec des équipes américaines par exemple, elles ont des plateformes de données avec archivage à long terme, elles ont des protocoles, des plans de gestion des données justement, donc c'est une différence qu'on voit et on voit que les demandes des revues sont calées plus sur ce que font les équipes américaines que sur ce qu'on fait, nous. Donc je pense que c'est un retard qu'on est en train de prendre aussi, qui n'est pas encore très remarqué pour le moment ; ça n'empêche personne de faire son travail.*

Sans tomber dans la boule de cristal, je vois la différence entre 5 ans et maintenant et ça ne va faire qu'évoluer. ».

- **Des positions favorables au partage, mais par obligation ou sur fond de questionnement, de réticences personnelles :**

- « *Sur le partage, en-dehors du labo (sur la mise en open access), on nous encourage à y aller ; après, dans quelques années, on ne nous encouragera plus mais on nous obligera ; donc mieux vaut s'y préparer en interne » ;*
- « *Est-ce qu'il faut ouvrir tout de suite, je ne sais pas mais je pense que la question sera vite résolue puisque ça « fuitera » de toute façon. Donc autant le faire de manière encadrée. » ;*
- « *Il y a toujours une ambivalence entre les deux. Favorable scientifiquement, défavorable dans le sens où ça permet aux autres d'avoir un regard sur ce que je fais et donc de me voler des idées sur le contenu de mes expériences. La bonne réponse, c'est favorable, forcément. » ;*
- « *Sur le principe, j'y suis plutôt favorable, l'idée paraît séduisante, ouvrir ses données et récupérer des données ailleurs et travailler dessus. Après c'est la manière dont ça va se faire, se mettre en place qui m'interroge un peu. » ;*
- « *Un des principes, c'est que, quand les données sont partageables, c'est très bien qu'elles soient partagées, mais après la question, c'est : partagées avec qui, comment ? C'est les modalités du partage. » ;*
- « *Oui, moi je distingue vraiment les deux (publications et données brutes) ; pour moi, à partir du moment où les recherches ont été validées, publiées, elles doivent être libres d'accès. Il n'y a aucune limite à avoir, on peut les mettre sur HAL, sur les sites de laboratoire. J'ai plus de réticence par rapport aux données brutes, en tout cas par rapport à certaines données brutes, et encore plus pour les données qualitatives. » ;*
- « *Sur les données qualitatives, j'ai plus de mal à voir comment les chercheurs pourraient s'en emparer, qu'est-ce qu'ils pourraient en faire. » ;*
- « *Je trouve que quand c'est bien (le partage), c'est bien, mais qu'il ne faut pas penser que c'est bien en tout et pour tout. » ;*
- « *Le positionnement absolu du tout disponible, tout en ligne, tout public me paraît être utopique. Comme toutes les utopies, il a des effets pervers qui peuvent être dangereux. On brasse des données très confidentielles » ;*
- « *A priori, ce n'est pas une idée que je défends dans l'absolu, je pense qu'il faut nuancer en fait cette position du partage absolu, quoi. Il y a des types de données, et du coup je pense qu'on ne peut pas tout partager. » ;*
- « *Ça dépend du cadre légal du partage des données ; ça c'est essentiel, parce qu'on sait que de toute façon, il y a de la concurrence sur l'acquisition des données et des choses comme ça ».*

3.2 Les facteurs d'incitation au partage des données

- **L'obligation faite par le financeur de la recherche :**

- « *L'obligation par le financeur de notre recherche, ça concerne assez peu les recherches en psychologie en général » ;*
- « *Je n'ai jamais travaillé sur des projets de recherche où on nous a obligés », « mais parfois cela pose problème, le libre accès aux données a des avantages et des inconvénients » ;*
- « *Le financeur non, ça c'est incitatif (...), c'est à dire que c'est incitatif pour faire avancer, mais je pense que quand on est chercheur, c'est d'emblée pour mettre à disposition ses travaux, ses résultats... » ;*
- « *Ni l'obligation faite par le financeur de la recherche, parce qu'il n'a pas le droit de m'y obliger ».*

- **La reconnaissance de la communauté scientifique :**

- « *À partir du moment où on exige de vous de rendre publics vos résultats de recherche, oui, en effet on est bien obligé de le faire, donc ça, maintenant c'est vrai que ce serait bien que ça devienne un peu plus spontané, un peu plus naturel ; la reconnaissance de la communauté scientifique, oui, c'est un*

argument, mais ça je pense que ça rejoint la question de la diffusion, sauf si vous entendez par reconnaissance les enjeux de carrière, là je ne sais pas trop comment entendre, comment comprendre ce terme de reconnaissance » ;

- « La reconnaissance de la communauté scientifique, ben oui, évidemment. » ;

- « Je ne sais pas trop. Je ne suis pas sûr que ça nous fasse beaucoup en reconnaissance ou en visibilité, sur le principe. » ;

- « La reconnaissance de la communauté scientifique : pas du tout, vraiment, ça ne me trouble pas » ;

- « Ça ne passe pas forcément par là » ;

- « Ce n'est pas par la capacité à accumuler des données, c'est par l'article. ».

- **Une meilleure visibilité de vos travaux de recherche :**

- « Oui une meilleure visibilité aussi, parce que l'intérêt c'est que ce soit vu. » ;

- « D'autres incitations à rendre vos données accessibles ? Réponse : Oui, la visibilité » ;

- « Les blogs, c'est plus pour rendre une visibilité des travaux de recherche et informer, puisqu'on est dans une logique de réseau, il faut constituer des champs, des repères, donc ça me paraît évident » ;

- « Meilleure visibilité, non, parce que le genre de travaux de recherche que je fais ne se visibilise pas par ça, mais par les publications qu'on fait à partir de notre travail sur notre corpus » ;

- « "Meilleure visibilité de vos travaux de recherche", pour nous ça ne passe pas par les données, ça passe par les articles. Ce n'est pas parce que tu as plein de données que tu vas être plus visible, si tu es visible c'est parce que tu fais quelque chose de tes données ».

- **L'occasion de nouveaux contacts, de nouvelles coopérations scientifiques :**

- « Nouveaux contacts et nouvelles coopérations scientifiques, ça je pense que c'est extrêmement important pour le coup, et c'est comme ça effectivement, enfin je trouve que l'effet des blogs est impressionnant. On est tout de suite identifié, on nous sollicite pour les recherches que l'on conduit. Alors ça ne veut pas dire mettre en libre-accès ses données, mais c'est quand même informer les collègues des sujets qu'on est en train d'aborder, et en même temps on livre un peu de résultats. Donc c'est plus dans l'activité quotidienne et du chercheur que dans le cadre institutionnel » ;

- « Probablement l'occasion de nouveaux contacts, de nouvelles coopérations scientifiques, ça oui, enfin moi ce que je défends, c'est que les travaux doivent alimenter le débat social et le débat scientifique et c'est sûr que je préfère publier un papier qui sera lu, même si ça dépasse la communauté scientifique que de publier dans une revue scientifique qui sera très peu lue par quelques personnes spécialisées sur le domaine; ça, c'est évident, donc cette question de l'ouverture, elle est fondamentale pour moi » ;

- « Oui, l'occasion de nouveaux contacts aussi. » ;

- « L'occasion de nouveaux contacts, ça peut arriver. Ça m'est arrivé une ou deux fois. Ce n'est pas la norme mais ça peut arriver » ;

- « Alors, l'occasion de nouveaux contacts, de coopérations, oui, c'est pas une priorité, mais c'est vrai que grâce à internet, ça permet... on peut être sollicité » ;

- « Les données en soi, je ne suis pas certain que cela crée de nouveaux contacts ou génère de nouvelles coopérations scientifiques ; ça, c'est plus l'accessibilité de nos productions scientifiques à partir des enquêtes que l'on va mener » ; « c'est un enjeu car plus les gens peuvent avoir accès à nos publications, plus facilement on développe des réseaux et des collaborations, ça c'est intéressant, je ne suis pas persuadé que mettre nos données brutes telles quelles accessibles à tous va faire que l'on va avoir plus de visibilité, que l'on va se développer un réseau » ;

- « Occasion de nouveaux contacts : ça passe par les publications, pas par un corpus brut ».

- **Le souci de pouvoir faire valider ses résultats à partir de ses jeux de données :**

- « La question de la vérification : je pense que de fait, on est toujours obligé de garder ses jeux de données, ses logiques de démonstration, si on les demande ; mais de là à les rendre visibles, ça pour moi, c'est une question qui est beaucoup plus à la marge. » ;

- « Le souci de faire valider nos résultats : je n'ai pas ce souci-là » ;

- « Le souci de pouvoir faire valider vos résultats à partir de jeux de données, donc une validation par d'autres, c'est ça hein ? Là moins, parce qu'au fond, comme on va déjà être quatre à le faire, on aura notre validation par nous-mêmes » ;

- « Le souci de pouvoir faire valider nos résultats, non. Ce n'est pas une préoccupation que j'ai, plutôt l'inverse en fait quelque part. On a plutôt confiance en nos résultats quand on les publie par définition, si on n'a pas confiance, on ne les publie pas. Mais on a plutôt peur que les gens viennent contester ce qu'on a fait » ;

- « Alors le souci de pouvoir "faire valider ses résultats à partir des jeux de données" : je ne vois pas pourquoi on les ferait valider par d'autres. On les valide nous même et si l'article est pris par un comité de lecture, c'est que ça marche ».

- **Le souhait de voir se développer d'autres recherches à partir de ses jeux de données :**

- « Le souhait de voir se développer d'autres recherches à partir de vos jeux de données : oui, ça, ça me semble important » ;

- « Voir se développer d'autres recherches à partir de vos jeux de données, oui, oui, bien sûr. C'est toujours plaisant de voir qu'on peut « faire des petits » en science, dans les domaines, c'est-à-dire avoir une forme de filiation, créer, susciter de nouveaux projets, de nouvelles recherches » ;

- « Quand ça sera complètement ouvert, le but du jeu, c'est que tout le monde puisse utiliser et puis importer les données, parce que moi j'en ai besoin par ailleurs » ;

- « Scientifiquement, c'est intéressant, parce que justement, ça permet aussi d'avoir d'autres lectures de ce qu'on avait fait comme corpus et qu'on n'aurait pas eues » ; « le souhait de voir se développer d'autres recherches, oui. Plutôt oui par contre ; ça arrive régulièrement, c'est un peu ce que je vous disais, c'est à dire que, quand on travaille sur la même thématique de quelqu'un, justement on peut avoir des idées, on peut trouver des nouvelles choses dans leurs données à eux, donc oui c'est l'un des intérêts principaux, je pense. » ;

- « Le "souhait de voir se développer d'autres recherches à partir de vos jeux de données", on est plutôt d'accord nous, parce que quand on exploite des données, on n'exploite pas tout. Par exemple moi dans mes carottages, je ne fais pas de micro-palé, enfin de paléo-animo si tu veux, parce que je n'y connais rien donc si quelqu'un veut prendre la carotte et s'occupe des mini coquillages et des machins comme ça, moi ça ne me dérange pas. Pareil les collègues en climato, ils ne travaillent pas sur tout à la fois, quoi » ;

- « J'ai déjà sinon donné, envoyé par mel des données pour des questions complètement différentes, pour une chercheuse en mathématiques en Israël, qui avait des questions sur d'autres choses sur la marche, et moi j'avais ce gros jeu de données de gens qui tournent, Du coup je lui ai données, et elle a fait ses analyses, qui n'avaient rien à voir avec les questions que je me posais. » ;

- « Je pense que sur des bases de données qui sont très calibrées, avec plus des logiques de sondage, que ce soit sur des dossiers ou autres, là ça ne pose aucun problème ; par contre, sur des données qui sont beaucoup plus qualitatives, qui sont liées à la relation d'enquête, je trouve que ça n'a pas de sens, on dénature en fait le matériau en le donnant comme ça sous la forme de jeu de données. Quand les catégories en fait sont déjà travaillées a priori, il n'y a pas de souci après pour les donner à voir, parce qu'elles sont définies, précises, etc. » ;

- « Le souhait de voir se développer d'autres recherches à partir de vos jeux de données : c'est une question qui est récurrente et de manière sous-jacente ; et le type de recherches qu'on fait nécessite des corpus tellement contextualisés que personne d'autre que la personne qui connaît le contexte ne peut les exploiter ; du coup ça ne marche pas ».

- **L'adhésion aux valeurs du Libre Accès aux résultats de la recherche scientifique :**

- « La question du libre accès, je trouve que c'est très important, (...) c'est quelque chose que je défends. Mais de l'autre côté, c'est que dans ce libre accès, on ne protège pas forcément du coup nos sources (...) ça expose aussi les situations et on doit garder la maîtrise de cette exposition en lien avec l'accès libre. C'est important de mixer les différentes possibilités de publication, d'avoir à la fois de l'accès libre, mais aussi de l'accès restreint et réservé. Parce que je pense qu'on n'écrit pas de la même manière, qu'on ne dit pas la même chose en fonction aussi du public à qui on s'adresse. » ;

- « L'adhésion aux valeurs du libre accès aux résultats de la recherche scientifique : alors là, il faudrait définir ce que l'on entend par valeurs du libre accès aux résultats de la recherche scientifique, il faudrait être peut-être plus explicite. » ;

- « L'adhésion aux valeurs du Libre Accès, alors c'est en plus intéressant, parce que vous ne dites pas quelles sont les valeurs du Libre Accès, donc on ne sait pas à quoi vous nous demandez si on veut y adhérer ».

- **De possibles retombées financières (crédits de recherche...) :**

- « De possibles retombées financières, c'est inexistant en psychologie, ça ne nous concerne pas. » ;
- « Les retombées financières, ça ne me concerne pas » ;
- « Et non malheureusement, pas de possibles retombées financières. C'est dommage mais non, non, non, ce n'est pas possible, je pense que personne ne m'achètera ce truc là, donc non ».

- **Autres raisons :**

- « Et il manque un item : par obligation. Ce qui est vraiment de plus en plus le cas chez nous. Je pense que dans les trois ou quatre derniers articles que j'ai publiés, il y avait une exigence de la revue qu'on rende les données accessibles. Donc par obligation tout simplement » ;
- « Je vois bien l'avantage que ça peut avoir, inversement. C'est-à-dire en termes d'enjeux de conservation : si j'avais déposé quelque part mes données, si quelqu'un avait géré la convertibilité de mes données, peut-être qu'aujourd'hui, je pourrais récupérer mes données. Il y a cet avantage, cet enjeu de conservation. Il y a un enjeu de traçabilité, même si j'y ai été peu confronté. » ;
- « Je pense que ce qu'on peut partager tout à fait bien, c'est tout ce qui a trait aux concepts fondamentaux d'identité, mémoire et territoire. Alors là on peut avoir quelque chose de tout à fait commun, où chacun peut aller... » ;
- « Q : Quelle aurait été l'autre raison possible qui t'aurait incité à partager les données ? R : Aucune ! Y'en a pas ».

3.3 Les facteurs de frein au partage des données

- **Sur les freins juridiques liés à la nature des données :**

- « J'ai souvent travaillé sur des dossiers de justice ; un certain nombre d'affaires ont été jugées, c'est des éléments importants, même si l'historien a droit à la citation, y compris les noms ; mais de là à les partager après, ça ne me serait pas venu à l'époque de mettre ces données en partage » ;
- « On les anonymise, mais ce n'est pas si simple que ça : les gens qui connaissent bien le sujet, avec les éléments du contexte, ils finissent par retrouver. L'anonymisation absolue, ça n'existe pas. Du coup, comment on fait pour partager ça, si on n'arrive pas à les anonymiser ? » ;
- « Vraiment, cela dépend des projets et des parties de projet, c'est-à-dire qu'il y a des projets sur lesquels on peut communiquer de manière grand public, etc., mais il y a d'autres choses en fait, des parties qui ne doivent pas sortir, parce que ça peut même être dangereux pour les personnes, pour les organismes, etc. » ;
- « Il peut y avoir des réticences sur l'ouverture de ces données, d'une part parce qu'il y a des données un petit peu confidentielles ; quand on est allés voir les ménages, on leur a dit que ce sera utilisé dans un cadre strictement scientifique, pas à des fins fiscales ou autres, mais on est quand même tenus par un secret » ;
- « Les fichiers mouvements que je vous ai montrés, bon c'est facile de les rendre anonymes, après c'est vrai que pour les supports de conférence, une vidéo de « comment ça s'est passé », c'est important pour comprendre ; mais qu'est-ce que je peux montrer, est-ce que je floute les visages ? c'est vrai qu'on fait signer aux personnes les consentements de participation, et on leur demande explicitement est-ce que vous nous autorisez à faire des films et des vidéos, on leur dit dans quel cadre on va les diffuser, ce sera dans le cadre de conférences, de divulgation scientifique, donc ils acceptent ou pas, en général ça pose pas de problème.. » ;
- « C'est sûr que les freins juridiques, ça pourrait être quelque chose sur lequel je ne suis pas vraiment au point. C'est quand il y a des usages de correspondances ou de fonds. C'est toujours la question que je me pose. J'ai eu la chance par exemple d'aller dans les fonds des éditions Gallimard, ce qui est très rare, il y a très peu d'accès, mais en fait on ne sait pas juridiquement derrière comment c'est utilisable.

Pareil pour les correspondances, les entretiens privés, ou les choses comme ça. C'est vrai que c'est des choses sur lesquelles je n'ai pas vraiment d'informations. Mais c'est une méconnaissance personnelle » ;

- « Freins juridiques, non. Non, je ne pense pas que ça nous concerne beaucoup. (...) épreuves cognitives, tests psycho-techniques, c'est deux façons de dire la même chose. Donc freins juridiques, pas tellement » ;

- « Des freins juridiques liés à la nature de vos données : je pense que tant qu'il n'y a pas les textes, non, parce c'est des données qu'on a établies nous-mêmes. » ;

- « Elles sont sensibles au moment où je les collecte, mais elles ne le sont plus au moment où je les diffuse. C'est à dire que ça m'arrive de plus en plus de devoir rendre mes données accessibles mais elles sont anonymisées et dépourvues de toute information d'identification, donc il n'y a plus rien de sensible à la fin ».

- **Sur les raisons économiques :**

- « Les raisons économiques, en psychologie ça nous concerne très peu aussi. »

- « Des raisons économiques : ben non »

- « Les données payantes, à partir du moment où il faut payer pour les avoir, la notion de les mettre à la disposition de tout le monde pose problème » ; « Q : c'est le labo qui achète les données ? R : alors là ça dépend, si le labo n'a pas beaucoup de sous, c'est surtout nos crédits de recherche personnels, pour l'essentiel. Les données gratuites, je ne vois pas de problème. ».

- **Sur les raisons épistémologiques et scientifiques :**

- « Vous n'allez pas à la cueillette, il n'y pas des pâquerettes qui sont déjà là, c'est vous qui les produisez. C'est important, parce que du coup ça veut dire qu'on oriente, etc. » ; « Les données, c'est toujours des productions, moi c'est quelque chose à laquelle je tiens et j'enseigne ça à mes étudiants, et c'est pour ça qu'on ne les appelle pas des données, on les appelle des observables, et on dit "les observables, vous les construisez". » ; « ça veut dire aussi qu'on a un droit intellectuel sur ces productions, parce que c'est un vrai travail, c'est pas juste : j'ai été ramasser des choses qui étaient là ; c'est : j'ai mis en place une méthode, j'ai passé du temps, je les ai élaborées, j'ai travaillé, j'ai produit mes données avec des acteurs sociaux, dont j'observe un certain nombre de faits, de comportements ; et du coup, ça pose le problème du droit intellectuel sur les observables qui ont été produits, de ce point de vue, pour moi il y a un droit » ; « Dans ma discipline, les sciences du langage, les corpus partagés ça fonctionne pour ce qu'on appelle de la linguistique de corpus, d'ailleurs le nom existe, où on considère qu'on va étudier les formes linguistiques d'une certaine façon en elles-mêmes et pour elles-même, en les sortant d'un contexte ; et effectivement, n'importe qui peut avoir accès au même corpus et puis le travailler ; mais dès lors qu'on travaille sur l'ancrage social et sur les interactions entre les pratiques langagières et les pratiques sociales dans lesquelles elles sont produites et qu'elles contribuent à produire, tu ne peux plus partager. » ; « ça se partage mais avec des gens qui partagent le même contexte ; ou alors quand on présente nos travaux, y compris au séminaire du labo, on partage avec les gens qui vont poser des questions ; ils demandent toujours de la contextualisation. » ;

- « Il faut qu'on sache qui a fait quoi, et puis aussi qu'il y ait un vademecum, qu'on sache comment on peut utiliser ces données, comment elles ont été collectées » ;

- « Oui, il y a quand même des freins à ce que le matériel ne soit pas dépouillé en quelque sorte de son objectif premier, même si c'est pour une autre utilisation à laquelle on n'a pas forcément pensé au départ, il y a forcément des utilisations potentielles mais ça reste quand même des données construites par une équipe. » ;

- « Je trouve que cela n'a pas forcément de sens de mettre à disposition ce genre de matériau, parce qu'en fait, il y a toute l'interaction sociale, l'historicité sociale qu'il y a dans cette relation. Seul celui qui a conduit l'entretien peut maîtriser et il peut le retransmettre s'il y a des questions, mais redonner tout ce contexte, c'est un lourd travail, pour après le partager. » ;

- « Quand on n'a pas participé au montage du projet, à la conception, à la définition de la problématique, récupérer derrière des données, ça me semble compliqué. Je ne dis pas que c'est impossible, mais il faudrait prendre contact avec les porteurs du projet initial, pour que au-delà des métadonnées, on ait les coulisses, le contexte ; ça me semble compliqué de réutiliser des données produites par d'autres, en SHS » ;

- « C'est surtout que c'est inexploitable, il y a un côté y compris un peu démagogique dans l'idée de faire croire que ce que l'on a intéresse tout le monde, c'est exploitable, parce que ce n'est pas le cas. C'est très compliqué » ;

- « Les biais sont multiples, en statistiques ou en géographie, on est toujours sensibles aux biais, c'est-à-dire de mal exploiter un corpus de données en ne tenant pas compte des effets de contexte qui font que voilà, telle corrélation on ne peut pas l'établir parce que les données ne permettent pas d'établir... Les calculs, on peut toujours calculer tout ce qu'on veut, l'âge du capitaine, la taille de sa moustache mais ça n'a pas forcément de sens. » ; « une donnée n'a de sens que dans le contexte où elle a été donnée et qu'en maîtrisant bien les métadonnées c'est à dire les éléments qui renseignent sur la création et la valeur de la donnée. » ;

- « C'est que, quand bien même je donnerais un corpus à quelqu'un, j'imagine l'effort qu'il me faudrait à moi pour le structurer pour le rendre appréhendable en fait par quelqu'un d'extérieur. En fait, je pense que c'est un métier à part entière » ;

- « Je trouve que la question est complexe, je ne pense pas qu'on puisse avoir une règle qui serait la même en tout et pour tout, et qu'il faut imaginer plusieurs cas d'espèce, etc. » ;

- « C'est vraiment des situations, je ne pense pas qu'il y ait de règle générale, en tout cas universelle, et qu'il faut garder des spécificités propres à des choix méthodologiques ».

- **Sur les raisons institutionnelles :**

- « Raisons institutionnelles, non pas du tout. Mon labo n'a aucun problème avec le fait de diffuser les données. Je pense qu'on le ferait si on avait un support de stockage pratique en fait. » ;

- « Des raisons institutionnelles : ben non, je ne pense pas que mon labo ne veuille pas diffuser les données ».

- **Les facteurs liés à l'écosystème de publication de la discipline**

- « Donc à partir du moment où on diffuse trop facilement, là je pense à des données non seulement produites, mais produites sous forme de livres, donc faciles d'accès, encore plus que les articles, on se pose la question de la survie de tout ce secteur de l'édition scientifique, semi-scientifique. » ;

- « Plus il y a de diffusion en ligne d'extraits de ces revues, moins il y a d'abonnés. » ;

- « Je pense que toutes les productions ne doivent pas ou ne méritent pas d'être mises à disposition, parce qu'après il y aura peaufinement, récupération ou bien rassemblement d'un certain nombre de travaux pour que ce soit mis par exemple sous forme officielle à la disposition du public, mais au moins que ce soit légalisé sous forme d'un ouvrage ou sous forme d'un ouvrage collectif au nom d'une équipe ».

- **Les pratiques scientifiques disciplinaires : l'artisanat de la recherche :**

- « On peut avoir encore un rapport très artisanal à la recherche. Donc c'est-à-dire aussi très individualiste, on est encore dans la culture du livre, voilà. Qui s'écrit seul, de bout en bout, ça reste encore le nec plus ultra de la profession » ; « chez nous, il y a encore cette idée que la part de travail individuel est très très déterminante » ; « je pense que nos collègues ont tous dans leurs cartons chez eux, dans leurs greniers en fait, enfin maintenant dans leurs disquettes, dans leurs disques durs, les résultats de leurs dépouillements ; ça tient au fait qu'on mène peu de projets collectifs. Je crois que c'est en grande partie ça. » ; « C'est-à-dire qu'on n'est pas financé pour notre recherche. Donc on ne sent pas le devoir de partager puisqu'au fond, on fait nos recherches à nos frais, pour une grande partie » ; « je trouve important d'essayer de placer des textes dans des revues exigeantes. Et chez nous, c'est clair, dans les revues exigeantes, on ne peut placer de papiers que si on a un terrain vraiment original » ; « parce que se constituer un terrain vraiment original, c'est un effort. Moi je ne sais pas si j'en ferai un troisième, peut-être. Mais c'est vraiment lourd. Quand on se constitue un terrain comme ça, c'est quelque chose de très précieux car c'est aussi la condition du maintien dans un niveau d'échange scientifique assez élevé » ; « Donc en fait, quand moi j'accumule pendant des années de l'information comme ça, ça va permettre la publication d'une grosse publication. Mais ça me donne aussi des billes pour publier pendant 15 ans. » ; « c'est la raison pour laquelle la plupart des gens qui publient dans ces bonnes revues sont soit des gens qui viennent de soutenir une thèse, soit des gens qui viennent de soutenir une habilitation. » ; « C'est un terrain que vous vous êtes constitué, vous n'avez pas forcément envie de partager car vous perdriez un peu ce qui fait votre spécificité » ;

- « J'avais noté le monde des disciplines, parce que je pense que toutes les SHS sont comme ça, mais dans notre discipline, c'est peut-être encore plus accentué, il y a quand même un attachement très fort à l'indépendance du chercheur et surtout ne pas trop s'immiscer dans ses pratiques et surtout ne pas lui imposer des choses ; c'est pas forcément dramatique, mais ça peut expliquer une certaine tendance à l'inertie quand on essaie d'introduire des nouveaux protocoles, souvent de remuer les troupes, c'est difficile. Il y a une capacité de résistance assez forte. ».

- **La crainte du plagiat :**

- « Si l'on met des retranscriptions d'entretiens brutes, se les faire détourner, voler, c'est des choses que j'ai déjà eues sans qu'il y ait de données en libre accès. On n'est pas à l'abri de ce genre de pratique, ça peut être un frein. » ;

- « Crainte du plagiat, ce serait plutôt ça. » ;

- « Après, l'autre frein, c'est vraiment la crainte du plagiat. Surtout ça » ;

- « C'est la crainte, primo du plagiat de la part des étudiants et des jeunes chercheurs, qu'ils prennent, qu'ils utilisent à leur manière, ou d'autres chercheurs ou d'autres collègues, peu importe » ;

- « Il y a une espèce de tiraillement là, entre peut-être notre crispation sur les droits d'auteur, la propriété intellectuelle, c'est-à-dire que si on met tout ça en ligne, où est-ce que ça va finir, tout le monde va nous voler nos idées, nos résultats et puis en même temps la nécessité de rendre les choses plus accessibles et ça je crois que tout le monde sait bien que c'est une nécessité aussi et c'est même un devoir. » ;

- « Le libre accès, faut faire attention ; c'est vrai qu'on est en concurrence entre guillemets, quand on reste sur un sujet de thèse, ou de recherche, ou un bouquin, faut qu'on soit quand même assez prudent sur nos réflexions, sur l'avancée. Faut arriver à trouver un système qui nous permette à la fois de partager pour avancer sans se faire plagier ou piquer une partie de la réflexion. » ;

- « Je n'ai pas un avis tranché dessus : c'est à la fois intéressant, séduisant, l'idée de partage en science devrait couler de source. Et en même temps, cela pose un certain nombre de problèmes à la fois pour protéger ses propres données et pour pas non plus se faire piller son propre travail. La recherche est malheureusement un univers où les chercheurs sont en concurrence entre eux, pas qu'au niveau d'un labo mais au niveau national ou international. Il y a des collègues peu scrupuleux ailleurs qui pourraient s'emparer facilement de données récupérées d'un travail produit dans un autre laboratoire de recherche » ;

- « Après, se faire voler les données, oui c'est toujours un risque ; le problème, c'est que tous les enseignants-chercheurs ne sont pas logés à la même enseigne, c'est-à-dire vous avez des enseignants-chercheurs comme moi, qui sont responsables de choses, et qui croulent sous les charges administratives, et leur temps de recherche est très faible par rapport à ce qu'il devrait être ; et d'autres qui eux, pour le coup, en fait refusent toute activité administrative et qui ont tout le temps pour faire ce qu'il faut » ;

- « Ces données-là ne sont pas à plagier, cela n'a pas d'intérêt particulier, en revanche on sait qu'elles sont plébiscitées par des entreprises, par ce monde qu'on connaît moins, Google par exemple ; tout ce qui est multilingue intéresse aussi le monde civil ; est-ce que ce monde peut se réapproprier ces données sans même énoncer quelque part que ça vient de l'Université Rennes 2 ? ça, ça me dérangerait par exemple, que ce soit réexploité sans le dire ou sans bénéfice pour l'université Rennes 2. » ;

- « La seule raison qui pourrait être un frein, c'est plus le côté scientifique, crainte du plagiat. Mais là encore je ne suis pas sûr que ce soit une vraie bonne raison en pratique. A partir du moment où on fait des publications sur une thématique, c'est logique qu'on donne accès aux données qui vont avec. Normalement, quand on publie un jeu de données, on est censé être exhaustif dans la publication de ce jeu de données. Si on faisait bien notre travail, on n'aurait même pas à avoir peur du plagiat puisqu'on serait censé les décrire à fond de toute façon. C'est ça je pense, ce qui freinerait le plus mais ce n'est pas à mon sens une bonne raison en tout cas. » ;

- « La crainte du plagiat, non je ne crois pas, le but du jeu c'est que ce soit utilisé. ».

- **Les raisons « personnelles » :**

- « Des raisons personnelles "je ne veux pas montrer mon arrière-cuisine", "mes données sont à moi" : oui un peu » ;

- « On est toujours un peu gênés par le fait de rendre nos données accessibles parce qu'on a toujours un peu peur que les autres viennent mettre le nez dans ce qu'on a fait et dire que peut-être on a mal fait notre travail. Et on a toujours un peu peur d'être passé à côté de quelque chose d'important ou que quelqu'un ait des idées qu'on n'a pas eues. » ;

- « Non, l'arrière-cuisine ça me paraît hyper important de la montrer, justement, ce que je trouve c'est qu'on ne la montre pas assez, en fait elle est hyper importante, le résultat ne veut rien dire sans l'arrière-cuisine, donc toute la base est construite pour que l'arrière-cuisine soit visible, utilisable, exportable et qu'on puisse faire une autre recette avec les mêmes ingrédients » ;

- « "Je ne veux pas montrer mon arrière-cuisine"... Non non, ça je pense que au contraire, c'est important que les gens aient accès finalement à cette arrière-cuisine. C'est-à-dire qu'eux pourront peut-être faire autre chose avec les données que ce que moi je peux apporter comme traitement, donc il faut avoir accès au matériel, aux matières premières. ».

- **Le manque de temps :**

- « Ça prend forcément du temps à vraiment formaliser, à décrire, de façon très fine les données, sans pour autant savoir si un jour ça va être utilisé par quelqu'un ; donc y'a quand même le rapport coût-bénéfice » ;

- « Les freins c'est plus une question de gestion du temps et donc c'est plus de la méthodologie en fait, être maître de son temps, enfin ce qu'on appelle les fictions de méthode peut-être, de dire voilà on fait quelque chose, après on fera autre chose, on s'impose des limites volontairement, parce que la recherche ça avance doucement et méthodiquement, palier par palier. » ;

- « Le manque de temps : je ne pense pas que ce soit très coûteux en temps pour nous honnêtement. Normaliser un fichier avec des intitulés de colonnes un peu clair, éventuellement diffuser un petit fichier avec pour dire la colonne 1 veut dire ça, la colonne 2 veut dire ça, c'est l'affaire de 15 minutes. Je pense que les bénéfices sont largement plus importants que le coût en temps. » ;

- « Le manque de temps : ben non à partir du moment où je fais la base de données, c'est ça qui est long, ce n'est pas l'accessibilité ».

- **Le manque de connaissances dans la description des jeux de données, les modalités de dépôt :**

- « Non c'est pareil, je ne pense pas qu'il y ait d'enjeux très complexes par rapport à nos données. » ;

- « Ben non je ne pense pas, parce qu'après, ce n'est pas moi qui gère (...), donc il n'y a pas de problème. » ;

- « Le fait est que, en tout cas à l'université, on n'est pas tellement au courant de tout cela, loin de là, et des évolutions très rapides » ;

- « En fait, les freins c'est plus le temps à passer pour faire ça bien, c'est à dire encore une fois je pense qu'on manque d'outils et de soutien pour le faire bien, dans les règles de l'art. Me concernant par exemple, moi j'ai déjà exprimé plusieurs fois que j'aurais besoin d'être formé sur tout ce qui relève de la propriété intellectuelle, parce que je sens qu'on n'est pas non plus armés pour en parler en connaissance de cause et être efficace ».

- **Le manque de compétences :**

- « Le manque de compétences, pareil. » ;

- « Et le manque de compétences non plus puisque je travaille avec (...). Alors toute seule, je n'aurais fait rien du tout, évidemment ».

- **Autres raisons :**

- « (Je suis) réfractaire à toute mise en scène, à la visibilité. J'aime bien contrôler ce qui est en ligne » ;

- « Après, le frein, ça peut être aussi finalement notre secteur de recherche. C'est-à-dire est-ce que ça intéressera... est-ce que (...) des choses comme ça peuvent intéresser d'autres collègues ? ».

3.4 Témoignages, interrogations, souhaits

- **Témoignages sur les pratiques de partage des données :**

- « *Moi, l'expérience que j'ai (du partage), c'est au sein des équipes ; au sein des programmes de recherche, on trouve une espèce de partage, mais autrement au sein du laboratoire, non.* » ;

- « *Oui, je l'ai déjà fait, dans le cadre d'un projet ANR, et sur le site du projet, on avait mis à disposition des jeux de données brutes. Question : et c'était à disposition, totalement public ? Réponse : oui, ça reste accessible à un public spécialisé* » ;

- « *“Vous-même avez-vous déjà stocké des données en libre accès sur le web... seriez-vous d'accord pour diffuser vos données de recherche en libre accès à condition qu'elles soient diffusables et protégées juridiquement” : encore une fois, ça dépend, oui, mais en fait ça dépend des projets ; il y a des projets, par exemple, on travaille sur l'expression de la souffrance sociale, ça, ça sortira jamais ; il y a d'autres projets qui sont plus sur ce qui se passe dans l'espace public, etc, là c'est moins problématique, ça dépend du type de sujet aussi.* » ;

- « *Nous, le problème qu'on a n'est pas de partager nos données vers l'extérieur, c'est de trouver des moyens de partager nos données à l'intérieur d'une équipe qui mène un même projet. On a déjà du mal actuellement à avoir des outils qui nous permettent de mettre en commun nos données, même à petite échelle, à échelle intersite* ».

- **Le besoin de garanties pour la diffusion des données :**

- « *Ça ne me pose pas de problème de les partager, à condition d'en être crédité et que ça fasse partie aussi, au même titre que la bibliographie papier, de l'évaluation, de l'estimation de la charge de travail réalisé, etc. C'est pour ça qu'il faut que cela apparaisse dans un cadre réglementaire et sur des sites qui ne soient pas privés, particuliers. Et il faudrait probablement aussi pour l'utilisation de ces données une forme de... un cadre légal, un copyright, enfin.* » ;

- « *Moi, je serais même assez partisan de mettre en ligne des choses raisonnées, qui sont plus proches des résultats de recherche que des données elles-mêmes. Et si c'est fait, ça doit être fait avec prudence parce que sinon... Et puis il y a quand même un élément de confidentialité, c'est-à-dire des heures et des heures qui ont été passées à monter une donnée, je ne sais pas où commence et où s'arrête le droit d'auteur mais il y a toute cette question, quand même* » ;

- « *En même temps, on est vraiment sur des données communes ; qui a la propriété des données ? et là, je ne parle pas des résultats scientifiques, je parle des données brutes. Là on avait une masse monstrueuse de données (entretiens, archives, etc.), on s'est retrouvés bloqués pour les utiliser parce qu'on n'avait pas de réponse sur ce qu'on avait le droit de faire, avec ces données.* ».

- **Ouverture des données aux étudiants :**

- « *Je crois quand même que l'université c'est le lien entre l'enseignement et la recherche ; j'aimerais bien que certaines choses qu'on fait, qui sont faites par nous en recherche, puissent servir de support au niveau enseignement ; moi je travaille sur les images, sur la transition, que les étudiants puissent les utiliser aussi pour les cours ; qu'il y ait une partie de ce travail-là, qui puisse être accessible, parce que les étudiants, c'est une partie du travail qu'ils ne voient pas* » ;

- « *Par contre ce qui serait bien, ce serait de pouvoir les rendre accessibles. Enfin, je me vois bien travailler avec des étudiants en leur disant, voilà vous avez des données, qu'est-ce que vous arrivez à produire à partir de ça ?* ».

- **Interrogations sur le Libre Accès des publications :**

- « *En particulier si tout chercheur commence à mettre le contenu de ses publications de façon intégrale sur internet, ça pose le problème de la survie des revues, qui vivent, elles sont subventionnées évidemment* », « *Je n'ai pas de solution toute faite, mais je sais que l'open access, ça pose une série de problèmes.* » ;

- « *Ensuite sur les questions des textes et des variantes, c'est des problématiques nouvelles qui se posent à moi. Je travaille sur les dépôts des textes sur les archives ouvertes et en préparant plusieurs textes plutôt anciens, je me rends compte qu'il y a quand même des problématiques complexes ; donc la*

version, est-ce que c'est la version qui a été déposée ou est-ce qu'au bout de 15 ans, il faut plutôt revoir la version parce qu'on n'a pas toujours la même... Ce n'est pas tout à fait les données mais en tout cas je découvre une série de problématiques qui sont liées, c'est quand même à la fois le stockage et la visibilité. Moi je travaille sur ces dépôts, j'ai plusieurs textes qui sont préparés mais il y a toujours quelque chose qui m'empêche de les déposer ».

4. Besoins, attentes

- « On a besoin de soutien du point de vue logistique, matériel et humain, tout simplement. Et puis aussi je pense au niveau de l'établissement ou des établissements, avoir une politique claire » ;
- « Il y a cette dimension formation, prise de connaissance, et puis soutien technique, je pense que c'est vraiment les trois choses »

• **Besoins techniques :**

- « Il faut qu'il y ait aussi une forme de souplesse, d'ergonomie, c'est à dire que si à chaque fois que l'on veut modifier quelque chose, il faut passer par un ingénieur d'étude, je sais pas quoi, un informaticien qui travaille tel jour, de telle heure à telle heure parce qu'il est sur un certain nombre d'autres projets, que quand c'est les vacances, c'est les vacances, que quand il finit son contrat c'est terminé, c'est pour ça que je dis qu'il faut pouvoir avoir la main dessus, pour soi-même » ; « idéalement, il faudrait qu'on puisse intervenir à tous les niveaux de la chaîne » ;
- « Moi, je pense que pour tout le monde, il faudrait qu'on ait des espaces de stockage pour pouvoir sauver les données » ;
- « Un soutien technique qui manque quand même cruellement, enfin, dans le labo ou à l'université [...]. Et après c'est des questions aussi personnelles, c'est-à-dire qu'il faut trouver la personne qui soit et compétente et disponible » ;
- « C'est vraiment très important qu'il y ait des informaticiens formés en sciences dures, quoi, vraiment en informatique, mais qui aient envie de travailler là » ;
- « Je travaille avec des collègues de Rennes 1 et de l'EHESP, on n'arrivait pas à trouver un outil qui soit satisfaisant. Peut-être que si on ne reste qu'à l'échelle de Rennes 2, on peut utiliser les outils qu'on trouve sur Coursus, qui permettent peut-être de créer un espace de partage qui soit satisfaisant, mais dès qu'on veut décloisonner, on est un peu en peine » ;
- « Voilà, il y a ça, des outils pour réussir à travailler ensemble sur un programme de recherche » ;
- « Moi je pense que les besoins sont des besoins techniques, des besoins d'ingénierie informatique, vraiment très technique » ; « l'archivage, l'entretien d'une année sur l'autre, la réactualisation des outils informatiques, là ça dépasse un peu les compétences et le temps que les chercheurs ont à consacrer à ça » ;

• **Besoins de formation :**

- « Je pense qu'il y a un gros travail de pédagogie à faire, sur l'open access, archives ouvertes, tous ces termes là, il faudrait peut-être mieux les définir » ;
- « C'est ce qui m'incite à penser qu'il faut absolument que l'on cale une formation continue, enfin, je ne sais plus comment ça s'appelle ici : on reçoit toutes les semaines une heure de formation et tous les ans parce qu'on risque de perdre la main, d'oublier un truc auquel on a eu accès mais qu'on n'a pas encore mis en pratique ; et les questions juridiques » ;
- « Ce serait une formation continue en termes juridiques et techniques. Je pense que tous les ans, il faudrait peut-être prévoir, je ne sais pas, une journée ou une demi-journée parce que je pense que les choses vont évoluer très vite » ; « Et puis la deuxième chose, une formation vraiment technique là pour avoir la main sur les données. » ;
- « Avant même l'exploitation des données, je pense que pour beaucoup d'entre nous en SHS, c'est la formation aux méthodes, aux outils mêmes. Enfin je vais être attentive à cela de plus en plus, mais je connais deux outils, on va dire, et il y en a plein d'autres que je ne connais pas. Et souvent je pense qu'il y a plein d'autres choses que je pourrais utiliser si je les connaissais. » ;

- « Besoin de sensibilisation auprès des collègues, sur cette incitation actuelle, sur le fait d'archiver correctement les données de la recherche, de les documenter correctement, et de les mettre à disposition sur des archives ouvertes. » ;

- « Pour moi il y a toute cette question de formation, qui est déjà un frein à l'utilisation de tout un tas de choses » ;

- « Il est très possible que si je prenais le temps pendant des mois de me former, j'aurais des idées supplémentaires qui viendraient enrichir mon travail intellectuel. Mais aujourd'hui, je suis intuitivement convaincu que ce ne serait pas considérable. Du coup, vu le peu de temps qu'on a, c'est un effort que je ne fais pas » ;

- « On a un besoin de formation : comment on travaille ensemble sur un projet collaboratif et comment on rationalise ces données car maintenant les enquêtes sont beaucoup plus volumineuses qu'auparavant » ; « il faut qu'on apprenne à travailler ensemble et à traiter ensemble cette masse de données. Là aussi, on a besoin de formations parce que très clairement les pratiques changent et si l'on ne veut pas avoir un train de retard, il faut s'adapter à cela » ; « Sur les données c'est plus les nouvelles méthodologies qui apparaissent, tout ce qui va être humanités numériques, on va avoir de plus en plus besoin de travailler avec des logiciels qui vont nous aider à traiter nos données, du fait de la masse de données à analyser » ; « on n'a pas tous les mêmes besoins ; si on recrute quelqu'un d'hyperspécialiste qui serve qu'à 2 ou 3 personnes du labo c'est pas très utile » ; « j'ai pas forcément le temps d'aller à la formation, et des fois en bidouillant, tout seul, ça prend un peu de temps mais je vais m'en sortir » ;

• **Besoins d'accompagnement :**

- « Accompagnement pour nous forcer à le faire, ça oui. Je pense que la caractéristique principale des enseignants-chercheurs, c'est l'inertie en général. Le temps et l'inertie, oui, ça va un peu ensemble. [...] il faudra vraiment nous forcer, moi et mes collègues, pour qu'on s'y mette » ;

- « Il faut que la collaboration se fasse dans les deux sens » ; « pas seulement un soutien technique en fait, une collaboration » ;

- « C'est le type d'informaticien dont on rêve dans un labo parce que tout ça ça l'intéresse en fait, presque plus que la dimension purement informatique, la capacité à créer un outil qui soit en adéquation avec les besoins du chercheur » ;

- « Il faut du personnel qui est là pour rappeler cette réalité aux chercheurs : voilà, ils ont des données, qu'il faut penser à organiser, penser à sauvegarder, mais ça ne peut pas être le travail du chercheur » ; « c'est un référent d'une certaine manière qui vient vers vous plutôt que le chercheur aille vers lui, qui vient faire un point sur l'état des données qu'on a récoltées sauvagement ou méthodiquement » ; « je trouve que c'est un métier qui manque mais il ne faut absolument pas qu'il mette la main à la pâte. Il n'est pas là pour être au service du chercheur. C'est plus un expert peut-être » ; « si on a quelqu'un qui nous conseille sur l'organisation, sur comment je dois nommer mes fichiers » ; « c'est plus dans la méthode, dans l'organisation des documents [...] on a appris à faire des bibliographies, OK, ça on sait le faire quand on écrit des articles, mais classer ses documents dans l'ordinateur, non » ;

- « S'il y avait un service qui réfléchissait comment rendre publiques les données de recherche, ce serait à échanger » ;