

Les données qualitatives, que peut-on en FAIR? : Les recherches qualitatives au défi de l'open data

Autor(en): **Amiotte-Suchet, Laurent**

Objektyp: **Article**

Zeitschrift: **Traverse : Zeitschrift für Geschichte = Revue d'histoire**

Band (Jahr): **29 (2022)**

Heft 1: **Publizieren in den Geisteswissenschaften : Vergangenheit, Gegenwart und Zukunft = Publier en sciences humaines : passé, présent et avenir**

PDF erstellt am: **17.07.2024**

Persistenter Link: <https://doi.org/10.5169/seals-981260>

Nutzungsbedingungen

Die ETH-Bibliothek ist Anbieterin der digitalisierten Zeitschriften. Sie besitzt keine Urheberrechte an den Inhalten der Zeitschriften. Die Rechte liegen in der Regel bei den Herausgebern.

Die auf der Plattform e-periodica veröffentlichten Dokumente stehen für nicht-kommerzielle Zwecke in Lehre und Forschung sowie für die private Nutzung frei zur Verfügung. Einzelne Dateien oder Ausdrucke aus diesem Angebot können zusammen mit diesen Nutzungsbedingungen und den korrekten Herkunftsbezeichnungen weitergegeben werden.

Das Veröffentlichen von Bildern in Print- und Online-Publikationen ist nur mit vorheriger Genehmigung der Rechteinhaber erlaubt. Die systematische Speicherung von Teilen des elektronischen Angebots auf anderen Servern bedarf ebenfalls des schriftlichen Einverständnisses der Rechteinhaber.

Haftungsausschluss

Alle Angaben erfolgen ohne Gewähr für Vollständigkeit oder Richtigkeit. Es wird keine Haftung übernommen für Schäden durch die Verwendung von Informationen aus diesem Online-Angebot oder durch das Fehlen von Informationen. Dies gilt auch für Inhalte Dritter, die über dieses Angebot zugänglich sind.

Les données qualitatives, que peut-on en FAIR?

Les recherches qualitatives au défi de l'*open data*

Laurent Amiotte-Suchet

Le terme «Open Science» désigne un vaste mouvement d'ouverture des milieux scientifiques, qui a pris sa source en 2001 avec l'initiative de Budapest (*Budapest Open Access Initiative*)¹ et en 2003 avec la déclaration de Berlin (*Berlin Declaration on Open Access to Knowledge in the Sciences and Humanities*).² Ces deux déclarations militaient alors pour une science ouverte à toutes et, tous, non réservée aux seul·e·s spécialistes et mise au service de l'intérêt commun. Elles sont à l'origine de deux stratégies dès lors formalisées au niveau international, puis déclinées au niveau national:³ l'*open access to publications* et l'*open research data*. L'*open access to publications* désigne le fait de rendre accessible et consultable gratuitement, via internet, l'ensemble des publications scientifiques (articles et ouvrages). Des négociations ont donc été menées avec les maisons d'édition. Moyennant des subventions, ces dernières ont, pour une partie d'entre elles, accepté d'ouvrir l'accès aux publications, selon différents modèles.⁴ La question de l'*open research data* soulève d'autres enjeux. Dans ce mouvement, ce ne sont pas les résultats des analyses que les chercheur·e·s sont encouragé·e·s à mettre à disposition, mais les données de la recherche elles-mêmes, c'est-à-dire la matière première à partir de laquelle les analyses ont été réalisées, les hypothèses formulées et les résultats obtenus. L'*open research data*, largement effectif dans les sciences expérimentales, est rarement pratiqué dans d'autres disciplines (notamment en sciences humaines et sociales). Cette injonction portée par les bailleurs de fonds et les organisations scientifiques internationales⁵ s'inscrit dans la même philosophie de départ que celle de l'*open access*: puisque la recherche scientifique est essentiellement financée par des fonds publics, tout ce qu'elle produit (les données primaires comme les analyses élaborées à partir de ces données) doit être restitué à la société civile qui la finance, sans délai et sans restriction. Et c'est là que la question des spécificités disciplinaires et des approches méthodologiques fait son apparition, mettant la communauté scientifique en garde contre une tendance à la standardisation qui pose des problèmes éthiques et épistémologiques de fond.

Injonction au bien FAIR

Les conférences et les séminaires consacrés à l'*open science* se sont largement, ces dernières années, appuyés sur les principes FAIR, souvent présentés comme une convention solide et acquise au sein des milieux scientifiques. Les formations proposées aux chercheur·e·s ont pour objectif de leur montrer comment préparer et déposer leurs données de recherche dans des *data repositories* certifiés FAIR, c'est-à-dire garantissant que les données de recherche stockées seront faciles à trouver (*findable*), accessibles (*accessible*), interopérables (*interoperable*) et réutilisables (*reusable*). À travers ces principes, l'*open science* cherche à soutenir une vision coopérative de la recherche scientifique. Si des chercheur·e·s publient des résultats à l'issue d'une étude, il est en effet pertinent que d'autres équipes de recherche puissent accéder aux données sur lesquelles se sont fondées leurs conclusions. Elles pourraient alors soit contrôler la fiabilité des résultats en reproduisant l'analyse, soit les approfondir en la poussant plus loin, soit croiser ces données avec celles d'autres recherches pour mener de nouvelles explorations. Dans tous les cas, il serait regrettable, coûteux et chronophage qu'une équipe de recherche doive reproduire un set de données existant parce que ce dernier n'est pas accessible. La mise à disposition des données de la recherche entend donc lutter contre les stratégies contre-productives, liées aux compétitions entre laboratoires et entre États, en particulier lorsque les recherches ont été financées par des fonds publics et poursuivent en théorie un intérêt commun.⁶ Or, pour les recherches privilégiant une approche qualitative, quelle que soit la discipline concernée, la mise à disposition des données primaires de la recherche pose un certain nombre de questions renvoyant à ce que l'on entend par le terme même de «données».

Peut-on donner des données qui ne nous ont pas été données?

S'il est difficile de ne pas promouvoir cette vision démocratique et ouverte de la science, il est essentiel que cette démarche se déploie de manière prudente afin que le niveau d'accessibilité des données puisse être adapté aux situations et ne vienne pas s'imposer arbitrairement et sans distinction à tous les domaines et disciplines de recherche. La formule régulièrement énoncée – «*As open as possible, as closed as necessary*»⁷ – entend en effet rappeler que la fin ne justifie pas les moyens et que, si les résultats de la recherche scientifique devraient idéalement être partagés sans restriction (*open access*), les données de la recherche, elles, sont de nature variée et ne peuvent pas être livrées en pâture à tous les appétits.

compte: le respect de l'anonymat des participant·e·s et le maintien d'un certain niveau de contextualisation des données. Or, ces deux aspects s'avèrent souvent difficilement conciliables si les données primaires de la recherche doivent être mises à la disposition de la communauté scientifique. Dans le cadre de recherches fondées sur des entretiens biographiques, sur des documents personnels ou confidentiels ou sur des observations de terrain, le codage des données⁸ est un exercice complexe et chronophage qui ne garantit que partiellement l'anonymat des participant·e·s, si les données doivent quitter les serveurs sécurisés des institutions pour être mises à la disposition de celles et ceux qui souhaiteraient les consulter et/ou les réexploiter. Les exigences des commissions d'éthique concernant la protection des données dites sensibles (données identitaires, de santé, d'orientation politique...) encadrent d'ailleurs ces pratiques en réclamant l'utilisation de formulaires de consentement permettant aux participant·e·s de donner un accord éclairé. Certaines recherches portent sur des publics fragilisés dont la prise de parole constitue aussi une prise de risque. Par ailleurs, les consentements oraux ou écrits donnés par les participant·e·s se négocient généralement à l'aune d'une problématique de recherche spécifique qui encadre leurs motivations fondées sur un rapport de confiance, ce qui pose évidemment problème si les données produites peuvent potentiellement être réexploitées par d'autres chercheur·e·s ou citoyen·ne·s dans un tout autre cadre de questionnement.

En sciences humaines et sociales, les données ne sont jamais «données», elles sont «produites» par les chercheur·e·s et, donc, de fait, «obtenues»⁹ dans le cadre d'un travail de négociation avec les participant·e·s, de mise en place d'un protocole, de réalisation de l'expérimentation... Les chercheur·e·s qui les ont produites, sous le patronage de leur institution et de leurs bailleurs de fonds, en sont dès lors les dépositaires.¹⁰ S'ils ou elles en sont responsables, ils ou elles n'en sont pas les propriétaires mais seulement les usufruitiers.¹¹ En ce sens, les données ne sont pas seulement «obtenues», elles nous sont surtout «confiées», de manière souvent implicite, dans le cadre de ce que l'on pourrait appeler «un pacte de bienveillance».¹² Les chercheur·e·s s'engagent en effet à respecter les données obtenues, c'est-à-dire à les faire parler dans le cadre spécifique d'une problématique énoncée au départ, même si les hypothèses et les orientations de la réflexion seront amenées à évoluer durant l'enquête sous l'influence des données elles-mêmes.

«La phase de production des données peut être ainsi analysée comme une restructuration incessante de la problématique au contact de celles-ci, et comme un réaménagement permanent du cadre interprétatif au fur et à mesure que les éléments empiriques s'accumulent.»¹³

Il n'y a donc pas de correspondance directe entre des données et leurs interprétations. Ces dernières sont également situées et dépendantes des ancrages théo-

riques/disciplinaires, voire des variables de genre, socioéconomiques et biographiques des chercheur·e·s qui doivent faire l'objet d'une approche réflexive.

Battre le FAIR quand il est faux

Les principes FAIR tentent d'encadrer l'ambition légitime d'une recherche coopérative dont les productions (données primaires et résultats) puissent profiter à l'ensemble de l'humanité, au-delà des frontières locales et nationales, à l'encontre des rétentions stratégiques qui peuvent se jouer entre chercheur·e·s, entre équipes, entre laboratoires... dans une course au primo-découvrant et aux brevets monnayables. Mais encore une fois, si l'intention de départ est unanimement saluée, sa traduction pratique pose passablement de questions. Le F (*findable*) et la A (*accessible*) ne soulèvent pas de problèmes majeurs. Dans un grand nombre de recherches (y compris qualitatives), un travail de nettoyage et d'anonymisation est possible en vue d'un partage des données respectant les principes éthiques qui encadrent la recherche scientifique. Mais les deux autres lettres de l'acronyme sont plus difficiles à généraliser. La réutilisation (R) de données n'est pas problématique. En anthropologie par exemple, elle se pratique dans le comparatisme et la revisite.¹⁴ Mais lorsque *reusable* est associé à *interoperable*, la réutilisation d'un set de données par une autre équipe de recherche, dans un nouveau cadre problématique et interprétatif, présuppose un travail de standardisation des données. Or, dans le cadre des recherches qualitatives, toute standardisation des données est à la fois un exercice de formalisation et de décontextualisation qui, de fait, transforme les données (et donc les appauvrit, voire les falsifie) pour en faire des sources stables et datées, totalement indépendantes de leur contexte de production et de la position singulière des chercheur·e·s qui s'y sont engagé·e·s et investi·e·s, en colorant ainsi les données produites (statut adopté face aux interlocuteurs·trices, méthodologie choisie, niveau d'empathie et d'engagement sur le terrain...). C'est à ce stade que l'interopérabilité, rêvant à des données suffisamment standardisées pour être téléchargées, croisées et analysées par des algorithmes, se confronte à la spécificité des données issues des recherches qualitatives.¹⁵

«Cet effort constant d'uniformisation (des outils, des rythmes, des rapports...) se heurte donc à l'hétérogénéité inévitable des regards. [...] Formalisation et autonomie de l'enquêteur sont donc de mise, avec toutes les tensions et résistances que présuppose ce couple antagoniste.»¹⁶

Dans de nombreuses disciplines, l'injonction à l'*open research data* représente donc tout autant un risque qu'une chance. Avant que ce mouvement n'emporte un peu trop vite avec lui notre légitimité à défendre nos spécificités discipli-

naires, il faut que les chercheur·e·s statutaires interpellent leurs autorités de tutelle et prennent position afin que la mutualisation des données soit d'abord une plus-value (quand elle est possible), avant d'être une condition *sine qua non* au financement et/ou à la publication. Mais aussi afin que les raisons juridiques ne soient pas sous peu les seules raisons légitimes pour refuser le partage des données de la recherche sans en subir de conséquences professionnelles.

S'il est donc légitime de poursuivre la réflexion sur l'adaptabilité des principes FAIR à la diversité des approches, domaines et disciplines, il est également essentiel que les institutions de recherche et les associations scientifiques revendiquent et acquièrent leur légitimité à statuer sur ces nouvelles pratiques pour les encadrer. On peut notamment citer ici en exemple la récente prise de position de la *Swiss Anthropological Association* (SAA) concernant le *data management plan* exigé par le Fonds national suisse de la recherche scientifique.¹⁷ Les ethnologues helvétiques expriment leurs réserves concernant le partage des données qualitatives dans un certain nombre de recherches et insistent sur la nécessaire contextualisation des données, au cœur de la démarche ethnographique. C'est bien ce type de démarche qu'il faut encourager et promouvoir. Car il nous appartient à nous, chercheur·e·s, de participer à l'élaboration d'une standardisation disciplinaire non décontextualisante¹⁸ et donc, de fait, in-interopérable.

Notes

- 1 www.budapestopenaccessinitiative.org (5. 1. 2022).
- 2 <https://openaccess.mpg.de/Berlin-Declaration> (5. 1. 2022).
- 3 Cf. *The Digital Switzerland Strategy* adopté par le Conseil fédéral suisse en 2018: www.bakom.admin.ch/bakom/en/homepage/digital-switzerland-and-internet/strategie-digitale-schweiz/digitale-schweiz.html (5. 1. 2022).
- 4 On distingue principalement la voie dorée (*gold road*), la voie verte (*green road*) et la voie de diamant (*diamond open access*). Toutes les questions que soulève le développement de l'*open access* – et notamment son impact sur la qualité de l'offre éditoriale – sont largement discutées dans d'autres contributions à ce numéro. Je ne les traiterai donc pas dans cet article.
- 5 OCDE, *Recommandation du Conseil concernant l'accès aux données de la recherche financée sur fonds publics*, 2021, <https://legalinstruments.oecd.org/fr/instruments/OECD-LEGAL-0347> (5. 1. 2022).
- 6 Comment ne pas penser aux récentes recherches sur la vaccination de la Covid-19 et à cette systématique tension entre les intérêts privés des laboratoires lancés dans la course au vaccin et les enjeux de santé publique à l'échelle internationale qui incitent à la coopération et au partage des données.
- 7 Cf. European Commission, *H2020 Programme Guidelines on FAIR Data Management in Horizon 2020*, 26. 7. 2016, https://ec.europa.eu/research/participants/data/ref/h2020/grants_manual/hi/oa_pilot/h2020-hi-oa-data-mgt_en.pdf (5. 1. 2022).
- 8 Pour que des données soient réellement anonymisées, il faut que l'équipe de recherche, au terme de la phase de récolte des données, détruise toutes possibilités (même pour elle-même) de retrouver l'identité des participant·e·s à partir des données. Les informations identifiantes doivent donc être supprimées, mais aussi toutes les correspondances (courriers, *e-mails*), tou-

tes les informations notées sur des agendas, etc. Dans ce cas, le partage des données ne pose alors théoriquement plus de problème (théoriquement, car il reste toujours, dans un entretien biographique par exemple, des données semi-identifiantes qui pourraient permettre, avec des technologies adaptées, de retrouver l'identité des personnes). Dans les recherches qualitatives, l'anonymisation des données est extrêmement rare. On parle plutôt de «codage» ou de «pseudonymisation», car les équipes de recherche conservent, dans des documents séparés et sécurisés, les clefs de correspondance permettant de réidentifier les données en cas de besoin.

- 9 Cf. Bruno Latour, *Petites leçons de sociologie des sciences*, Paris 1996, 188.
- 10 Le sujet de la propriété intellectuelle des données (et donc de la gouvernance de leur potentiel partage) pose néanmoins question. La légitimité des chercheur·e·s à partager ou non leurs données et à en définir les conditions ne semble pas aujourd'hui mise en cause. Mais comment ces thématiques seront-elles gérées par les institutions quand les chercheur·e·s ne seront plus en poste? Qui aura autorité pour réguler les conditions d'accès aux données déposées sur des *data repositories*? Ces interrogations me semblent encore bien peu clarifiées sur le plan juridique.
- 11 Même s'il vient du jargon juridique, le terme «usufruitiers» est ici tout à fait approprié puisqu'il évoque la possibilité de jouir d'un bien (ici les données) sans en être propriétaire, ce qui implique de fait un accord contractuel avec le propriétaire (ici les participant·e·s à l'étude) pour un usage convenu du bien.
- 12 Cf. Laurent Amiotte-Suchet, «L'enquête de terrain en sciences sociales des religions», *Études théologiques et religieuses* 93 (2018), 529–541.
- 13 Cf. Jean-Pierre Olivier de Sardan, «La politique du terrain», *Enquête* 1 (1995), 71–109, ici 95.
- 14 L'utilisation de matériaux produits par les chercheur·e·s sur différents terrains est en effet courante dans le cadre d'analyses comparées. Il existe également une longue tradition de revisite des terrains, afin de mettre en évidence des variations dans le temps ou de changer de points de vue en faisant varier les regards, cf. Laurent Amiotte-Suchet et al., «Enquêtes collectives. Histoires et pratiques contemporaines», *ethnographiques.org* 32 (2016), www.ethnographiques.org/2016/Amiotte-Suchet-Laferte-Lauriere-Renahy (5. 1. 2022).
- 15 Voir notamment Sophie Duchesne (éd.), «La réanalyse des enquêtes qualitatives à l'épreuve de l'expérimentation», *Recherches qualitatives* 21 (2017), www.recherche-qualitative.qc.ca/documents/files/revue/hors_serie/HS-21/rq-hs-21-numc-ro-complet.pdf (5. 1. 2022); Pablo Diaz, «Archiving Qualitative Data in Practice. Ethical Feedback», *Bulletin of Sociological Methodology* 150 (2021), 7–27.
- 16 Amiotte-Suchet et al. (voir note 14).
- 17 Swiss Anthropological Association, *Open Science and Data Management in Anthropological Research. Position Paper of the Swiss Anthropological Association (SAA)*, 2021, www.sagw.ch/en/seg/about-the-saa/swiss-anthropological-association/position-papers (5. 1. 2022).
- 18 Difficile d'approfondir en quelques lignes une idée qui mériterait, de fait, un long développement. En ce qui concerne la mise à disposition des données observationnelles, on peut citer la contribution de Daniel Bizeul (Avec ceux du FN. 1995–2003, Centre de données sociopolitiques, beQuali (2020), https://bequali.fr/fr/les-enquetes/lenquete-en-bref/cdsp_bq_s11 (5. 1. 2022)), mais sans oublier de mentionner qu'un tel travail n'a été possible qu'à partir d'une longue collaboration entre le chercheur et des professionnels de l'archivage (et donc du temps et des moyens dont peu d'entre nous disposent). Pour rendre des jeux de données qualitatives compréhensibles et potentiellement réutilisables par d'autres chercheur·e·s, les équipes de recherche se tournent généralement vers des outils de codage (ATLAS-ti, NVivo, MAX-QDA). Mais en codant et catégorisant les données, ces outils les transforment. Après le travail de codage, elles ne sont plus des données mais des informations catégorisées associées à des métadonnées. Or, si ces démarches peuvent s'avérer heuristiques, elles ne sont pas pertinentes pour toutes questions de recherche et représentent un travail rigoureux, sur les plans technique, éthique et analytique, qui ne peut pas être réalisé sans s'inscrire dans un objectif précis (comparaison, croisement avec des données quantitatives, analyse de cohortes, etc.) qui délimite déjà les conditions de la réutilisation.