

FORMATION

CONFÉRENCES-ATELIERS | 2019-2020

Publications et données (en) SHS sous l'angle de l'éthique et du droit

18 oct 2019 | 9h30-12h30

Éthique et traitement de données web

Camille Alloing (IAE Poitiers), Mariannig Le Béhec (Univ. Lyon 1)

17 janv 2020 | 14h-17h

Réseaux sociaux numériques de la recherche : entre visibilité et pratiques éthiques

Aline Bouchard (Urfist Paris), Mariannig Le Behec (Univ. Lyon 1)

4 mai 2020 | 14h-17h > *reporté (date à préciser)*

Enquêtes enregistrées sur le terrain : collecter, archiver et diffuser des données avec le droit et l'éthique comme alliés

Véronique Ginouvès (CNRS, MMSH)

9 juill 2020 | 14h30-17h > *en visioconférence*

L'éthique des données de la recherche en sciences humaines et sociales

Bernard Jacquemin et Joachim Schöpfel (Univ. Lille, GERiCO)

■
Lieu : MSH Lyon St-Étienne 14 av. Berthelot, Lyon 7^e
(salle Elise Rivet, Ennat Léger) ou en visioconférence

■
Infos et inscription : [www.msh-lse.fr/evenements/
publications-donnees-en-shs-sous-angle-ethique-droit](http://www.msh-lse.fr/evenements/publications-donnees-en-shs-sous-angle-ethique-droit)

L'ÉTHIQUE DES DONNÉES DE LA RECHERCHE EN SHS

9 juillet 2020

Cycle de formations co-organisées par la
MSH Lyon Saint Etienne et l'Urfist de Lyon
2019/2020

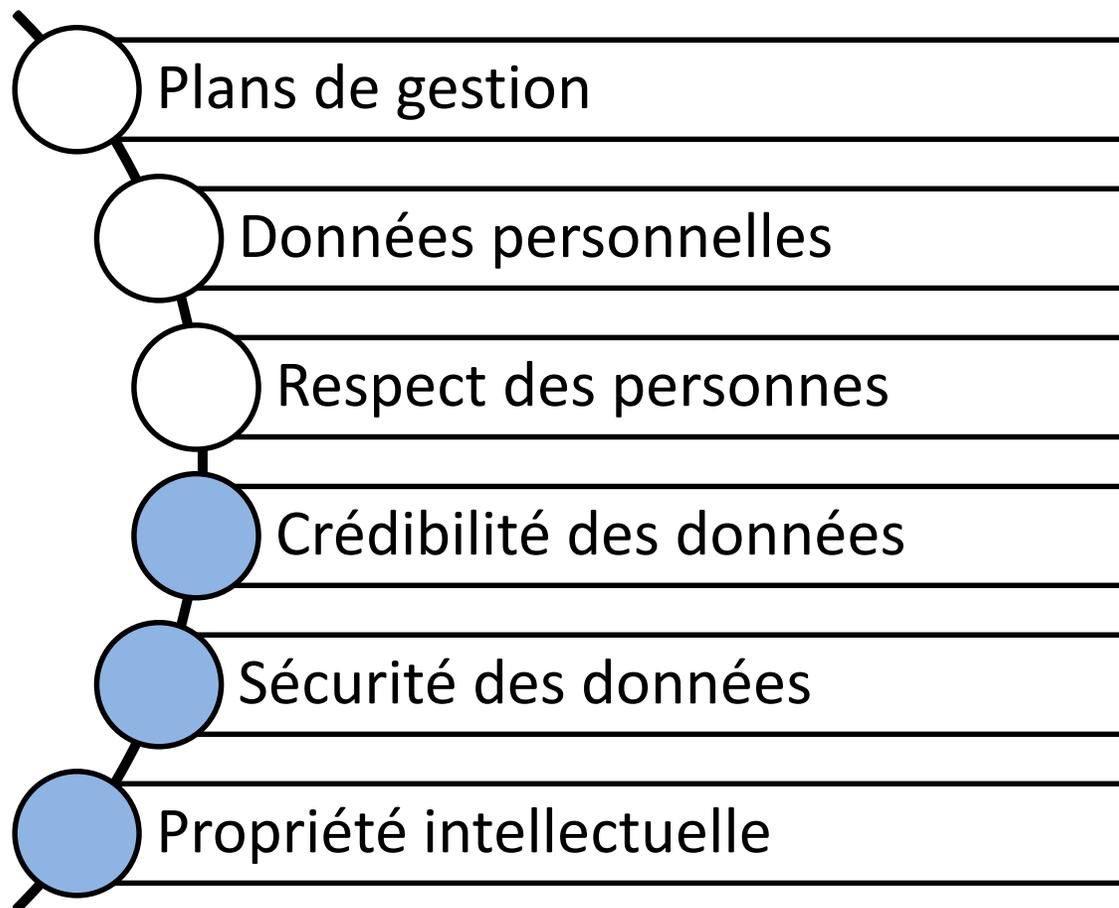
Ethique des données de la recherche en SHS

Joachim Schöpfel



- *Groupe d'Études et de Recherche Interdisciplinaire en Information et Communication*
- Equipe d'accueil de l'Université de Lille (EA 4073)
- 37 chercheurs de plusieurs départements
- 28 doctorants
- <https://geriico-recherche.univ-lille3.fr/>
- Centré sur la question des médiations des connaissances, des savoirs et des cultures dans la société contemporaine
 - Axe 3 : Innovation par l'usage et dispositifs numériques
 - Axe 4 : Circulation de l'information et organisation des connaissances
- <https://hal.archives-ouvertes.fr/GERIICO>

Six facettes de la dimension éthique



Au préalable

- Ethique à plusieurs niveaux
 - général
 - domaine / discipline
 - métier
- Dynamique
 - ce qui était accepté hier, ne l'est plus aujourd'hui
 - ce qui n'était pas possible hier, l'est maintenant
- Ethique : une contrainte ou un obstacle ?
 - Mais : qui veut savoir quoi, pourquoi ? Qui veut les données, pour quel usage ? Légitimité ? Est-ce bien ?

4^e facette

CRÉDIBILITÉ DES DONNÉES

Crédibilité des données

- *Les bibliothécaires et les autres professionnel(le)s de l'information recherchent l'excellence professionnelle (et) visent la qualité de service la plus élevée (Code d'éthique de l'IFLA).*
- *Les chercheurs et les personnels du monde de la recherche doivent être formés aux dimensions éthiques de la gestion des données, en particulier au respect (...) de la qualité et de l'intégrité des données (Comité d'éthique du CNRS, avis 2015-30).*

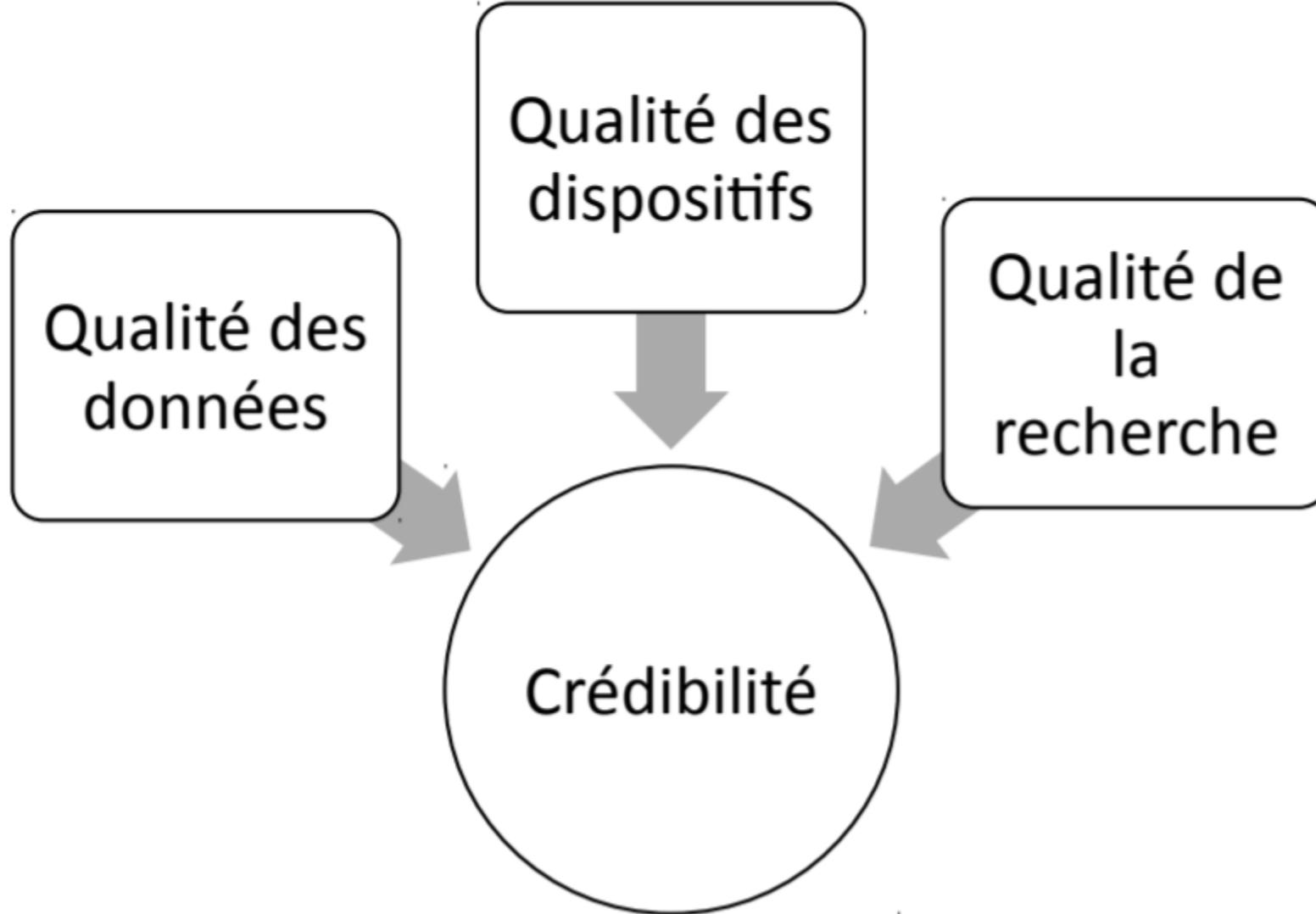
La qualité est cruciale...

- La qualité des données est cruciale
 - Pour la validation des résultats scientifiques (intégrité etc.)
 - Pour la réutilisation des résultats (réanalyse etc.)
- Mais : comment garantir la qualité ?
- Réponse : (aussi) en assurant la crédibilité du dispositif
 - Dispositif = contenu, système, fonctionnalités, pratiques...
- Pas (uniquement) une question de compétence et de technologie mais (aussi) de valeur

Le terme de crédibilité

- Crédibilité = digne de confiance (*trustworthiness*)
- Dispositif & contenu (données, métadonnées)
- Pour une plateforme de données (entrepôt), la crédibilité est importante
 - En amont, pour déposer
 - En aval, pour réutiliser
- Plateforme biface : double enjeu
- Pour développer les infrastructures de données et pour inciter les chercheurs à stocker et publier leurs données via des entrepôts de données, il faut rassurer sur la qualité, intégrité, traçabilité etc. des données et de leur description mais surtout, sur la pérennité, la cohérence, l'accessibilité et l'interopérabilité des entrepôts.

Trois variables de la crédibilité d'un service de données



Qualité des données

- L'éthique des métiers de l'information souligne l'engagement de fournir une information de qualité, sans censure, sans discrimination, sans biais.
- Dans le cadre de la gestion des données, ce terme recouvre plusieurs significations, aussi bien de la véracité (pertinence) de l'information que de la cohérence entre plusieurs sources (bases de données, etc.), la conformité avec l'usage (utilité, finalité), l'absence de problème d'accès, etc.
- Du point de vue informatique, il s'agit d'une question technique (*cleansing* etc.).
- Mais du point de vue scientifique, la question technique se transforme en problème éthique : comment éviter d'exploiter et surtout de publier des données caduques, erronées, falsifiées, non pertinentes, inexploitable ?

Qualité des dispositifs (1)

- En 2019, l'association Science Europe a publié un « Guide pratique pour une harmonisation internationale de la gestion des données de recherche ». Science Europe y présente une liste de critères minimaux permettant de savoir comment sélectionner un entrepôt de confiance.
- Ces critères sont organisés autour de quatre thèmes majeurs, que tous les entrepôts de confiance devraient satisfaire :
 - Attribution d'identifiants uniques et pérennes ;
 - Métadonnées ;
 - Accès aux données et licences d'utilisation ;
 - Préservation.
- Pour chaque thème, Science Europe décrit des recommandations plus ou moins précises, comme l'utilisation de métadonnées repérables et communautaires (standards) ou la garantie de l'intégrité et de l'authenticité des données (Science Europe 2019).

Qualité des dispositifs (2)

En 2020, un groupe de travail a proposé un catalogue de cinq principes TRUST qui rendent les entrepôts de données de recherche “trustworthy”, crédibles et dignes de confiance (Lin et al. 2020).

T	Transparency	Transparence (conditions d'utilisation, durée de conservation, promesse de service...)
R	Responsibility	Responsabilité (normes, standards, droits...)
U	User focus	Centré sur l'utilisateur (pratiques communautaires, processus de recherche, cycle de vie des données...)
S	Sustainability	Durabilité, pérennité (gestion des risques, modèle d'affaires)
T	Technology	Technologie, matériel, logiciel, services

Qualité des dispositifs (3)

- La réalisation d'audits réguliers voire d'une certification est un moyen pour créer de la crédibilité.
- Actuellement, surtout un certificat avec 16 critères : CoreTrustSeal (CTS 2019)

1	Mission/Scope	L'entrepôt a une mission explicite de diffusion et de conservation de données.	T, R
2	Licenses	L'entrepôt diffuse les données avec des licences appropriées.	T
3	Continuity of access	L'entrepôt est en mesure d'assurer une accessibilité continue et la conservation pérenne des données.	S
4	Confidentiality/Ethics	L'entrepôt garantit, dans la mesure du possible, une production et gestion de données en conformité avec des standards disciplinaires et éthiques.	R, U
5	Organizational infrastructure	L'entrepôt dispose de ressources humaines et financières suffisantes, avec une gouvernance transparente, pour remplir sa mission avec efficacité.	T, S
6	Expert Guidance	L'entrepôt a mis en place un conseil scientifique ou une autre forme d'orientation et de conseil d'expert.	R, U
7	Data integrity and authenticity	L'entrepôt garantit l'intégrité et l'authenticité des données.	R
8	Appraisal	L'entrepôt met en oeuvre une politique explicite d'évaluation et de sélection des données à intégrer et de leurs métadonnées.	T, R
9	Documented storage procedures	L'entrepôt applique des procédures documentées pour le stockage des données et pour leur conservation à long terme.	T, S
10	Preservation plan	L'entrepôt assume la responsabilité pour la conservation pérenne des données, et il en a les capacités techniques.	R, S
11	Data quality	L'entrepôt mène une politique de qualité au niveau des données et métadonnées (contrôle, nettoyage...).	R
12	Workflows	L'entrepôt met en oeuvre des flux de production (ingestion, gestion, diffusion, conservation...) transparents, documentés et efficaces (standard OAIS).	T, Tech
13	Data discovery and identification	L'entrepôt propose des métadonnées riches et normalisées (y compris des identifiants pérennes) indispensables au référencement, à l'identification et à la découverte des données.	R
14	Data reuse	L'entrepôt fait le nécessaire pour faciliter la réutilisation à plus long terme (formats, métadonnées).	S
15	Technical infrastructure	L'entrepôt fonctionne sur des infrastructures de base ("coeur) fiables et stables, appropriées par rapport aux pratiques et attentes de la communauté cible.	R, S, Tech
16	Security	L'infrastructure technique de l'entrepôt assure la protection de la structure et de ses données, produits, services et utilisateurs.	S, Tech

Qualité des dispositifs (4)

- Rendre les entrepôts dignes de confiance est une démarche qui vise avant tout les utilisateurs humains.
- Mais il n'y a pas que l'utilisation humaine.
- Pour développer l'interopérabilité des dispositifs de données de recherche, un comité international a rédigé les FAIR Guiding Principles (Wilkinson et al. 2016).

Findability	F1	Les (méta)données se voient attribuer un identifiant unique et persistant au niveau mondial.
	F2	Les données sont décrites avec des métadonnées riches.
	F3	Les métadonnées comprennent clairement et explicitement l'identifiant des données qu'elles décrivent.
	F4	Les (méta)données sont enregistrées ou indexées dans une ressource consultable.
Accessibility	A1	Les (méta)données peuvent être récupérées par leur identifiant en utilisant un protocole de communication standardisé.
	A2	Les métadonnées sont accessibles, même lorsque les données ne sont plus disponibles.
Interoperability	I1	Les (méta)données utilisent un langage formel, accessible, partagé et largement applicable pour la représentation des connaissances.
	I2	Les (méta)données utilisent des vocabulaires qui suivent les principes FAIR.
	I3	Les (méta)données contiennent des références qualifiées à d'autres (méta)données.
Reusability	R1	Les méta(données) sont richement décrites avec une pluralité d'attributs précis et pertinents.

Qualité de la recherche

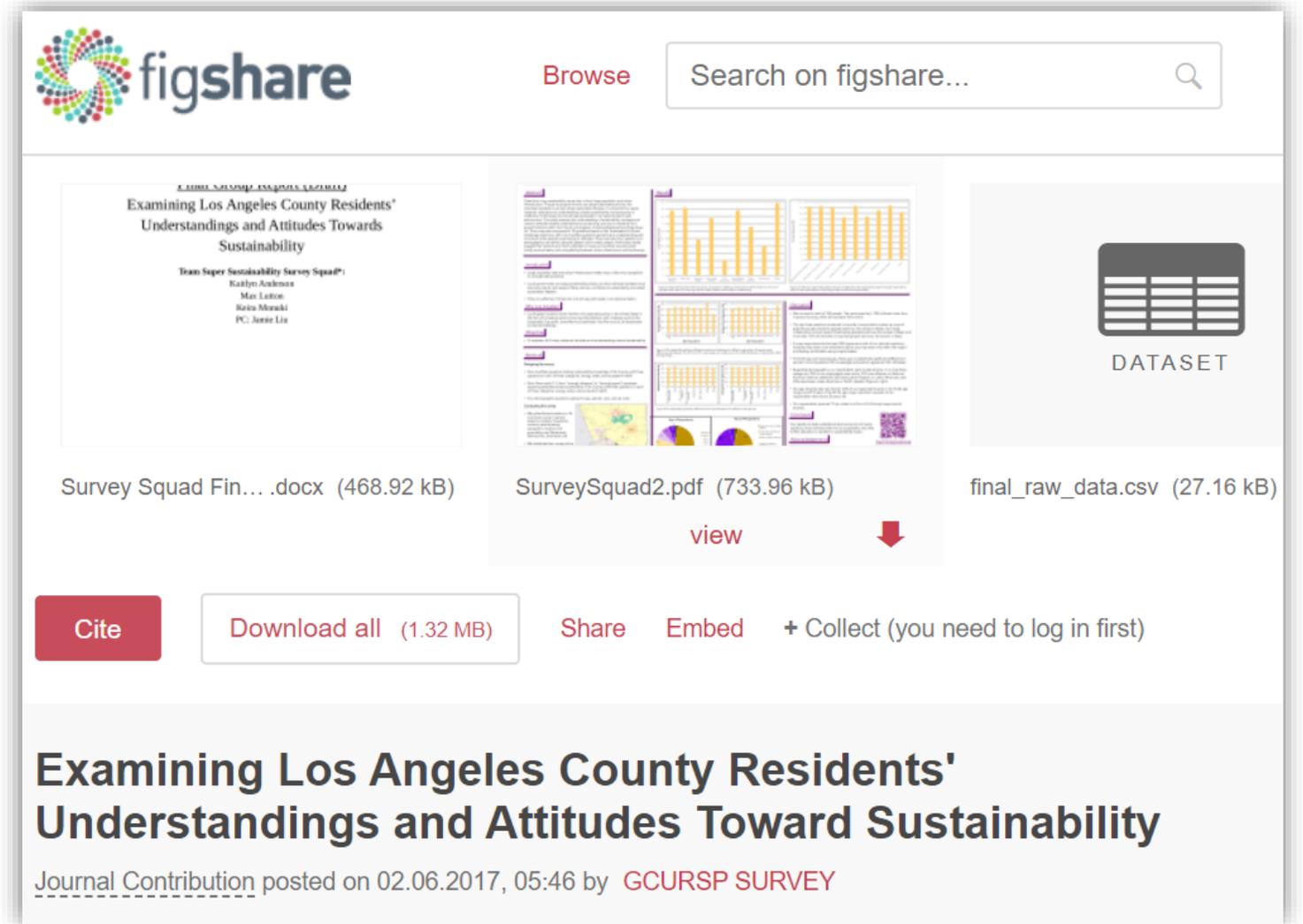
- Cf. recommandations COMETS du CNRS
 - Institution & équipe/individu
- La publication des données de la recherche est censée faciliter la reproductibilité de la recherche, d'augmenter la transparence du travail scientifique, de prévenir la falsification des données et de contribuer ainsi à l'amélioration des résultats et des connaissances.
- Regardé de près, ce lien entre la diffusion des données et l'intégrité et la responsabilité du chercheur est double :
 - d'une part, l'ouverture des données renforce l'intégrité et la valeur du travail scientifique ;
 - d'autre part, la responsabilité et l'intégrité du chercheur (de l'équipe ou de la structure) contribuent à rendre les données déposées et diffusées crédibles et potentiellement réutilisables.

Crédibilité des données

Quelle plateforme choisir pour déposer et partager des résultats d'une enquête ?

Réutiliser les données d'un sondage trouvées sur une plateforme ?

Comment garantir des données non falsifiées (responsabilité de la plateforme) ?



The screenshot shows the Figshare interface for a dataset. At the top, there is the Figshare logo, a 'Browse' button, and a search bar containing 'Search on figshare...'. Below the header, three items are displayed in a grid:

- On the left, a document titled 'Examining Los Angeles County Residents' Understandings and Attitudes Towards Sustainability' by the 'Team Super Sustainability Survey Squad'. Below it is the file 'Survey Squad Fin... .docx (468.92 kB)'.
- In the center, a PDF preview of 'SurveySquad2.pdf (733.96 kB)' showing various charts and text. Below the preview is a 'view' button and a red downward arrow.
- On the right, a 'DATASET' icon with the file 'final_raw_data.csv (27.16 kB)'.

Below the grid, there are several action buttons: a red 'Cite' button, a 'Download all (1.32 MB)' button, and buttons for 'Share', 'Embed', and '+ Collect (you need to log in first)'. At the bottom of the page, the title 'Examining Los Angeles County Residents' Understandings and Attitudes Toward Sustainability' is repeated in large bold text, followed by the text 'Journal Contribution posted on 02.06.2017, 05:46 by GCURSP SURVEY'.

5^e facette

SÉCURITÉ DES DONNÉES

Sécurité des données

- Comme pour tout dispositif de stockage et de conservation de données en ligne, la sécurisation des données est un enjeu de première importance.
- D'un point de vue technique, il s'agit à la fois de veiller à l'intégrité de la conservation des données, et de sécuriser leur accès en lecture et en écriture.
- La valeur des données conservées réside en effet dans leur intégrité, et il n'est pas acceptable qu'elles soient altérées, de quelque manière que ce soit.
- La perte, la modification, l'adjonction incontrôlée de toute valeur est susceptible non seulement de fausser une action de recherche menée à partir d'un jeu de données corrompues, mais également d'invalider une publication déjà diffusée si l'éditeur demande à avoir accès aux données qui la portent, comme cela commence à se produire.

Sécurité ?

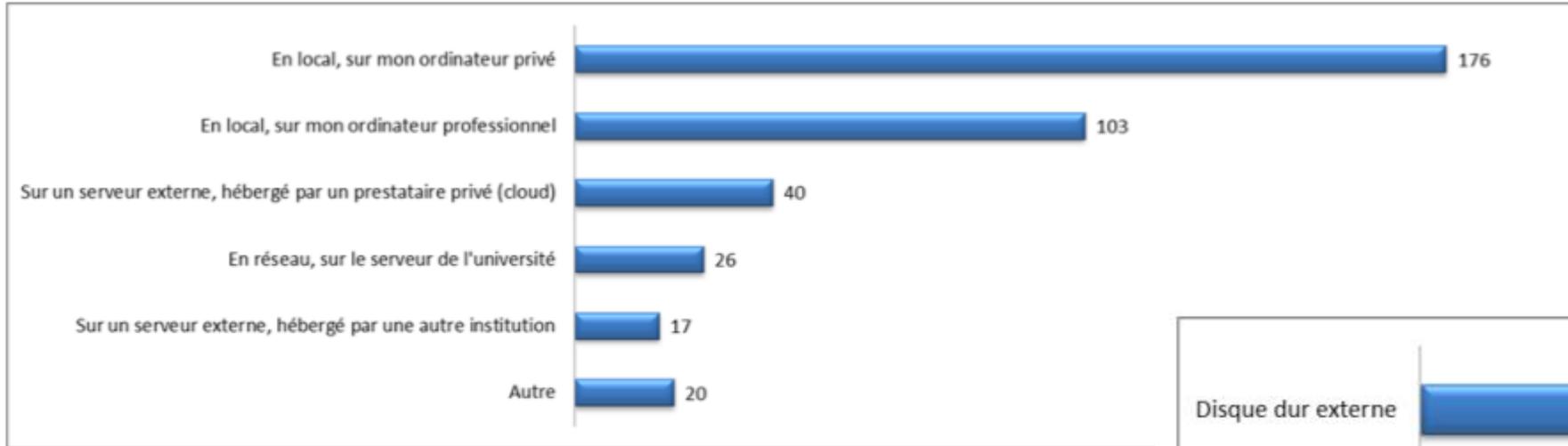


Figure 6 : Endroits de stockage (N=212)

Source : Hélène Prost, Joachim Schöpfel. *Les données de la recherche en SHS. Une enquête à l'Université de Lille 3.* : Rapport final. [Rapport de recherche] Lille 3. 2015. [hal-01198379](https://hal.archives-ouvertes.fr/hal-01198379)

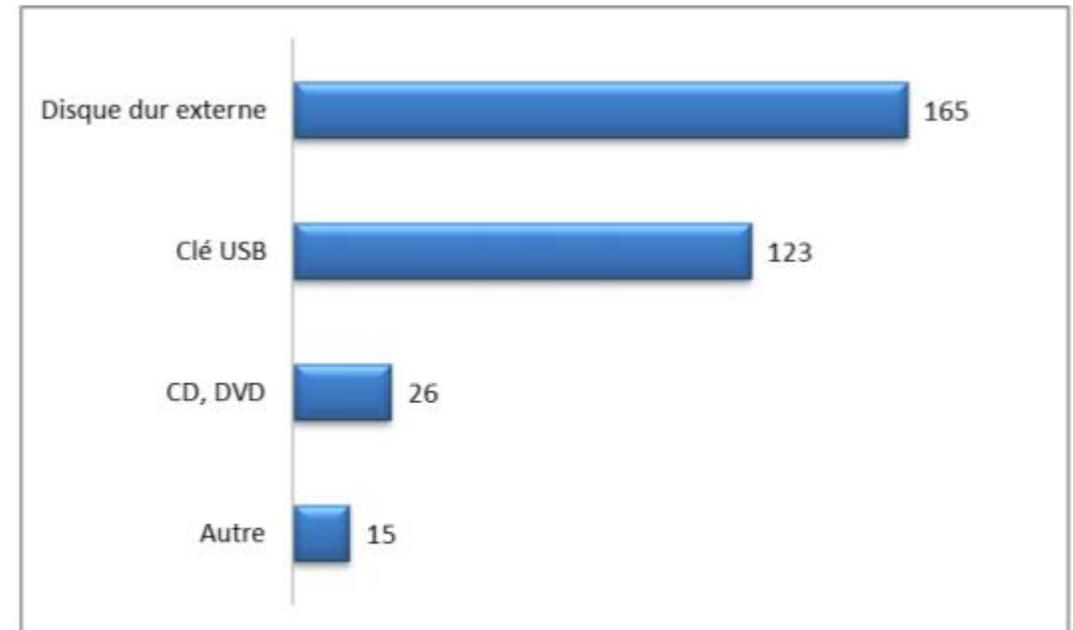


Figure 7 : D'autres supports de stockage (N=204)

Sécuriser

TRUST

R	Responsibility
S	Sustainability
T	Technology

CoreTrustSeal

4	Confidentiality/Ethics	L'entrepôt garantit, dans la mesure du possible, une production et gestion de données en conformité avec des standards disciplinaires et éthiques.
7	Data integrity and authenticity	L'entrepôt garantit l'intégrité et l'authenticité des données.
12	Workflows	L'entrepôt met en oeuvre des flux de production (ingestion, gestion, diffusion, conservation...) transparents, documentés et efficaces (standard OAIS).
15	Technical infrastructure	L'entrepôt fonctionne sur des infrastructures de base ("coeur) fiables et stables, appropriées par rapport aux pratiques et attentes de la communauté cible.
16	Security	L'infrastructure technique de l'entrepôt assure la protection de la structure et de ses données, produits, services et utilisateurs.

Sécuriser (suite)

- Sécuriser les accès
- Différents niveaux d'accessibilité
 - Authentification
 - Contrôle des accès et de la réutilisation
- Sécuriser les données d'accès
 - Fichiers log

Sécurité des données

Comment stocker les transcriptions d'entretiens avec des délinquants mineurs ?

Comment faciliter l'accès (et l'usage) des données ?



6^e facette

PROPRIÉTÉ INTELLECTUELLE

Propriété intellectuelle

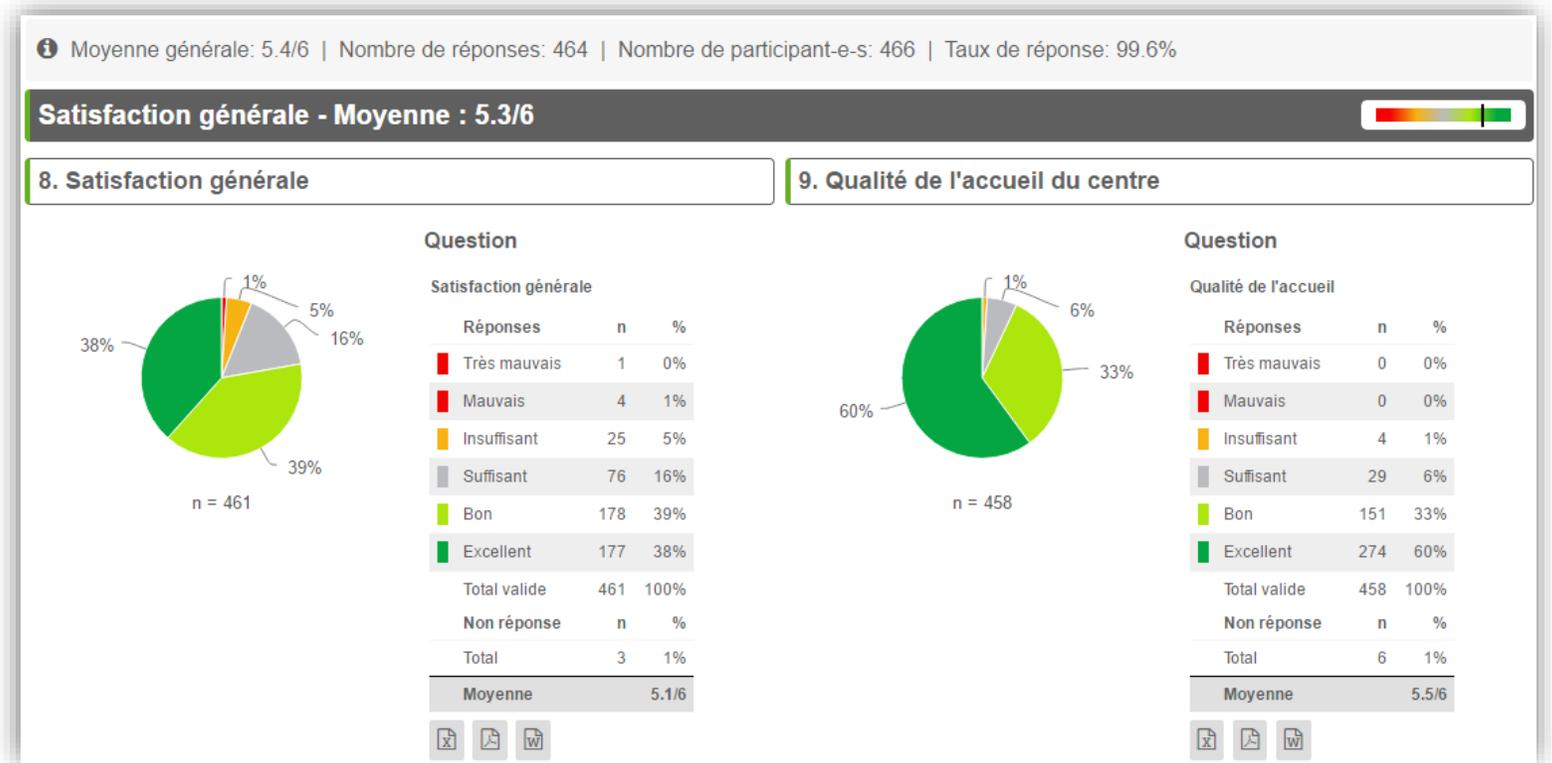
- *Les bibliothécaires et les autres professionnel(le)s de l'information ont pour but de fournir un accès équitable, rapide, économique et efficace à l'information (...)*
- *Les bibliothécaires et les autres professionnel(le)s de l'information sont les partenaires des auteurs, des éditeurs et de tous les créateurs d'œuvres protégées. Les bibliothécaires et les autres professionnel(le)s de l'information reconnaissent le droit à la propriété intellectuelle des auteurs et des créateurs et cherchent à s'assurer que ces droits sont respectés (Code d'éthique de l'IFLA).*

Trois aspects

- La préservation à la fois de l'intégrité des données et de leur lien avec les responsables de leur collecte, dans un univers numérique dans lequel chacun semble faire son marché librement et sans respect pour la propriété.
- La distinction entre propriété intellectuelle et propriété industrielle, souvent mal comprise ou interprétée, où l'éthique et la légalité ne semblent pas toujours aller de concert.
- Une dissolution du lien de paternité des données lorsque la recherche est financée sur fonds publics.

A qui appartiennent les données ?

- Employeur ?
- Chercheur ?
- Domaine public ?



Une question spécifique SHS

- La particularité des SHS est la construction et **l'utilisation de données complexes** (« agrégations spécialisées ») comme des éditions scientifiques, des corpus de textes, des textes structurés, des collections scientifiques thématiques, des données avec analyses et/ou annotations, ou encore des outils de découverte (bibliographies...).
- Une autre particularité est l'importance de **la dimension interprétative** pour l'encodage des données (dans une édition critique numérique, ce qui est retenu comme faisant partie du texte qu'il s'agit d'établir à partir de divers manuscrits par exemple), pour les métadonnées descriptives (qui dérivent des principes théoriques sur ce qu'est une édition ou des principes éditoriaux), pour les annotations, etc.
- Une autre particularité est l'importance de **la contextualisation des données** en question, c'est-à-dire la connaissance précise de leur acquisition, production ou construction, la contribution et responsabilité intellectuelle des personnes à l'origine des données, l'historique des données (éditions, versions...) etc.

Schöpfel, J. (2020). À propos des données de recherche en SHS. *Lettres de l'INSHS*, (63), 23–25.

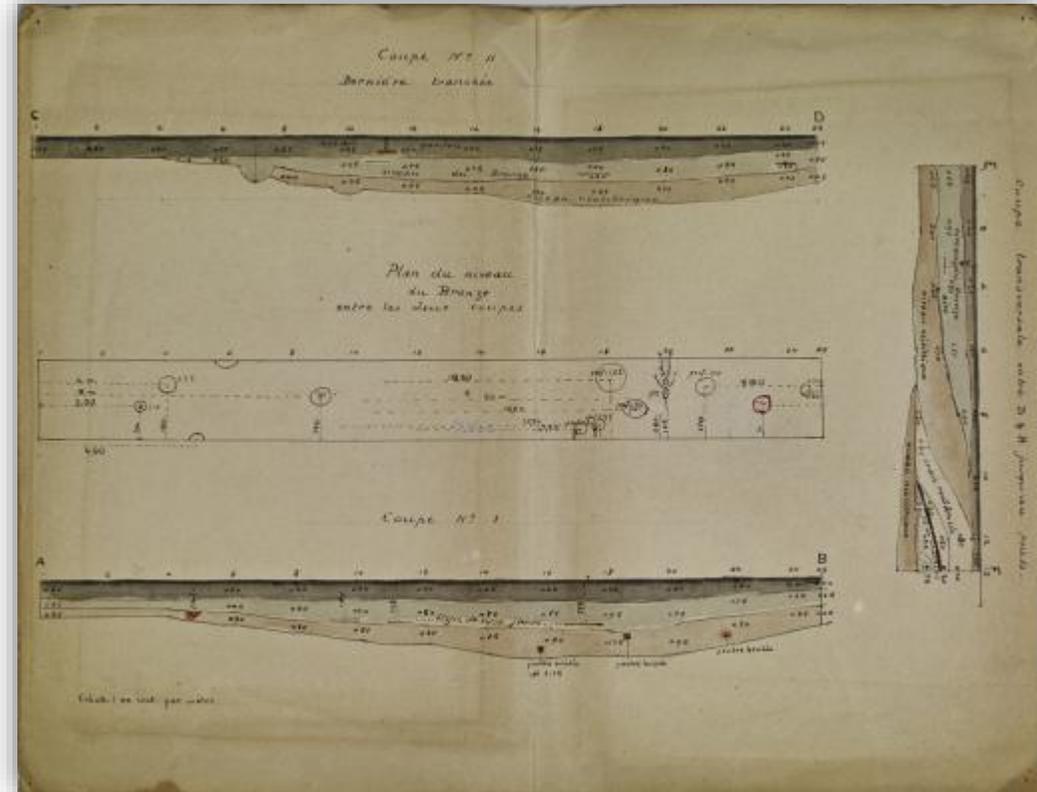
https://www.inshs.cnrs.fr/sites/institut_inshs/files/download-file/lettre_infoINSHS_63v3_0.pdf

Propriété intellectuelle

Comment diffuser les plans de fouille établis par un IGE ?

Que faire des données d'un(e) doctorant(e) ?

Comment (et pourquoi) dire à un(e) doctorant(e) que ses données ne lui appartiennent pas et qu'il (elle) doit les rendre accessibles et exploitables par n'importe qui et n'importe comment et pour n'importe quelle finalité, sans pouvoir en décider ?



Références

- CTS (2019). *CoreTrustSeal Trustworthy Data Repositories Requirements 2020–2022* (Version v02.00-2020-2022). CoreTrustSeal Standards and Certification Board. <https://doi.org/10.5281/zenodo.3638211>
- Lin, D., Crabtree, J., Dillo, I., Downs, R. R., Edmunds, R., Giaretta, D., ... Westbrook, J. (2020). The TRUST Principles for digital repositories. *Scientific Data*, 7(1), 144. <https://doi.org/10.1038/s41597-020-0486-7>
- Science Europe (2019). *Practical Guide to the International Alignment of Research Data Management*. Bruxelles : Science Europe Working Group on Research Data https://www.scienceeurope.org/media/jezkhnoo/se_rdm_practical_guide_final.pdf
- Wilkinson, M. D. et al. (2016). The FAIR Guiding Principles for scientific data management and stewardship. *Scientific Data*, 3(1), 160018. <https://doi.org/10.1038/sdata.2016.18>

Jacquemin, B., Schöpfel, J., Chaudiron, S., & Kergosien, É. (2018). L'éthique des données de la recherche en sciences humaines et sociales. In L. Balicco et al. (coord.), *L'éthique en contexte info-communicationnel numérique*. Bruxelles : De Boeck (pp. 71–86).

<https://hal.archives-ouvertes.fr/hal-01958472>

Contact

joachim.schopfel@univ-lille.fr