# Un data paper en SHS : pourquoi, pour qui, comment ?

Victor Gay

# Un *data paper* en SHS : pourquoi, pour qui, comment ?

Victor Gay[1]

*Destinée aux chercheurs en sciences humaines et sociales souhaitant se lancer dans l'écriture d'un data paper, cette intervention propose un retour d'expérience prenant appui la production récente d'un tel article (Gay 2021). Celle-ci aborde les enjeux auxquels un producteur de data paper est souvent confronté.*

Pourquoi écrire un *data paper ?* Alors que les sciences humaines et sociales connaissent un tournant quantitatif depuis une dizaine d'années, la valeur scientifique de la production de données reste peu reconnue : les comités d'évaluations privilégient encore l'article de recherche traditionnel tandis les pairs réutilisent volontiers les données créées par d'autres sans pour autant leur accorder citation. Ce manque de reconnaissance semble pourtant incompatible avec le travail chronophage que demandent la documentation du processus de production des données ainsi que leur mise en conformité avec les principes FAIR – deux éléments nécessaires à la reproductibilité des travaux de recherche. Dans ce contexte, le *data paper* constitue un outil qui peut permettre aux producteurs de données de faire reconnaître leur contribution scientifique en rendant leurs données facilement citables, mais aussi en améliorant la pertinence ainsi que le périmètre de la réutilisation de leurs données.

Ensuite, pour qui écrire un *data paper* ? Aujourd'hui, seules quelques revues en sciences humaines et sociales acceptent le format *data paper*. Cette rareté peut paradoxalement constituer une chance pour les producteurs de données dans la mesure où cela peut leur permettre d'atteindre un lectorat relativement large et interdisciplinaire. Une telle portée demande cependant une adaptation de l'écriture à une audience non spécialiste. Cela peut aussi avoir des implications sur les formats adéquats à appliquer aux données ainsi qu'à ses canaux de diffusion.

Enfin, comment écrire un *data paper* afin qu'il constitue une véritable clé d'accès pour la compréhension et la réutilisation des données décrites ? Ici, il semble opportun de s'inspirer de modèles éprouvés issus des sciences dures, tout en les adaptant aux spécificités des sciences humaines et sociales.

Il ressortira que de nombreuses compétences sont nécessaires afin de produire un écosystème cohérent « base de données-*data paper* ». Heureusement, un certain nombre d'appuis institutionnels existent pour permettre aux chercheurs d'acquérir ces compétences ou de se faire accompagner au cours du processus de production – en particulier le réseau national des MSHS.

Tout au long de cette intervention, je prendrai appui sur mon expérience issue de la rédaction récente d'un *data paper :* « *Mapping the Third Republic. A Geographic Information System of France (1870-1940) »* (Gay 2021). Ce *data paper* décrit un système d'information géographique de la France de la Troisième République – la base TRF-GIS[2]. Cette base de données met à disposition nomenclatures et *shapefiles* annuels correspondant aux circonscriptions administratives de France métropolitaine de 1870 à 1940. Elle décrit les circonscriptions administratives générales (départements, arrondissements, cantons) ainsi que les circonscriptions militaires, judiciaires, pénitentiaires, électorales, académiques, ecclésiastiques et les inspections du travail. Elle met aussi à disposition des nomenclatures annuelles établissant une correspondance entre chaque commune contemporaine et les circonscriptions auxquelles elle appartenait[3].

---

[1] Victor Gay est enseignant-chercheur à l'École d'Économie de Toulouse (TSE) et à l'Institut d'Études Avancées de Toulouse (IAST) de l'Université Toulouse 1 Capitole, Toulouse, France. Adresse électronique : victor.gay@tse-fr.eu.

[2] TRF-GIS signifie *Third Republic France Geographic Information System.*

[3] Le *data paper* est librement disponible sur HAL à l'adresse suivante : https://hal.archives-ouvertes.fr/hal-02951461. Les données sont accessibles sur le Harvard Dataverse à l'adresse suivante : https://dataverse.harvard.edu/dataverse/TRF-GIS.

## Un *data paper* en SHS : pourquoi ?

L'usage des données a pris une place prépondérante dans la recherche en sciences humaines et sociales au cours des dix dernières années, en partie grâce à la production sans précédent de statistiques portant sur les faits sociaux et leur disponibilité via les catalogues de données en ligne tels que PROGEDO-ADISP[4]. C'est par exemple le cas en sociologie, une discipline au cœur des SHS : une analyse des 400 articles parus entre 2000 et 2020 dans la *Revue française de sociologie* révèle une nette tendance vers le quantitatif, si bien que depuis une décennie, plus de la moitié des articles dans cette revue contient au moins une table ou un graphique présentant des statistiques (figure 1). C'est aussi le cas en histoire, comme en témoigne le récent numéro des *Annales. Histoire, sciences sociales* consacré à l'histoire quantitative (Karila-Cohen *et al.* 2018), ainsi que le succès de l'ouvrage *Méthodes quantitatives pour l'historien* (Zalc et Lemercier 2008), réédité en version anglaise il y a peu (Zalc et Lemercier 2020). De même, ce tournant quantitatif concerne les milieux anglo-saxons, aussi bien en histoire social qu'en histoire économique, où elle est plus ancienne : alors que la proportion d'articles avec au moins une table ou un graphique présentant des statistiques restait stable autour de 90 % dans les principales revues en histoire économique entre 2005 et 2020 (Cioni *et al.* 2021, 24), celle-ci est passée de 5 à 13 % dans l'*American History Review* sur la même période (Ruggles 2021, 14)[5].

**Figure 1. Pourcentage d'articles parus dans la *Revue française de sociologie* contenant au moins une table ou un graphique présentant des statistiques (2000-2020)**


La courbe représente une moyenne mobile sur cinq ans. Les articles ont été récoltés par l'auteur via les portails Cairn.info pour les volumes 44 à 62 (2004-2021) et Persée pour les volumes 39 à 43 (1998-2002). Les éditoriaux, notes, critiques, débats, articles traduits, commentaires, et les articles *in memoriam* n'ont pas été retenus.

C'est dans ce contexte de renouveau quantitatif que la crise de la reproductibilité, partant de la psychologie et de la médecine, a rattrapé les sciences sociales, en commençant par l'économie (Maniadis, Tufano et List 2017) : à cause des divers biais affectant le processus de publication – manipulation des valeurs *p*, faible puissance statistique, biais de confirmation des auteurs et relecteurs, etc. – une majorité des résultats de recherche publiés constitueraient en réalité des « faux positifs », autrement dit, une illusion statistique (Ioannidis 2005). Plusieurs réponses ont été apportées pour palier à ce problème : les plans de gestion de données, les plans

[4] Le catalogue PROGEDO-ADISP diffuse enquêtes et bases de données issues de la statistique publique française. Il est accessible à l'adresse suivante : http://www.progedo-adisp.fr/.

[5] Les journaux considérés par Cioni *et al.* (2021) sont l'*Economic History Review*, le *Journal of Economic History*, *Explorations in Economic History*, la *European Review of Economic History* et *Cliometrica*.

de pré-analyse, ou encore les méta-analyses (Maniadis et Tufano 2017 ; Christensen et Miguel 2018). Il semble cependant que la condition *sine qua non* de sortie de crise est la reproductibilité en tant que telle, c'est-à-dire la capacité à reproduire les résultats des travaux de recherche publiés. Tel n'est pas encore le cas : par exemple, Chang et Li (2022) montrent que seule la moitié d'un échantillon de 67 articles de macro-économie publiés dans des revues de renom sont reproductibles. La première étape qui découle de ce programme consiste donc en la mise en place des conditions pour la réutilisation (adéquate) des données de la recherche.

Comment y parvenir ? C'est ici qu'interviennent les principes FAIR, selon lesquels la reproductibilité requiert en premier lieu des données trouvables, accessibles, interopérables, et réutilisables (Wilkinson *et al.* 2016). Le respect de ces principes est depuis quelques années déjà au cœur de la politique nationale de la recherche française au travers du plan national pour la science ouverte ainsi que celle du financement de la recherche par l'ANR (CoSO 2019). Cependant, quel que soit le support des établissements et des infrastructures de recherche, le poids de la documentation et de la mise en conformité des données aux principes FAIR revient *in fine* en grande partie aux producteurs de données eux-mêmes, les chercheurs – qui sont par ailleurs déjà débordés par des tâches administratives qui ne cessent de s'accumuler (Ali et Rouch 2013). Ce rôle est pourtant en contradiction avec les incitations auxquelles ils font face. En effet, la production de données n'est généralement pas valorisée par les comités d'évaluation, qui se fient encore bien plus aux publications traditionnelles dans des revues à comité de lecture (Gozlan 2016). La production de données n'est pas plus valorisée par les pairs, qui en général ne citent pas les données qu'ils réutilisent. Par exemple, Robinson-García *et al.* (2015, 2970) montrent que seules 18 % des données réutilisées par les articles en sciences sociales publiés entre mai et juin 2013 et disponibles dans *Web of Science* étaient citées explicitement dans leur bibliographie[6].

Face à ces défis, les *data papers* offrent des perspectives intéressantes. En effet, en décrivant le processus de production des données dans un article publié dans une revue, ils offrent aux réutilisateurs de données un moyen simple pour citer les données, permettant par-là la reconnaissance scientifique du travail des producteurs de données par les comités d'évaluation, mais aussi par les pairs – au-delà, bien entendu, d'une utilisation à la fois plus adéquate grâce à la documentation et plus accessible grâce à la conformité aux principes FAIR[7]. De plus, dans la mesure où les *data papers* passent au crible du comité de lecture, ils génèrent des incitations pour les producteurs de données eux-mêmes afin de parfaire leurs données ainsi que leur description, ce qui *in fine* ne peut que favoriser leur diffusion et le périmètre de leur réutilisation (Walters 2020). Dans ce sens, le *data paper* peut contribuer à résoudre le problème classique du passager clandestin qui caractérise la production de biens publics purs – ici, la donnée documentée et FAIR.

Le *data paper* « *Mapping the Third Republic* » (Gay 2021) répond à ces problématiques dans un contexte de recherche particulier, celui de la Troisième République (1870-1940). En effet, cette période historique est caractérisée par une production sans précédent de statistiques (Desrosières 2010 [1993]) par des administrations opérant à des niveaux d'agrégation hétérogènes, dans un contexte de profonds changements socio-économiques tels que la Seconde Révolution Industrielle (1870-1914) ou la Première Guerre Mondiale (1914-1918). Ainsi, grâce aux nouveaux moyens de numérisation des archives statistiques, la recherche en histoire

[6] Une étude plus récente portant sur les pratiques de citation des données en biodiversité en 2019 constate que sur un échantillon aléatoire de 100 articles réutilisant de telles données, seuls 27 % les mentionnent explicitement dans leur bibliographie (Khan *et al.* 2021).

[7] Même si les effets des *data papers* sur la réutilisation des données n'ont pas encore atteint leur plein potentiel, en tout cas en sciences dures (Jiao et Darch 2020).

quantitative sur cette période est en forte expansion. L'analyse et la visualisation de données historiques géolocalisées requiert cependant un cadre de référence ou un système d'information géographique (SIG) commun. Celui-ci n'existant pas à ce jour pour la France de la Troisième République, chaque chercheur doit s'atteler individuellement à cette tâche. Cela implique une perte de temps considérable, des approximations à cause de la difficulté de l'entreprise, mais aussi un manque d'interopérabilité avec d'autres programmes de recherche et donc *in fine* un manque de reproductibilité. En effet, ces systèmes d'information sont rarement mis à disposition du public et plus rarement encore conformes aux principes FAIR[8]. La base de données TRF-GIS offre donc une solution à ces problèmes en proposant des données FAIR dont le processus de construction est documenté avec précision dans un *data paper*.

## Un *data paper* en SHS : pour qui ?

Pour qui écrire un *data paper* ? Les utilisateurs potentiels des données décrites, bien sûr. Cependant, le lectorat d'un *data paper* est souvent bien plus large et interdisciplinaire qu'initialement envisagé par le producteur de données. En effet, les revues acceptant ce genre d'article restent rares en sciences humaines et sociales. À ma connaissance, il existe deux revues dédiées aux *data papers* – des *data journals* – dans ce domaine : le *Journal of Open Humanities Data* et le *Research Data Journal for the Humanities and Social Sciences* – bien que certains *data journals* publient en sciences sociales mais aussi dans d'autres disciplines, comme *Data in Brief* ou *F1000 Research*[9]. De plus, quelques revues en sciences humaines et sociales acceptent des *data papers* en plus d'articles de recherche traditionnels, comme *Cybergeo* pour la géographie, *Historical Methods : A Journal of Quantitative and Interdisciplinary History* pour l'histoire ou *Frontiers in Sociology* pour la sociologie. Enfin, un certain nombre de revues offrent la possibilité de publier des *data papers* dans leur section pour articles courts sans qu'ils explicitent la distinction de ce genre d'article avec un article de recherche traditionnel. Pour le domaine de l'histoire, c'est par exemple le cas pour *Explorations in Economic History* ou encore *Histoire & mesure*.

Le producteur de *data paper* doit donc composer avec une offre réduite qui implique mécaniquement un lectorat plus large que son propre champ de recherche. Une telle portée demande cependant une adaptation de l'écriture à une audience non spécialiste : un travail d'exposition du contexte qui évite le jargon disciplinaire, une explicitation des présupposés ayant présidé à la collecte de données et aux catégorisations, et une clarification des usages potentiels des données au-delà de son propre champ de recherche. Ces considérations ont aussi des implications sur la forme des données : on peut bien entendu mettre à disposition ses données dans le format le plus courant dans sa discipline (SPSS, Stata ou SASS) mais il est important de donner la possibilité à d'autres champs disciplinaires de s'en saisir en proposant aussi un format ouvert et universel comme le TXT ou le CSV. Le choix de l'entrepôt de données doit lui aussi être réfléchi : on peut ainsi se concentrer sur un entrepôt français comme Nakala ou PROGEDO-ADISP, ou bien viser un public plus international en utilisant le Harvard Dataverse, Figshare, ou Zenodo – avec le risque de perdre en visibilité dans le paysage hexagonal, leur interface restant en anglais[10].

---

[8] Une exception toutefois sont les *shapefiles* de cantons pour 1884 et 1925 publiés par le LARHRA (2011) et construits selon la méthode de vectorisation manuelle de cartes historiques géoréférencées. Voir Gay (2021, 14-15) pour une analyse critique de cette méthode.

[9] Une liste des revues publiant des *data papers* est proposée par Laurence Dedieu (Cirad) à l'adresse suivante : https://doi.org/10.18167/coopist/0057.

[10] Il n'existe pas à ma connaissance d'entrepôt spécifiquement dédié aux sciences humaines et sociales, mis à part Didómena (https://didomena.ehess.fr/), l'entrepôt de données de l'EHESS dont l'usage est limité à ses membres.

Enfin, il convient de clairement délimiter les contours du *data paper* en tant que descripteur de données et non comme procédant de l'analyse ou de la généalogie de projet, surtout si celui-ci est publié dans une revue n'ayant pas de section dédiée aux *data papers*. En effet, ce format n'est pas encore ancré dans les usages et nombre d'articles dont le but principal est de décrire une base de données comportent aussi des analyses des données elle-même, ce qui contribue à brouiller les pistes[11].

Le *data paper* « *Mapping the Third Republic* » (Gay 2021) propose nomenclatures et *shapefiles* annuels pour différentes administrations de la Troisième République (1870-1940). Dans la mesure où cette base de données permet de réaliser aussi bien de la cartographie de données historiques géolocalisées que d'appareiller des données issues de ces administrations pour réaliser des analyses statistiques, elle s'adresse à l'ensemble des sciences sociales avec une composante historique, et en premier lieu l'histoire économique et la démographie historique. C'est pourquoi mon choix s'est porté sur le journal *Historical Methods : A Journal of Quantitative and Interdisciplinary History*, une revue relativement reconnue, avec une diffusion internationale large, et dont le lectorat reste interdisciplinaire en sciences sociales avec une forte composante historique et une dominante quantitative. *Explorations in Economic History* aurait été une autre candidate, mais son lectorat reste trop restreint à l'économie historique anglo-saxonne. A contrario, le lectorat de *Journal of Open Humanities Data* ou de *Research Data Journal for the Humanities and Social Sciences* est relativement moins tourné vers la cartographie ou l'analyse économétrique.

En termes de format, la base de données TRF-GIS est disponible dans le format dominant du champ de l'histoire économique : le Stata *data format* (DTA). Mais pour permettre à d'autres disciplines de se saisir des données, j'ai aussi rendu celles-ci disponibles en format TXT, facilement importable dans n'importe quel logiciel statistique ou cartographique. Enfin, je me suis porté sur l'entrepôt Harvard Dataverse. Cet entrepôt propose un espace de stockage gratuit conséquent pour les chercheurs individuels, des facilités de dépôts, une gestion automatique des métadonnées, ainsi qu'une bonne ergonomie d'utilisation pour le téléchargement et la navigation. C'est aussi un entrepôt avec une très forte diffusion à l'international. Afin de ne pas manquer de toucher les communautés francophones, j'ai cependant pris soin de diffuser les données sur la plateforme data.gouv.fr ainsi que dans le catalogue PROGEDO-ADISP[12].

Enfin, pour ce qui est de la forme du *data paper*, le passage du comité de lecture s'est passé sans encombre. Cependant, la remarque suivante de la part d'un relecteur anonyme montre bien le chemin qu'il reste à parcourir quant à ce type d'article : « *The question is to determine if this work constitutes an article. I do not think so. Thus, I do not think that the review could accept this text which does not ask a clear historical problem, but is only a kind of guideline and a description of an end product.* »

## Un *data paper* en SHS : comment ?

Comment écrire un *data paper,* pour qu'il ne soit pas qu'un simple *codebook*, mais constitue une véritable clé d'accès pour la compréhension et la réutilisation des données décrites ? Une tendance actuelle est la génération automatique de *data papers* à partir des métadonnées de la base de données (Schöpfel *et al.* 2019). Bien que ce type de format reste bien moins chronophage qu'un *data paper* standard et permette une diffusion rapide via le moissonnage des catalogues de

[11] Par exemple, à la fin d'un article décrivant une nouvelle base de données sur les esclaves émancipés de la ville du Cap en 1834, Ekama *et al.* (2021, 9) analysent les déterminants des prix des esclaves par une régression linéaire.

[12] Le dépôt sur data.gouv.fr est disponible à l'adresse suivante : https://www.data.gouv.fr/fr/datasets/systeme-dinformation-geographique-de-la-france-de-la-troisieme-republique-1870-1940/. Les données seront prochainement disponibles dans le catalogue PROGEDO-ADISP dans la section « Données historiques » à l'adresse http://www.progedo-adisp.fr/enquetes_donhist.php.

données, il ne semble pas permettre de réutilisation adéquate à cause du manque de description des données dont il fait preuve – cette méthode est donc à proscrire. La description des méthodes de construction et des choix de typologie doit être au cœur du *data paper*, chose que seul un humain est capable de réaliser.

Dans ce contexte, la première difficulté à laquelle est confronté un producteur de *data paper* lors de la rédaction est celle de la structuration du texte, car celle-ci ne correspond pas tout à fait à celle d'un article traditionnel. Un excellent modèle à suivre est celui proposé par le *data journal* en sciences dures *Scientific Data* du groupe *Nature*[13]. Bien que ce modèle s'adresse aux sciences dures, sa structure peut tout à fait s'adapter aux sciences humaines et sociales. La structure proposée commence par une première section (« Contexte et résumé ») qui décrit succinctement les données produites, leur contexte scientifique ainsi que leurs réutilisations potentielles. Une seconde section (« Méthodes ») décrit avec précision toutes les procédures utilisées dans le processus de production des données afin que celui-ci soit reproductible. Ensuite, une section « Fichiers de données » décrit chaque jeu de données associé avec le *data paper*. Cela comprend variables, noms de fichiers, localisation, ainsi que formats et poids numérique. Une quatrième section (« Validation technique ») présente les analyses ou procédures ayant permis de confirmer la validité des données décrites (en sciences humaines et sociales, il peut s'agir d'une confrontation avec différentes sources ou avec des données comparables). Une cinquième section (« Notes d'usage ») permet à l'auteur de décrire en plus de détails les procédures de réutilisation des données ainsi que de développer quelques exemples. Enfin, une dernière section (« Disponibilité du code ») permet le cas échéant de décrire les procédures d'accès au code de reproduction de la base de données.

Une seconde difficulté à laquelle est confronté le producteur de *data paper* concerne la mise à disposition des données. En effet, contrairement à l'article de recherche classique où les fichiers de données et de reproduction ne sont possiblement requis que lors de la phase de publication proprement dite, ceux-ci sont ici requis lors de la soumission de l'article à la revue avec des procédures d'accès qui doivent être explicitées dans le texte. En effet, dans la mesure où c'est aussi l'intégrité des données qui est évaluée par les relecteurs, ceux-ci doivent impérativement y avoir accès dès l'étape de relecture. Les données doivent donc déjà avoir un identifiant pérenne à ce stade, par exemple un DOI, ce qui permet de rattacher les données au *data paper* et de créer un écosystème « données-*data paper* ». Cela peut cependant poser problème si les relecteurs suggèrent des modifications pour publication. Heureusement, la plupart des entrepôts de données offrent la possibilité aux producteurs d'attribuer un identifiant pérenne qui ne sera accessible (anonymement) que par l'intermédiaire d'un code qu'il faudra fournir au comité de lecture.

La rédaction du *data paper* « *Mapping the Third Republic* » (Gay 2021) suit rigoureusement le modèle proposé par *Scientific Data* décrit plus haut. En effet, après une introduction présentant le contexte scientifique dans lequel la base de données s'insère, le corps de l'article consiste en une section « Méthode » qui explique en détail non seulement la méthodologie technique de construction ainsi que ses limites, mais aussi les éléments institutionnels qui sous-tendent les variations temporelles de chaque administration au cours de la Troisième République. J'y aborde par exemple les changements territoriaux impliqués par la perte et le retour de l'Alsace-Lorraine,

[13] Les instructions de soumission à *Scientific Data* sont disponibles à l'adresse suivante : https://www.nature.com/sdata/publish/submission-guidelines. D'autres modèles existent, comme celui proposé par *Data in Brief*, disponible à l'adresse suivante : https://www.elsevier.com/journals/data-in-brief/2352-3409/guide-for-authors. Plus généralement, Kim (2020) propose un tour d'horizon des formats de *data papers* proposés par un certain nombre de *data journals*.

la réforme des arrondissements de 1926, les réformes militaires de 1873-1874, ou encore les différentes lois de redécoupage électoral (les éléments précis sont étayés par des tableaux situés dans un appendice électronique d'une centaine de pages). S'ensuit une description des quinze jeux de données de la base TRF-GIS (variables, espace de stockage, formats, licence), puis une exposition de la validation technique de l'ensemble, sous deux formes : une confrontation des sources secondaires utilisées avec un ensemble de 175 sources primaires (individuellement listées en appendice et disponibles en PDF dans l'entrepôt de données) et une validation de la méthode de construction des *shapefiles* par comparaison avec des données similaires (le *shapefile* des cantons de 1884 du LARHRA). L'article se termine par la description de l'emplacement du code et des données, ainsi qu'un exemple de réutilisation (la cartographie de l'abstention lors des élections législatives de 1914 au niveau des circonscriptions électorales).

## Conclusion

La production d'un *data paper* demande un certain apprentissage dans la mesure où il s'agit d'un format nouveau en sciences humaines et sociales. Il est cependant en pleine expansion, et les chercheurs désireux de se lancer dans l'écriture d'un *data paper* ont à leur disposition de nombreux retours d'expérience à travers séminaires, journées spéciales, ou colloques dédiés (souvent mis en ligne, comme celui-ci). J'ai par exemple pu présenter le *data paper* décrivant la base TRF-GIS dans le cadre de la semaine DATA SHS 2020 organisée par la Plateforme Universitaire de Données de Toulouse (PUD-T) en décembre 2020, au webinaire dédié aux *data papers* organisé par le Groupe de Travail inter-réseaux « Atelier Données » du CNRS en février 2021, au séminaire « Des sources aux SIG : des outils pour la cartographie dans les humanités numériques » en mai 2021, ou encore au séminaire « Histoire et numérique » du Centre d'histoire de Science Po (CHSP) en mai 2021[14].

De nombreuses ressources sont aussi mises à disposition par différents acteurs institutionnels comme DoRANum et CoopIST (Cirad), qui proposent des dossiers sur le sujet, les URFIST ou l'INRAE, qui proposent des formations, ou encore les Plateformes Universitaires de Données (PUD) de PROGEDO, qui proposent des journées sur ce thème dans les MSHS lors de la semaine annuelle DATA SHS[15].

## Références

Ali, Nawel Aït et Jean-Pierre Rouch. 2013. « Le « je suis débordé » de l'enseignant-chercheur ». *Temporalités de la recherche* 18 : 1-25. https://doi.org/10.4000/temporalites.2632.

Chang, Andrew C. et Phillip Li. 2022. « Is Economics Research Replicable ? Sixty Published Papers From Thirteen Journals Say « Often Not » ». *Critical Finance Review* 11 : 1-22. http://dx.doi.org/10.1561/104.00000053.

Christensen, Garret et Edward Miguel. 2018. « Transparency, Reproducibility, and the Credibility of Economics Research ». *Journal of Economic Literature* 56 (3) : 920-980. https://doi.org/10.1257/jel.20171350.

[14] La vidéo de ma présentation dans le cadre du séminaire « Des sources aux SIG » est disponible à l'adresse suivante : https://youtu.be/mBAlRdWR41k, de la minute 2 à 45.

[15] Les ressources de DoRANum sur les *data papers* sont disponibles à l'adresse suivante : https://doranum.fr/data-paper-data-journal/. Celles de CoopIST du Cirad sont disponibles à d'adresse suivante : https://doi.org/10.18167/coopist/0057. Un exemple de support de cours proposé par l'INRAE est disponible à l'adresse suivante : https://dx.doi.org/10.15454/1.478247389988942E12. Voir aussi Reymonet (2017).

Cioni, Martina, Giovanni Federico et Michelangelo Vasta. 2020. « The Long-Term Evolution of Economic History : Evidence from the Top Five Field Journals (1927-2017) ». *Cliometrica* 14 : 1-39. https://doi.org/10.1007/s11698-019-00186-x.

CoSO. 2019. « Pour un politique des données de la recherche : guide stratégique ». https://www.ouvrirlascience.fr/pour-une-politique-des-donnees-de-la-recherche-guide-strategique-a-lusage-des-etablissements/.

Desrosières, Alain. 2010 [1993]. *La politique des grands nombres. Histoire de la raison statistique.* Paris : La Découverte. https://doi.org/10.3917/dec.desro.2010.01.

Ekama, Kate, Johan Fourie, Hans Heese et Lisa-Cheree Martin. 2021. « When Cape Slavery Ended : Introducing a New Slave Emancipation Dataset ». *Explorations in Economic History* 81 : 101390. https://doi.org/10.1016/j.eeh.2021.101390.

Gay, Victor. 2021. « Mapping the Third Republic : A Geographic Information System of France (1870-1940) ». *Historical Methods : A Journal of Quantitative and Interdisciplinary History*. À paraître. https://doi.org/10.1080/01615440.2021.1937421.

Gozlan, Clémentine. 2016. « Les sciences humaines et sociales face aux standards d'évaluation de la qualité académique ». *Sociologie* 7 (3) : 261-280. https://doi.org/10.3917/socio.073.0261.

Ioannidis, John P. A. 2005. « Why Most Published Research Findings Are False ». *Plos Medicine* 2 (8) : e124. https://doi.org/10.1371/journal.pmed.0020124.

Jiao, Chenyue et Peter T. Darch. 2020. « The Role of the Data Paper in Scholarly Communication ». *Proceedings of the Association for Information Science and Technology* 57 : e316. https://doi.org/10.1002/pra2.316.

Karial-Cohen, Karine, Claire Lemercier, Isabelle Rosé et Claire Zalc. 2018. « Nouvelles cuisines de l'histoire quantitative ». *Annales. Histoire, sciences sociales* 73 (4) : 771-783. https://doi.org/10.1017/ahss.2019.90.

Khan, Nushrat, Mike Thelwall et Kayvan Kousha. 2021. « Measuring the Impact of Biodiversity Datasets : Data Reuse, Citations and Altmetrics ». *Scientometrics*, 126 : 3621-3639. https://doi.org/10.1007/s11192-021-03890-6.

Kim, Jihyun. 2020. « An Analysis of Data Papers Templates and Guidelines : Types of Contextual Information Described by Data Journals ». *Science Editing* 7 (1) : 16-23. https://doi.org/10.6087/kcse.185.

LARHRA. 2011. *Les cantons français de 1884 à 1966*. Lyon : Laboratoire de Recherche Historique Rhône-Alpes. http://geo-larhra.ish-lyon.cnrs.fr/?q=atlas-historique/territoires-d-etat/evolution-des-cantons-en-france.

Maniadis, Zacharias et Fabio Tufano. 2017. « The Research Reproducibility Crisis and Economics of Science ». *The Economic Journal* 127 (605) : F200-208. https://doi.org/10.1111/ecoj.12526.

Maniadis, Zacharias, Fabio Tufano et John A. List. 2017. « To Replicate or Not To Replicate ? Exploring Reproducibility in Economics through the Lens of a Model and a Pilot Study ». *The Economic Journal* 127 (605) : F209-235. https://doi.org/10.1111/ecoj.12527.

Reymonet, Nathalie. 2017. « Améliorer l'exposition des données de la recherche : la publication de data papers. » https://archivesic.ccsd.cnrs.fr/sic_01427978.

Robinson-García, Bicolas, Evaristo Jiménez-Contreras et Daniel Torres-Salinas. 2015. « Analyzing Data Citation Practices Using the Data Citation Index ». *Journal of the Association for Information Science and Technology* 67 (12) : 2964-2975. https://doi.org/10.1002/asi.23529.

Ruggles, Steven. 2021. « The Revival of Quantification : Reflections on Old New Histories ». *Social Science History* 45 (1) : 1-25. https://doi.org/10.1017/ssh.2020.44.

Schöpfel, Joachim, Dominic Farace, Hélène Prost et Antonella Zane. 2019. « Data Papers as a New Form of Knowledge Organization in the Field of Research Data ». *Knowledge Organization* 46 (8) : 622-638. https://doi.org/10.5771/0943-7444-2019-8-622.

Walters, William H. 2020. « Data Journals : Incentivizing Data Access and Documentation Within the Scholarly Communication System ». *Insight* 33 (1) : 1-18. http://doi.org/10.1629/uksg.510.

Wilkinson, Mark D, *et al.* 2016. « The FAIR Guiding Principles for Scientific Data Management and Stewardship ». *Scientific Data* 3 : 160018. https://doi.org/10.1038/sdata.2016.18.

Zalc, Claire et Claire Lemercier. 2008. *Méthodes quantitatives pour l'historien*. Paris : La Découverte. https://doi.org/10.3917/dec.lemer.2008.01.

Zalc, Claire et Claire Lemercier. 2019. *Quantitative Methods in the Humanities. An Introduction*. Charlottesville (Virg.) : University of Virginia Press. https://doi.org/10.2307/j.ctvbqs963.