

NUMEREV

16 octobre 2018, Montpellier

**« Partage des données de recherche en SHS :
quelles pratiques
des enseignants-chercheurs ? »**



**Alexandre Serres,
Réfèrent à l'intégrité scientifique
de l'Université Rennes 2,
ex-coresponsable de l'URFIST de Rennes**



En guise de plan

- L'enquête de Rennes 2
- Complexité(s) des données de recherche en SHS
- La diversité des données
- Partage des données : opinions, représentations et pratiques
- Conclusion

Présentation de l'enquête de Rennes 2

3

L' ENQUÊTE : PAR QUI ET SUR QUI ?

4



► Acteurs :

- Co-organisateurs : URFIST de Rennes, MSHB et SCD Rennes 2
- Soutiens de l'université

► Cible, périmètre :

- Université Rennes 2 : les SHS
- Les enseignants-chercheurs, chercheurs et ingénieurs ; pas les doctorants.

► Trois objectifs :

- Faire un état des lieux
- Elaborer une offre de service
- Alimenter la recherche sur les pratiques informationnelles

L'ENQUÊTE : QUELLE MÉTHODOLOGIE ?

5

► Double méthodologie :

► Une enquête quantitative : questionnaire statistique en ligne :

► Sources du questionnaire : [Les données de recherche en SHS. Une enquête à l'Université de Lille 3](#) et deux autres enquêtes ; questions retravaillées

► Un long processus d'élaboration : de mai 2016 à février 2017

► 32 questions structurées en cinq axes :

- Informations personnelles
- Typologie des données de recherche
- Pratiques de stockage et d'archivage
- Pratiques de partage et de diffusion
- Besoins

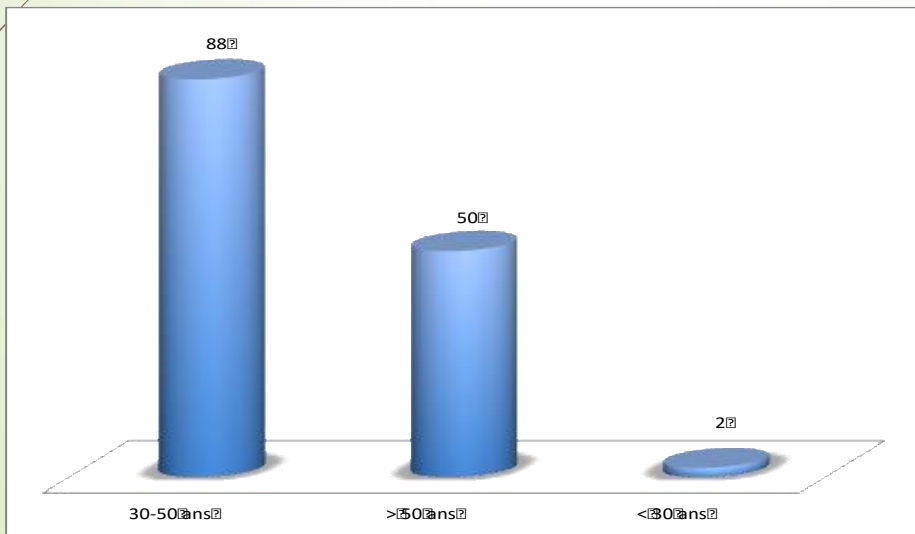
► Une enquête qualitative : 21 entretiens (d'environ 30 mns)

- Chercheurs volontaires
- VP, Directeurs de labos volontaires

> Un matériau très riche

RÉSULTATS DE L'ENQUÊTE STATISTIQUE : TAUX DE RÉPONSE

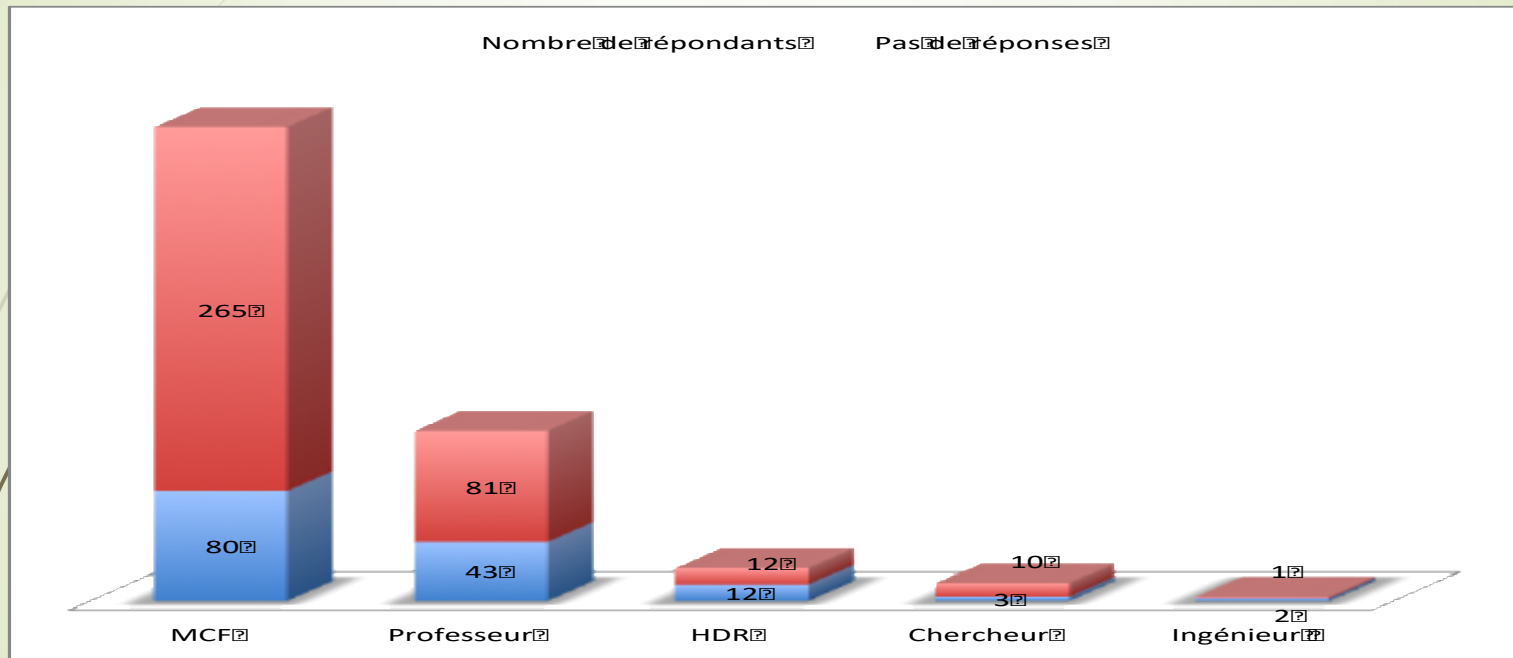
- **496 invitations** envoyées par mel, via l'outil LimeSurvey ;
- **143 participants** : soit **28%** de taux de participation
- 109 réponses complètes ; 34 réponses incomplètes



Répartition des
répondants par tranche
d'âge (N=140)

TAUX DE RÉPONSES AU QUESTIONNAIRE

7



Effectifs de répondants par statut

Complexité(s) des données de recherche en SHS :

8

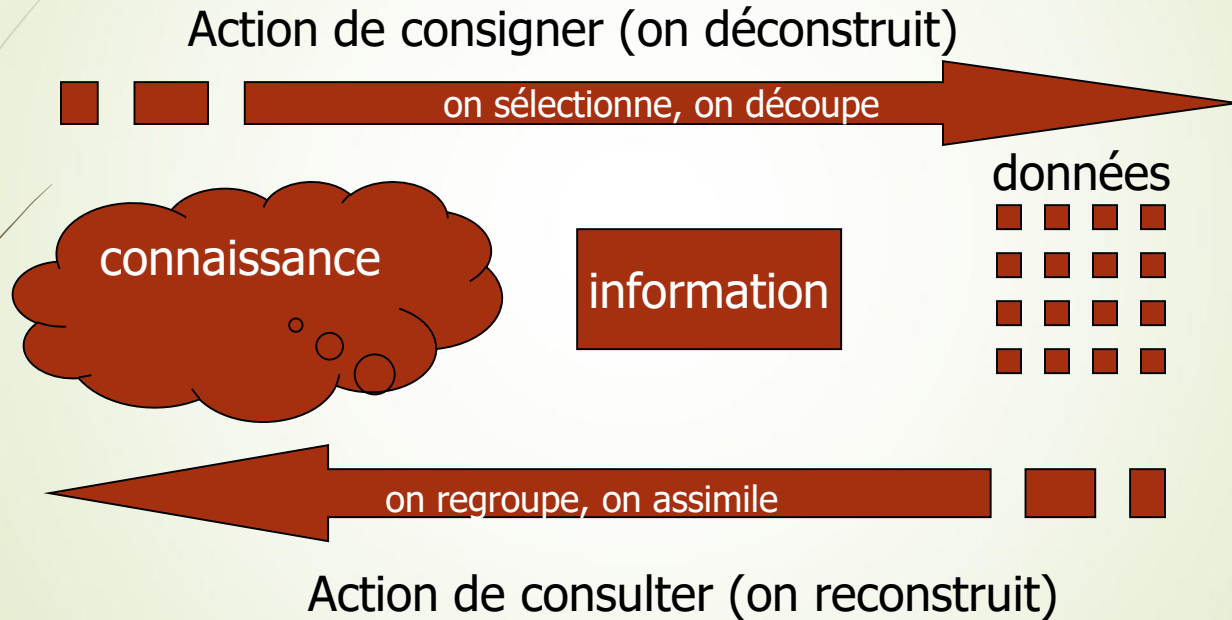
- Théorique
- Épistémologique
- Sémantique
- Complexité du partage

Bref retour aux fondamentaux

- **Qu'est-ce qu'une donnée ? Quelle différence avec l'information ?**
 - Définition de Serge Abiteboul
 - **« La donnée serait la traduction la plus immédiate, la plus brute, d'un fait. Elle n'est pas le fait, mais l'unité minimale d'observation qui permet de le caractériser. »** (Bruno Teboul, « Vers une philosophie de la donnée »)
- **Les données jamais vraiment « données » :**
 - **« Les données sont toujours sélectionnées, transformées, car on ne peut accéder à la totalité des données passées, présentes et à venir. Les données procèdent d'une différence ou d'un manque d'uniformité dans un contexte et surgissent du fait d'un certain cadrage »** (Sylvie Leleu-Merviel)
- **Les données comme matrice de l'information**

Trois grandes catégories de l'information

Connaissance \leftrightarrow Information \leftrightarrow Données



Complexités épistémologique et scientifique

- ▶ **Statut épistémologique de la donnée en SHS :**
 - ▶ Toujours une construction
 - ▶ Perspective constructiviste des jeux de données
 - ▶ Exemples des corpus de textes, d'images, collections de liens, bases de données, etc.
 - ▶ **Une conséquence juridique importante :**
 - ▶ La question du droit d'auteur et de la protection de la paternité des données
- ▶ **Question scientifique de la contextualisation des données :**
 - ▶ Objection de plusieurs chercheurs (géographie sociale, socio-linguistique...) sur le partage :
 - ▶ Données produites toujours en contexte
 - ▶ Contexte pas toujours « descriptible »
 - ▶ Données pas toujours réutilisables

Complexités épistémologique et scientifique

► Les données sont une construction :

« Le problème, c'est que l'on parle de "données" mais en fait, pour moi, ce ne sont pas des données, ce sont des construits. [...] Le mot "donnée", concrètement cela m'embête. Cela veut dire que les choses sont données. Si vous avez construit tout un appareil méthodologique qui permet de prélever des éléments, en aucun cas c'est donné, mais c'est construit, prélevé... Donc cela 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. »

D'après Morgane Mignon, MSHB

• C'est le travail de recherche qui confère une valeur scientifique aux données :

« Ma collecte de données, ce sont [...] tous les échanges avec les artistes, mais qui ne sont pas méthodiques, ce ne sont pas des formes d'enquêtes ou d'interviews, mais 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. »

D'après Morgane Mignon, MSHB

Complexités épistémologique et scientifique

► L'objection de la contextualisation :

► En sciences du langage :

« Dans ma discipline, les sciences du langage (...), 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. »

► En sociologie :

« 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. » ;

Complexités épistémologique et scientifique

- **Rapport mutuel d'influence entre les données et les pratiques de recherche**

« Et puis après effectivement, tous ces logiciels comme Gephi, Paraview [...] cela 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 cela a bouleversé ma façon de travailler, cela a radicalisé ma façon de travailler. »

UNE NOTION COMPLEXE

15

« 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 »

« C'est difficile de savoir où commence le traitement. »

« La donnée que je manipule est diverse »

« 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 [...] en géographie sociale, [...] ou certains sociologues ou ethnologues, etc. bien qu'on soit pas dans la même discipline »

Complexité sémantique

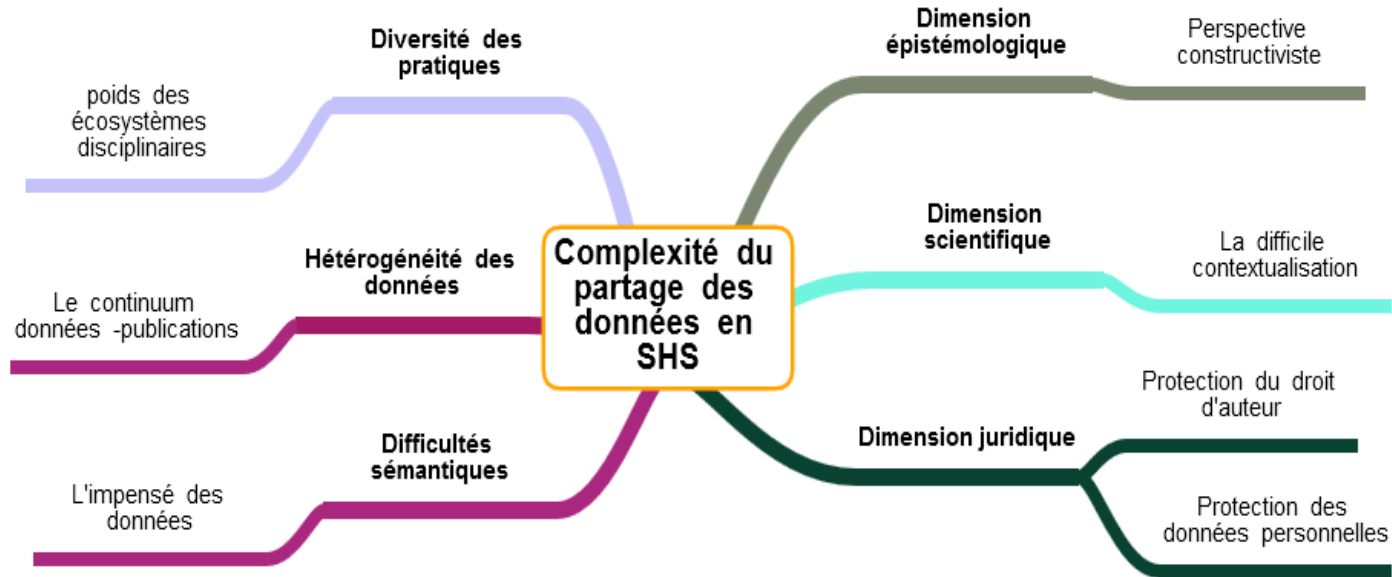
- ▶ Deux exemples de définitions officielles :
 - ▶ OCDE, 2007 :
 - ▶ « les « données de la recherche » sont définies comme des **enregistrements factuels** (chiffres, textes, images et sons), qui sont utilisés comme **sources principales pour la recherche scientifique** et sont généralement reconnus par la communauté scientifique comme **nécessaires pour valider des résultats** de recherche. » (souligné par nous)
 - ▶ « Proposition de DIRECTIVE DU PARLEMENT EUROPÉEN ET DU CONSEIL concernant la réutilisation des informations du secteur public (refonte) », 2018 :
 - ▶ « données de la recherche », des **documents** se présentant **sous forme numérique, autres que des publications scientifiques**, qui sont recueillis ou produits au cours d'activités de recherche scientifique et utilisés comme **éléments probants** dans le processus de recherche, ou dont la communauté scientifique admet communément qu'ils sont **nécessaires pour valider** des conclusions et résultats de la recherche ; »

Complexité sémantique

- Pour les chercheurs, difficulté à **nommer les données** :
 - Les données comme *impensé* de la recherche en SHS
- **Pas de délimitation claire entre données et publications**, entre données « primaires » et données produites :
 - Idée de continuum
- **Difficultés des chercheurs par rapport à la « terminologie IST »** et à la typologie des données :
 - Notions de « données sources », « données résultats », amont-aval...
- **Très grande hétérogénéité** des données de recherche

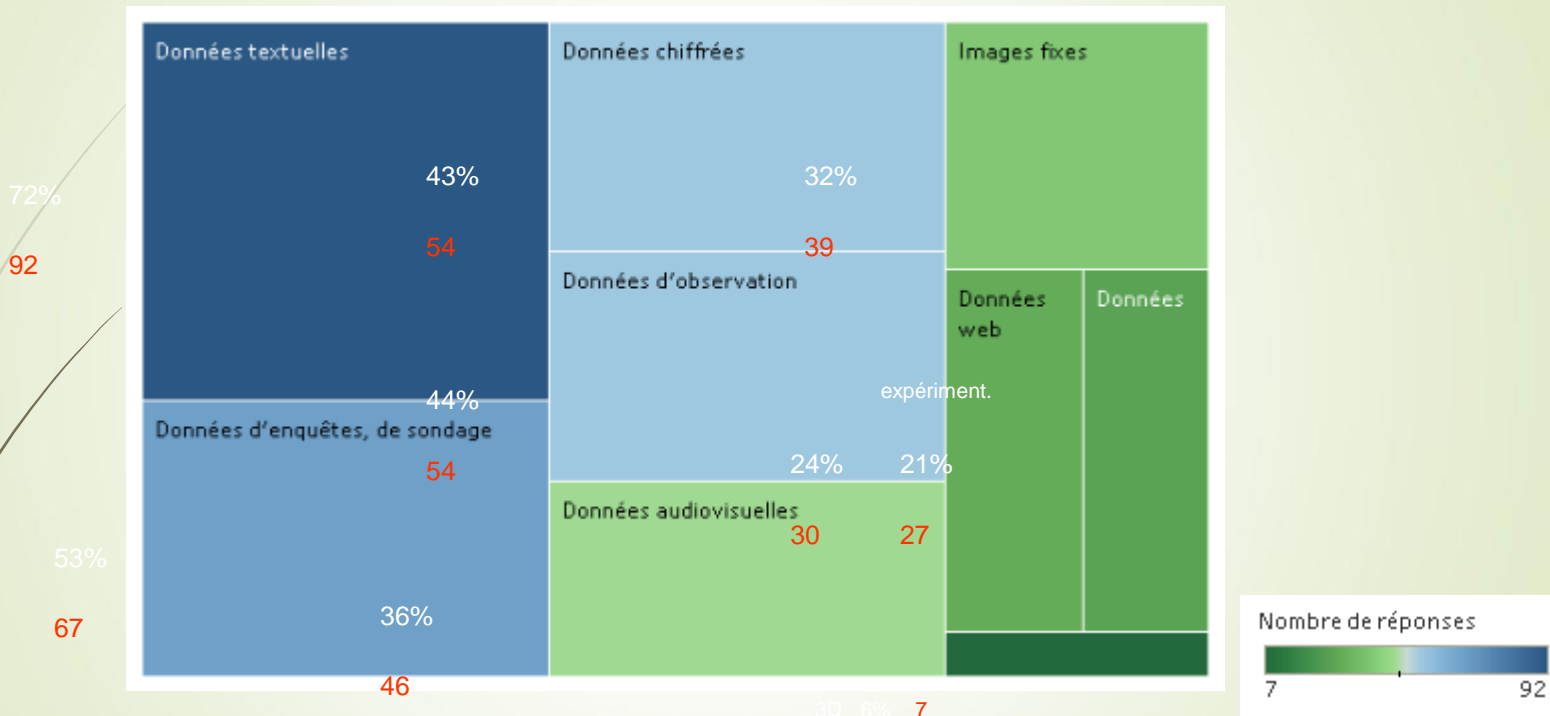
Complexité du partage

18



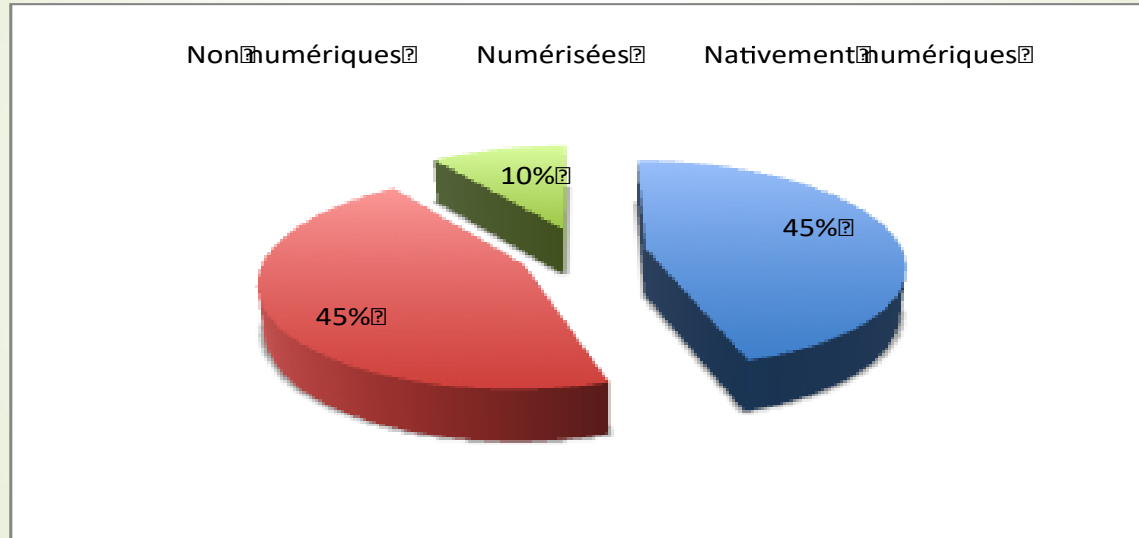
Diversité des données de recherche en SHS

QUELS TYPES DE « DONNÉES SOURCES » ?



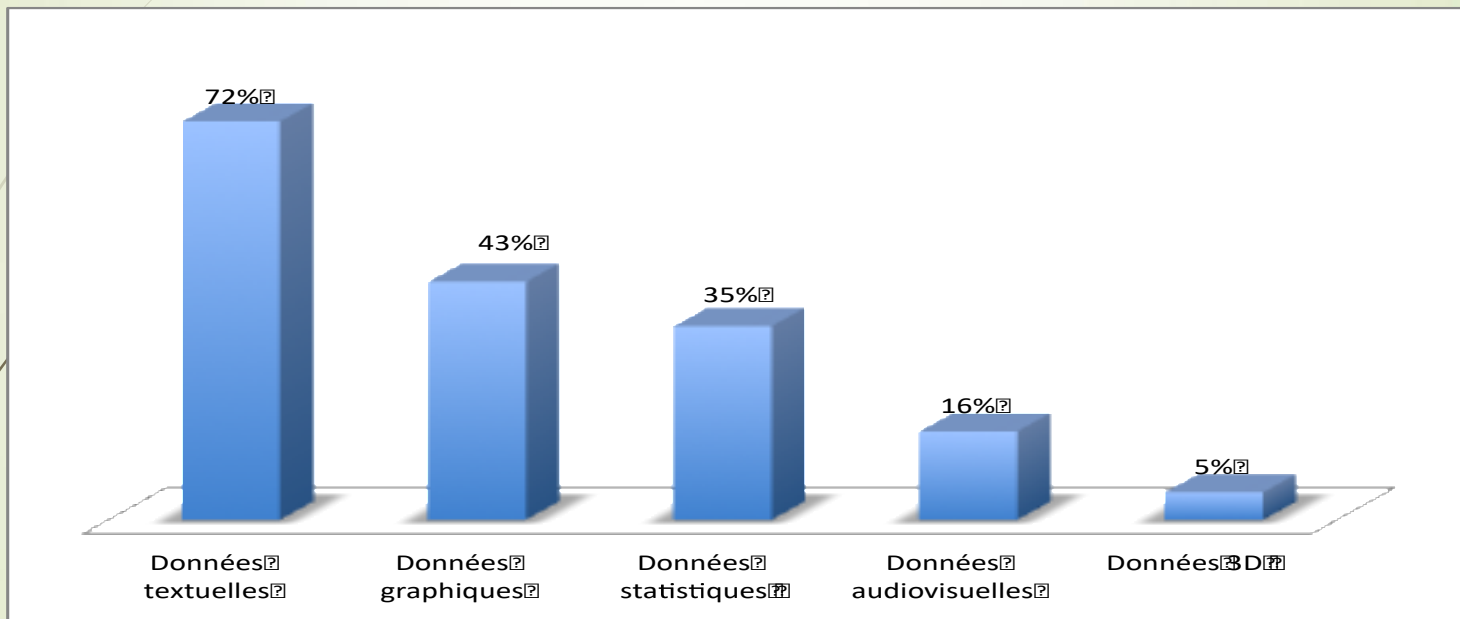
Catégories des données sources (N = 127)

DONNÉES NUMÉRIQUES OU NON



Support matériel des données sources (N = 127)

QUELS TYPES DE DONNÉES PRODUITES ?



Catégories de données produites (N = 127)

Hétérogénéité des données de recherche collectées

« du matériau langagier [extrait d'] un terrain ethnographique [investi] sur la durée, sur la profondeur, avec une compréhension très contextualisée de la situation ; des conversations ; du témoignage ethnographique ; des entretiens ; des biographies langagières orales ou écrites ; des documents d'archive ; des documents audiovisuels ; des émissions radiophoniques; des corpus de presse; des programmes, les discours des inspecteurs généraux, des manuels ; des corpus de dossiers d'aide à l'enfance ; des corpus de textes de loi ; des échanges avec les artistes ; des interviews avec des collègues ; des relevés GPS ; des données d'enquête ; des images ; des textes ; des corpus de textes ; des corpus visuels ; des adresses ; des ouvrages ; des articles ; des facsimiles d'ouvrages ; des documents ; des brochures ; des photos ; des fragments de pierres, des inscriptions ; de la donnée géographique numérique ; des cartes papier ; des données de type raster, matriciel, tout ce qui est image, photographie aérienne, image satellite ; de la donnée vectorielle, avec une composante géométrique (la donnée c'est de l'objet) ; de la donnée attributaire (ce qui vient décrire l'objet ; des tables) ; des information informelles glanées lors de colloques ; des données récupérées sur Internet ; des SMS ; des enregistrements de webcams ; des vidéos (youtube ou autres plateformes) ; des pages de forums Internet, de discussion, sites web, ; des pages de tchat SOS ; des pages ou échanges de réseaux sociaux, comptes twitter, Facebook, ; des PDF d'œuvres sur Internet ; des index et listes ; des photos aériennes. ; des données numériques, statistiques ; des cartes ; des données archéologiques ; des captures d'écran ; des traces GPS ; des trajectoires de marche ; des indicateurs liés à de petites tâches cognitives ; des données territorialisées... »

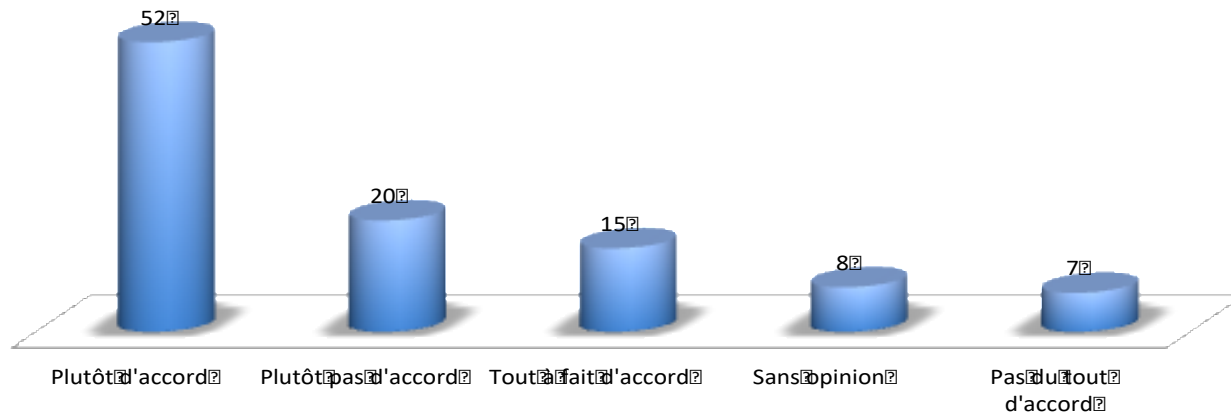
Hétérogénéité des données de recherche produites

- « des données textuelles ; des bases de données ; des enregistrements audio ou vidéo d'entretiens ; des notes sur des entretiens ; des questionnaires ; des formulaires papier remplis ; des tableaux Excel ; des fichiers Word ; plusieurs versions de textes ; des fiches à partir de documents d'archive ; des sites ou plateformes internet ; des retranscriptions d'entretiens ; des analyses statistiques lexicales ; des dossiers PDF, des textes, essentiellement des textes ; des livres ; des revues ; des données excel ; des données chiffrées ; des données qualitatives ; des études de cas cliniques ; des logiciels ; des tâches ; des petits programmes ; de petites épreuves cognitives ; des tests de raisonnement ; des fichiers texte avec reconnaissance de caractères, des transcriptions ; des documents numérisés ; des photographies de documents originaux ; des enregistrements vidéo ; des rushes ; des documentaires ; des catalogues prosopographiques ; des arbres généalogiques ; de recensements d'archives ; des campagnes de photos ; des fiches thématiques ; des bases de données géographiques ; des traces GPS ; des indicateurs ; des données biologiques, des dosages ; des électrocardiogrames ; des consentement de participation (papier) ; des protocoles ; des fichiers Latex ; des fichiers Matlab ; des fichiers partagés ; des bibliographies ; des fichiers statistiques ; des nuages de points, des captures de mouvement ; des photographies aériennes ; des cartes ; des croquis ; des enregistrements sonores suivis sur plusieurs années ; des photographies suivies sur plusieurs années ; des corpus multilingues ; des carnets papier... »

Partage des données : opinions, représentations et pratiques

PARTAGE DES DONNÉES : LES OPINIONS

Je suis prêt à partager mes données de recherche avec les autres



N = 102

PARTAGE DES DONNÉES : LES OPINIONS

« Sur le principe, j'y suis plutôt favorable. L'idée paraît séduisante, ouvrir ses données et récupérer des données ailleurs puis travailler dessus. Après c'est la manière dont cela 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 ? »

« Cela dépend du cadre légal du partage des données : ça c'est essentiel, parce qu'on sait de toute façon qu'il y a de la concurrence sur l'acquisition des données et des choses comme ça... »

« Ça ne me pose pas de problème de partager [mes données], à condition d'en être crédité et que cela 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. »

PARTAGE DES DONNÉES : LES REPRÉSENTATIONS

28

Facteurs de frein
au libre accès
de ses données
de recherche
N = 108

	Nb de réponses	Pourcentage
Des freins juridiques liés à la nature de vos données	58	54%
Le manque de temps	55	51%
La crainte du plagiat	40	37%
Des raisons scientifiques	29	27%
Le manque de connaissances dans la description des jeux de données	28	26%
Le manque de compétences	20	19%
Des raisons personnelles	13	12%
Des raisons économiques	8	7%
Des raisons institutionnelles	3	3%

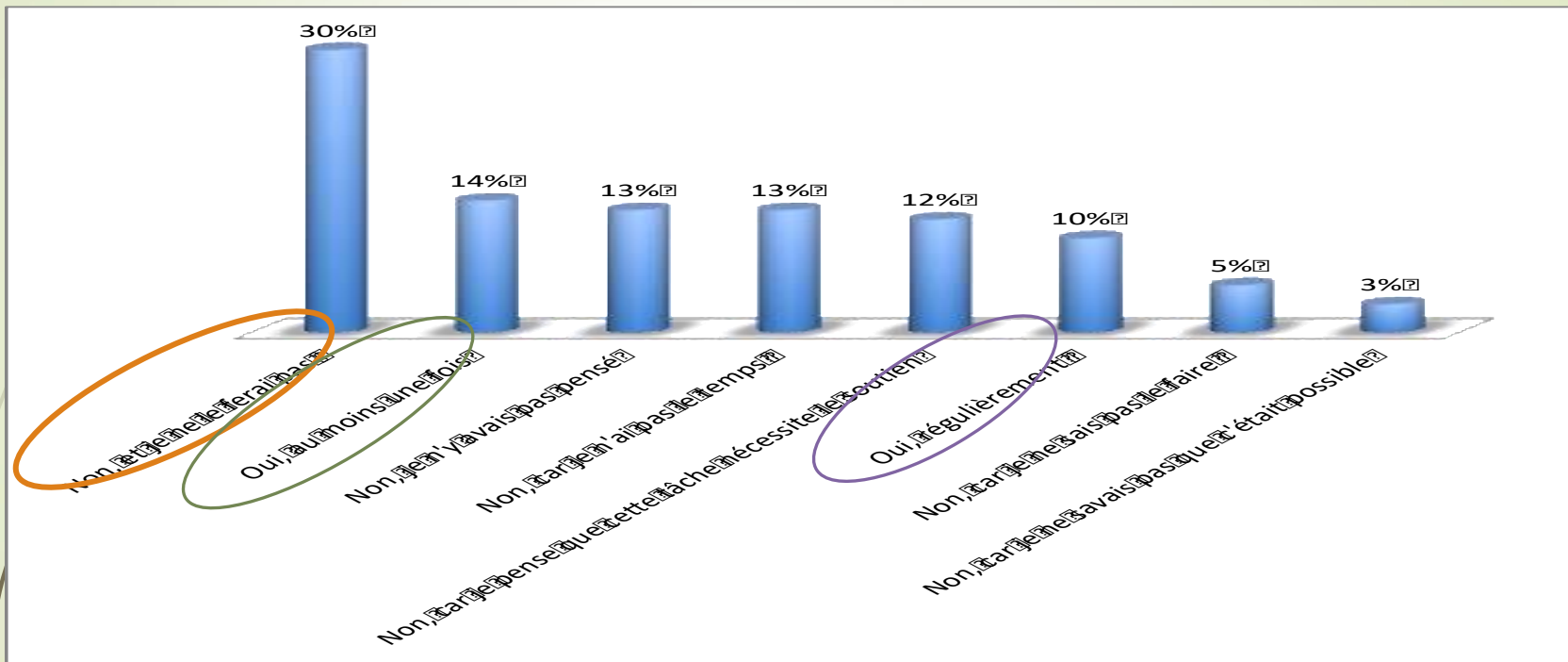
PARTAGE DES DONNÉES : LES REPRÉSENTATIONS

29

**Facteurs
d'incitation au
libre accès de
ses données de
recherche
(N = 107)**

	Nb de réponses	%
Une meilleure visibilité de vos travaux de recherche	70	65%
L'occasion de nouveaux contacts, de nouvelles coopérations scientifiques	58	54%
Le souhait de voir se développer d'autres recherches à partir de vos jeux de données	43	40%
L'adhésion aux valeurs du Libre Accès aux résultats de la recherche scientifique	43	40%
Une obligation faite par le financeur de votre recherche	35	33%
La reconnaissance de la communauté scientifique	34	32%
Le souci de pouvoir faire valider vos résultats à partir des jeux de données	12	11%
De possible retombées financières	8	7%

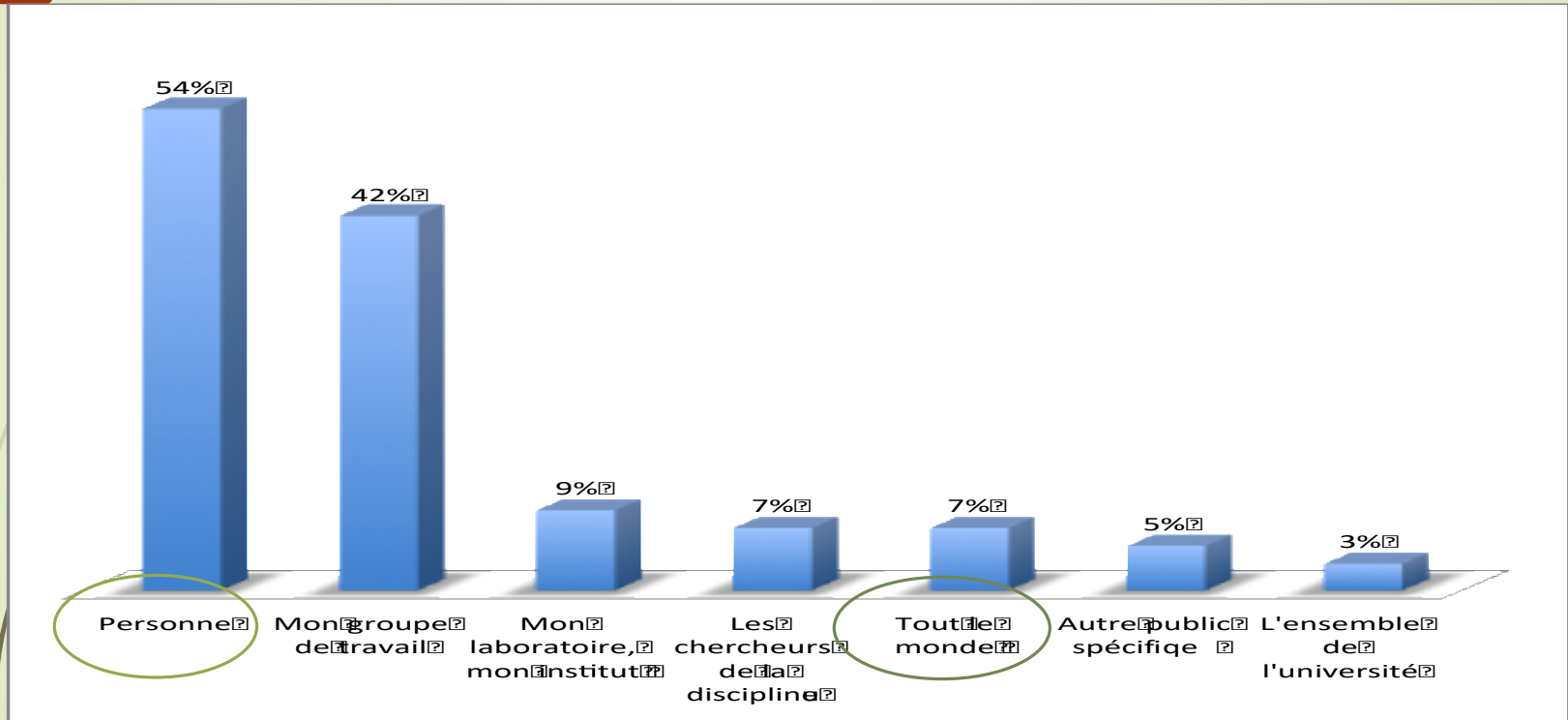
PARTAGE DES DONNÉES : LES PRATIQUES DU LIBRE ACCÈS



« Avez-vous déjà stocké vos données en libre accès sur le web ? » N = 116

PARTAGE DES DONNÉES : LA MISE EN ACCÈS

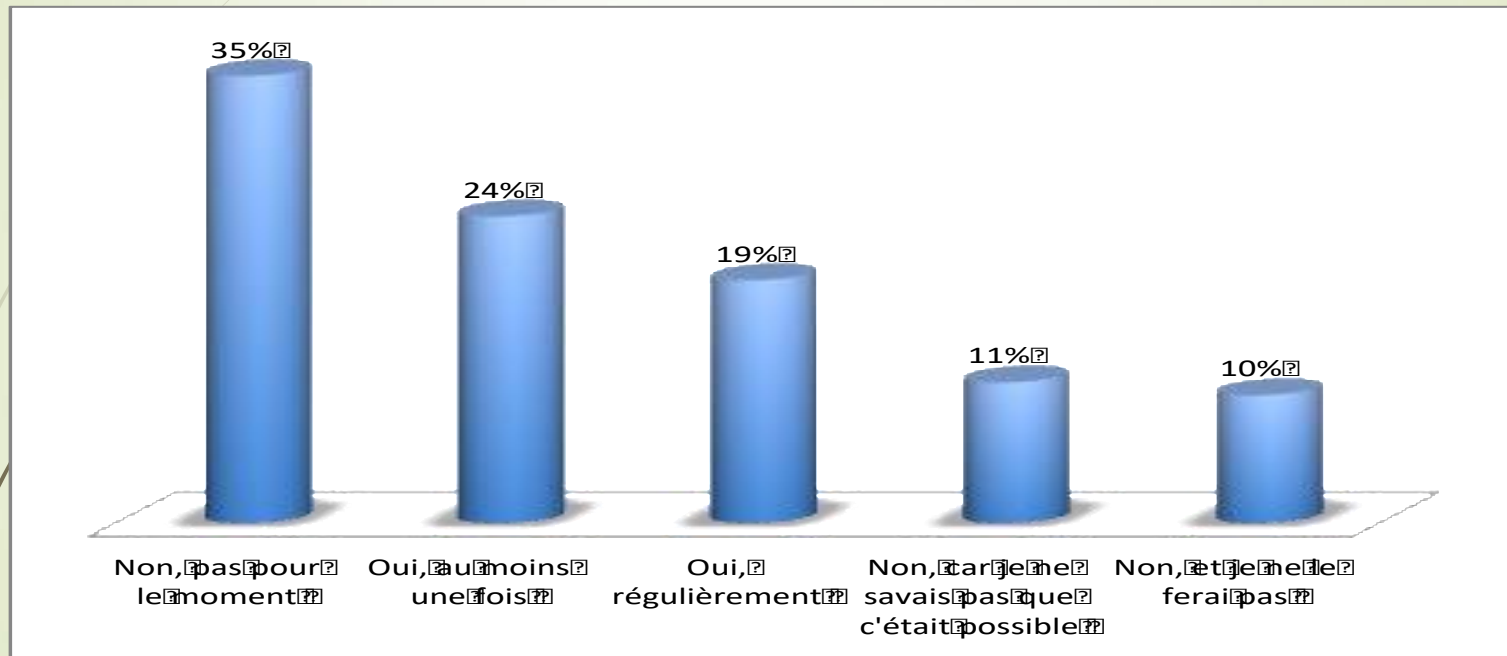
31



« Qui peut accéder à vos données ? » N = 121

PARTAGE DES DONNÉES : LES PRATIQUES DE COLLECTE

32



« Avez-vous déjà téléchargé les données d'autres chercheurs ? »

N = 121

PARTAGE DES DONNÉES : LES SUPPORTS DE DIFFUSION

33

	Nombre de réponses	Pourcentage
Une archive ouverte (de type HAL)	43	45%
Une plateforme spécialisée dans l'archivage et la diffusion des données de recherche	39	41%
Un réseau social académique	26	27%
Le site ou le blog du laboratoire	25	26%
Votre blog ou votre site personnel	20	21%
Un serveur de l'université	19	20%
Un data paper	8	8%

Si vous souhaitez diffuser vos données en libre accès, quel est (ou serait) votre support principal ? N = 96

PARTAGE DES DONNÉES ET INTÉGRITÉ SCIENTIFIQUE

- ▶ **Deux situations symétriques :**
 - ▶ **D'un côté :**
 - ▶ Faiblesses méthodologiques, PRC (Pratiques Contestables de Recherche, fraude, et
 - ▶ Inaccessibilité des données de recherche
 - ▶ **De l'autre côté :**
 - ▶ « bonnes pratiques » de recherche, respect de l'intégrité scientifique, méthodologies robustes... et
 - ▶ Accès et partage des données de recherche

PARTAGE DES DONNÉES ET INTÉGRITÉ SCIENTIFIQUE

- Lien fondamental « SO-IS » (Science Ouverte-Intégrité Scientifique) :
 - *« L'accès aux données brutes des produits de la recherche est essentiel pour plusieurs raisons : la possibilité de reprise et de réutilisation des données et des corpus par d'autres chercheurs du secteur public pour d'autres fins ; le fondement de la preuve en cas de revendication de propriété intellectuelle et de prise de brevet ; l'établissement de la preuve de la fraude ou du manquement à l'intégrité de la recherche ; la possibilité de déposer des données négatives qui sont souvent difficiles à publier et parfois plus importantes que les données confirmatoires. » (Rapport Corvol, p. 36)*

PARTAGE DES DONNÉES ET INTÉGRITÉ SCIENTIFIQUE

- Lien SO-IS :
 - *« Enfin, la science ouverte constitue un levier pour l'intégrité scientifique et favorise la confiance des citoyens dans la science. Elle constitue un progrès scientifique et un progrès de société. »* (Plan National pour la Science Ouverte, 4 juillet 2018)
- Caractère désormais obligatoire de l'ouverture des données de recherche :
 - Dans les revues (Psychologie)
 - Dans les projets ANR
 - ...
 - En lien avec l'intégrité scientifique

EN CONCLUSION : QUELLES LEÇONS DE L'ENQUÊTE ?

37

- **Confirmation de la complexité des données de recherche en SHS**
- **Transversalité de la question du stockage, de l'archivage et de la sécurisation des données**
- **Écarts entre pratiques et représentations, notamment sur la question du partage des données**

EN CONCLUSION : QUELLES RECOMMANDATIONS ?

- **Partir de la complexité de la question des données**
- **Adapter les discours aux préoccupations des chercheurs (pas de "discours d'injonction" sur le partage des données)**
- **Décider du "bon niveau" d'une politique des DR**

EN CONCLUSION : LA QUESTION DES CULTURES DE L'INFORMATION

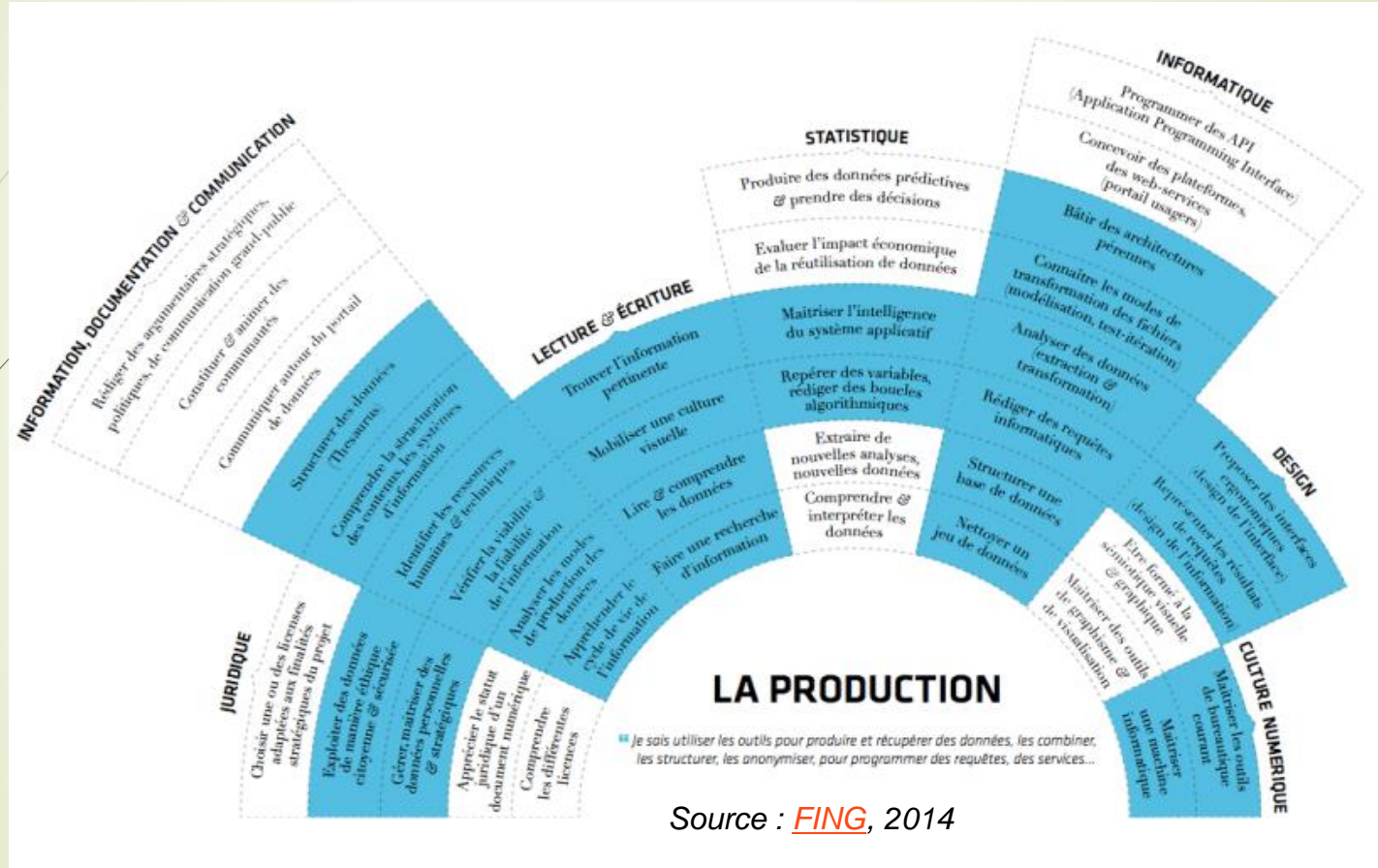
- Confirmation par l'enquête de Rennes 2 de l'insuffisance de cultures de l'information chez les chercheurs
- Deux acceptions de la notion de culture :
 - anthropologique : les pratiques spontanées
 - perfective : les savoirs, la réflexion...
- Comprend à la fois :
 - ensemble de compétences
 - savoirs
 - savoir faire
 - savoir être
 - un regard, une réflexion, un recul...

EN CONCLUSION : LA QUESTION DES CULTURES NÉCESSAIRES

- **Diversité des compétences nécessaires mobilisées pour la gestion et le partage des données :**
 - compétences numériques
 - compétences informatiques
 - compétences informationnelles
 - compétences disciplinaires
 - compétences juridiques
- **Plusieurs cultures de l'information liées aux données :**
 - Culture informationnelle « de base » :
 - Transversale, non spécifique aux chercheurs.
 - Culture de l'IST : englobe toutes les dimensions de l'IST : production, traitement, publication, diffusion... de l'IST
 - spécificité de la culture de l'Open Access

EN CONCLUSION : LA CULTURE DES DONNÉES

41



Source : [FING](#), 2014

Références

- Corvol, Pierre. « Bilan et propositions de mise en œuvre de la charte nationale d'intégrité scientifique ». Paris: Secrétariat d'Etat chargé de l'Enseignement supérieur et de la Recherche, juin 2016. http://cache.media.enseignementsup-recherche.gouv.fr/file/Actus/84/2/Rapport_Corvol_29-06-2016_601842.pdf.
- Leleu-Merviel, Sylvie. « Le sens aux interstices, émergence de reliances complexes ». In *Colloque international francophone "Complexité 2010"*, http://www.trigone.univ-lille1.fr/complexite2010/programme/B2.4-T_Leleu.pdf. Lille, France, 2010. <https://hal.archives-ouvertes.fr/hal-00526508>.
- Ministère de l'Enseignement Supérieur, de la Recherche et de l'Innovation. « Plan national pour la science ouverte », juillet 2018. <http://www.bibliothequescientifiquenumerique.fr/plan-national-pour-la-science-ouverte/>.
- Prost, Hélène, et Joachim Schöpfel. « Les données de la recherche en SHS. Une enquête à l'Université de Lille 3. » Report. Lille 3, 2014. <https://hal.univ-lille3.fr/hal-01198379/document>.

POUR ACCÉDER À L'ENQUÊTE

► Rapport déposé sur Hal

<https://hal.archives-ouvertes.fr/hal-01635186>

- rapport de synthèse
- annexes (avec tous les tris à plat de l'enquête statistique)
- données brutes des croisements statistiques

► Visualisation des données sur [Tableau public](#)

Merci de votre attention !

Contact : alexandre.serres@univ-rennes2.fr