



### A panorama of self-deposit practices in European and North American archives

Paul Colin, Alina Danciu, Guillaume Garcia IASSIST AND CARTO 2024, Halifax, Nova Scotia

Creative Commons Attribution 4.0 International (CC BY 4.0)

## **Open science movement**

- In recent years, growing institutional incentives for the development of open science, particularly the opening of research data, have been strengthened in many countries (Horton and Katsanidou, 2018)
- New data repositories have been created to answer to this incentives
- Self-deposit practices not studied enough



#### Self-deposit practices, an emerging trend

- Very few studies tackled this subject head-on
- How are **curation practices** actually carried out?
- What are the **various costs** involved?
- What types of **guidelines** are provided to depositors?
- How are the archive's recommendations and instructions followed by depositors?
- To what extent is the **data re-use potential** taken into account when it comes to self-deposit?

## **Research method**



- Research based on 20 interviews (an in-depth questionnaire and/or guided interviews conducted remotely) conducted between 2021 and 2022, with CESSDA archives or SSH repositories referenced by the Dataverse network
- In the case of some of the repositories, we sent them a second round of questions after the initial interview/questionnaire to clarify certain questions

## **Theoretical sampling**

The idea was not to have a representative sample in the statistical sense of the term of the various categories of existing repositories, but to have a sample as diverse as possible of existing cases/situations.



## Interview questions



- Context in which the repository was created
- Technical, legal and practical aspects
- Deposit procedures & the challenges encountered by depositors and data curators themselves
- Data reuse potential
- Past and future **developments** of the infrastructure

#### **Repositories typology**

- Data centres, with a national or international focus : <u>AUSSDA</u>, <u>DANS</u>, <u>DataverseNO</u>, <u>Datice</u>, <u>DRYAD</u>, <u>FORS</u>, <u>FSD</u>, <u>GESIS</u>, <u>IQDA</u>,
- Institutional repositories at first, whose area of action is now national: <u>QDR</u>, <u>UNC DV</u>
- Library & research institutions networks (Johnston et al., 2017; Johnston et al., 2018), such as <u>Boréalis</u>
- Institutional repositories (Tillman, 2017; Lee, Stvilia, 2017): Edinburgh DataShare, the French CIRAD, INRAE, IRD
- Entities that act as or are backed up by national state archives : <u>SOHDA</u>, <u>RIGSARKIVET</u>

#### Social sciences & humanities or beyond

- Data centres tend to specialise in the social sciences (<u>GESIS</u>, <u>FSD</u>, <u>UNC DV</u>, <u>QDR</u>, etc.), and even extend their scope to the humanities (<u>AUSSDA</u>, <u>FORS</u>, <u>IQDA</u>, etc.)
- Others, however, are **open to all kinds of disciplines beyond the social sciences**, such as <u>DANS</u> or <u>DRYAD</u>.
- Institutional repositories tend to cover a wider range of disciplines those of their institution - whether STM (science, technology, medicine) or STEM (science, technology, engineering, and mathematics), as is the case with <u>Boréalis</u>, <u>IRD</u>, <u>CIRAD</u>, <u>INRAE</u>, <u>DataverseNO</u>, <u>Edinburgh DataShare</u>
- Repositories linked to national archives (<u>SOHDA</u>, <u>RIGSARKIVET</u>) are more in the SSH niche, although other disciplines may be covered from time to time.

# What are the general tendencies when it comes to self-deposit practices?

#### **Preliminary conclusions (1)**

- Very few data centres or repositories offer a "genuine" self-deposit service
- The most commonly adopted solution is to offer a support service with a relatively **high** level of intervention by data curators
- Qualitative data deposits, particularly ethnographic surveys, is rare, and most of the datasets deposited come from questionnaire surveys
- Recommendations and guidelines are not standardised:
  - some repositories propose very elaborate guidelines
  - others only propose very concise/straightforward deposit forms and guidelines, often relying on the expertise accumulated by the data curators that assist depositors

#### **Preliminary conclusions (2)**

- For the vast majority of the centres and repositories surveyed, there is a (significant) difficulty in achieving a high level of standardisation in self deposit/curation practices
- A human intervention is often required to perform a quality check of the deposit, in particular to check the quality of metadata and anonymisation aspects - without which datasets cannot be FAIR
- In rare cases, self-deposits stick to metadata publication, as the depositor has no intention to publish data (especially true in France, in the case of new repositories)

### How about data reuse?

- Most of the data processing is carried out with users in mind (Thomer et al., 2022)
- Integration of user needs into curation practices needs improvement, but there are not enough studies are carried out to collect user feedback, which is paradoxical

#### What strategy to adopt for quality checks?

- Set of **formal self-deposit rules**: opening selfdeposit only to partner research institutions and offering training before the actual self-deposit
- Developing an expert or non expert correspondents network in every partner institution (data ambassadors)
- Post-deposit control performed by data managers

#### What strategy to adopt in the future?

- Empirical examination of curators' work also shows what all curation professionals know: there is often a gap between what is planned in terms of curation, and what is actually done in practice
- Data curation and documentation activities are interdependent, and operate in a circular way (Perry, Netscher, 2022)
- Flexible process of curation and documentation and flexible guidelines for self-deposit? Or other strategies like a fee-based data curation or training service?

## Share your repository's self-deposit practices with us!

Data professionals invite researchers to give access to their research data and research work in general. This invitation can also be made to data professionals themselves! (Plantin, 2019).

## Get in touch!

alina.danciu@sciencespo.fr guillaume.garcia@sciencespo.fr

#### **Bibliography (1)**

- Andrea K. Thomer, Dharma Akmon, Jeremy York, Allison R. B. Tyler, Faye Polasek, Sara Lafia, Libby Hemphill, & Elizabeth Yakel. (2022). *The craft and coordination of data curation: complicating "workflow" views of data science*. https://arxiv.org/pdf/2202.04560.pdf
- Annika Valaranta. (2021, May 19). *How to guide anonymisation?* IASSIST CONFERENCE 2021. https://doi.org/https://doi.org/10.5281/zenodo.51810
- Antes, A. L., Walsh, H. A., Strait, M., Hudson-Vitale, C. R., & DuBois, J. M. (2018). Examining Data Repository Guidelines for Qualitative Data Sharing. *Journal of Empirical Research on Human Research Ethics: JERHRE*, *13*(1), 61–73. https://doi.org/10.1177/1556264617744121
- Bizeul, D. (2021). Should Everything be Exposed about an Investigation at the Front National? Reflections on Data Sharing and Ethical Duty in Sociology. Bulletin of Sociological Methodology/Bulletin de Méthodologie Sociologique, 150(1), 70-105. https://doi.org/10.1177/0759106320960887
- Borgman, C. L., Scharnhorst, A., & Golshan, M. S. (2019). Digital data archives as knowledge infrastructures: Mediating data sharing and reuse. *Journal of the Association for Information Science and Technology*, 70(8), 888–904. https://doi.org/10.1002/asi.24172
- Borgman, Christine L. Big Data, Little Data, No Data: Scholarship in the Networked World. First MIT Press paperback edition. Cambridge, Massachusetts London, England: The MIT Press, 2016.
- Bouchet, C. (2023). How I submitted my research data (not knowing what was awaiting me). Genèses, 133, 113-128. https://doi.org/10.3917/gen.133.0113

#### **Bibliography (2)**

- Demgenski, R. & Karcher, S. & Kirilova, D. & Weber, N., (2021) "Introducing the Qualitative Data Repository's Curation Handbook", Journal of eScience Librarianship 10(3): 8. doi: https://doi.org/10.7191/jeslib.2021.1207
- Denis, J. (2018). Le travail invisible des données: Éléments pour une sociologie des infrastructures scripturales. Presses des Mines. https://doi.org/10.4000/books.pressesmines.3934
- Fink, A.S., Kjeldgaard, A. S. F., Priisholm, S., & Jensen, B. G. (2005). Preservation of Knowledge- Data processing in the Danish Data Archives. IASSIST Quarterly, 27(4), 9. <u>https://doi.org/10.29173/iq865</u>
- Hayslett, M., & Jansen, M. (2022). Factors contributing to repository success in recruiting data deposits. *IASSIST Quarterly*, *46*(2). https://doi.org/10.29173/iq1037
- Horton, L., & Katsanidou, A. (2012). Purposing your survey: archives as a market regulator, or how can archives connect supply and demand? *IASSIST Quarterly*, *35*(4), 18. https://doi.org/10.29173/iq894
- Johnston, L. R., Carlson, J., Hudson-Vitale, C., Imker, H., Kozlowski, W., Olendorf, R., Stewart, C., Blake, M., Herndon, J., McGeary, T. M., & Hull, E. (2017). Data Curation Network: A Cross-Institutional Staffing Model for Curating Research Data. *International Journal of Digital Curation*, *13*(1), 125–140. https://doi.org/10.2218/ijdc.v13i1.616
- Kjeldgaard, A. S. F., Priisholm, S., & Jensen, B. G. (2005). Preservation of Knowledge- Data processing in the Danish Data Archives. *IASSIST Quarterly*, 27(4), 9. https://doi.org/10.29173/iq865
- Kudrnáčová, M., & Trtíková, I. (2020). Sustainability through the liaison with data archive users. *IASSIST Quarterly*, *44*(4). https://doi.org/10.29173/iq976
- Lafferty-Hess, S., & Christian, T.-M. (2017). More Data, Less Process? The Applicability of MPLP to Research Data. *IASSIST Quarterly*, *40*(4), 6. https://doi.org/10.29173/iq907

#### **Bibliography (3)**

- Lee, D. J., & Stvilia, B. (2017). Practices of research data curation in institutional repositories: A qualitative view from repository staff. *PLOS ONE*, *12*(3), e0173987. https://doi.org/10.1371/journal.pone.0173987
- Michele Hayslett, & Matt Jansen. (n.d.). Factors Affecting Deposits in Data Repositories. IASSIST CONFERENCE 2022.
- Perry, A., & Netscher, S. (2022). Measuring the time spent on data curation. *Journal of Documentation*, 78(7), 282–304. https://doi.org/10.1108/JD-08-2021-0167
- Plantin, J.-C. (2019). Data Cleaners for Pristine Datasets: Visibility and Invisibility of Data Processors in Social Science. Science, Technology, & Human Values, 44(1), 52-73. https://doi.org/10.1177/0162243918781268
- Poole, A. H. (2016). The conceptual landscape of digital curation. *Journal of Documentation*, 72(5), 961–986. https://doi.org/10.1108/JD-10-2015-0123
- Stewart, E., & Shaffer, M. (2021). The Ethical and Practical Challenges of Archiving Refugee Accounts: Reflections from Two Research Projects in the UK. Bulletin of Sociological Methodology/Bulletin de Méthodologie Sociologique, 150(1), 51-69. https://doi.org/10.1177/0759106321995707
- Thomer, A. K., Akmon, D., York, J., Tyler, A. R. B., Polasak, F., Lafia, S., Hemphill, L., & Yakel, E. (2022). *The craft and coordination of data curation: complicating "workflow" views of data science*. <u>https://doi.org/10.7302/4017</u>
- Tillman, R. K., (2017) "Where Are We Now? Survey on Rates of Faculty Self-Deposit in Institutional Repositories", Journal of Librarianship and Scholarly Communication 5(1), eP2203. doi: <u>https://doi.org/10.7710/2162-3309.2203</u>
- Yoon, A., & Tibbo, H. (2011). Examination of Data Deposit Practices in Repositories with the OAIS Model. IASSIST Quarterly, 35(4), 6. https://doi.org/10.29173/iq892

#### **Photo credits**

Slide 2:

https://media.istockphoto.com/id/1060814150/photo/elearning-book-as-laptop-electronic-bookconcept.jpg?s=1024x1024&w=is&k=20&c=8uOdOC2uFRIgFgconLa7JIOaO\_fL2ynrgb1MU-u4UXI=

Slide 4:

https://media.istockphoto.com/id/871450678/vector/technology-business-research-and-learning-and-people-business-team-workingconcept.jpg?s=612x612&w=0&k=20&c=R1PWk4BS8WIx7VC03BSzAxZyVykt321Z6P7NGWCZ444=

Slide 5:

https://media.istockphoto.com/id/1499696527/fr/vectoriel/programmation-et-ing%C3%A9nierie-informatique-informatique-illustration-vectorielle-r%C3%A9paration.jpg?s=612x612&w=0&k=20&c=ape8tso24cAAucz2Pw94aKz1WNriNQ5CWkHWFWEEBNA=

Slide 6:https://media.istockphoto.com/id/1277975670/fr/vectoriel/ic%C3%B4ne-de-ligne-plate-de-microphone-de-fixation-de-mainillustration-de-contour-de-vecteur.jpg?s=612x612&w=0&k=20&c=hkJSKfJOiyoaOWZSQmc2uRtJMCVssrst7UfhjtZEgNA=