

This is an electronic reprint of the original article. This reprint may differ from the original in pagination and typographic detail.

Supporting open research data practice through data curation and discovery: A global perspective

Liu, Ying-Hsang; Chen, Hsin-Liang (Oliver); Kato, Makoto P.; Wu, Mingfang; Huvila, Isto

Published in:
Proceedings- Association for Information Science and Technology

DOI:
[10.1002/pr2.291](https://doi.org/10.1002/pr2.291)

Published: 22/10/2020

Document Version
Accepted author manuscript

Document License
Publisher rights policy

[Link to publication](#)

Please cite the original version:
Liu, Y.-H., Chen, H.-L., Kato, M. P., Wu, M., & Huvila, I. (2020). Supporting open research data practice through data curation and discovery: A global perspective. *Proceedings- Association for Information Science and Technology*, 57(1), e291. <https://doi.org/10.1002/pr2.291>

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.

Take down policy

If you believe that this document breaches copyright please contact us providing details, and we will remove access to the work immediately and investigate your claim.

This is the peer reviewed version of the following article:
Liu, Y.-H., Chen, H.-L. (Oliver), Kato, M. P., Wu, M., & Huvila, I. (2020). Supporting open research data practice through data curation and discovery: A global perspective. *Proc Assoc Inf Sci Technol*, 57(1), e291. which has been published in final form at doi.org/10.1002/pr2.291. This article may be used for non-commercial purposes in accordance with Wiley Terms and Conditions for Use of Self-Archived Versions.

Supporting Open Research Data Practice through Data Curation and Discovery: A Global Perspective

Ying-Hsang Liu¹, Hsin-Liang (Oliver) Chen², Makoto P. Kato³, Mingfang Wu⁴, & Isto Huvila⁵

¹ University of Southern Denmark, Denmark; ² Missouri University of Science and Technology, USA; ³ University of Tsukuba, Japan; ⁴ Australian Research Data Commons, Australia; ⁵ Uppsala University, Sweden

Abstract

This panel will address the issues associated with the practice and service of open research data curation and discovery from a global perspective. The sub-fields of information science such as information retrieval, information curation, information practices and human-centered data science have approached the open research data initiatives from multiple lenses. The issues of data creation, capturing, curation, sharing, discovery and reuse of cut across the sub-fields. We will identify and discuss the emerging themes in open data curation and discovery drawing on active research projects, repository practices and research data capturing and reuse in a selection of disciplines from health domain to archaeology and cultural heritage.

Introduction

The transparency of research through data sharing has received much more attention from researchers, funding agencies and general public recently. Research data originally locked in a cabinet in research institutes now can be accessible through institutional repositories and research data portals. Over the past decade, we have seen an increasing number of public and domain specific data repositories appear. For example, re3data.org, the Registry of Research Data Repositories, had 23 repositories when it went online in 2012, the number quickly increased to more than 2450 repositories from across the globe¹. While there are more and more data open and available through data repositories, it becomes challenging for researchers to find relevant data, especially when required data are from several repositories.

Some recent studies have explored the issues of research data curation and discovery services from the perspectives of researchers and data curation and discovery service providers (Faniel et al., 2019; Kacprzak et al., 2019; Walsh et al., 2019). Yet a relatively small part of the research data has been shared outside the research community; the roles played by the

¹ <https://blog.datacite.org/german-research-foundation-to-fund-new-services-of-re3data/>

different stakeholders and how to facilitate the practices of data sharing are still emerging (Cox et al., 2019; Khalsa et al., 2018; Polona, 2019; Wu et al., 2019).

This panel brings together information science researchers and practitioners internationally to push the boundaries of sub-fields in information science regarding open research data curation and discovery services. Panellists have established their research programs in user- and system-oriented information retrieval and data curation, as well as practical experiences in academic libraries and national data services. Specifically, this panel aims to address the issues of open research data services as follows:

- 1) What are the current practices of capturing, sharing and use of research data in various disciplines?
- 2) How can we enhance user search experience through studies of user behaviour?
- 3) How can we build robust information retrieval systems for open data?
- 4) What are the barriers to open research data curation and discovery services from institutional perspectives?

The 90-minute panel is composed of an overview of the research issues associated with open research data (10 minutes), followed by 15-minute presentation by four panellists addressing the proposed questions. Participants will have the opportunity to contribute to the discussions at the Q&A session.

Panellists and their Contributions

Ying-Hsang Liu, Ph.D. is associate professor at the Department of Design and Communication, University of Southern Denmark. His research program lies in the areas of information retrieval, information search, information interaction and information architecture. In the past few years, he has conducted collaborative research within these areas by user experiments using the techniques of eye tracking, think-aloud protocols and usability testing. He will provide an overview of the research issues regarding open research data services and moderate the panel.

Hsin-Liang (Oliver) Chen, Ph.D. is Dean of the Library, Missouri University of Science and Technology. His research interests focus on the application of information and communication technologies (ICTs) to assist users in accessing and using information in different environments. Dr. Chen and his collaborators are developing and evaluating a series of ontology-aware software prototype modules to produce phenotypic data that can be readily harvested for scientific computations. He will discuss the barriers to open research data services from institutional perspectives.

Makoto P. Kato, Ph.D. is associate professor at the Faculty of Library, Information and Media Science, University of Tsukuba in Japan. His research interests include information retrieval, web mining and machine learning and is concerned with information retrieval system design by eliciting information needs from users, using sensors, brain signals and mouse clicks for personalization. He is organizer of the NTCIR-15 Data Search (<https://ntcir.datasearch.jp>), a shared task on ad-hoc retrieval for governmental statistical data. He will discuss how to build robust information retrieval systems for open data.

Mingfang Wu, Ph.D. is senior research data specialist at the Australian Research Data Commons (ARDC). She has conducted research in the areas of question-answering,

interactive information retrieval, interfaces supporting exploratory search and enterprise search. Her recent research focuses on the data discovery paradigms as part of the Research Data Alliance (RDA) initiative and for improving data discovery service of Australian national research data catalogue. As an experienced information retrieval researcher and practitioner of data management, she will discuss the barriers to open research data services from institutional perspectives and how to support data discovery effectively.

Isto Huvila, Ph.D. is professor chair in information studies at the Department of ALM (Archival Studies, Library and Information Studies and Museums and Cultural Heritage Studies) at Uppsala University in Sweden. His primary areas of research include information and knowledge management, information work, knowledge organization, documentation, research data, and social and participatory information practices. Current research project investigates the creation, capturing, use and reuse of research data in the domain of archaeology. He will discuss on the basis of his recent and on-going evidence-based research can help to understand current practices of capturing, sharing and using research data in a selection of disciplines from health domain to archaeology and cultural heritage.

References

- Cox, A. M., Kennan, M. A., Lyon, L., Pinfield, S., & Sbaffi, L. (2019). Maturing research data services and the transformation of academic libraries. In *Journal of Documentation* (Vol. 75, Issue 6). <https://doi.org/10.1108/JD-12-2018-0211>
- Faniel, I. M., Frank, R. D., & Yakel, E. (2019). Context from the data reuser's point of view. *Journal of Documentation*, 75(6), 1274–1297. <https://doi.org/10.1108/JD-08-2018-0133>
- Kacprzak, E., Koesten, L., Ibáñez, L.-D., Blount, T., Tennison, J., & Simperl, E. (2019). Characterising dataset search: An analysis of search logs and data requests. *Journal of Web Semantics*, 55, 37–55. <https://doi.org/https://doi.org/10.1016/j.websem.2018.11.003>
- Khalsa, S., Cotroneo, P., & Wu, M. (2018). *A survey of current practices in data search services*. Mendeley. <https://doi.org/10.17632/7J43Z6N22Z.1>
- Polona, V. (2019). Research data management and research data literacy in Slovenian science. *Journal of Documentation*, 75(1), 24–43. <https://doi.org/10.1108/JD-03-2018-0042>
- Walsh, D., Clough, P., Hall, M. M., Hopfgartner, F., Foster, J., & Kontonatsios, G. (2019). Analysis of transaction logs from National Museums Liverpool. In A. Doucet, A. Isaac, K. Golub, T. Aalberg, & A. Jatowt (Eds.), *Digital libraries for open knowledge. TPDL 2019* (pp. 84–98). Springer International Publishing. https://doi.org/10.1007/978-3-030-30760-8_7
- Wu, M., Psomopoulos, F., Khalsa, S. J., & de Waard, A. (2019). Data discovery paradigms: User requirements and recommendations for data repositories. *Data Science Journal*, 18(1). <https://doi.org/10.5334/dsj-2019-003>