

How Data Papers Present a Unique Contribution to Open Research in the Humanities and Social Sciences

Mandy Wigdorowitz University of Cambridge

Barbara McGillivray King's College London

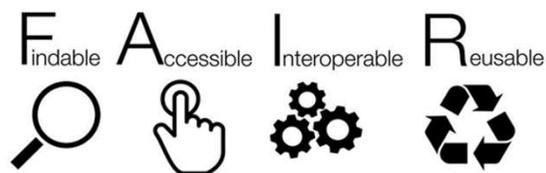
Marton Ribary Royal Holloway, University of London

Edinburgh Open Research Conference

16/05/2023

Content or **data** that is free to use, reuse, and redistribute

Open
research



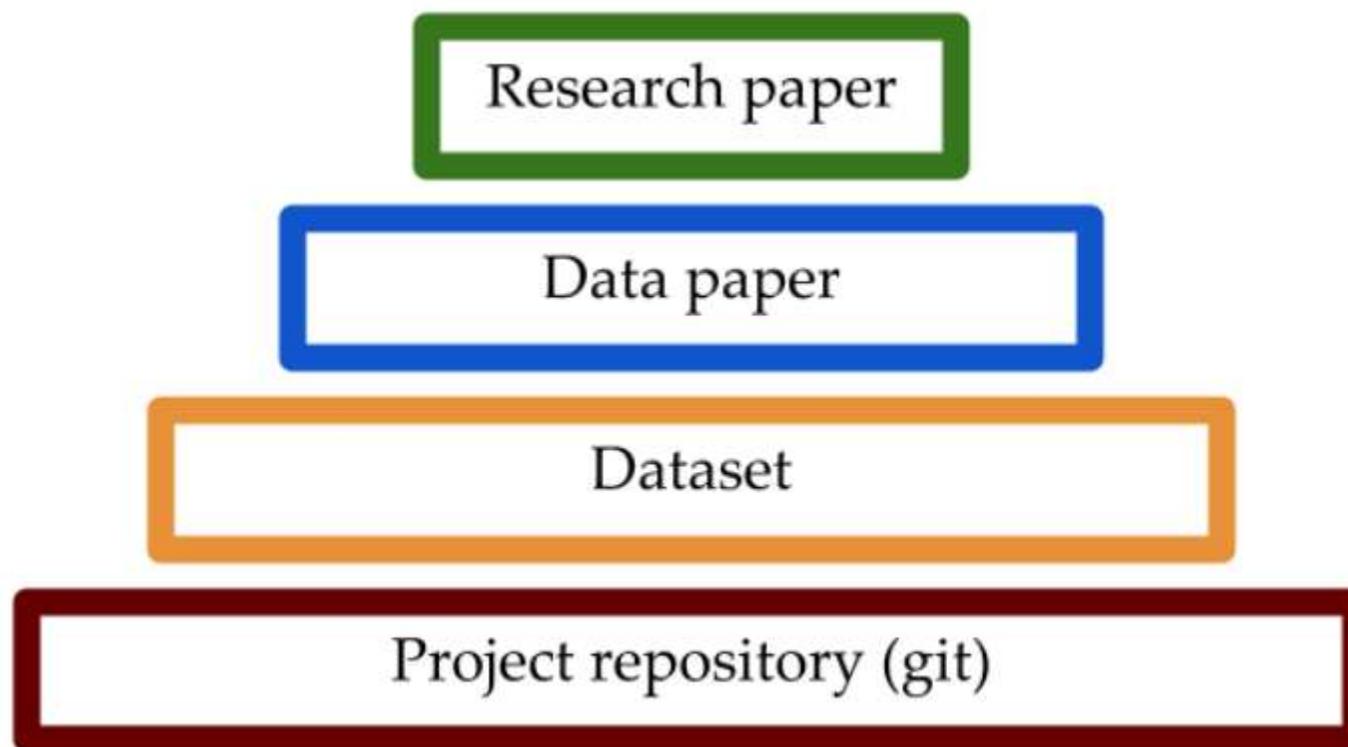
 **Plan S**
Making full & immediate
Open Access a reality

Open and shared data

- Largely driven by STEM and Health research
- Over recent years, data sharing practices have grown in HSS
 - increased availability of **digital collections**, the development of **new data-intensive methods**, an increasingly **solid infrastructure**, increased **pressure from funders**, the requirement of **data management plans** for preservation purposes, and the involvement of **research libraries in data curation**
 - how is data produced?
 - how it is openly and transparently shared?
 - is data shared in a reputable outlet?
 - how can data be reused?
 - what impact does data sharing generate?

The impact pyramid

Open research and dissemination structure of data-driven projects



adapted from Marongiu et al. (2022)

Data journals and data papers

- Avenue of implementing open research practices
- Incentivise the open sharing of resources according to best practices

Data paper

Research articles that describe one or more proficiently curated datasets

Data journal

Peer reviewed academic journals dedicated to the publication of data papers

Data journals in the Humanities and Social Sciences (HSS)

Data journals in HSS are slowly starting to grow but there are specific barriers

HSS encompass many disciplines compared to STEM and Health

Data in HSS exists in many forms



Data journals in HSS

Name of Journal	Year Since Publications	Field	Publications so Far	Publisher
(CODATA) Data Science Journal	2002 (relaunch 2014)	Science; Technology; Humanities; Arts	823	Ubiquity Press
Data in Brief	2014	All including Humanities	>6000	Elsevier
F1000Research	2012	All including Humanities	>5000	F1000 Research
Journal of Cognition	2017	Cognitive psychology	222	Ubiquity Press
Journal of Cultural Analytics	2016	Cultural analytics	109	Department of Languages, Literatures, and Cultures at McGill University
Journal of Open Archaeology Data	2012	Archaeology	57	Ubiquity Press
Journal of Open Humanities Data	2015	Humanities	67	Ubiquity Press
Journal of Open Psychology Data	2013	Psychology	48	Ubiquity Press
Research Data Journal for the Humanities and Social Sciences	2016	Humanities; Social Sciences	46	Brill

JOHD

- The **Journal of Open Humanities Data** features peer reviewed publications describing humanities research objects or techniques with high potential for reuse
- Open access
- Openly shared dataset in a repository with persistent identifier (DOI)
- Paper types
 - **Short data papers** - description of a humanities research object; short (1000 words) highly structured narratives
 - **Full length papers** - discuss and illustrate methods, challenges, and limitations in humanities research data creation, collection, management, access, processing, or analysis; longer narratives (3,000 - 5,000 words) which give authors the ability to contribute to a broader discussion regarding the creation of research objects or methods
- Social media outreach: @up_johd #johdnews #johdCfP #showmeyourdata

JOHD content

Collections

Collection

New perspectives and practices in Digital Humanities Data

We have created this special collection to celebrate 10 years of the Digital Humanities at Oxford Summer School, and to focus on the unique expertise of its community.

[View Articles](#)

Collection

Humanities Data in the time of COVID-19

In this special collection, we present data papers that span various areas of enquiry about the COVID-19 pandemic, through the lens of the Humanities. The wide range of data presented, from narrative to meta-data, has far-reaching and impactful reuse potential. It has been edited by Mandy Wigdorowitz and Sahba Besharati.

[View Articles](#)

Collection

Language Documentation Collections: Assessment and Recognition

This Special Collection aims to develop a detailed outline of what an effective peer-review process for documentary linguistic materials might look like, and how such a system would foster better recognition for these materials in academic evaluation systems such as in hiring, promotion, and tenure.

[View Articles](#)

Collection

Computational Humanities Research Data

Research in computational and quantitative approaches to humanities data is a fast growing interdisciplinary area. The first Computational Humanities Research workshop (CHR2020) took place online from 18 to 20 November 2020, organized by the DHLab of the KNAW Humanities Cluster in Amsterdam and The Alan Turing Institute

[View Articles](#)

	RESEARCH PAPERS Reception Reader: Exploring Text Reuse in Early Modern British Publications David Rosson, Eetu Mäkelä, Ville Vaara, Ananth Mahadevan, Yann Ryan, Mikko Tolonen	17 Apr 2023 9: 5
	RESEARCH PAPERS CREMMA Medii Aevi: Literary Manuscript Text Recognition in Latin Thibault Clérice, Malamatenia Vlachou-Efstathiou, Alix Chagué	12 Apr 2023 9: 4
	DATA PAPERS MultiHATHI: A Complete Collection of Multilingual Prose Fiction in the HathiTrust Digital Library Sil Hamilton, Andrew Piper	8 Feb 2023 9: 3
	DATA PAPERS A Global Lexical Database (GLED) for Computational Historical Linguistics Tiago Tresoldi	2 Feb 2023 9: 2
	RESEARCH PAPERS Critical Reflections on Cinema Belgica: The Database for New Cinema History in Belgium Vincent Ducatteeuw, Daniel Biltreyst, Philippe Meers, Christophe Verbruggen, Dries Moreels, Julia Noordegraaf, Sally Chambers, Pieterjan De Potter, Tamar Cachet, Nicolas Franck, Florian Deroo	13 Jan 2023 9: 1
	DATA PAPERS The TRANSCOMP Dataset of Literary Translations from 120 Languages and a Parallel Collection of English-language Originals Matt Erlin, Andrew Piper, Douglas Knox, Stephen Pentecost, Allie Blank	26 Dec 2022 8: 29
	DATA PAPERS Forming the <i>Corpus Criticorum</i> (1450-1640): Bibliography, Title Pages, Dataset Goran Gaber	14 Dec 2022 8: 28
	DATA PAPERS Oregon Theater Project: A Dataset of Oregon Cinemas from the Silent Era Michael Aronson, Elizabeth Peterson, Gabriele Hayden	12 Dec 2022 8: 27

Measuring and tracking data impact and reuse

Standard data citation practices

- Joint Declaration of Data Citation Principles (JDDCP; CODATA/ICTI)
- Scholix (Burton et al., 2017)
- Make Data Count (Cousijn et al., 2019)

Data citations counts can underestimate true reuse (Federer, 2020)



Altmetric (social media, news, blogs)



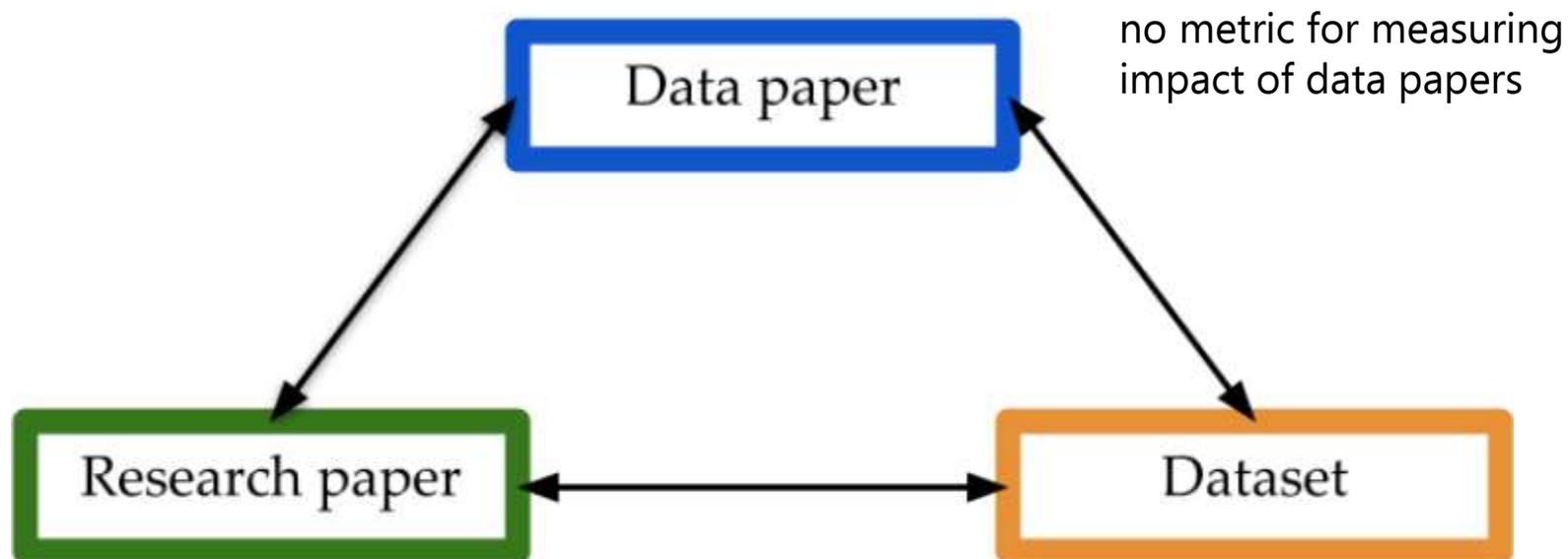
Altmetric + usage stats: necessary for data papers



HSS represent a particular challenge

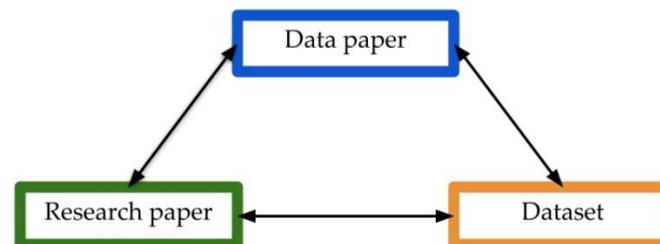
The impact triangle

- Move away from a **single metric** (citation of research paper)
- We should take the **three DOI-stamped elements of the pyramid** where the performance of each element affects the others



no standard practice to link elements of the triangle → relies on manual and semi-automatic linking

Evidence of data paper impact and future directions



Data papers have a positive impact on both the metrics of research papers associated with them and on data reuse (McGillivray et al., 2022)

- Find a programmatic way to identify these links by extracting information from available metadata of datasets and verifying their accuracy
- Create a “ground truth” in a manual and/or machine-assisted way which would enable the training of more sophisticated NLP-based methods

Conclusion

- Data papers present a unique contribution to addressing global challenges within the open research arena, and HSS specifically
- Data papers should be included in conversations about open research dissemination





Thank you

mw738@cam.ac.uk

[@Mandy_Wig](https://twitter.com/Mandy_Wig)