Revue COSSI

N° 2 | 2017
L'information, la communication et les organisations au défi de l'altérité

Distance informationnelle scientifique : le risque d'une altérité informationnelle ?

Christian MARCON *Professeur des universités*
IAE
CEREGE
University of Poitiers

Édition électronique :
URL :
https://revue-cossi.numerev.com/articles/revue-2/51-distance-informationnelle-scientifique-le-risque-d-une-alterite-informationnelle
DOI : numerev_1585
Date de publication : 12/05/2017

Cette publication est sous licence **CC BY-NC-ND** (Attribution - No commercial - No derivatives).

Pour **citer cette publication** : MARCON, Ch. (2017) Distance informationnelle scientifique : le risque d'une altérité informationnelle ?. *Revue COSSI*, (2). https://doi.org/10.34745/numerev_1585

A partir de l’hypothèse selon laquelle les chercheurs et laboratoires qui ne développent pas une politique de mise en ligne de leurs publications et données de recherche se mettent à l’écart du mouvement international d’open data scientifique en accroissant la distance informationnelle avec leurs travaux, cette communication présente les conclusions de l’étude des pratiques des laboratoires en sciences humaines de l’université de Poitiers en matière de données de recherche.

We assume that researchers and laboratories must develop a policy of digital diffusion of publications and research data in order to fit into the international movement for open scientific data. If then don’t, they increase the informational distance with their work and develop a risk of being isolated. This paper presents the findings of the study of the practices of social science laboratories of the University of Poitiers in research data.

Mots-clés :
Information, Données de la recherche, Open data, Distance informationnelle, Information, Research data, Open data, Informational distance

---

Dans un article paru en 2010, Bernhard Rieder a défini la notion de « distance informationnelle » comme « la probabilité qu’un internaute se trouve confronté à une certaine information (un document, une idée, etc.). Pensée comme relationnelle, cette distance dépend à la fois de l’usager, de l’information et de l’ensemble des couches médiatrices. »
Cette définition, de portée générale, nous interpelle pour ce qui concerne le champ particulier des chercheurs, devenus/devenant internautes. De même qu’il n’est pas suffisant qu’un jeune adulte soit *digital native* pour inférer qu’il maitrise évidemment ces usages, repères, savoir-faire et outils d’internet que l’on regroupe sous l’expression *digital litteracy*, il n’est pas certain qu’un chercheur soit, parce qu’il est apte à traiter de gros volumes de données en milieu complexe dans son domaine d’expertise, habile à réduire la distance informationnelle qui le sépare des productions scientifiques qui lui seraient utiles.
La question qui nous intéresse dans cette communication est celle de la démarche entreprise par des laboratoires de recherche afin de pousser vers les chercheurs ou, à

tout le moins, ne pas soustraire aux chercheurs, l'accès à des données de recherche qui pourraient trouver un second usage scientifique par leur réemploi par d'autres chercheurs ou équipes. Plus directement posée, la question est la suivante : les laboratoires de recherche ont-ils une démarche visant à réduire la distance informationnelle avec les données de recherche ?
Nous introduirons dans une première partie la notion de distance informationnelle scientifique, posant plus particulièrement la question de la distance à laquelle se trouvent les données de recherche. La seconde partie présentera l'étude menée auprès de sept laboratoires en sciences humaines de l'université de Poitiers quant à leurs pratiques de partage des données de recherche.

1- DISTANCE INFORMATIONNELLE SCIENTIFIQUE DES DONNÉES DE RECHERCHE

1-1- Introduction de la notion de la distance informationnelle scientifique

L'accès par Internet aux publications de recherche est désormais devenu d'un usage courant. Une quinzaine de plateformes, gratuites ou payantes, permettent cet accès (Froissard, 2010). Les universités mettent usuellement à disposition des étudiants et des chercheurs des offres qui articulent tout ou partie de ces plateformes professionnelles, quoi que s'observe un mouvement fort de désabonnement des revues numériques en ligne pour des raisons de tarifs (Larousserie, 2014 ; Bianchini, 2014). Malgré ce mouvement qui pose question à terme, la *distance informationnelle scientifique* que nous proposons d'approcher par la *probabilité qu'un chercheur n'accède pas via Internet à une certaine ressource numérique utile à son travail de recherche* est aujourd'hui relativement faible concernant les publications. Dans bien des cas, seule la question du coût d'acquisition du document constitue la limite à l'accès.
Cette probabilité dépend évidemment (1) des habiletés numériques du chercheur, lesquelles s'inscrivent dans le contexte plus global de ses « pratiques informationnelles » que Ihadjadene et Chaudiron (2008, p. 14) décrivent comme « la manière dont un ensemble de dispositifs, de sources formelles ou non, d'outils, de compétences cognitives sont effectivement mobilisées, par un individu[…] dans les différentes situations de production, de recherche, d'organisation, de traitement, d'usages, de partage et de communication de l'information.» Meilleure est l'habileté du chercheur, plus faible est la distance informationnelle scientifique (DIS).
Cette probabilité dépend également (2) des informations recherchées (données de recherche, résultats de recherche, textes scientifiques) et des « couches médiatrices » qui favorisent ou non l'accès à celles-ci, notamment les dispositions mises en œuvre par le collectif des chercheurs et les structures de recherche pour favoriser la structuration numérique des données, leur accessibilité et leur partage.
La distance informationnelle scientifique dépend enfin aussi, sans doute, (3) des normes

et règles en usage dans le groupe social d'appartenance du chercheur (Savolainen, 2008) dont les pratiques se « saisissent comme des phénomènes de signification dans des lieux sociaux spécifiques » (Souchier, Jeanneret et Le Marec, 2003).

1-2- Le cas des données de recherche

Le questionnement sur la distance informationnelle scientifique trouve un terrain d'analyse particulièrement sensible sur le sujet des données de recherche.
Les définitions et l'appréhension de l'étendue de ce que l'on peut nommer « données de recherche » sont assez variables. Le Groupe de travail sur la stratégie des données de recherche canadien définit comme des « *éléments factuels utilisés comme principales sources de recherche et communément acceptés au sein du milieu de la recherche comme étant nécessaires à la validation des résultats de recherche.* » (Rapport sur le sommet 2011, p. 4) La Ligue des Bibliothèques Européennes de Recherche (LIBER) les résume à la matière première des recherches quand Simukovic et al (2014, cités par Prost et Schöpfel, p. 10) retiennent à la fois ladite matière première qualifiée de « données sources » sur la base desquelles le chercheur construit son travail et les données « résultats » des travaux conduits par le chercheur.
Le questionnement sur le partage des données de recherche par leur publication, que les anglo-saxons désignent par l'expression *research data publishing* définie comme « *the release of (research) data for (re)use by others* » (Assante et al, 2016) n'est pas nouveau en soi. Si l'on en croît Mochmann (2007), les premières initiatives fortes en vue du partage des données pour la recherche comparée internationale datent de 1946. Caporali *et al* (2015) exposent de manière synthétique les multiples intérêts du partage des données pour les sciences sociales : production d'analyses secondaires, vérification et réplicabilité des études, progression des connaissances, découragement de la fraude scientifique, enseignement des méthodes d'analyse, amélioration de la connaissance du travail des chercheurs, justification du coût des enquêtes…
Ces avantages, ainsi que les exigences nouvelles dans le cadre des appels à projet H2020 par exemple, expliquent sans doute l'émergence de projets visant à favoriser le partage de données et l'accès à ces données, ce qui réduit d'autant la distance informationnelle scientifique. Dans le champ des sciences humaines, citons l'exemple du projet ADONIS visant à offrir aux chercheurs un « espace de navigation unifié » pour explorer les documents et les données numériques des sciences humaines et sociales. « Dans cette vision, les données correspondent aux matériaux bruts et les publications aux résultats de l'exploitation des données. » (Aussant, 2009). Plus largement, le projet Science Commons aspire également « à aller au-delà de l'accès ouvert aux résultats de la recherche, en facilitant un partage non seulement des données et de l'information mais des données brutes et des outils logiciels. » (Bourcier, 2010) Les réseaux sociaux de la recherche tels que Researchgate, Academia ou ou MyScienceWork se situent, nous semble-t-il dans la même logique, même s'ils recueillent essentiellement des travaux publiés.
Les exemples mentionnés ci-dessus s'inscrivent dans le mouvement émergent de l'open data scientifique, un mouvement qui ne va pas sans questionner profondément les chercheurs. Caporali et al ont expliqué l'inquiétude des chercheurs à l'égard de cette

dernière idée, citant De Moor et Van Zanden (2008) : « les chercheurs ne sont pas à l'aise avec l'idée que d'autres puissent bénéficier de jeux de données élaborés à grand-peine et, peut-être plus important encore, ils craignent qu'en rendant leurs données publiques, des erreurs ne soient découvertes dans la collecte et le traitement des données et dans les résultats qu'ils ont publiés ou qu'ils publieront ».
Il faut bien reconnaître que la mise en œuvre d'un projet de partage de données est d'une grande complexité et peut mettre en tension des pratiques des chercheurs. Ainsi, pour le Comité Consultatif Commun d'Ethique pour le Recherche Agronomique (2016, p.5):

> « Le partage des données suppose un travail rigoureux qui commence dès la conception d'un projet de recherche et peut se prolonger bien au-delà de son achèvement. En effet, la mise en ligne de données doit être précédée d'une délicate étape de préparation et d'organisation pour les rendre non seulement accessibles, mais aussi intelligibles en dehors de leur contexte d'origine. Produire des données de grande qualité représente une exigence déontologique majeure pour chaque chercheur. Cependant, le partage des données implique aussi de pouvoir fournir la preuve de leur fiabilité, et par conséquent de développer un contrôle qualité exhaustif. Enfin, la réutilisation des données ne sera possible que si elles sont « portables », « inter-compatibles » et « interopérables », ce qui implique un considérable effort de standardisation qui ne pourra être mené à bien qu'en s'appuyant sur des collaborations internationales. »

Une étude menée auprès d'environ 3000 chercheurs autrichiens publiée en 2015 (*Researchers and their Data. An Austrian Survey*), montre que dans 93% des cas, c'est le chercheur lui-même qui archive ses données de recherche, à 71% sur son propre ordinateur, dans 64 % des cas sur un disque dur externe. Seuls 27 % des répondants de l'étude affirment assurer de manière systématique cet archivage et 17% seulement selon des standards qui permettent un réemploi. Autre information importante : 39 % des chercheurs stockent leurs données de recherche sur le serveur d'une institution. Ces pratiques, dont nous faisons l'hypothèse qu'elles ne sont pas fondamentalement différentes en Autriche de ce qui se pratique ailleurs, assurent de manière peu rassurante la pérennité des données en les exposant à l'obsolescence du matériel, son renouvellement non maîtrisé en termes de transfert des données, une difficile réutilisation… Surtout, ce type de pratique individualisé ne favorise pas le partage des données de recherche.
Il semble par conséquent que la distance informationnelle scientifique des données de recherche est relativement forte, au contraire de la DIS des publications scientifiques : une large fraction des acteurs de la recherche rechigne à partager ses données ou ne sait comment faire, alors qu'elle aspire à voir ses publications lues et s'active en ce sens. Cela n'est pas sans poser question quant au risque de creusement d'une *altérité informationnelle scientifique* des laboratoires et chercheurs qui se tiendraient à l'écart du mouvement en faveur du partage des données de recherche. Les démarches entreprises par les universités de Lille, Paris Diderot… pour accélérer la réflexion et l'inflexion des pratiques en ce sens sont significatives d'une prise de conscience des enjeux.

2- RISQUE D'ALTÉRITÉ INFORMATIONNELLE SCIENTIFIQUE : UNE ÉTUDE DE L'ÉPAISSEUR SOCIALE DES PRATIQUES DANS LES LABORATOIRES DE SHS À L'UNIVERSITÉ DE POITIERS

2-1- Origine et conséquences de l’altérité informationnelle scientifique

L’altérité informationnelle scientifique *nait d’une mise à l’écart de la communauté scientifique internationale* aux multiples causes : non insertion dans les mouvements scientifiques (non participation aux colloques, non insertion dans des projets de recherche), difficulté de publication, difficulté de connexion à Internet, manque de moyens pour effectuer des recherches… Elle *consiste en une mise à l’écart de la dynamique de la recherche* qui a pour conséquence une moindre visibilité des travaux produits, une perte de contact avec l’actualité des découvertes qui peut conduire à une déconnexion des travaux par rapport aux dernières connaissances élaborées et une sorte d’oubli ou de mise à l’écart par les pairs.
Compte tenu du mouvement naissant évoqué plus haut, cette altérité pourrait bien aussi à brève échéance être causée par *le maintien durable hors du mouvement de transformation du rapport aux données de recherche qui s’opère dans le cadre de la transformation numérique de nos sociétés*. Celui-ci réduirait la probabilité que des productions scientifiques, ou des données sources élaborées ou collectées par des chercheurs, soient consultées et contribuent à de nouvelles recherches.
En résumé, l’altérité informationnelle scientifique se déduit d’un allongement de la distance informationnelle scientifique. Elle place le chercheur à l’écart du mouvement, l’isole, le marginalise. Il en va de même, probablement, pour les laboratoires de recherche qui ne se préoccupent pas du mouvement en cours.

2-2- Une étude de l’épaisseur sociale des pratiques

Notre étude du risque d’altérité informationnelle scientifique laisse de côté la question des habiletés individuelles des chercheurs en matière de numérique et celle des couches intermédiaires que représentent les plateformes d’accès en ligne aux publications et données, dispositifs de médiation qui mériteraient d’être questionnés par ailleurs. Ces aspects, et d’autres encore sont déjà largement abordés dans le livre blanc *Une science ouverte dans une république numérique* (CNRS, avril 2016). Nous nous centrons ici sur l’exploration de ce que Souchier, Jeanneret et Le Marec (2003) appellent « l’épaisseur sociale des pratiques », une épaisseur profondément marquée par le contexte de mise en œuvre des pratiques, notamment dans le cadre de systèmes normés. Les pratiques ont une épaisseur sociale si elles sont marquées par des normes

peu évolutives, ancrées dans la culture d'un groupe.
Plus précisément, nous avons étudié un aspect de l'épaisseur sociale des pratiques des laboratoires de recherche en sciences humaines et sociales de l'université de Poitiers s'agissant de la conservation, de l'organisation et de la mise en partage des données de recherche. Notre conviction est en effet que c'est au niveau de la direction d'un laboratoire que se trouve l'une des clés de l'évolution des pratiques de partage des données de recherche, donc une des clés de l'intégration à la communauté scientifique numérique ou du développement d'une altérité scientifique. Cette position peut être discutée. La Ligue Européenne des Bibliothèques de Recherche (LIBER) estime par exemple que les bibliothèques ont vocation à jouer ce rôle, mais avec prudence : « *Libraries should be careful not to place too much emphasis on 'open' (as in 'open data policies') because especially in the early stages of a research project most researchers do not want their data to be 'open', certainly not without their consent and outside their control.* »[1]
L'étude a pris la forme d'une série d'interviews semi qualitatives des directeurs des laboratoires de recherche. Il s'agissait, dans une perspective de recherche compréhensive, de saisir le regard qu'ils portent sur les données de recherche produites par leur laboratoire, d'appréhender la démarche de celui-ci en matière de conservation de ces données et le rôle attribué à chacun en cette matière, d'identifier l'institution d'une politique précise de gestion des données de recherche (ou non) et enfin de comprendre la réceptivité même des personnes interrogées au principe du partage des données de recherche.
Les sept entretiens ont duré en moyenne entre trente minutes et une heure et demie. Ils ont été réalisés, après formation, par des étudiants en première année de master dans le cadre d'un cours *d'introduction à la recherche*, qui a vu les étudiants participer à l'élaboration même de la démarche de recherche. Les entretiens ont été enregistrés puis entièrement reportés sur papier, hésitations et digressions comprises. C'est ce matériau que nous avons analysé pour être présenté.
En termes de disciplines, sept des neuf laboratoires en sciences humaines de l'université de Poitiers ont été questionnés. Il s'agit des laboratoires d'économie (CRIEF), de gestion (CEREGE), de sociologie (GRESCO), de géographie (RURALITE) et de langues et littérature (FORELL), d'histoire - histoire de l'art et musicologie (CRIHAM), histoire et archéologie (HERMA). Il n'a pas été possible de questionner les autres responsables sur la période consacrée à la collecte des résultats.

2-3- Les résultats de l'étude

√ Pour la conservation des données, la pratique commune est le stockage individuel, dans son propre ordinateur, avec les risques que cela implique qui semblent une sorte de fatalité.

- « Chacun est responsable de ses données. »
- « C'est souvent lié effectivement à une personne »
- Les recherches sont en général solitaires et chacun a son propre système, ce qui est une « catastrophe » car il n'y a aucune procédure de stockage

- « Chacun ici gère ses données. Enfin, c'est le milieu universitaire. »
- « Les gens on une vision du stockage individuelle des données donc c'est limité, quand c'est conservé ! »
- « Très franchement, je ne suis pas sûr que mes collègues voient un intérêt à ce que le laboratoire archive les données de recherche, mais il faudrait que je leur pose la question pour savoir. Chaque chercheur est persuadé du fait qu'il faut archiver les données de recherche, il n'y a pas le moindre doute par rapport à cela. Mais le fait de proposer à nos collègues de leur réserver un espace d'archivage de leurs données de recherche qui soit protéger avec un accès limité, pourquoi pas ? Il faudrait que mes collègues y trouvent un intérêt, et je ne suis pas sûr du tout qu'ils y trouvent un intérêt, donc je me pose la question et c'est une question que je pourrais poser à l'occasion, par exemple, du conseil scientifique du laboratoire. »

Deux directeurs posent la question des risques générés par cette conservation individuelle : la perte des données lorsque le chercheur cesse ses activités d'une part et le risque technique lié à la défaillance du matériel d'autre part. L'idée d'offrir des espaces de stockage via l'université ou une bibliothèque est assez peu évoquée, ou marquée par un fort doute quant à sa réalisation, en raison notamment de l'investissement en temps nécessaire pour classer les données. Sur ce point, le fait qu'aucun directeur de laboratoire ne soit issu de la documentation nous semble expliquer l'absence de questions relatives à la gestion de contenu et à la documentation des archives.
Seul le laboratoire d'archéologie est habitué à un stockage administratif extérieur, au niveau du ministère de la culture.

√ La question des types de données qui doivent être conservée et de la durée de leur conservation suscite un quasi consensus. Quasiment toutes les données de recherche méritent d'être conservées, sur la période la plus longue possible, en cohérence avec les usages propres à chaque domaine scientifique.

- « infinie » car, en archéologie on détruit le site que l'on explore.
- « Une donnée n'est pas disqualifiée, notamment par rapport à cette perspective socio-historique qui est de plus en plus présente en sociologie. Et il y aurait une autre raison […] qui serait une contribution à l'histoire de la discipline. »
- « On considère qu'on peut garder au moins 10 ans, travailler au moins 10 ans sur les même données » […] Donc il y a l'obligation dans nos disciplines de faire des séries longues et il y a l'obligation de garder aussi des données assez longtemps, considérant que ces données peuvent toujours illustrer le propos parce que c'est des choses qui ne changent pas aussi vite. »
- « Il est intéressant de conserver les données très longtemps pour pouvoir avoir une appréhension longitudinale des choses, c'est-à-dire une appréhension dans la durée. Il n'y a pas de règles universelles. »

√ La réticence au partage des données est très forte. Elle est principalement expliquée par le sentiment de propriété de ses données de recherche que confère l'habitude du travail individuel. Le partage n'est envisagé qu'à raison d'une relation étroite avec ceux avec qui l'on accepte(rait) de partager.

- Il faut « que tout le monde joue le jeu » ; je suis « ouverte, mais ça dépend des gens […] Il faut connaître les gens pour partager et d'une certaine manière avoir confiance »
- « On est quand même encore très marqué par la recherche individuelle du chercheur qui porte son projet, qui récupère son matériel. […] Et donc la mise en commun (cherche ses mots / hésitant) y'a le rapport à à son propre matériau qui fait que l'archiver ou le stocker, ça veut dire le rendre public […] Je ne dis pas que ça bloque, mais […] ça peut provoquer des blocages. »
- En sont responsables « tous ceux qui participent à une enquête donnée et une étude donnée. Ce n'est pas forcément une personne qui dirige les choses mais ça appartient à tous ceux qui participent au projet. C'est l'équipe des chercheurs qui disposent des données et qui a « droit de propriété » de ces données. C'est eux qui gèrent si on a le droit de les diffuser ou pas. »
- « On donne rarement bénévolement. Les enquêtes ont un coût. Il est difficile de dire « moi je ne supporte pas les coûts seulement les bénéfice ».
- « Les gens ne sont pas réticents du tout à travailler de manière collective. […] on a l'occasion aujourd'hui de le faire plus. Je vous parle pour mon labo, mais je pense que c'est quand même une tendance générale pour les Sciences Humaines et Sociales. Il y a plus d'occasions de travailler de manière collective. Ça n'empêchera pas que les gens continuent à travailler de manière individuelle. Pour une raison simple, c'est que l'évaluation dans la recherche en Sciences Humaines et Sociales se fait à partir d'articles. Ils sont plus souvent écrits seuls dans le domaine des Sciences Humaines et Sociales qu'en sciences. »
- « On considère, c'est un usage chez nous, que ces données appartiennent aux chercheurs. Et je pense que dans beaucoup de champs disciplinaires c'est le cas. On ne va pas nécessairement mettre en commun des données, sauf peut-être sur des programmes de recherche où l'on travaille à plusieurs, des programmes de recherche collectifs. […] Tout travail mérite salaire. A partir du moment où l'on fait une recherche où l'on fait des efforts pour recueillir des informations et les analyser soi-même, on n'a pas envie que la Terre entière y ait accès. Pas même les autres membres du laboratoire. Je vois difficilement comment les chercheurs accepteraient comme ça de partager leurs données de recherche avec tout le monde, même après les avoir exploitées. »
- Ne pas donner ses données de recherche par peur d'être jugé sur ce que l'on a fait ? « Alors, non ça ne rentre pas vraiment en ligne de compte. »

Plusieurs directeurs de thèses considèrent différemment le cas des doctorants à qui l'on demande de joindre leurs données de recherche en annexe de leurs thèses, ce qui revient à leur imposer de les partager, à tout le moins avec les membres du jury. Pour l'un d'entre eux, en contrepartie, « les collègues seraient complètement prêts à donner leurs sources de base à des doctorants, s'ils veulent faire des comparaisons. »

Il faut signaler également que la diversité des champs disciplinaires couverts par les directeurs de laboratoire interrogés a mis en évidence une difficulté triviale, celle due à la grande diversité des données de recherche : enregistrements d'entretiens, prises de notes, journal, bases de données quantitatives, statistiques, archives, matériaux archéologiques, sources littéraires anciennes, gravures, inscriptions, archives, presse, sculptures, œuvres d'art, partitions de musique… Tout matériau de recherche doit être considéré non seulement dans sa dimension *donnée* mais également dans sa matérialité, laquelle rend substantiellement différente la problématique du stockage et du partage selon les types d'objet supports. Si des photos ou des images numériques peuvent, dans bien des cas, constituer des empreintes satisfaisantes en terme de contenu informationnel, le contact avec l'objet lui-même reste, dans certaines disciplines, difficilement remplaçable, ce qui fait du partage de données numériques une opération appauvrie dans sa substance dans certaines disciplines.

√ Le questionnement sur les risques associés à la mise à disposition d'autres chercheurs de certaines données, notamment les risques déontologiques liés à la confidentialité, est fort dès que le matériau de travail du laboratoire comporte une forte proportion d'études qualitatives par entretiens.

- « On peut diffuser les analyses que l'on peut faire à partir d'une enquête. Mais livrer les chiffres eux même, les informations, c'est plus compliqué. Il y a des problèmes de déontologie qui se posent. A mon sens dès l'instant où c'est un peu confidentiel, c'est toujours difficile de mettre sur la place publique. »
- « Il y a quand même une frontière qui est une frontière déontologique de la profession. Quand on fait des entretiens par exemple, on garantit l'anonymat. »
- « Il peut y avoir des données confidentielles, parce qu'il y a beaucoup d'enquêtes qualitatives. On pourrait cacher les noms, anonymer les enquêtes sans problèmes. Les thèmes que l'on traite, c'est évidemment des thèmes qui intéressent le grand public. Il faudrait arriver à faire de l'open data […] D'ailleurs le fait que l'on ait [ce] contrat avec la région va nous y obliger. Je pense que la région va se poser la question avec nous de savoir comment elle peut rendre accessible toutes ces données à la communauté humaine. »
- « Rendre les enquêtes anonymes c'est complètement possible. Alors après [le problème vient] des tous petits échantillons. Même si c'est anonymé, on reconnait tout de suite d'où ça vient. […] Si vous faites une enquête sur la concentration des immigrés à Poitiers, si vous êtes avec des petits échantillons de blocs de maisons, de blocs de HLM par exemple, assez vite les gens du quartier qui consulteraient ces données pourraient faire le lien entre les réponses et les auteurs. »
- « On va se poser la question de savoir si l'accès à cette connaissance qui sera mise à disposition du public peut entrainer des réactions discriminatoires. Je crois que dans 99,9 % des cas ce ne sera pas le cas. Il faudra regarder quand même les 0,01 % restant. Qu'est ce que ça veut dire discriminatoire ? Qu'on ne peut pas utiliser ces données contre les gens qui sont concernés par cette enquête là. »

√ Lorsque la question du partage des données avec le grand public est abordée, elle

suscite de nombreuses appréhensions. L'un évoque des problèmes sémantiques (interprétation des mots plus connotée par le grand public que par les chercheurs), l'aspect disparate des données difficile à comprendre, la confidentialité dans le domaine du qualitatif, la difficulté technique d'accès aux données, une nouvelle fois le sentiment de la propriété des données, enfin le risque de perturbation des sites de fouille et de pillage si leur localisation devait être tout à fait précise.

> Je serais réticent sur les cahiers d'observations ethnographiques parce que c'est noté […] à l'état brut. Alors faut pas non plus que ce soit mésinterprété […]ces archives pour qu'elles soient intéressantes, [il faut qu'elles] soient reliées aux productions scientifiques réalisées à partir de ces archives. Parce que si elles sont à l'état brut, ça peut être difficile de comprendre dans quel contexte elles ont été obtenues et à quoi elles ont servi en fin de compte. »

√ Notre hypothèse était que le directeur du laboratoire a un rôle central dans l'impulsion d'une réflexion ou d'une action en matière de conservation et/ou de partage des données. Or les directeurs interrogés ne se considèrent pas responsables dans ce domaine et n'ont rien entrepris de particulier en matière de conservation sécurisée des données ou de partage de données. Deux groupes se distinguent : les plus nombreux sont ceux qui semblent considérer l'état actuel comme non réformable (pour les raisons évoquées ci-dessus) ; deux par contre reconnaissent se sentir questionnés dans leurs pratiques par l'interview à laquelle ils ont répondu.

- « Pas directement en tant que directeur, mais ce sont les responsables des études qui sont responsables de leurs données. Ils les archivent eux-mêmes. Il n'y a pas de centralisation au niveau du laboratoire. Chacun est propriétaire de ses données dès l'instant ou ça résulte d'enquête. »
- « Il faudrait qu'on progresse là-dessus. On n'a pas systématiquement versé dans un pot commun quand les personnes s'en vont, […] la partie base de données de leurs recherches. Moi je vais le faire par exemple [à propos de la matière réunie par une doctorante], enfin on le mettra quelque part. Mais on ne sait pas où est le quelque part justement. […]. Quand les gens partent ou quand les gens ont terminé une recherche, c'est un vrai problème. C'est une question qu'il faudra que l'on se pose. Votre venue elle est utile pour moi parce quelle me fait prendre conscience de ça. »
- « Ce sont les chercheurs qui conservent par devers eux les données de recherche. Le laboratoire n'y a pas accès. Pourquoi le laboratoire n'y a pas accès, parce que nous considérons quelque part que ces données de recherche relèvent de la propriété intellectuelle de chaque chercheur.
- « A l'échelle de l'équipe on ne sait jamais vraiment posé la question de mettre en commun des données »
- « Il va falloir qu'on y pense »

√ Notons pour terminer que l'un des entretiens a évolué vers une réflexion, non recherchée initialement, sur l'hypothèse de la vente de données de recherche, forme

envisageable de partage des dites données. La réaction du directeur de laboratoire confronté à cette question a été celle d'une opposition très vive et ferme à ce qui lui est apparu comme de la sous-traitance, voire de la « prostitution intellectuelle.

> « Est-il envisageable pour les chercheurs de partager leurs données de recherche en échange d'une contrepartie économique ? Personnellement j'y suis hostile. […] Je pense que d'autres ne seraient pas contre, mais j'y suis hostile parce qu'à ce moment-là, si les données de recherche deviennent des données commercialisables… de toute manière rassembler des données qui n'ont pas été recueillies de la même manière, avec le même protocole et essayer d'exploiter des données plus larges, d'un point de vue scientifique cela va poser des problèmes méthodologiques. Imaginons qu'on essaye d'utiliser les données collectées sur une recherche de type entretien semi-directif mené chez un constructeur automobile, une recherche de type ethnographie menée chez un fabricant d'électroménager, le mode de collecte des données n'a pas nécessairement été le même et donc ensuite essayer de rapprocher ces données va poser un problème méthodologique énorme. Je dis bien d'un point de vue méthodologique de recherche. D'un point de vue économique, je vois un problème d'éthique qui se pose : c'est que les équipes de recherche les mieux dotées en moyens financiers seront capables d'acheter les données des autres et donc produiront des recherches avec des données à plus large échelle, ainsi de suite, et dans le contexte académique actuel où plus vous faites d'études de cas, plus vous multipliez les données, plus on pense que c'est fiable, [alors que] ce n'est pas toujours le cas, alors ils en tireront un mérite supérieur, ils en garderont une reconnaissance supplémentaire et récolteront des fonds plus importants. Donc en définitive, si vous faites ça, vous aller mettre en place un système où l'on ne va prêter qu'aux riches. Et ce sont les gros laboratoires avec de gros moyens financiers qui vont récupérer les données. » […] Nous [pourrions] peut-être gagner de l'argent en faisant ça, mais en termes de reconnaissance de notre qualité de chercheur, il n'y aurait plus rien. Je vais utiliser un terme très fort : c'est de la prostitution intellectuelle, c'est-à-dire que dans un domaine où nous devons valoriser notre capacité à fournir des résultats de recherche et construire des théories originales, certains d'entre nous choisiraient de ne plus être reconnus pour ça, simplement faire de l'argent en vendant leurs services à d'autres qui cherchent la reconnaissance. »

3- CONCLUSION

Notre étude montre bien l'existence *d'une forme d'épaisseur sociale commune* aux différents laboratoires de sciences humaines étudiés, celle de pratiques de conservation des données avant tout personnelles accompagnées d'une forte réticence au partage des données, partage acceptable avec les membres d'un même projet de recherche, voire avec un tout petit nombre de chercheurs proches mais sans plus. Le partage dans un esprit *d'open data* est quasiment exclu. Si l'expression de ces habitus varie, le fond

reste le même. Certains interrogés estiment ces usages communs aux sciences humaines et sociales qui seraient moins habituées aux pratiques collectives que les sciences de la matière ou la médecine par exemple. Nous entendons cette explication mais n'y souscrivons pas.
Il serait tout à fait prétentieux d'affirmer quoi que ce soit pour l'avenir, mais nous estimons que cette épaisseur sociale peut être à l'origine d'un risque réel d'altérité informationnelle scientifique des laboratoires de l'étude et, plus largement, des laboratoires qui resteront dans cette démarche. Il est à craindre que les laboratoires ne se comportent pas en « couches médiatrices » du partage des données mais en filtres protecteurs de leurs chercheurs.
La question mérite d'être posée stratégiquement. Le portail européen pour la recherche et l'innovation Horizon 2020 suggère une volonté d'aller vers un partage des données de recherche, dès lors qu'elles ont fait l'objet d'un financement européen. Les équipes et laboratoires qui entrent dans ce processus prennent l'habitude de ce partage. Ce faisant, ils obtiennent des financements et ne sont pas dépouillés de leurs possibilités de publication par les voies usuelles scientifiques. En termes d'intelligence économique, l'influence de ces laboratoires promet de s'accroitre alors que leurs pratiques évolueront dans le sens de l'*open data* aujourd'hui fortement préconisé. Partager des données qui ont été exploitées, ce n'est pas se priver de la possibilité de les réexploiter soi-même. Ce peut être aussi une manière d'affirmer un travail réalisé. Selon le Knwoledge Exchange and Science Europe Briefing Paper de mai 2016 (p. 8), « *Research Data Management is undeniably part of good scientific practice ; RDM implies specific tasks and responsibilities which require adequate support and provision to be properly undertaken by researchers.* »
Les règles du jeu scientifique sont en train de changer. Il sera peut-être difficile à des scientifiques qui aspirent, comme citoyens, à un large partage des connaissances dans le contexte de l'open data administratif par exemple, de maintenir la privauté de données de recherche lorsqu'elles ne sont sous-tendues par aucune une obligation de confidentialité. Si les arguments techniques, méthodologiques, documentaires demandent un examen attentif, il nous semble que certains arguments deviendront assez vite obsolètes dans les années à venir, notamment l'argument de la propriété individuelle des données de recherche par un chercheur appointé par la nation. Les nombreuses initiatives prises au cours des derniers mois, évoquées plus haut, semblent donner le sens de l'histoire dans ce domaine.

4- BIBLIOGRAPHIE

Assante M. et al, (2016) Are Scientific Data Repositories Coping with Research Data Publishing, *Data Science Journal*, 15 :6, pp. 1-24
Aussant D., (2009), Le nouveau visage des données de recherche, *Documentaliste - Sciences de l'information*, vol. 46, n°1, p. 15-17
Bianchini L. (2014), Dans les universités, désabonnements en masse aux revues scientifiques, *L'Obs-Rue 89*, mis en ligne le 3 février. Disponible à l'adresse suivante : http://rue89.nouvelobs.com/2014/02/03/les-universites-desabonnements-masse-revues-

scientifiques-249590
Bourcier Danièle, (2010), Science commons : nouvelles règles, nouvelles pratiques, *Hermès, La Revue* 2010/2 (n° 57), p. 153-160.
Caporali A. et al., (2015), La mise à disposition des enquêtes quantitatives en sciences sociales : l'exemple de l'Ined, *Population* 2015/3 (Vol. 70), p. 567-597.â¨DOI 10.3917/popu.1503.0567
Chaudiron S. & Ihadjadene M., (2009), Des processus aux pratiques : quels modèles informationnels pour analyser l'accès à l'information en contexte professionnel ?, GRESEC. *Evolutions technologiques et information professionnelle : pratiques, acteurs et documents*, pp.1-12. Accessible en ligne sous la référence : hal-00468728
Chaudiron S. & Ihadjadene M., (2010), *De la recherche de l'information aux pratiques informationnelles*, [En ligne] Consulté le 11 février 2016. URL : http://edc.revues.org/2257
Chaudiron S., Maignant C., Schöpfel J., Westeel I., (2015) Les données de la recherche dans les thèses de doctorat - Livre blanc. [Rapport de recherche] Université de Lille 3.
Connehaye E., coord, (2016), *Avis sur les enjeux éthiques et déontologiques du partage et de la gestion des données de recherche*, INRA Science & Impact et CIRAD, 32 p.
Direction de l'Information Scientifique et Technique, CNRS, (2016) Une science ouverte dans une république numérique – Livre blanc, avril, 195 p.
E-infrastruture Austria, (2015), *Researchers and their Data. An Austrian Survey*, 206 p.
Froissart P., (2010), Les principales plates-formes de revues scientifiques, *Hermès, La Revue* 2010/2 (n° 57), p. 119-120
Gardiès C., Fabre I. & Couzinet V., (2010), Re-questionner les pratiques informationnelles, *Études de communication* [En ligne] 35 Consulté le 24 janvier 2016. URL : http://edc.revues.org/2241
Groupe de travail canadien sur la stratégie des données de recherche (2011) *Rapport sur le Sommet 2011 sur les données de recherche canadiennes*, http://rds-sdr.cisti-icist.nrc-cnrc.gc.ca
Knowledge Echange Research Data Expert Groupe et al, (2016) *Funding research data management and related infrastructures*, 28 p.
Larousserie D., (2014) Des universités de désabonnent des revues scientifiques, *Le monde sciences et techno*, 2 février. Disponible en ligne : http://abonnes.lemonde.fr/sciences/article/2014/02/10/des-universites-se-desabonnent-des-revues-scientifiques_4363717_1650684.html
LIBER working group on E-Science / Research Data Management (2012), *Ten recommendations for libratires to get started with research data management*, 4p. http://libereurope.eu/wp-content/uploads/The%20research%20data%20group%202012%20v7%20final.pdf
Mochmann E., (2007), Un meilleur corpus de données pour la recherche comparée internationale, *Revue internationale des sciences sociales* 2007/3 (n° 193-194), p. 547-567. DOI 10.3917/riss.193.0547
Rieder B., (2010), Pratiques informationnelles et analyse des traces numériques : de la représentation à l'intervention, *Études de communication*, 35, Consulté le 25 janvier 2016. URL : http://edc.revues.org/2249
Savolainen R, (2008), *Everyday Information Practices : A Social Phenomenological*

Perspective, Plymouth : The Scarecrow Press
Souchier, E., Jeanneret, Y. & Le Marec, J., (2003), *Lire, écrire, récrire : objets, signes et pratiques des médias informatisés*, Paris, Bibliothèque publique d'information, 350 p.

[1] Final report of the LIBER working group on E-science / Research Data Management, juillet 2012, p. 2