

Findable and reusable?

Citation for published version (APA):

Gregory, K. (2021). *Findable and reusable? Data discovery practices in research*. Maastricht University. <https://doi.org/10.26481/dis.20210302kg>

Document status and date:

Published: 01/01/2021

DOI:

[10.26481/dis.20210302kg](https://doi.org/10.26481/dis.20210302kg)

Document Version:

Publisher's PDF, also known as Version of record

Please check the document version of this publication:

- A submitted manuscript is the version of the article upon submission and before peer-review. There can be important differences between the submitted version and the official published version of record. People interested in the research are advised to contact the author for the final version of the publication, or visit the DOI to the publisher's website.
- The final author version and the galley proof are versions of the publication after peer review.
- The final published version features the final layout of the paper including the volume, issue and page numbers.

[Link to publication](#)

General rights

Copyright and moral rights for the publications made accessible in the public portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

- Users may download and print one copy of any publication from the public portal for the purpose of private study or research.
- You may not further distribute the material or use it for any profit-making activity or commercial gain
- You may freely distribute the URL identifying the publication in the public portal.

If the publication is distributed under the terms of Article 25fa of the Dutch Copyright Act, indicated by the "Taverne" license above, please follow below link for the End User Agreement:

www.umlib.nl/taverne-license

Take down policy

If you believe that this document breaches copyright please contact us at:

repository@maastrichtuniversity.nl

providing details and we will investigate your claim.

9.3. Valorization statement

The synthesis and discussion presented in the previous section has the potential to inform the development of both systems and policy. There is also evidence that the work presented in this dissertation is already being taken up and applied within various areas of research. The papers presented in Chapters 3 through 5 have been cited within a variety of journals and conference proceedings within information science (i.e. JASIST, the Journal of Documentation and the ACM Joint Conference on Digital Libraries); computer science (i.e. the VLDB Journal, the ACM SIGIR conference, and the Web conference); and STS (i.e. Big Data & Society).³¹ It has also been referenced in outlets which cater to practitioners in library and information science, i.e. LIBER Quarterly and eLucidate, the journal of the UK electronic information group, as well as in doctoral dissertations and masters' theses. Citing authors are affiliated with institutions primarily across Europe and the United States, but also more globally, i.e. with institutions in Brazil and Malaysia.

The work has been cited in interdisciplinary (i.e. PLOS ONE) and discipline-specific journals (i.e. the Journal of Mechanical Design) in examinations of disciplinary practices of data discovery, sharing, analysis and management. It has also been referenced within the context of work focusing on systems development, data stewardship, linked open data and the development of FAIR metrics.

³¹ Analysis was conducted using GoogleScholar on 14 October, 2020. The articles stemming from this dissertation have accrued a total of 52 citations, excluding self-citations, during the period of 2017-2020.

Some of these citations are to pre-prints of the published articles. However, It is likely that some of the observed citations stem from collaborations formed during the PhD thesis.

This diversity reflects various engagements and collaborations developed over the course of four years. During this time, I have been an active member of the Data Discovery Paradigms Interest Group within the Research Data Alliance (Research Data Alliance, n.d.). Here, I was involved in a collaboration which resulted in a practical guide to help individuals locate data for research (Gregory, Khalsa, et al., 2018). This article, which has been downloaded over 2000 times, has since been integrated into online training material published by CESSDA, the Consortium for European Social Science Data Archives (CESSDA Training Team, 2020).

As a result of my research, I was invited to participate in various projects where data discovery was implicated, i.e. the Dagstuhl Perspectives Workshop on Implementing FAIR Data Infrastructures (Schloss Dagstuhl, 2018) and the GO FAIR implementation network focusing on semantic interoperability (GO Inter, n.d.). Building on this interest, a team of colleagues and I worked together to create a best practices guide for creating linked open data to enhance interoperability (Siebes et al., 2019), as part of the annual Mozilla-Library Carpentry Global Sprint series (Library Carpentry, 2019).

Citations and collaborations provide one way to think about the reach of this research; another is to consider how it has been integrated into the development of data search systems, particularly DataSearch, Elsevier's data search engine. I regularly presented the results of my research to the DataSearch team at internal project meetings as well as to Elsevier's broader research data management team. During these presentations, we discussed the recommendations summarized in Box 9.1, particularly suggestions to index diverse data supplies, to support social interactions and to rethink disciplinary classifications for data.

Since the beginning of this project in 2017, DataSearch has increased the diversity of data repositories which it indexes. It has also been integrated into an existing research data tool, Mendeley Data, and work is proceeding to link it with the Scopus literature database. Although considering how to support social interactions around data does not appear to be a high priority for development, assigning meaningful disciplinary classifications for data is a known problem that the design team is continuing to address. Although these changes align with some of my recommendations, it is difficult to say whether these developments were always planned, or whether my research directly influenced the design of the system.

This research used a theoretical framework developed, in part, to facilitate communication with systems designers. Although the recommendations made in Box 9.1 may not be realized in DataSearch (immediately), the communication of my findings to different

stakeholders has been made easier using this approach. It has also become apparent that systems development and design is itself a sociotechnical process. It requires time and adjustment and is shaped by factors such as changes in organizational structure, staffing and competition - all of which influenced the DataSearch team at some point in the project.

The course of my research was also impacted by larger changes in the world, as my final year of work was conducted primarily during the COVID-19 pandemic. Fortunately, I had finished collecting data before the pandemic began. However, speaking engagements and new collaborations, such as a planned research stay at the Institute for Quantitative Social Sciences at Harvard University, were postponed or cancelled.

The collaborations, both realized and potential, that this research has stimulated are one of its most valuable outcomes. This research has brought together individuals with diverse expertise, motivations and approaches to studying data practices and for designing practical solutions to support them. Despite current global challenges, I remain optimistic that such collaborations will continue to stem from and build on this research, providing the opportunity to further trace the relation of data discovery and reuse with other scholarly practices, technologies, and communities.