

Science ouverte et intégrité scientifique

Victor Gay

TSE, IAST, UT1

Journées MATE-SHS



Toulouse
School of
Economics

- Contexte : un tournant quantitatif en SHS depuis 15 ans.
- La crise de la reproductibilité.
- La science ouverte comme solution à la crise :
 - Les données ouvertes et les principes FAIR.
 - Les data papers.
 - Les répliquions.

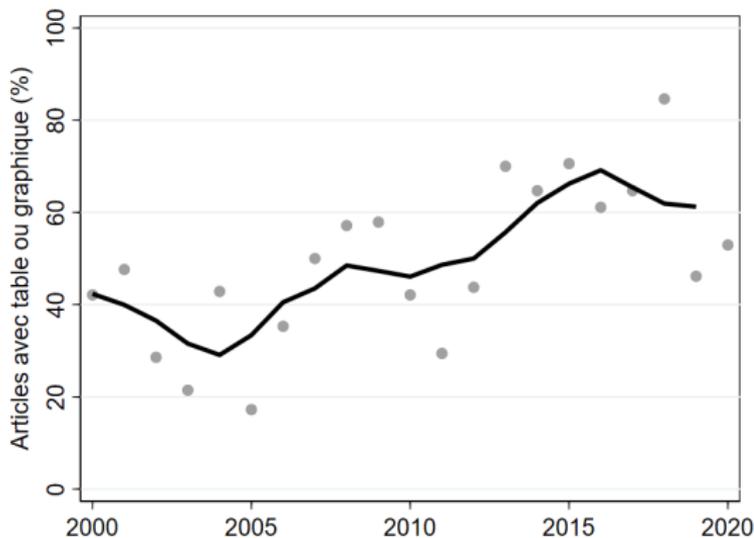
- **Contexte : un tournant quantitatif en SHS depuis 15 ans.**
- La crise de la reproductibilité.
- La science ouverte comme solution à la crise :
 - Les données ouvertes et les principes FAIR.
 - Les data papers.
 - Les répliquions.

Le tournant de la quantification

- Tournant quantitatif en SHS depuis 15 ans.

Le tournant de la quantification

- Tournant quantitatif en SHS depuis 15 ans.
- En sociologie.



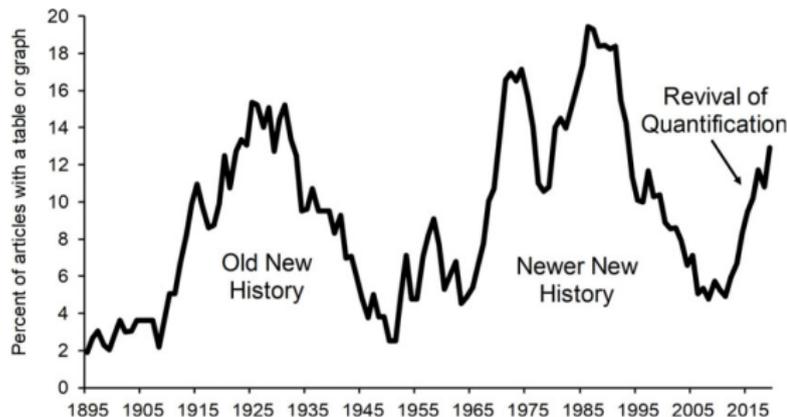
Pourcentage d'articles de la *Revue française de sociologie* avec au moins une table ou graphique de statistiques. Moyenne mobile sur 5 ans. Source : Gay (2021, 2).

Le tournant de la quantification

- Tournant quantitatif en SHS depuis 15 ans.
- En histoire (française).
 - Numéro « Histoire quantitative » des *Annales. Histoire, sciences sociales*. 2018.
 - Zalc, Claire et Claire Lemerrier. 2008. *Méthodes quantitatives pour l'historien*. Paris : La Découverte.
 - Zalc, Claire et Claire Lemerrier. 2020. *Quantitative Methods in the Humanities. An Introduction*. Charlottesville (Virg.) : University of Virginia Press.

Le tournant de la quantification

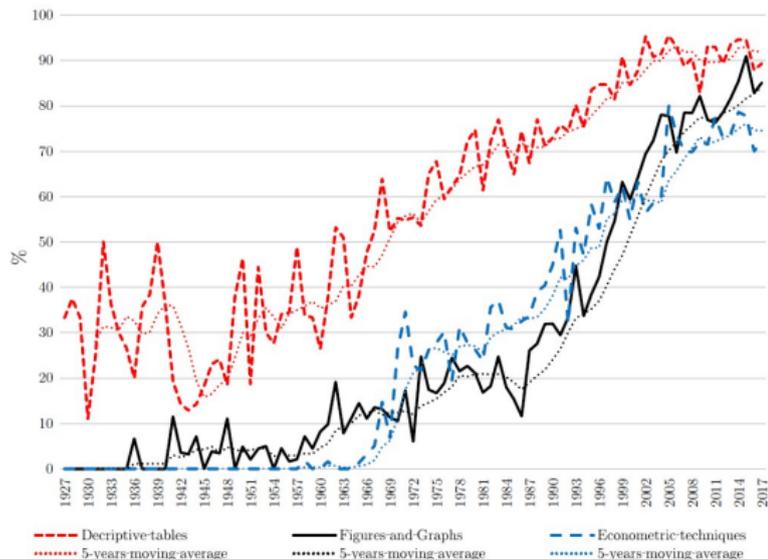
- Tournant quantitatif en SHS depuis 15 ans.
- En histoire (anglo-saxonne).



Pourcentage d'articles de l'*American History Review* avec au moins une table ou graphique de statistiques. Moyenne mobile sur 7 ans. Source : Ruggles (2021, 14).

Le tournant de la quantification

- Tournant quantitatif en SHS depuis 15 ans.
- En histoire économique (anglo-saxonne).



Pourcentage d'articles avec au moins une table ou graphique de statistiques dans le *Economic History Review*, le *Journal of Economic History*, *Explorations in Economic History*, la *European Review of Economic History*, et *Cliometrica*. Moyenne mobile sur 5 ans. Source : Cioni et al (2021, 24).

- Contexte : un tournant quantitatif en SHS depuis 15 ans.
- **La crise de la reproductibilité.**
- La science ouverte comme solution à la crise :
 - Les données ouvertes et les principes FAIR.
 - Les data papers.
 - Les répliquions.

La crise de la reproductibilité

- Biais affectant le processus de publication :
 - Manipulation des valeurs p . [Brodeur et al. \(2022\)](#)
 - Faible puissance statistique. [Ioannidis et al. \(2017\)](#)
 - Biais de confirmation des auteurs et relecteurs. [Kerr \(1998\)](#)

⇒ Résultats publiés = faux positifs ?

[Ioannidis \(2005\)](#) [Maniatis et al. \(2017\)](#)

La crise de la reproductibilité

- Biais affectant le processus de publication :
 - Manipulation des valeurs p . [Brodeur et al. \(2022\)](#)
 - Faible puissance statistique. [Ioannidis et al. \(2017\)](#)
 - Biais de confirmation des auteurs et relecteurs. [Kerr \(1998\)](#)

⇒ Résultats publiés = faux positifs ?

[Ioannidis \(2005\)](#) [Maniatis et al. \(2017\)](#)

- Quelques réponses :
 - Plans de pré-analyse. [Banerjee et al. \(2020\)](#)
 - Méta-analyses. [Stanley et Doucouliagos \(2021\)](#)
 - Certification de reproductibilité (CASCAD). [Pérignon et al. \(2019\)](#)

La crise de la reproductibilité

- Biais affectant le processus de publication :
 - Manipulation des valeurs p . [Brodeur et al. \(2022\)](#)
 - Faible puissance statistique. [Ioannidis et al. \(2017\)](#)
 - Biais de confirmation des auteurs et relecteurs. [Kerr \(1998\)](#)

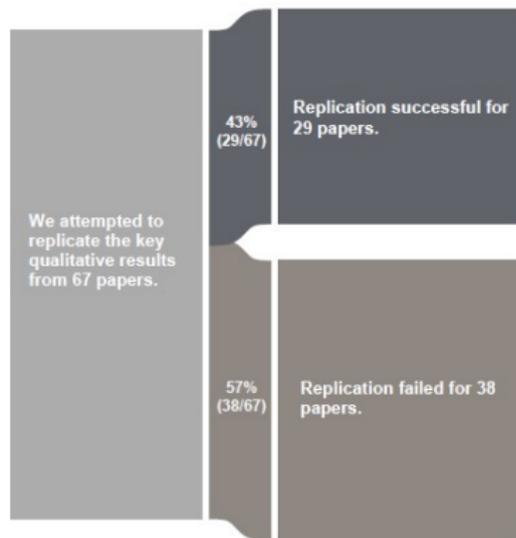
⇒ Résultats publiés = faux positifs ?

[Ioannidis \(2005\)](#) [Maniatis et al. \(2017\)](#)

- Quelques réponses :
 - Plans de pré-analyse. [Banerjee et al. \(2020\)](#)
 - Méta-analyses. [Stanley et Doucouliagos \(2021\)](#)
 - Certification de reproductibilité (CASCAD). [Pérignon et al. \(2019\)](#)
- Condition nécessaire = science ouverte et capacité à reproduire.

La crise de la reproductibilité

- Pas le cas par exemple en macro-économie



Résultats de reproduction de 67 articles en macro-économie. Source : Chang et Li (2022, 11)

- Contexte : un tournant quantitatif en SHS depuis 15 ans.
- La crise de la reproductibilité.
- **La science ouverte comme solution à la crise :**
 - **Les données ouvertes et les principes FAIR.**
 - Les data papers.
 - Les répliquions.

- Les principes FAIR pour la reproductibilité : [Wilkinson et al. \(2016\)](#)
 - Données trouvables (*Findable*).
 - Données accessibles (*Accessible*).
 - Données interopérables (*Interoperable*).
 - Données réutilisables (*Reusable*).

Les données ouvertes et les principes FAIR

- Les principes FAIR pour la reproductibilité : [Wilkinson et al. \(2016\)](#)
 - Données trouvables (*Findable*).
 - Données accessibles (*Accessible*).
 - Données interopérables (*Interoperable*).
 - Données réutilisables (*Reusable*).
- Des outils pour des données ouvertes :
 - Infrastructures de recherche publiques : Nakala, PROGEDO, Recherche data gov.
 - Infrastructures institutionnelles : Dataverse (INRAE, SciencePo), Didomena.
 - Infrastructures internationales : Harvard Dataverse, Figshare, Zenodo.

- Poids de la documentation et mise en conformité FAIR = chercheurs.
- Contradiction avec incitations pour les chercheurs.
- Production de données peu valorisée :
 - Par les comités d'évaluation pour promotion.
 - Par les pairs, qui citent peu les données réutilisées :
 - SHS 2013 : 18 %. [Robinson-García et al. \(2015\)](#)
 - Biodiversité 2019 : 27 %. [Kahn et al. \(2021\)](#)

- Contexte : un tournant quantitatif en SHS depuis 15 ans.
- La crise de la reproductibilité.
- **La science ouverte comme solution à la crise :**
 - Les données ouvertes et les principes FAIR.
 - **Les data papers.**
 - Les répliquions.

- L'outil *data paper* = outil d'alignement incitations :
 - Publication dédiée à des données = citation facilitée.
 - Publication dans une revue à comité de lecture = valorisation par comités d'évaluation.

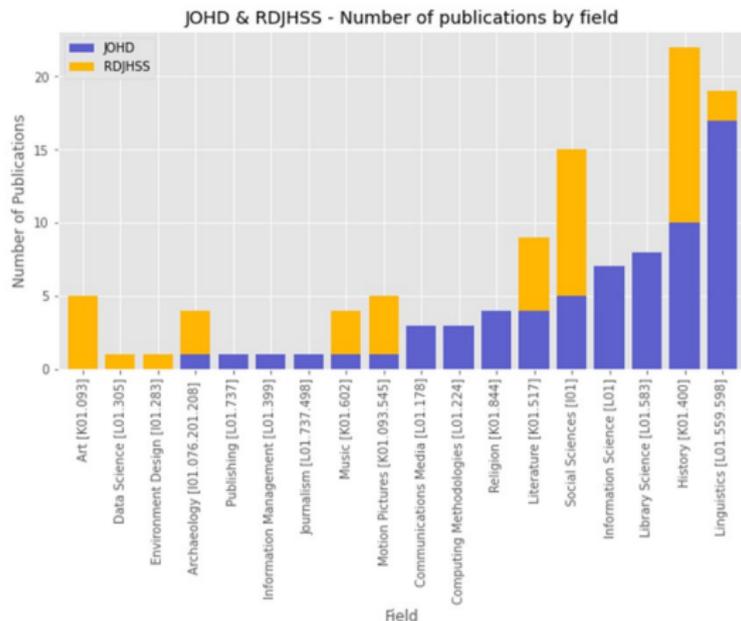
- L'outil *data paper* = outil d'alignement incitations :
 - Publication dédiée à des données = citation facilitée.
 - Publication dans une revue à comité de lecture = valorisation par comités d'évaluation.
- Un cercle vertueux :
 - Meilleure documentation = utilisation adéquate et facilitée.
 - Conformité FAIR = données plus accessibles et diffusées.
 - Crible du comité de lecture = meilleure documentation et conformité FAIR.

- L'outil *data paper* = outil d'alignement incitations :
 - Publication dédiée à des données = citation facilitée.
 - Publication dans une revue à comité de lecture = valorisation par comités d'évaluation.
- Un cercle vertueux :
 - Meilleure documentation = utilisation adéquate et facilitée.
 - Conformité FAIR = données plus accessibles et diffusées.
 - Crible du comité de lecture = meilleure documentation et conformité FAIR.
- Le *data paper* contribue à résoudre le problème du passager clandestin des biens publics purs : la donnée documentée et FAIR.

- Les opportunités de publication : [Schöpfel \(2022\)](#)
 - Data journals : *Journal of Open Humanities Data*, *Research Data Journal for the Humanities and Social Sciences*, *Data in Brief*, *F1000 Research*.
 - Journaux avec section dédiée : *Cybergeog*, *Historical Methods*, *Frontiers in Psychology*, *Demographic Research*.
 - Journaux pouvant les accepter : *Explorations in Economic History*, *Histoire & Mesure*.

- Les opportunités de publication : [Schöpfel \(2022\)](#)
 - Data journals : *Journal of Open Humanities Data*, *Research Data Journal for the Humanities and Social Sciences*, *Data in Brief*, *F1000 Research*.
 - Journaux avec section dédiée : *Cybergeog*, *Historical Methods*, *Frontiers in Psychology*, *Demographic Research*.
 - Journaux pouvant les accepter : *Explorations in Economic History*, *Histoire & Mesure*.
- Pour un exemple et un guide : [Gay \(2021\)](#)

Les data papers



Publications de data papers par discipline dans JOHD et RDJHSS (2015-2022). Source : McGillivray *et al.* (2022, 18).

- Contexte : un tournant quantitatif en SHS depuis 15 ans.
- La crise de la reproductibilité.
- **La science ouverte comme solution à la crise :**
 - Les données ouvertes et les principes FAIR.
 - Les data papers.
 - **Les réplifications.**

- Qu'est-ce qu'une réplique ? [Clemens \(2017\)](#)
 - Les tests de vérification :
 - Données originales, same échantillon, même méthode.
 - Permet de résoudre les erreurs de mesure et de code.

- Qu'est-ce qu'une réplique ? [Clemens \(2017\)](#)
 - Les tests de vérification :
 - Données originales, same échantillon, même méthode.
 - Permet de résoudre les erreurs de mesure et de code.
 - Les tests de reproduction :
 - Ré-échantillonnage de la même population, même méthode.
 - Permet de résoudre les erreurs d'échantillonnage et de faible puissance statistique

- Qu'est-ce qu'une réplique ? [Clemens \(2017\)](#)
 - Les tests de vérification :
 - Données originales, same échantillon, même méthode.
 - Permet de résoudre les erreurs de mesure et de code.
 - Les tests de reproduction :
 - Ré-échantillonnage de la même population, même méthode.
 - Permet de résoudre les erreurs d'échantillonnage et de faible puissance statistique
 - Les tests de solidité :
 - Réanalyse : même population, différente méthode.
 - Extension : ré-échantillonnage, différente méthode.

- Des taux de réplique faibles.

Paper	Number of replications as coded in paper (1)	Number of policing replications (our coding) (2)	Share of policing replications (3)	Replication rate according to paper (4)	Policing replication rate [(3)*(4)] (5)
Berry et al. (2017)	52	1	2%	60.0%*	1.2%
Sukhtankar (2017)	71	50	70%	6.2%†	4.4%
Mueller-Langer et al. (2019)	130	121	93%	0.1%	0.1%

Taux de réplique des articles en économie. Source : Fiala et al. (2022, 19).

- Plusieurs projets collectifs.
 - Institute for Replication : replication games et série de documents de travail.
 - Social Science Reproduction Platform : plateforme de répliques.
 - Replication Wiki : catalogue de répliques.
 - Many Economists project.
 - OSF's SCORE project.

- Des opportunités de publications avec comité de lecture :
 - Journal dédié : *Journal of Comments and Replications in Economics*.
 - Section pour les réplifications : *Journal of Political Economy*, *Microeconomics*, *Journal of Applied Econometrics*, *Spatial Economic Analysis*.
 - Journaux disciplinaires acceptant parfois des réplifications.
[Fiala et al. \(2022\)](#)

- Des opportunités de publications avec comité de lecture :
 - Journal dédié : *Journal of Comments and Replications in Economics*.
 - Section pour les répliques : *Journal of Political Economy*, *Microeconomics*, *Journal of Applied Econometrics*, *Spatial Economic Analysis*.
 - Journaux disciplinaires acceptant parfois des répliques.
[Fiala et al. \(2022\)](#)
- La science est-elle auto-correctrice ?
[Coupé et Reed \(2022\)](#) [Ankel-Peters et al. \(2023\)](#)
 - Les répliques (en économie) sont moins bien citées que les travaux originaux.
 - Les répliques (en économie) n'affectent pas l'influence des articles originaux dans la littérature.