

# Science ouverte, données ouvertes : Quels enjeux en sciences de gestion ?

Héloïse Berkowitz, LEST, CNRS, Aix Marseille Université

Hélène Delacour, Université de Lorraine

*Peer Community in Organization Studies*



# Contexte

- La science comme un bien public/commun global (Schiltz, 2007; Berkowitz & Delacour, 2020; Boulton, 2021)
- Transformations sur toute la chaîne de valeur, des données aux publications (Berkowitz & Delacour, 2022)
- Consensus croissant d'institutions, gouvernements et ONGs sur la nécessité de:
  - Mettre en œuvre des stratégies de science ouverte et des expérimentations d'alternatives (e.g. Plan S and cOAlition S, roadmap on open science, H2020 FAIR data principles, UNESCO 2021 Recommendation on Open Science)
  - Repenser l'évaluation de la recherche et des carrières et d'améliorer la bibliodiversité (e.g. DORA San Francisco Declaration on Research Assessment, COARA international coalition on research assessment)
- Transformer le système scientifique et de publication : d'un système commercial dominé par de grands éditeurs, à un système ouvert basé sur les communautés (Coalition S, 2023), Cf. Peer Community in Organization Studies

# Pourquoi ouvrir les données? 1/2

- Plus de **transparence, intégrité, accessibilité** et **rendu de compte**, selon certaines organisations publiques (e.g., European Commission, CNRS, etc.)
- Vérification des données, réduction des risques de duplication et réduction des coûts de collecte et de production de la données, mises en commun et réutilisation, **engagement** et confiance dans la science (OCDE, 2016)
- Meilleur accès à la recherche, partage d'information, promesses du “data-driven decision-making”, réduction des coûts gouvernementaux, services gouvernementaux plus efficaces (OpenAIRE)
- Mais ces derniers arguments reposent sur 2 hypothèses discutables:
  - La science devrait contribuer à améliorer la performance et l'efficacité
  - L'ouverture des données garantit des réutilisations et des gains d'efficacité

## Pourquoi ouvrir les données? 2/2

- Faciliter la recherche **multidisciplinaire sur des défis sociétaux complexes** (Gauquelin et al. 2017)
- Valoriser la collecte et la production de données, et surtout patrimonialiser les données
- >> Les données ont une valeur patrimoniale. Elles doivent être préservées, stockées, rendues accessibles - nous fabriquons aujourd'hui les archives de demain

# Défis et obstacles à l'ouverture des données en sciences sociales

- Des problèmes de perception des motivations pour l'ouverture, considérée comme un moyen de contrôle de la liberté académique
- Compétition et starification de la recherche en sciences de gestion, rankification (cf. Osterlof & Frey 2015, Brankovic et al. 2022), même si l'on sait que les équipes collaboratives sont plus innovantes et ont plus d'impact à long terme (Xu et al. 2022)
- Question de l'attribution de la propriété intellectuelle (Gauquelin et al. 2017, Millerand & Bowker 2008) et verrou de l'évaluation

# Contraintes et risques

- Coûts et manque de ressources et capacités, chronophage, épuisement professionnel
- Risques d'aggravation des inégalités entre chercheur-es, institutions, pays
- Manque d'habitude/de socialisation à l'ouverture des données dans nos communautés, rend difficile le changement de pratiques (cf. Longley Arthur & Hearn 2021)
- Eventail large de types et formats de données, méthodes mixtes, problèmes de normalisation
- Question de la protection/confidentialité/vie privée des données et des personnes interrogées (Bonneville et al. 2021)
- Risques de perdre l'accès au travail de terrain (réticence des participants, ou simplement coûts de réalisation des projets).

# Principes CARE

Parallèlement, réflexivité accrue sur la collecte de données et le risque de “subordination of the researched” (Vijay 2021)

Développement des principes CARE dans les protocoles de recherche avec les communautés indigènes (Carroll et al. 2021)



Source: <https://www.gida-global.org/care>

# Conditions opérationnelles pour l'ouverture des données de recherche : Quelques questions pour les chercheurs (1/2)

- Politique de données ouvertes à M@n@gement, réflexion autour des pratiques des chercheur-es et des journaux (Berkowitz & Delacour, 2022)
- Les données peuvent-elles être partagées et dans quelles conditions (éthiques) ?
- Comment préparer les données pour le partage ?
- Comment trouver et choisir un entrepôt ?
- Qu'est-ce qu'une « data availability statement » et que doit-on détailler dans l'article ?
- Comment lier vos ensembles de données (et d'autres) à votre article ?

<https://open-research-europe.ec.europa.eu/for-authors/data-guidelines>

Practice 'CARE' in data collection	Engage 'CARE' in data stewardship	Implement 'CARE' in data community	Use 'FAIR' with 'CARE' in data applications
Define cultural metadata Record provenance in metadata	Use appropriate governance models Make data 'FAIR'	Indigenous ethics inform access Use tools for transparency, integrity and provenance	Fairness, Accountability, Transparency Assess equity

Implementation of CARE principles (Carroll et al., 2021, p. 4)



# Conditions opérationnelles pour l'ouverture des données de recherche : Quelques questions pour les chercheurs (2/2)

Ces questions sont abordées dans les plans de gestion des données, qui se généralisent avec le fonctionnement de la recherche par projet

Section	Information and issues addressed
<i>General information</i>	Name of applicant, project number, funding program and number, version of DMP...
<i>Data description and collection or reuse of existing data</i>	How data will be collected or produced, what kind of data, formats and volumes, and whether and which data quality control measures will be used...
<i>Documentation and data quality</i>	Meta-data, standards, methodology describing data collection, measures for quality control...
<i>Storage issues and backup solutions (during research)</i>	Storage, backup, data security and protection of sensitive data...
<i>Legal and ethical requirements and codes of conduct</i>	How personal data is processed in compliance with the GDPR and other data privacy and protection laws, how intellectual property rights and ownership are managed, how ethical issues and codes of conduct arise and are managed in the research...
<i>Data sharing and long-term preservation mechanisms</i>	Possible restrictions or reasons for embargo, tools for data use and reuse, attribution of unique and persistent identifiers...
<i>Data management responsibilities and resources</i>	Who is responsible for data stewardship, what resources will be dedicated to it...

# Conclusion

- Tensions paradoxales : collaboration, ouverture, gratuité contre système starifié, métriques et concurrence ; bien public mondial contre commodification et exploitation de la science et du travail des chercheur-es.
- Ouverture des données de recherche nécessaire pour faire face aux urgences socio-écologiques et à l'aggravation des inégalités.
- Mais sa mise en œuvre prend du temps, nécessite de renforcer les capacités collectives et de veiller à ne pas reproduire ou aggraver les inégalités dans le monde universitaire
- Ouverture des données de recherche n'est pas seulement une question de transparence, de reproductibilité ou de patrimonialisation, mais aussi de garantie d'une accessibilité inclusive et équitable aux connaissances.
- « Aussi ouvertes que possibles, aussi fermées que nécessaires »

# Questions?



# Operational conditions for opening research data: Some questions for journals (1/2)

- Few journals in our fields with open data policy  
*American Economic Review* has been a standard-setter in economics using Harvard Dataverse as a repository
- Questions we asked ourselves when developing Management's policy:
  - What kind of guidelines can a journal develop?
  - Should open data sets be reviewed and by whom? (not by us!)
  - Who should bear the cost of reviewing data? (not us!)
  - What should we do with data papers?

# Operational conditions for opening research data: Some questions for journals (2/2)

Roadmap for defining a data policy for journals (based on Feret et al. 2021, Hrynaszkiwicz et al. 2020, see also Marlet et al. 2022).

General topic	Issues and questions to address
<i>Definition of Research Data and exceptions</i>	What are the access and embargo conditions? Which data are concerned by the policy? At which stage of the process? Submission? Acceptance?
<i>Data and metadata standards and formats</i>	Which standards should be used and what are the meta-data protocols?
<i>Data access, hosting and publishing</i>	What are the protocols and guidelines for depositing data and choosing a repository?
<i>Data availability procedures</i>	What are the data availability procedures (timeframe, stages, etc.)? Is data peer-reviewed, and by whom? Is there a data availability statement?
<i>Data accessibility</i>	How will data be connected to the publication? Are permanent identifiers used?
<i>Noncompliance</i>	What happens when authors do not comply with the journal's data policy?
<i>Support for authors, reviewers and editors</i>	Does the journal or its academic association offer some form of support?

# Concluding thoughts

- Paradoxical tensions: collaborative, open, free vs. star-system, metrics and competition; Global public good vs profit-making
- We use the Creative Commons CC BY-NC 4.0 (noncommercial purposes)
- Opening research data needed to address socio-ecological emergencies and widening inequalities
- But implementing it takes time, requires building collective capacities and ensuring that we do not reproduce or aggravate inequalities in academia. Still firmly believe in “sustainable academia” as open, slow and engaged (Berkowitz & Delacour 2020)
- CARE Principles invite us to be more reflexive and careful about risks of “subordination of the researched” (Vijay 2021)
- Open research data is not just about transparency, reproducibility or patrimonialization, but also about ensuring inclusive and fair accessibility to knowledge