

*Data Papers in the  
context of Open  
Research, with a focus  
on the Humanities*

Barbara McGillivray

*Training webinar Data Notes on Open  
Research Europe: maximizing the visibility  
& discoverability of research data*

*25 October 2022*

Journal of **Open**  
**Humanities Data**

**]u[** ubiquity press  
open scholarship

**KING'S**  
*College*  
**LONDON**



---

# Open Humanities Data

---

**Open Humanities Data**

# The humanities have produced and collected data for centuries



This Photo by Unknown Author is licensed under [CC BY](#)

# What is different now?



# Many words, many objects, all data

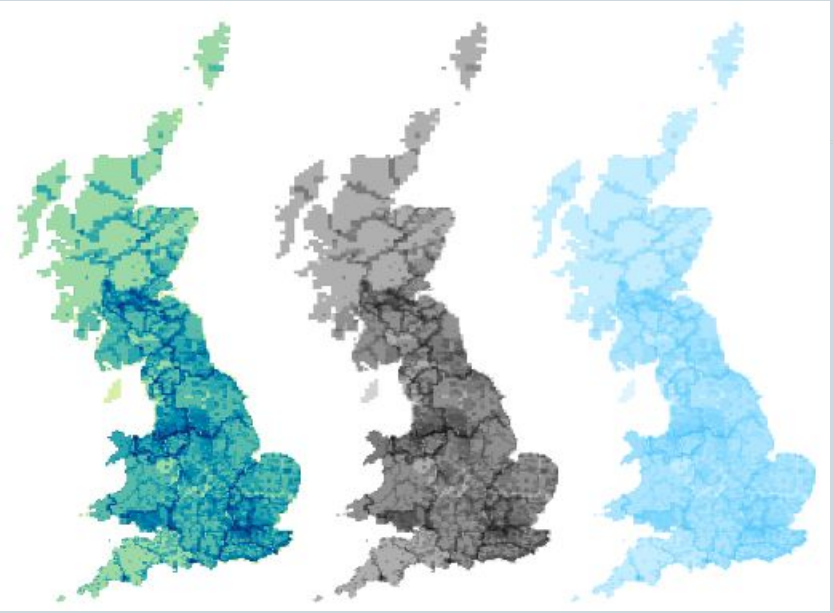
```

dvdrental=# select title, release_year, length, replacement_cost from film
dvdrental=# where length > 120 and replacement_cost > 29.50
dvdrental=# order by title desc;

```

title	release_year	length	replacement_cost
West Lion	2006	159	29.99
Virgin Daisy	2006	179	29.99
Uncut Suicides	2006	172	29.99
Tracy Cider	2006	142	29.99
Song Hedwig	2006	165	29.99
Slacker Liaisons	2006	179	29.99
Sassy Packer	2006	154	29.99
River Outlaw	2006	149	29.99
Right Cranes	2006	153	29.99
Quest Mussolini	2006	177	29.99
Poseidon Forever	2006	159	29.99
Loathing Legally	2006	140	29.99
Lawless Vision	2006	181	29.99
Jingle Sagebrush	2006	124	29.99
Jericho Mulan	2006	171	29.99
Japanese Run	2006	135	29.99
Gilmore Boiled	2006	163	29.99
Floats Garden	2006	145	29.99
Fantasia Park	2006	131	29.99
Extraordinary Conqueror	2006	122	29.99
Everyone Craft	2006	163	29.99
Dirty Ace	2006	147	29.99
Clyde Theory	2006	139	29.99
Clockwork Paradise	2006	143	29.99
Ballroom Mockingbird	2006	173	29.99

(25 rows)



- Concordance
- Word List
- Word Sketch
- Thesaurus
- Find X
- Sketch-Diff
- Sketch-Eval
- Corpus Info
- Save
- View options
- KWIC
- Sentence
- Sort
- Left
- Right
- Node
- References
- Shuffle
- Sample
- Filter
- Overlaps
- 1st hit in doc
- Frequency
- Node tags
- Node forms
- Doc IDs

Query colour 16,486 (147.0 per million)

Page 1 of 825 Go Next Last

**J2L** It would be tedious to list the types and colours of stone, ceramic etc. used at each site

**J2L** types of stone used for various shades of colour are predictable and limited in number.

**J2L** Birdcombe Avon. Here, sandstone furnished a buff colour, pennant stone a blue, liar the white for

**J2L** most mosaics comprise three to six basic colours, a work of good quality will include many

**J2L** therefore, to note ten or twelve different colours of tesserae in one pavement. In some, such

**J2L** the Woodchester Orpheus mosaic. <fp> 3.2 The colour of Tesserae <p> Sensitive use of shading

**J2L** 1976, 9). Elsewhere, intelligent use of colour is responsible for the blue shading which

**J2L** are notable. <fp><p> Whilst considering the colour of tesserae it is also pertinent to mention

**J2L** : 0.5 cm. sq. and 1.5 cm. sq. <fp><p> Like colour, the size of the tesserae affects the perspective

**J2L** fairly dark tesserae (deep red is a favourite colour), so producing a stronger" proximity effect

**J2L** panels (pl. 5b). At Leicester the rosettes - coloured (from the edges inwards) red, yellow and

**J2L** be cramped (although" loose"). There are colour contrasts however: the simple guilloche

**J2L** former. However, the-more subtle use of colour in the latter also produces a less contrived

**J2L** angular appearance. An overall poverty of colour, and the use of slightly larger (but still

**J2L** mosaic A). Although including the same basic colours, as well as tesserae of a similar size,

**J2L** blending of many tones of five or six basic colours, is notable in both designs. It is a sensitivity

**J2L** shows a generally consistent interface of colour, one in every four tongues of the latter

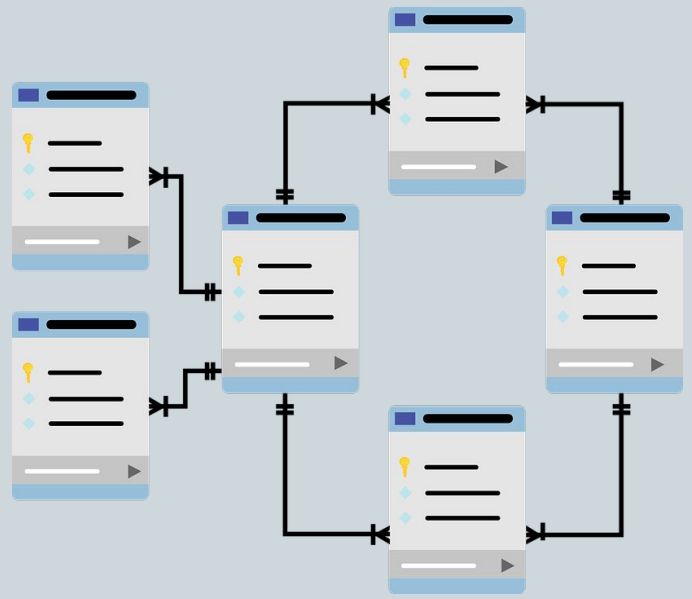
**J2L** Oceanus panel (contrast the confusion of colour around the heads of the lion and stag)

**J2L** However, on balance, the use here of similar colours (red, yellow, grey, pale-blue, brown) and

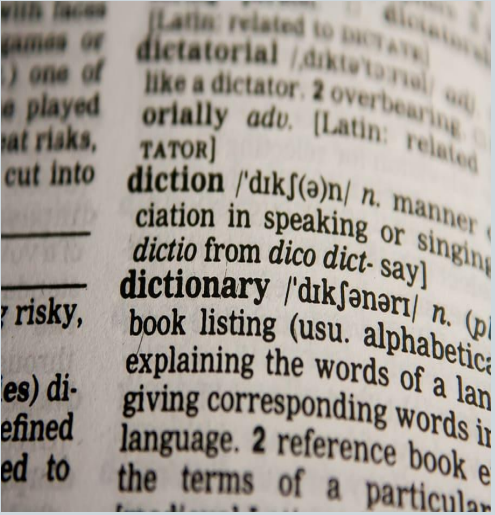
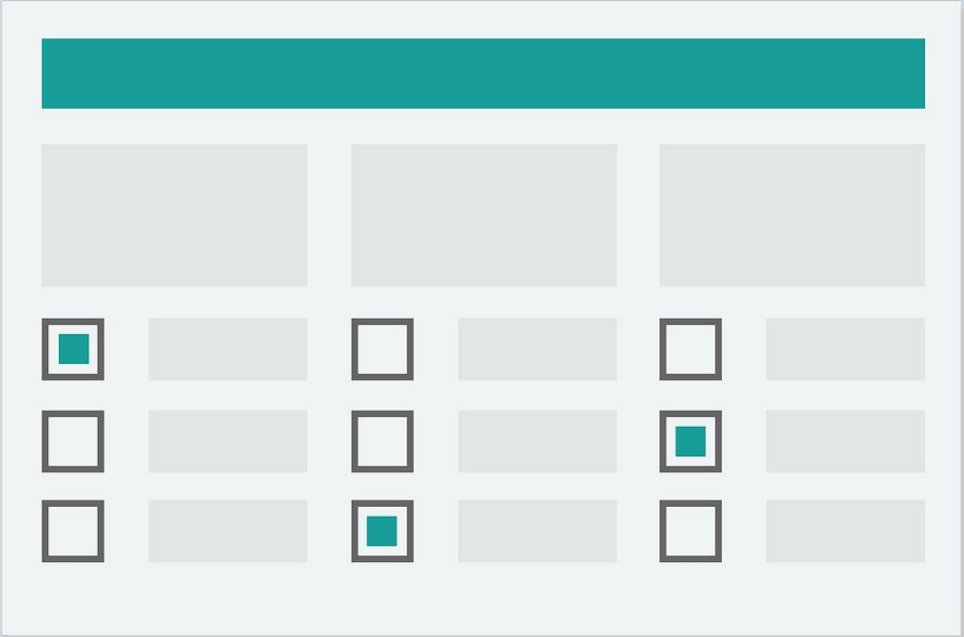
**J2L** Street mosaic, the presence there of a richly coloured figured panel (enclosed by a chain-guilloche

Page 1 of 825 Go Next Last

Lexical Computing



# Many words, many objects, all data



# What's special in humanities data?

- Manual curation
- Subjective/personal/individual interpretation, following standards/consensus
- Importance of context

Corpus of the Epigraphy of the Italian Peninsula in the 1<sup>st</sup> Millennium BCE (CEIPoM)

**Author:** Reuben J. Pitts 

## Abstract

The *Corpus of the Epigraphy of the Italian Peninsula in the 1st Millennium BCE* (CEIPoM) is a linguistic database which covers the Oscan, Umbrian, Old Sabellic, Messapic and Venetic languages, as well as epigraphic Latin up to 100 BCE. The database is hosted on GitHub and Zenodo, and provides manually annotated linguistic information on all levels of language structure, ranging from phonology to syntax. In providing a high-resolution digital dataset for language varieties that have until now been largely restricted to printed reference works, this corpus opens up new avenues for research into this unique ancient linguistic area.

**Keywords:** corpus linguistics, language contact, linguistic area, Italic, epigraphy

**How to Cite:** Pitts, R. J. (2022). Corpus of the Epigraphy of the Italian Peninsula in the 1<sup>st</sup> Millennium BCE (CEIPoM). *Journal of Open Humanities Data*, 8, 1. DOI: <http://doi.org/10.5334/johd.65>



---

**[[Open Humanities] Data]**

# The Open Definition

Availability and access

The **Open Definition** sets out principles that define “openness” in relation to **data and content**.

It makes **precise** the meaning of “open” in the terms “**open data**” and “**open content**” and thereby ensures **quality** and encourages **compatibility** between different pools of open material.

It can be summed up in the statement that:

Re-use and redistribution

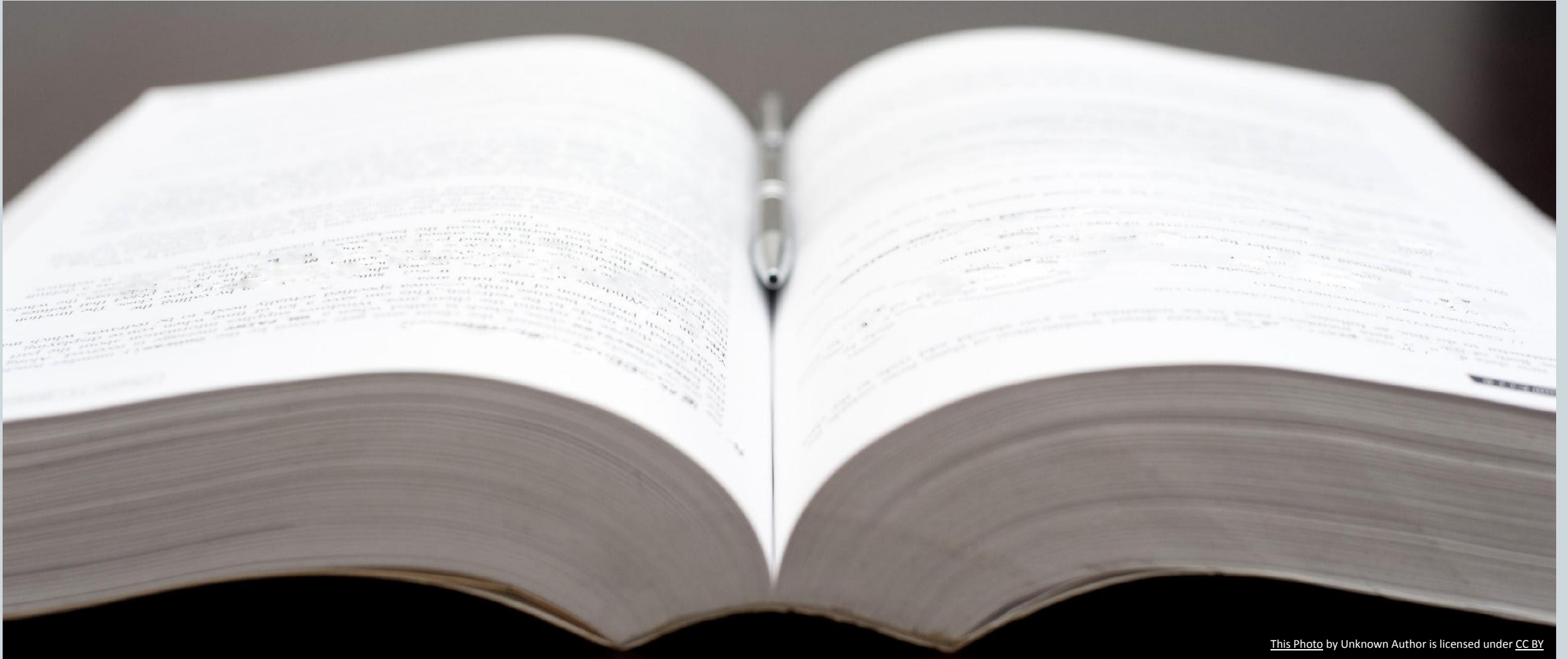
“Open means **anyone** can **freely access, use, modify, and share** for **any purpose** (subject, at most, to requirements that preserve provenance and openness).”

Put most succinctly:

Universal participation

“Open data and content can be **freely used, modified, and shared** by **anyone** for **any purpose**”

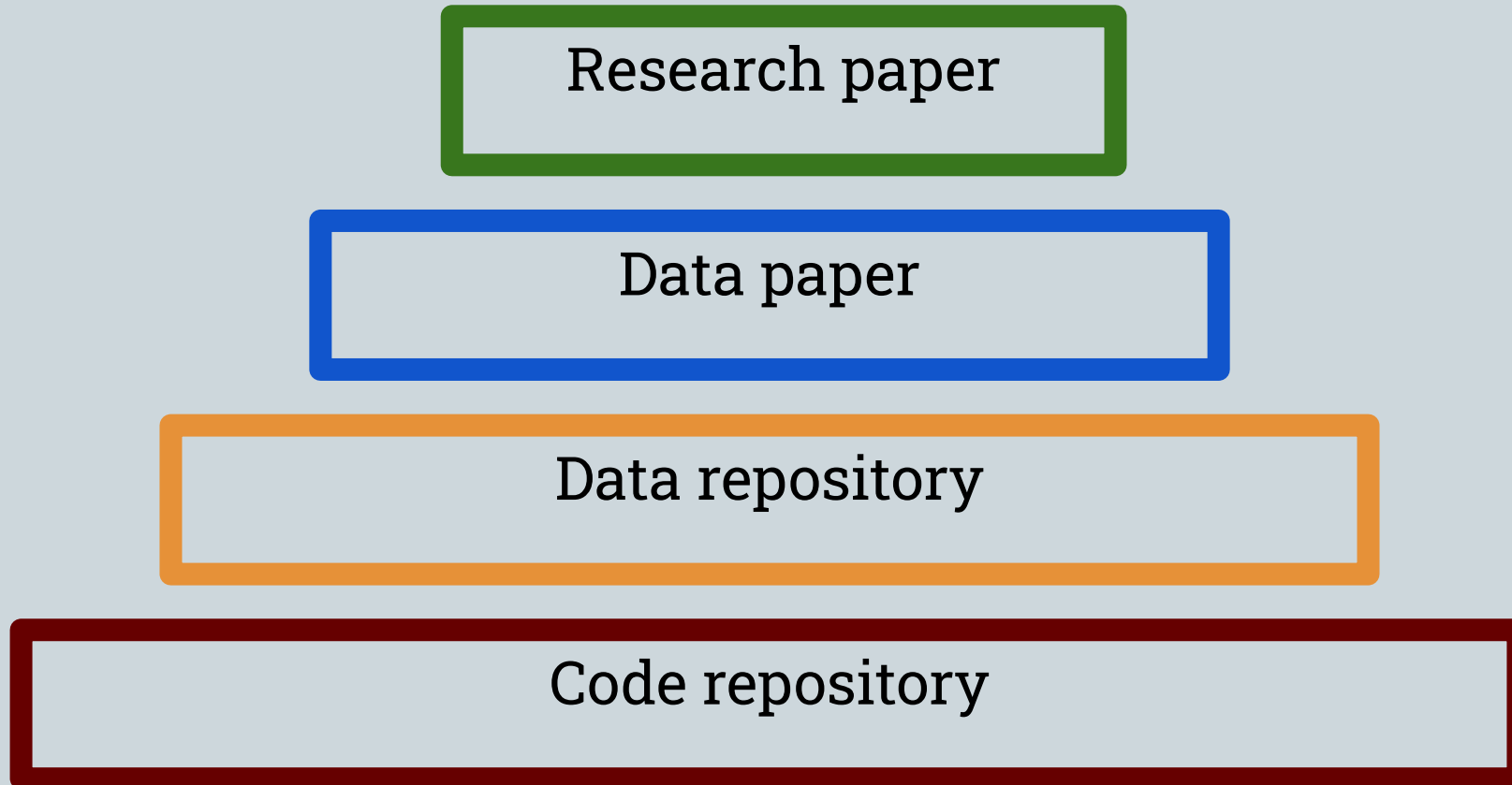
# « Extend the principles of openness to the whole research cycle »



This Photo by Unknown Author is licensed under [CC BY](#)

<https://www.fosteropenscience.eu/content/what-open-science-introduction>

# The open research pyramid

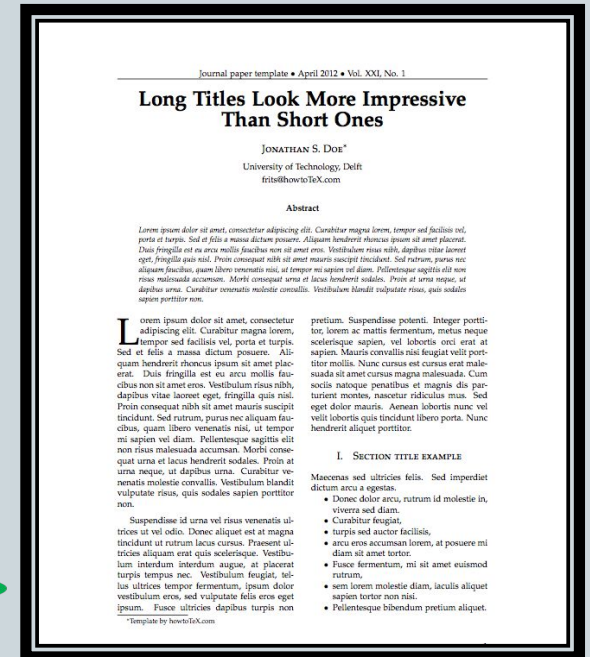
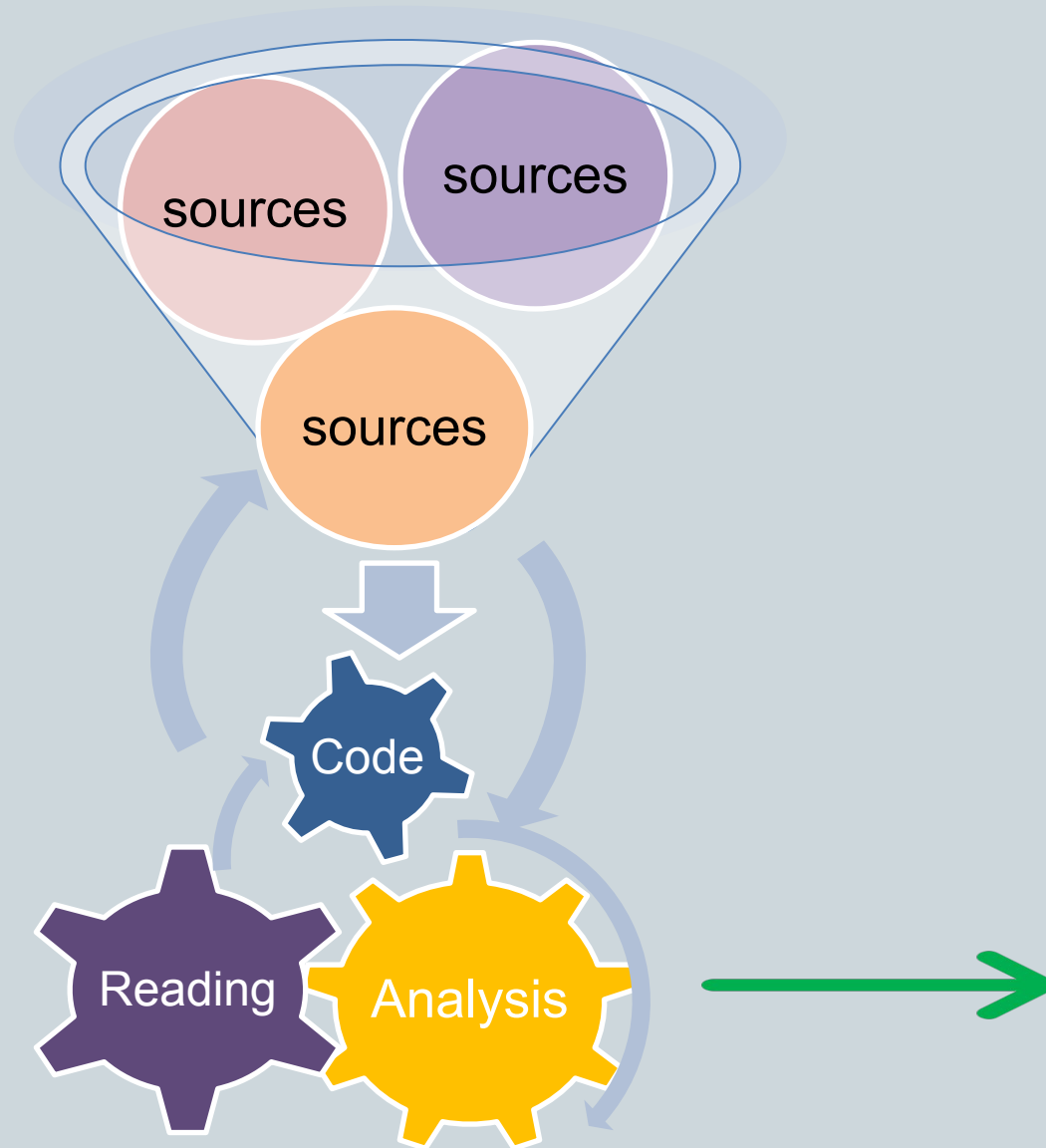


Goal: research is transparent, reproducible and impactful

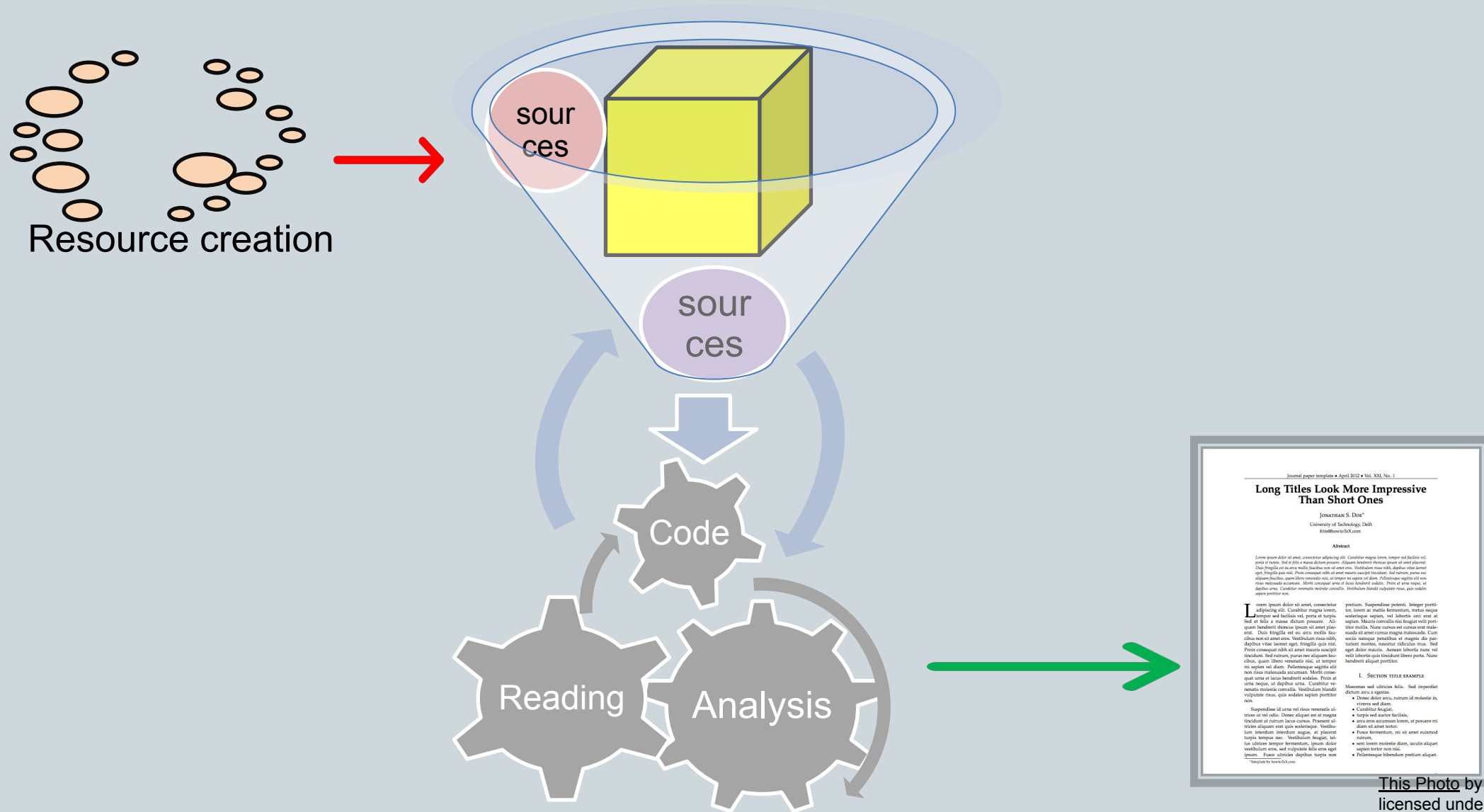
## Open Humanities and open science

“Open science commonly refers to efforts to make the output of publicly funded research more widely accessible in digital format to the scientific community, the business sector, or society more generally.”

# The traditional research process

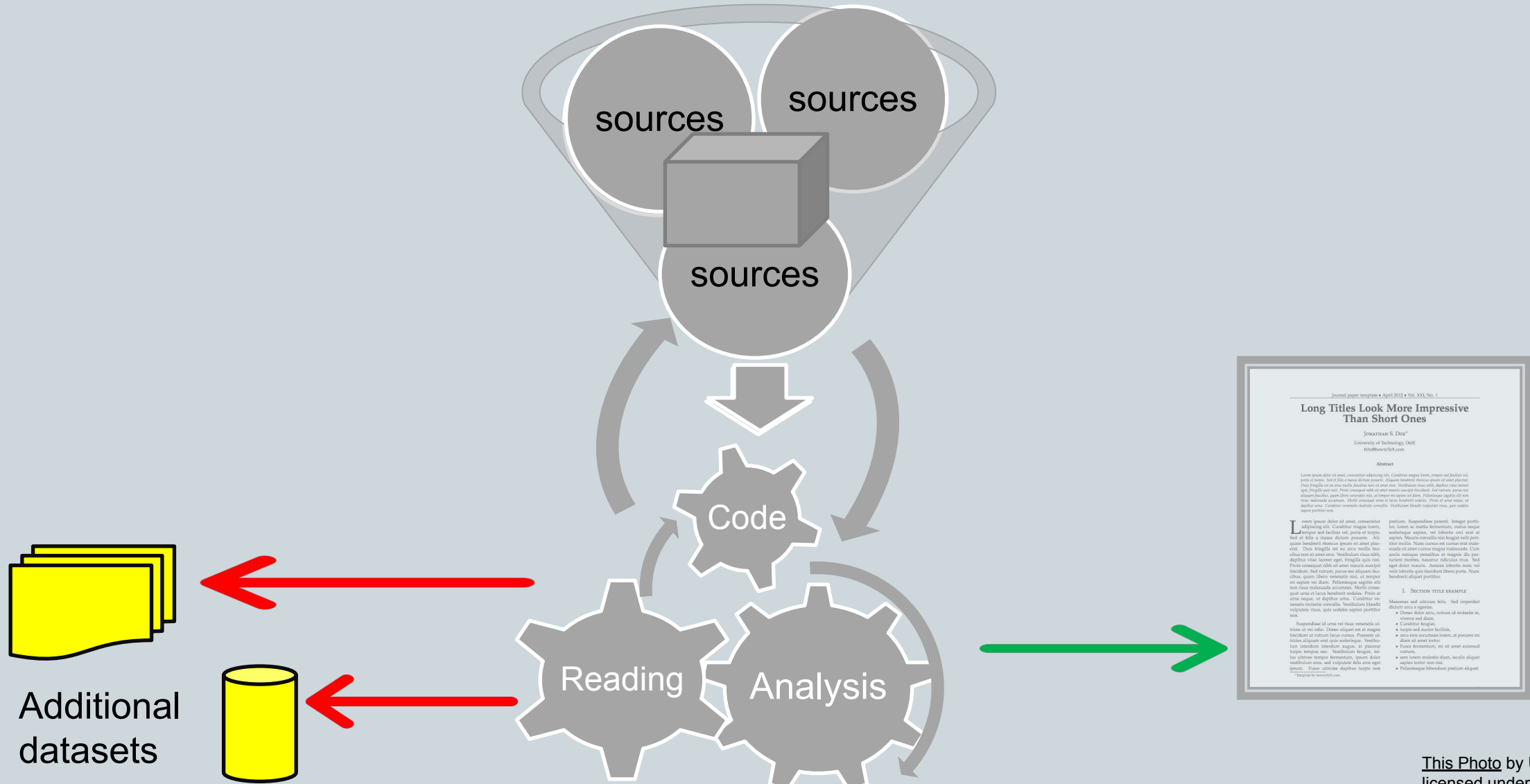


# What is left out?



This Photo by Unknown Author is licensed under CC BY-SA

# What more is left out?



This Photo by Unknown Author is licensed under CC BY-SA



# Contextualisation is critical

Data repository

The screenshot shows the Zenodo dataset page for the 'Corpus of the Epigraphy of the Italian Peninsula in the 1st Millennium BCE'. The page includes a search bar, navigation links for 'Upload' and 'Communities', and a 'Log in / Sign up' button. The dataset title is 'Corpus of the Epigraphy of the Italian Peninsula in the 1st Millennium BCE' by Reuben Pitts. It features a table with columns for Text\_ID, Sentence\_ID, Sentence\_position, Token\_ID, Token\_position, Analysis\_ID, Lemma, and another Lemma. The table contains 8 rows of data. To the right of the table, there are statistics: 143 views and 113 downloads. Below the statistics, it is noted that the dataset is indexed in OpenAIRE. Further down, the publication date is listed as April 21, 2022, with a DOI of 10.5281/zenodo.6475427. The keywords are 'Italic, linguistics, corpus'. The license is Creative Commons Attribution Share Alike 4.0 International. A 'Versions' section at the bottom shows version 1.3 from April 21, 2022.

Text_ID	Sentence_ID	Sentence_position	Token_ID	Token_position	Analysis_ID	Lemma	Lemma
1	1	1	161279	1	1468730	-	-
1	1	1	161280	2	1468809	-	-
2	2	1	165252	1	1467526	12970a	12970
2	2	1	165253	2	1468699	12444a	12444
2	2	1	165254	3	1469354	12495a	12495
2	2	1	165256	4	1467735	13184a	13184

Contextualisation + peer review + academic credit

Data paper

Corpus of the Epigraphy of the Italian Peninsula in the 1<sup>st</sup> Millennium BCE (CEIPoM)

**Author:** Reuben J. Pitts

## Abstract

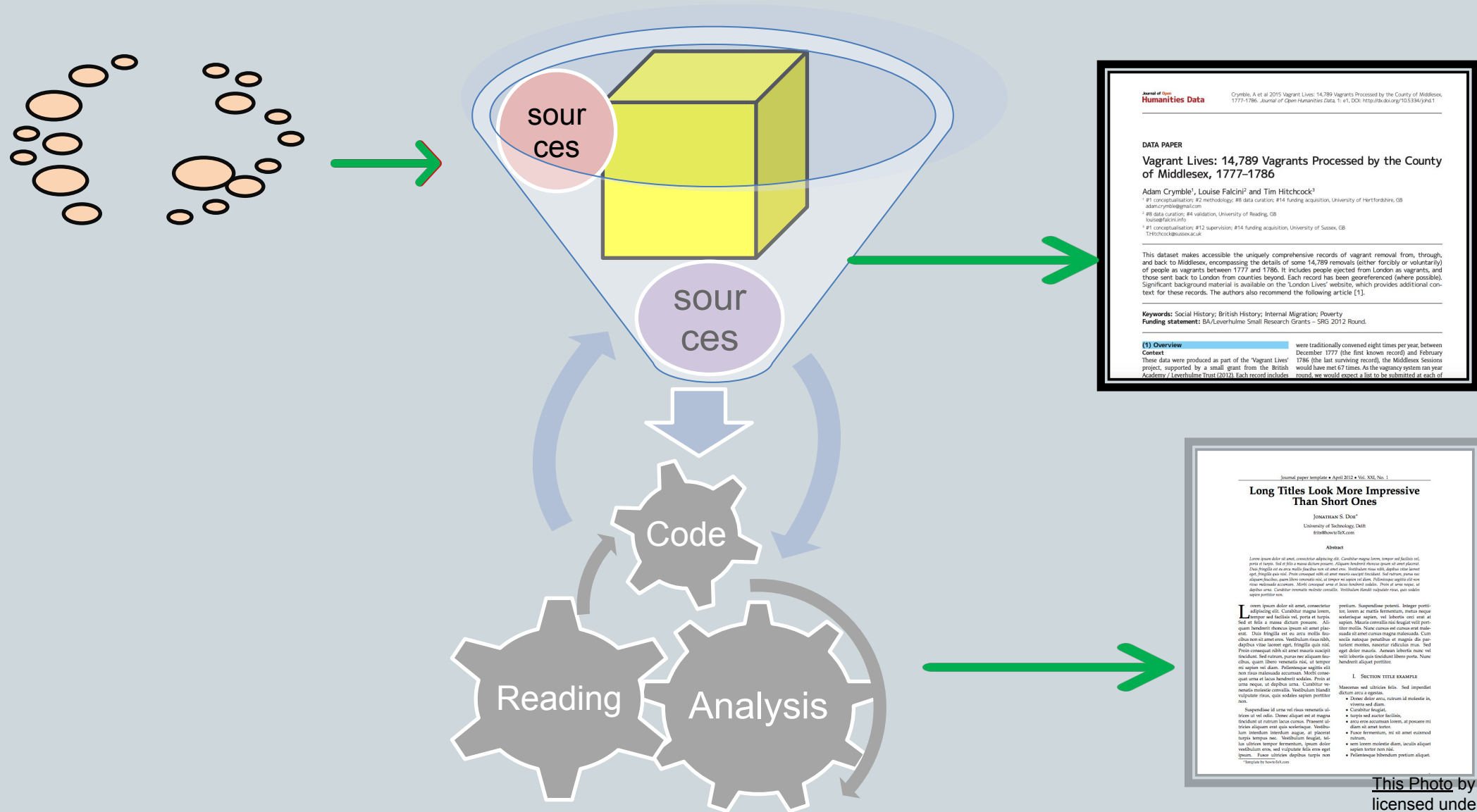
The *Corpus of the Epigraphy of the Italian Peninsula in the 1st Millennium BCE* (CEIPoM) is a linguistic database which covers the Oscan, Umbrian, Old Sabellic, Messapic and Venetic languages, as well as epigraphic Latin up to 100 BCE. The database is hosted on GitHub and Zenodo, and provides manually annotated linguistic information on all levels of language structure, ranging from phonology to syntax. In providing a high-resolution digital dataset for language varieties that have until now been largely restricted to printed reference works, this corpus opens up new avenues for research into this unique ancient linguistic area.

**Keywords:** corpus linguistics, language contact, linguistic area, Italic, epigraphy

**How to Cite:** Pitts, R. J. (2022). Corpus of the Epigraphy of the Italian Peninsula in the 1<sup>st</sup> Millennium BCE (CEIPoM). *Journal of Open Humanities Data*, 8, 1. DOI:

<http://doi.org/10.5334/johd.65>

# Crediting resource creation



The Data Paper

# Promoting reusable data and reproducibility

**Journal of Open Humanities Data**  
 Crymble, A. et al. 2015 Vagrant Lives: 14,789 Vagrants Processed by the County of Middlesex, 1777-1786. *Journal of Open Humanities Data*, 1, #1, DOI: <http://dx.doi.org/10.5334/johd.1>

**DATA PAPER**  
**Vagrant Lives: 14,789 Vagrants Processed by the County of Middlesex, 1777-1786**

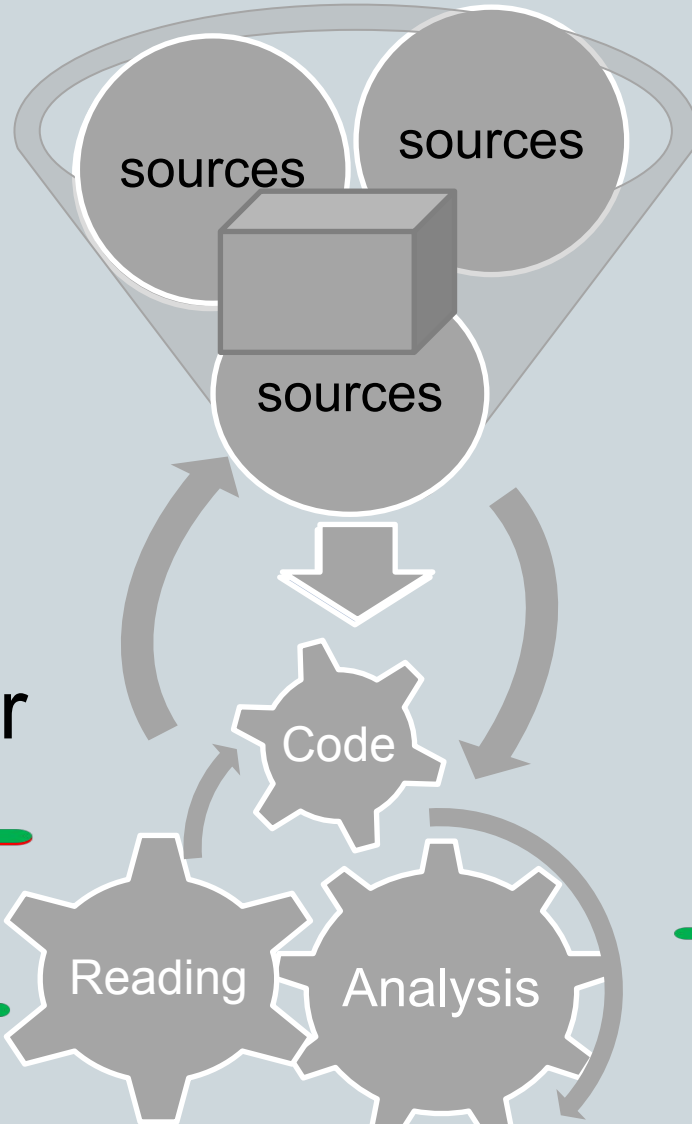
Adam Crymble<sup>1</sup>, Louise Falcin<sup>2</sup> and Tim Hitchcock<sup>3</sup>  
<sup>1</sup> #1 conceptualisation; #2 methodology; #8 data curation; #14 funding acquisition, University of Hertfordshire, GB  
 adamcrymble@gmail.com  
<sup>2</sup> #8 data curation; #4 validation, University of Reading, GB  
 louise@falcin.info  
<sup>3</sup> #1 conceptualisation; #12 supervision; #14 funding acquisition, University of Sussex, GB  
 T.Hitchcock@sussex.ac.uk

This dataset makes accessible the uniquely comprehensive records of vagrant removal from, through, and back to Middlesex, encompassing the details of some 14,789 removals (either forcibly or voluntarily) of people as vagrants between 1777 and 1786. It includes people ejected from London as vagrants, and those sent back to London from counties beyond. Each record has been georeferenced (where possible). Significant background material is available on the 'London Lives' website, which provides additional context for these records. The authors also recommend the following article [1].

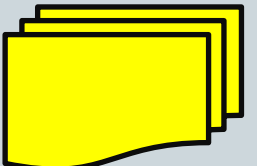
**Keywords:** Social History; British History; Internal Migration; Poverty  
**Funding statement:** BA/Leverhulme Small Research Grants - SRG 2012 Round.

**(1) Overview**  
 Context: These data were produced as part of the 'Vagrant Lives' project, supported by a small grant from the British Academy's Leverhulme Trust (L1017). Each record includes

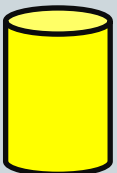
were traditionally convened eight times per year, between December 1777 (the first known record) and February 1786 (the last surviving record), the Middlesex Sessions would have met 67 times. As the vagrancy system ran year round, we would expect a list to be submitted at each of



The Data Paper



Additional datasets



*Journal paper template* • April 2012 • Vol. XXI, No. 1

### Long Titles Look More Impressive Than Short Ones

JONATHAN S. DOE\*  
 University of Technology, Dept  
 info@university.com

**Abstract**

*Lorem ipsum dolor sit amet, consectetur adipiscing elit. Curabitur magna lorem, tempus sed facilisis nisi, porta et turpis. Sed et sit a massa dictum pretium. Aliquam hendrerit rhoncus ut ipsum sit amet placerat. Donec fringilla est in porta nulla. Vestibulum non nisi, dapibus vitae diam. Ut, fringilla quis nisi. Proin consequat nibh sit amet mauris. Sed ipsum, posuere nec aliquam. Nullam, quam libero venenatis nisi, ut tempus non dignum. Proin ut urna, ut dapibus enim. Curabitur venenatis venenatis venenatis. Vestibulum hendrerit rhoncus, enim sed nisi, sapien porttitor nisi.*

*Inter ipsum dolor sit amet, consectetur adipiscing elit. Curabitur magna lorem, tempus sed facilisis nisi, porta et turpis. Sed et sit a massa dictum pretium. Aliquam hendrerit rhoncus ut ipsum sit amet placerat. Donec fringilla est in porta nulla. Vestibulum non nisi, dapibus vitae diam. Ut, fringilla quis nisi. Proin consequat nibh sit amet mauris. Sed ipsum, posuere nec aliquam. Nullam, quam libero venenatis nisi, ut tempus non dignum. Proin ut urna, ut dapibus enim. Curabitur venenatis venenatis venenatis. Vestibulum hendrerit rhoncus, enim sed nisi, sapien porttitor nisi.*

**1. SECTION TITLE EXAMPLE**

*Maecenas sed ultricies felis. Sed imperdiet dictum arcu a egetis.*

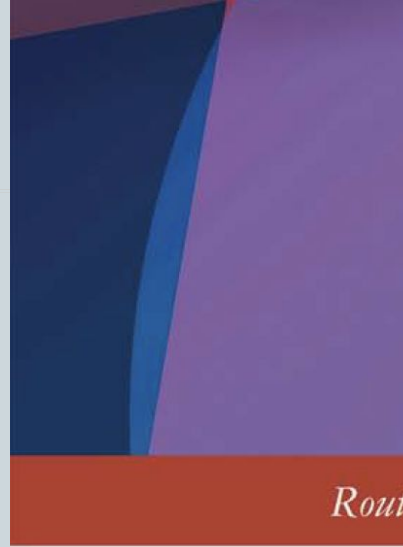
- Donec dolor arcu, curam id molestie in, viverra sed diam.
- Curabitur magna.
- Turpis sed auctor facilisis.
- Arcu enim accumsan lobortis, ac posuere nisi diam sit amet turpis.
- Fusce fermentum, nisi sit amet euismod rutrum.
- Nullam ipsum molestie diam, tunc sit amet sapien porttitor non nisi.
- Pellentesque nibhulum pretium aliquet.

\*Template by [www.latexstudio.net](http://www.latexstudio.net)

This Photo by Unknown Author is licensed under CC BY-SA

# An example from linguistics

Jenset, G B, McGillivray, B and Rundell, M. The English dative alternation revisited: Fresh insights from contemporary British spoken data. In: Brezina, V, Love, R and Aijmer, K (eds.), *Corpus approaches to contemporary British speech: Sociolinguistic studies of the spoken BNC2014*. London: Routledge; 2018; 185–207.  
DOI: <https://doi.org/10.4324/9781315268323-10>



**CORPUS  
TO  
BOLD**

Vaclav Brezina

## 10 The Dative Alternation Revisited

Fresh Insights From Contemporary British Spoken Data

*Gard B. Jenset, Barbara McGillivray and Michael Rundell*

### 10.1 Motivation

A well-known feature of English grammar is the dative alternation, whereby a verb may be used in a V-NP-NP construction (*Give me the money*) or with a prepositional phrase in the pattern V-NP-PP, typically with the preposition *to* (*Give the money to me*). In this study, we use data from the early access Sample of the Spoken British National Corpus 2014 (the Spoken BNC2014S) to investigate the behaviour of six high-frequency verbs whose argument structure preferences include the dative alternation. Given that speakers have both patterns available to them, our goal is to discover whether the choice of pattern is motivated rather than arbitrary—and if so, which factors influence that choice.

Although the dative alternation is a well-researched topic, most published work draws either on introspection or on data from written sources. Using contemporary unscripted spoken text from face-to-face conversations takes us into new territory, especially as the linguistic data in the Spoken BNC2014S are complemented by rich sociolinguistic information on participating speakers. By “sociolinguistic information” we mean the social phenomena that co-occur with linguistic variables (Bayley, 2002, p. 118). The corpus represents a powerful new research resource, and in this chapter we show how it yields new insights into the use of the dative alternation.



File info

BNCspoken2014\_dative\_dataset\_v1.csv (839.47 kB)

MD5 code:  
1c041dbcb635855f3d8f4be9f3e1fced

[download](#)

2	give	A9	VNN	you	3	you	TRUE	Z8	TRUE	a list
3	give	A9	VNN	you	3	you	TRUE	Z8	TRUE	a list of special offers
4	give	A9	VNN	you	3	you	TRUE	Z8	TRUE	a list of different types
5	give	A9	VNN	you	3	you	TRUE	Z8	TRUE	a few like oh these
6	give	A9	VNN	them	4	them	TRUE	Z8	FALSE	exactly what you want
7	give	A9	VNN	them	4	them	TRUE	Z8	FALSE	the makeup
8	give	A9	VNN	him	3	him	TRUE	Z8	TRUE	the choice of which or
9	give	A9	VNN	me	2	me	TRUE	Z8	TRUE	experience
10	give	A9	VNN	you	3	you	TRUE	Z8	TRUE	half an hour of being a
11	give	A9	VNN	me	2	me	TRUE	Z8	TRUE	all sorts of crap
12	give	A9	VNN	me	2	me	TRUE	Z8	TRUE	all sorts of crap
13	give	A9	VNN	them	4	them	TRUE	Z8	FALSE	the agency number

# ... and the data paper

Jenset, G.B. and McGillivray, B., 2019. Data from 'The Dative Alternation Revisited: Fresh Insights from Contemporary British Spoken Data'. *Journal of Open Humanities Data*, 5(1), p.1. DOI: <http://doi.org/10.5334/johd.11>

## License

CC BY 4.0

## Repository name

Figshare

## Publication date

2018-11-16

## (4) Reuse potential

There is a growing trend in linguistics for quantitative research, a trend which is not proceeding at the same pace in all branches of linguistics [6]. A natural corollary of this increasing quantitative research is a focus on replicable and reproducible research [7].

True replicability is difficult to achieve in many field-based disciplines and social sciences [7]. A more achievable goal is reproducibility. Reproducibility is clearly important for increasing scientific transparency and accountability. A move towards greater reliance on usage-based theory development can drive convergence in linguistic theory generally [8] as well as in specific sub-fields [6]. Despite some notable exceptions (such as second language acquisition), most linguistic sub-fields do not have a strong tradition for making research data available [7]. Publishing not only corpora and raw data, but also the annotated research datasets means that data can be compared quantitatively across

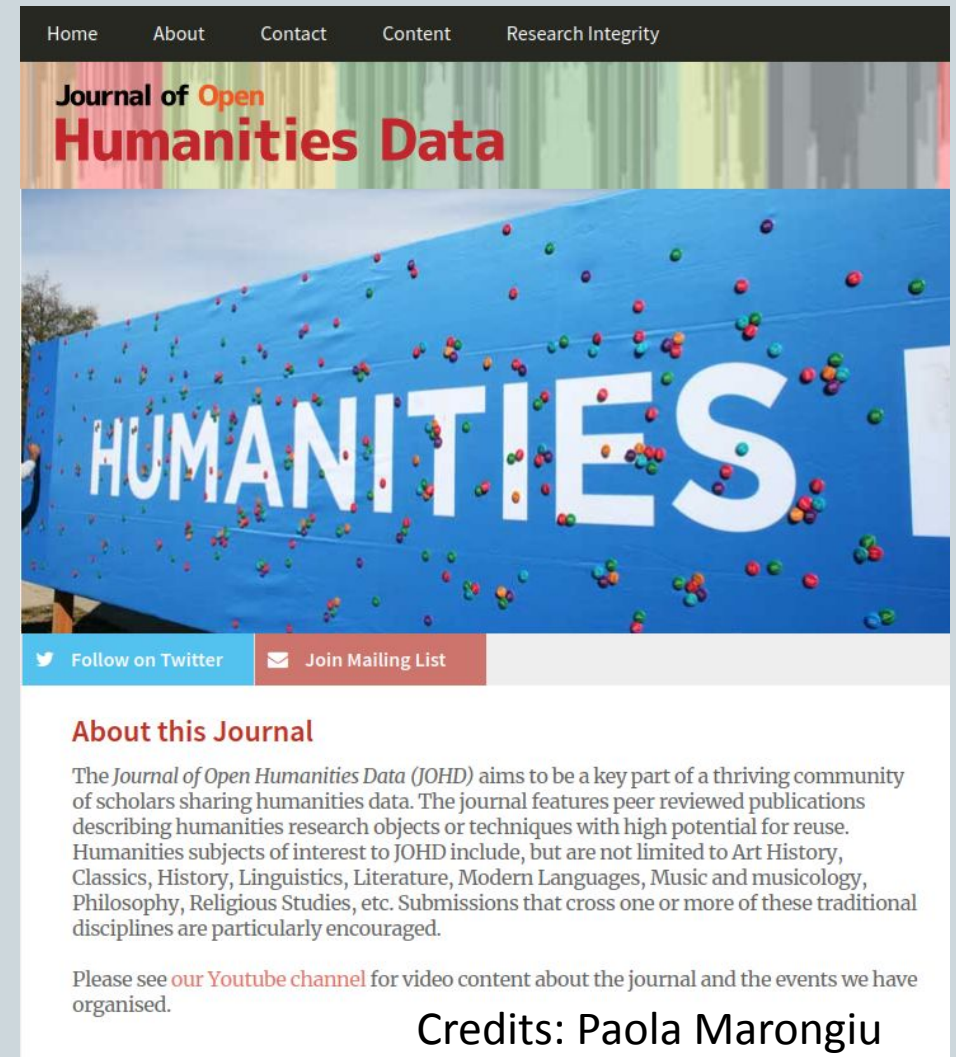
---

Journal of

**Open Humanities Data**

# The *Journal of Open Humanities Data*

- Launched in 2015
- It is part of the ‘*metajournals*’ family in Ubiquity Press. They publish papers about:
  - research data
  - software
  - hardware



Home About Contact Content Research Integrity

Journal of **Open**  
**Humanities Data**

HUMANITIES

Follow on Twitter Join Mailing List

### About this Journal

The *Journal of Open Humanities Data (JOHD)* aims to be a key part of a thriving community of scholars sharing humanities data. The journal features peer reviewed publications describing humanities research objects or techniques with high potential for reuse. Humanities subjects of interest to JOHD include, but are not limited to Art History, Classics, History, Linguistics, Literature, Modern Languages, Music and musicology, Philosophy, Religious Studies, etc. Submissions that cross one or more of these traditional disciplines are particularly encouraged.

Please see [our Youtube channel](#) for video content about the journal and the events we have organised.

Credits: Paola Marongiu



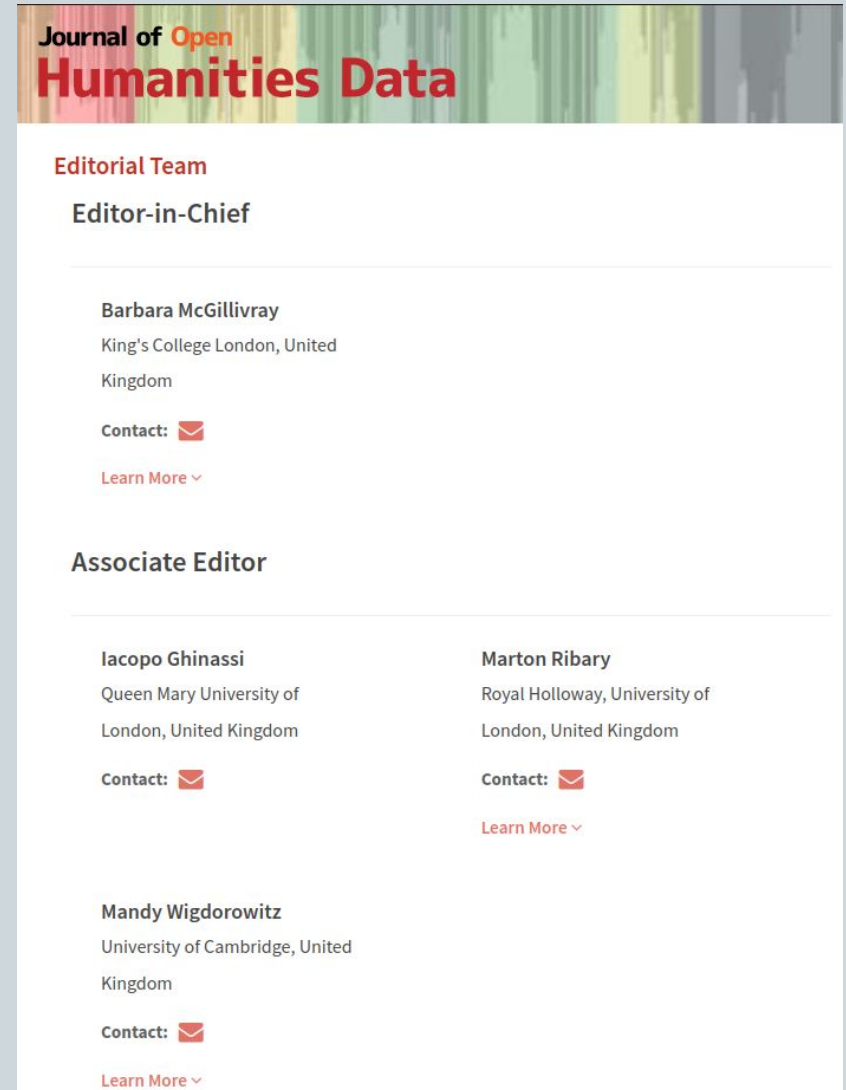
## Vision

Be a **key** part of a **thriving community** of  
scholars **sharing humanities data**

# JOHD's team

- 1 editor-in-chief
- 3 associated editor
- 2 editorial assistants
- 7 copy editors
- 9 members in the editorial board
- 2 social media editors

→ We are researchers in linguistics, classics, law and more





The screenshot shows the 'Journal of Open Humanities Data' website. The header features the journal title in red and black text against a colorful, abstract background. Below the header, the 'Editorial Team' section is displayed. It is divided into three sub-sections: 'Editor-in-Chief', 'Associate Editor', and 'Associate Editor'. Each sub-section lists the name of the editor, their affiliation, and a contact icon (an envelope). There are also 'Learn More' links with a downward arrow next to each editor's name.


**Journal of Open Humanities Data**



**Editorial Team**



**Editor-in-Chief**

Barbara McGillivray  
King's College London, United Kingdom  
Contact:   
[Learn More](#) 

**Associate Editor**

Iacopo Ghinassi  
Queen Mary University of London, United Kingdom  
Contact: 

Marton Ribary  
Royal Holloway, University of London, United Kingdom  
Contact:   
[Learn More](#) 

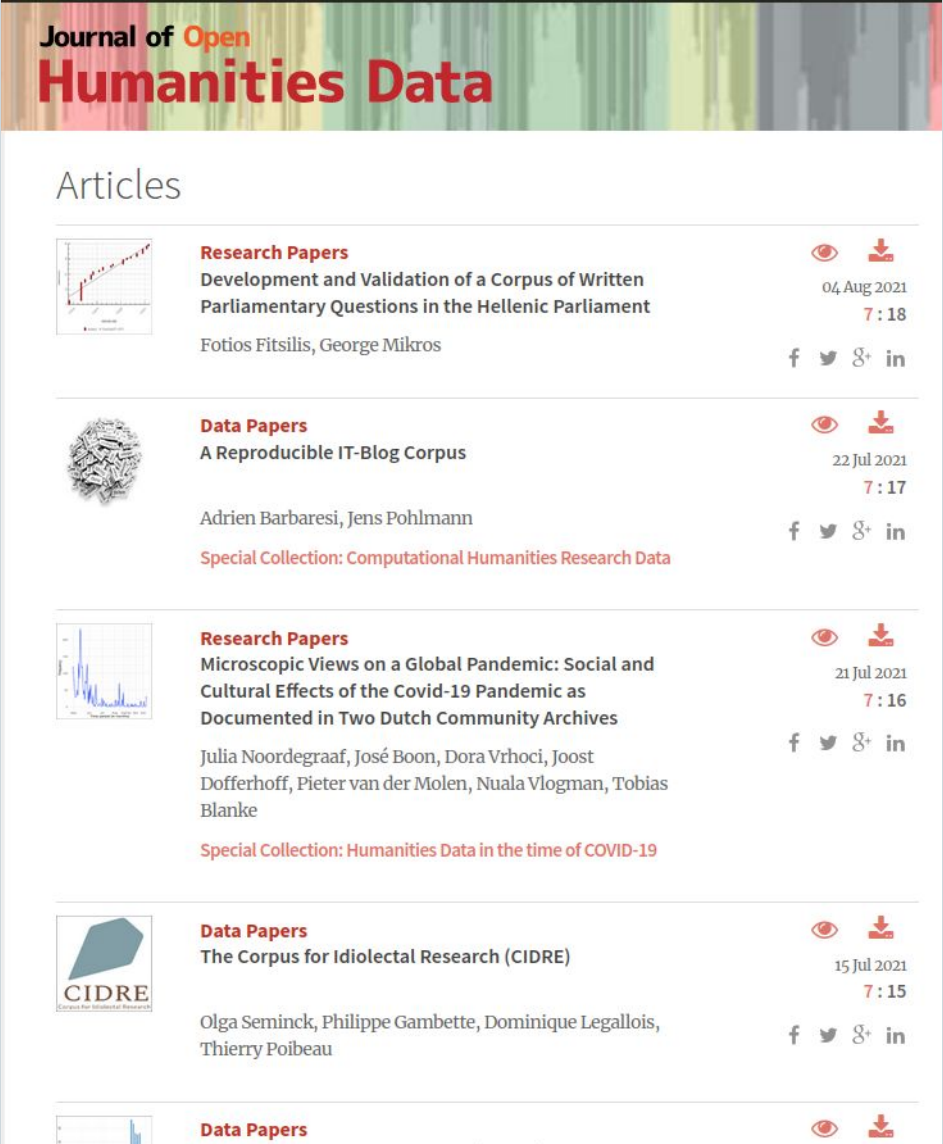
Mandy Wigdorowitz  
University of Cambridge, United Kingdom  
Contact:   
[Learn More](#) 

# JOHD's publications

Two types of publication, both data-focused:

- 1) data papers (1000 words)
- 2) research papers (3000-5000 words)

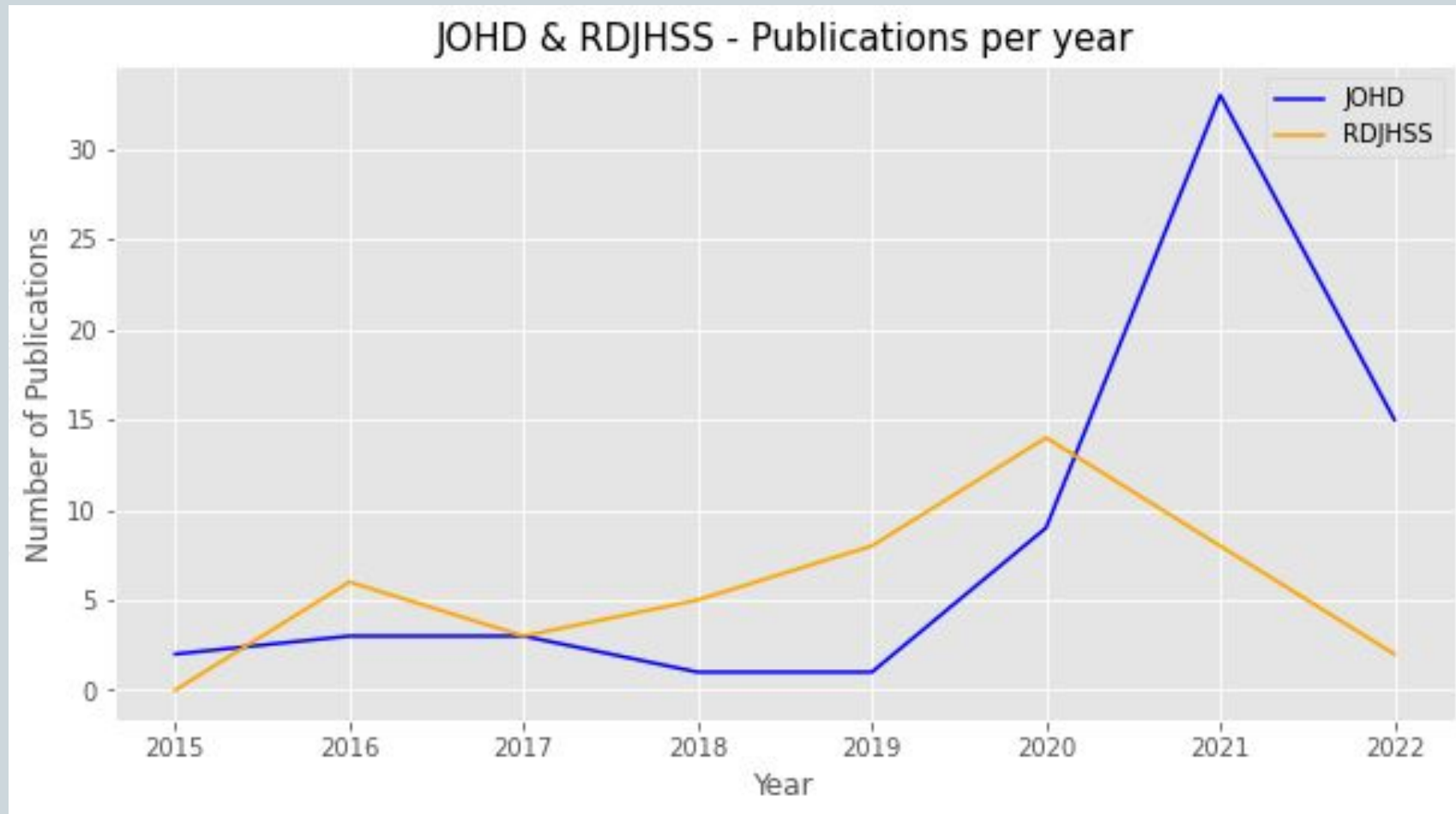
We have published **74** papers so far (October 2022)



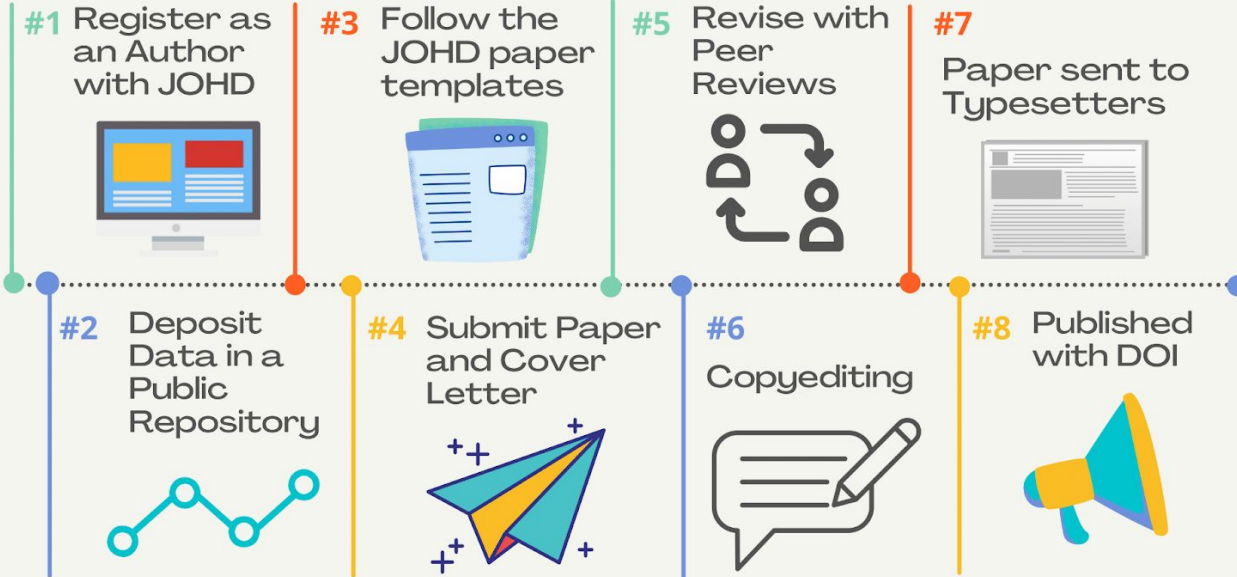
The screenshot displays the homepage of the Journal of Open Humanities Data. The header features the journal's title in a bold, sans-serif font. Below the header, the 'Articles' section lists several publications. Each entry includes a small thumbnail image, the article title, authors' names, the publication date, and the number of views and downloads. Social media sharing icons for Facebook, Twitter, Google+, and LinkedIn are also present for each article.

Article Type	Title	Authors	Date	Views	Downloads
Research Papers	Development and Validation of a Corpus of Written Parliamentary Questions in the Hellenic Parliament	Fotios Fitsilis, George Mikros	04 Aug 2021	7	18
Data Papers	A Reproducible IT-Blog Corpus	Adrien Barbaresi, Jens Pohlmann	22 Jul 2021	7	17
Research Papers	Microscopic Views on a Global Pandemic: Social and Cultural Effects of the Covid-19 Pandemic as Documented in Two Dutch Community Archives	Julia Noordegraaf, José Boon, Dora Vrhoci, Joost Dofferhoff, Pieter van der Molen, Nuala Vlogman, Tobias Blanke	21 Jul 2021	7	16
Data Papers	The Corpus for Idiolectal Research (CIDRE)	Olga Seminck, Philippe Gambette, Dominique Legallois, Thierry Poibeau	15 Jul 2021	7	15

# JOHD's publications



# Sharing your data and publishing your data paper



The dataset must be stored in an **open access** repository:

- ❖ JOHD Dataverse
- ❖ Zenodo
- ❖ Figshare
- ❖ SND
- ❖ DANS

Image created by Amanda Sorensen

# The special collections

- ❖ ***New perspectives and practices in Digital Humanities Data***: 3 articles
- ❖ ***Humanities data in the time of COVID-19***: 5 articles
- ❖ ***Language documentation collections: Assessment and Recognition***: 2 articles
- ❖ ***Computational Humanities Research Data***: 9 articles

JOHD accepts submissions to the general issue on a rolling basis

The screenshot displays the website for the Journal of Open Humanities Data. At the top, there is a navigation bar with links for Home, About, Contact, Content, and Research Integrity. Below this is the journal's title, "Journal of Open Humanities Data", in a stylized font. The main content area is titled "Special Collections" and features four distinct collection cards arranged in a 2x2 grid. Each card includes a title, a brief description, and social media sharing options. The top-left card is for "New perspectives and practices in Digital Humanities Data", the top-right for "Humanities Data in the time of COVID-19", the bottom-left for "Language Documentation Collections: Assessment and Recognition", and the bottom-right for "Computational Humanities Research Data".

Home About Contact Content Research Integrity

Journal of Open Humanities Data

### Special Collections

Special 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.

Share this: [f](#) [t](#) [g](#) [in](#)

Special 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.

Share this: [f](#) [t](#) [g](#) [in](#)

Language Documentation Collections: Assessment and Recognition

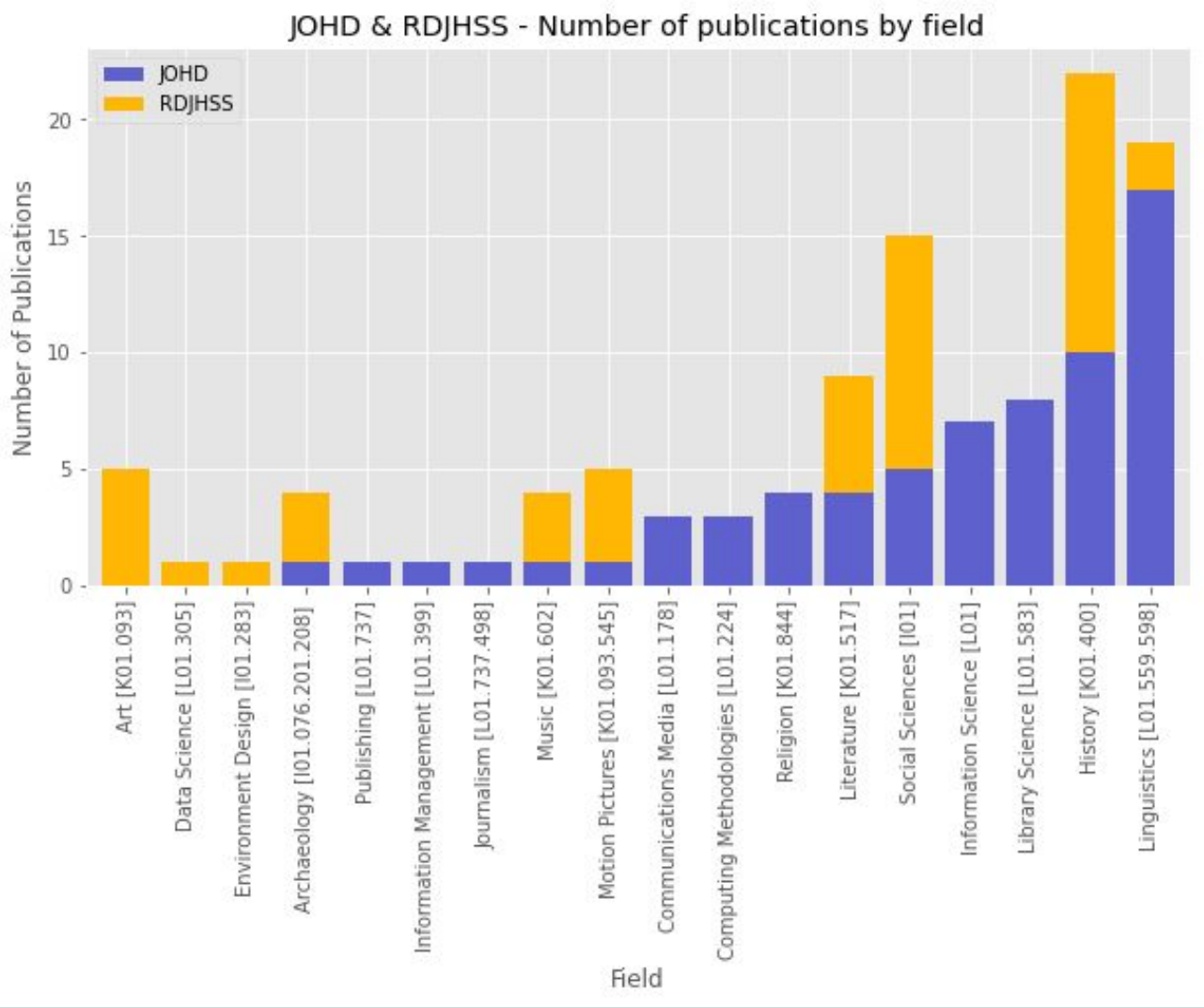
This Special Collection aims to develop a detailed outline of what an effective peer-review process for documentary linguistic materials might look like, and how such a system would foster better recognition for these materials in academic evaluation systems such as in hiring, promotion, and tenure.

Special 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](#)

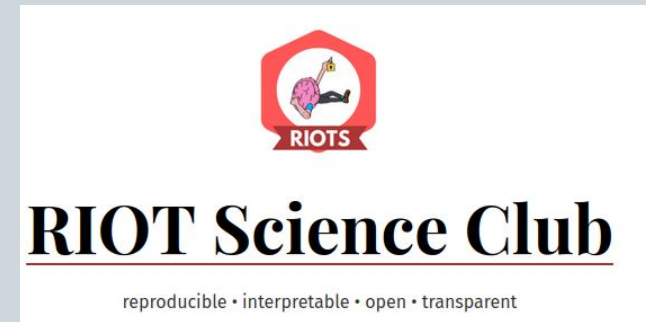
# Disciplinary scope



# Conferences and events

- ❑ LIBER conference (session #3 “Working with Software and Data”)
- ❑ round table of experts about data citation, organised by SSHOC (Social Sciences & Humanities Open Cloud)
- ❑ #dhnord2021 about the publication of data papers in the humanities
- ❑ DHASA conference 2021 (Digital Humanities Association of Southern Africa)

- ❑ SciDataCon-IDW Seoul 2022 about the impact of data papers and datasets Australian Text Analytics Platform
- ❑ Humanities and Data Science Discussion Group at the Alan Turing Institute
- ❑ RIOT science club
- ❑ Oxford Open Humanities Seminar Series





# The #showmeyourdata campaign

What is it?

- We ask our authors to post an **image** of their data and describe it. It can be the screenshot of (any type of) dataset, a visualisation etc.

What for?

- Our aim is to promote **open research** and **data sharing** values: we would like to encourage the conversation on this topic and demystify it

You Retweeted



LOC Crowdsourcing

@Crowd\_LOC

We're showing off our data(sets) today! Here's what a CSV file of transcriptions from the Rosa Parks collection looks like. What could you do with this data? [loc.gov/item/202044559...](http://loc.gov/item/202044559...)

Item	Itemid	Asset	AssetStatus	DownloadUri	Transcription
Notes Rosa Park	mss85943	mss859430232	completed	http://tile.loc	Wednesday 1/25/89, 1PM Vivian's deathFuneral service Sat. 10AM 1/28/
Notes Rosa Park	mss85943	mss859430232	completed	http://tile.loc	1/26/89 Star Makers Kelly & CoAgents Cheryl KaganMich Bank interest
Notes Rosa Park	mss85943	mss859430232	completed	http://tile.loc	Dr. Anderson24535 N. CarolinaSouthfield MI 48075met herat Farmer Jac
Notes Rosa Park	mss85943	mss859430232	completed	http://tile.loc	New Bride + GroomMr. & Mrs. Anderson BowlesRev. Bynum's grand dau
Notes Rosa Park	mss85943	mss859430232	completed	http://tile.loc	Varsity -Church'sJames Forman2615 EastlawnNear Mack & ConnerChris
Notes Rosa Park	mss85943	mss859430232	completed	http://tile.loc	Varsity -Church's[Mr. Will]Mr. M.C. GlennDet. Gen Hospital4th floor W-2
Notes Rosa Park	mss85943	mss859430232	completed	http://tile.loc	837-5007 68 Mike15 83[Jim Forman][NYC][212 663-2883]Ann Jones
Notes Rosa Park	mss85943	mss859430232	completed	http://tile.loc	728-0812Miss Kim Bond home after 4 or 5 pmInkster HighApr. 3. 10:30M
Notes Rosa Park	mss85943	mss859430232	completed	http://tile.loc	Katherine Wright13940 Sta[?eli?]836-8736Vets 962-4539Fred D. GrayP.O.
Notes Rosa Park	mss85943	mss859430232	completed	http://tile.loc	Katherine Wright13940 Sta[????]836-8736Vets 962-45391/2 " 5.90KC
Notes Rosa Park	mss85943	mss859430232	completed	http://tile.loc	Tom Bean2493 SeyburnDetroit Mich48214Telephone 92508 341/2 " 5.
Notes Rosa Park	mss85943	mss859430232	completed	http://tile.loc	Latin qtr.TR 5-6600Mr. Charles SandersStudent Teacher4th gradeParker E
Notes Rosa Park	mss85943	mss859430232	completed	http://tile.loc	for officeammoniadust cloths Bed - 39" w65" l
Notes Rosa Park	mss85943	mss859430232	completed	http://tile.loc	10 lbsSugar,lemon juice,cornflakesT. Tissue9631 MartindaleBostonJames
Notes Rosa Park	mss85943	mss859430232	completed	http://tile.loc	gov/image-services/iiif/service:mss:mss85943.0018:16.0031/full/pct:100
Notes Rosa Park	mss85943	mss859430232	completed	http://tile.loc	Hat MakerJessieWO 5-0573220 B[??] Mkt16th StRosa Parks
Notes Rosa Park	mss85943	mss859430232	completed	http://tile.loc	Mr. Wilson, TY 6.9578P.C.L. TY 4-1271Jackson TY 7-711-5Mrs. Montgom
Notes Rosa Park	mss85943	mss859430232	completed	http://tile.loc	922-7006Ethel AmbroseNAACP YouthTE 4-0177TE 1-0 600 or 1-4344
Notes Rosa Park	mss85943	mss859430232	completed	http://tile.loc	5-6-622.00 .501.00
Notes Rosa Park	mss85943	mss859430232	completed	http://tile.loc	Church 2-25-62 .10 3-11-62 3-18-623.00
Notes Rosa Park	mss85943	mss859430232	completed	http://tile.loc	Nebscough dropsPatrickpants 23 1/2 l w 21paj pattern - 4[Sh?]?r] &



Journal of Open Humanities Data (JOHD) @up\_johd · 18 Mar

It's #showmeyourdata time!

Today's #datapaper is "By the People Crowdsourcing Datasets from the Library of Congress", authored by the @Crowd\_LOC team [doi.org/10.5334/johd.67](https://doi.org/10.5334/johd.67)

Authors, if you'd like to participate in our campaign, please RT sharing an image of your data

2:24 pm · 18 Mar 2022 · Twitter Web App

10 Retweets 2 Quote Tweets 22 Likes

# Other activities

**Open Humanities Data Forums:** promoting values of data sharing and reusability in the humanities

The screenshot shows the YouTube channel page for 'Journal of Open Humanities Data', which has 16 subscribers. The channel is subscribed to. The navigation menu includes HOME, VIDEOS, PLAYLISTS, CHANNELS, DISCUSSION, and ABOUT. Under the 'Uploads' section, there are six video thumbnails:

- Second Open Humanities Data Forum - Q&A** (40:21, 4 views, 3 months ago)
- Second Open Humanities Data Forum - Richard...** (17:02, 5 views, 3 months ago)
- Second Open Humanities Data Forum - Lauren...** (6:39, 4 views, 3 months ago)
- Second Open Humanities Data Forum - Introduction** (23:19, 7 views, 4 months ago)
- Q&A** (30:06, 17 views, 7 months ago)
- First Open Humanities Data Forum - Introduction** (19:03, 32 views, 7 months ago)

The video thumbnails include various content: a woman speaking, a screenshot of a website titled 'The Endangered Languages Archive', a slide for 'Linguistics Data Interest Group (LDIG)' with the 'Research Data Alliance' logo, a diagram of 'The traditional research process', and two more women speaking in video call settings.

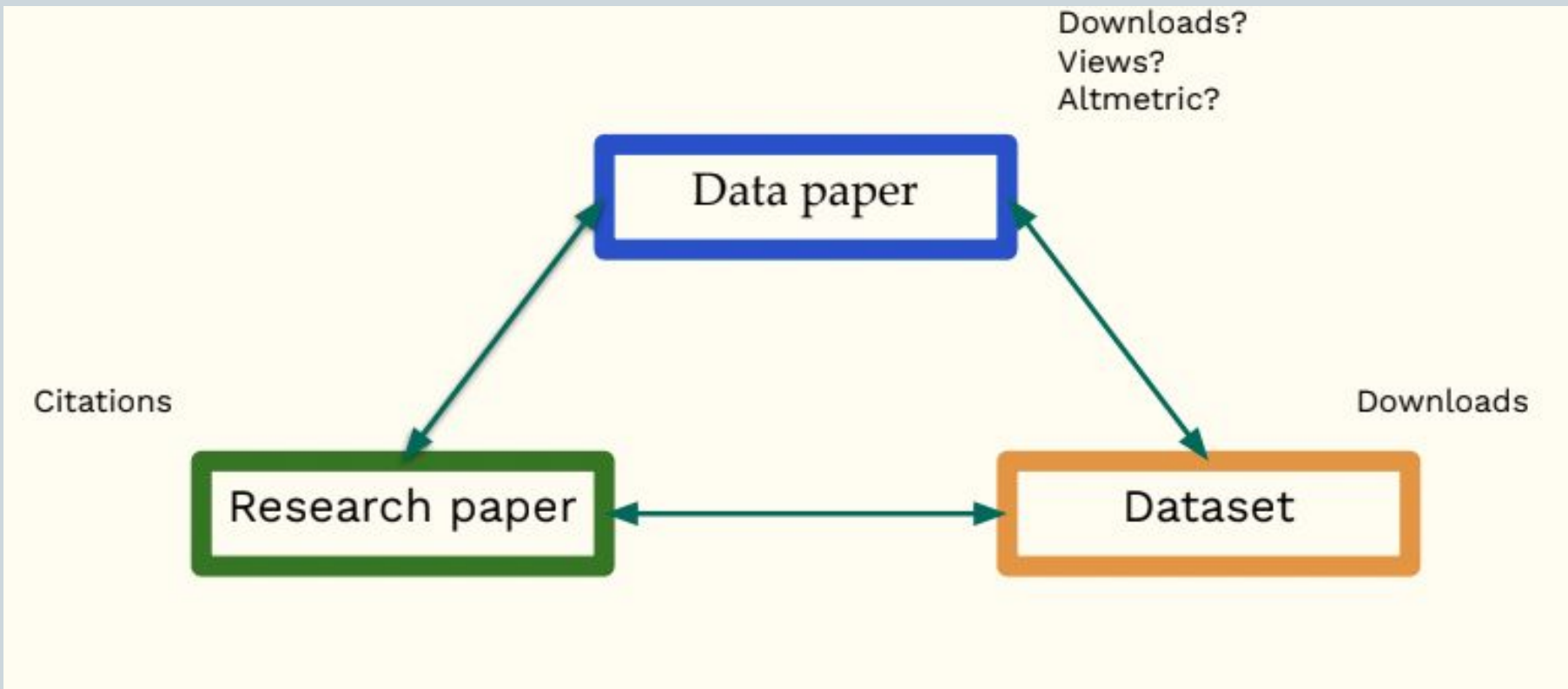
Credits: Paola Marongiu

# Papers

---

- ❑ “Le Journal of Open Humanities Data (JOHD) : enjeux et défis dans la publication de data papers pour les sciences humaines” (forthcoming). In *Publier, partager, réutiliser les données de la recherche: les data papers et leurs enjeux*, edited by Christine Kosmopoulos and Joachim Schöpfel. Presses Universitaires du Septentrion.
- ❑ “Deep Impact: A study on the impact of data papers and datasets in the Humanities and Social Sciences” (2022). *Research Data and Data Papers*, special issue of *Publications*, edited by Jorge Reverz.

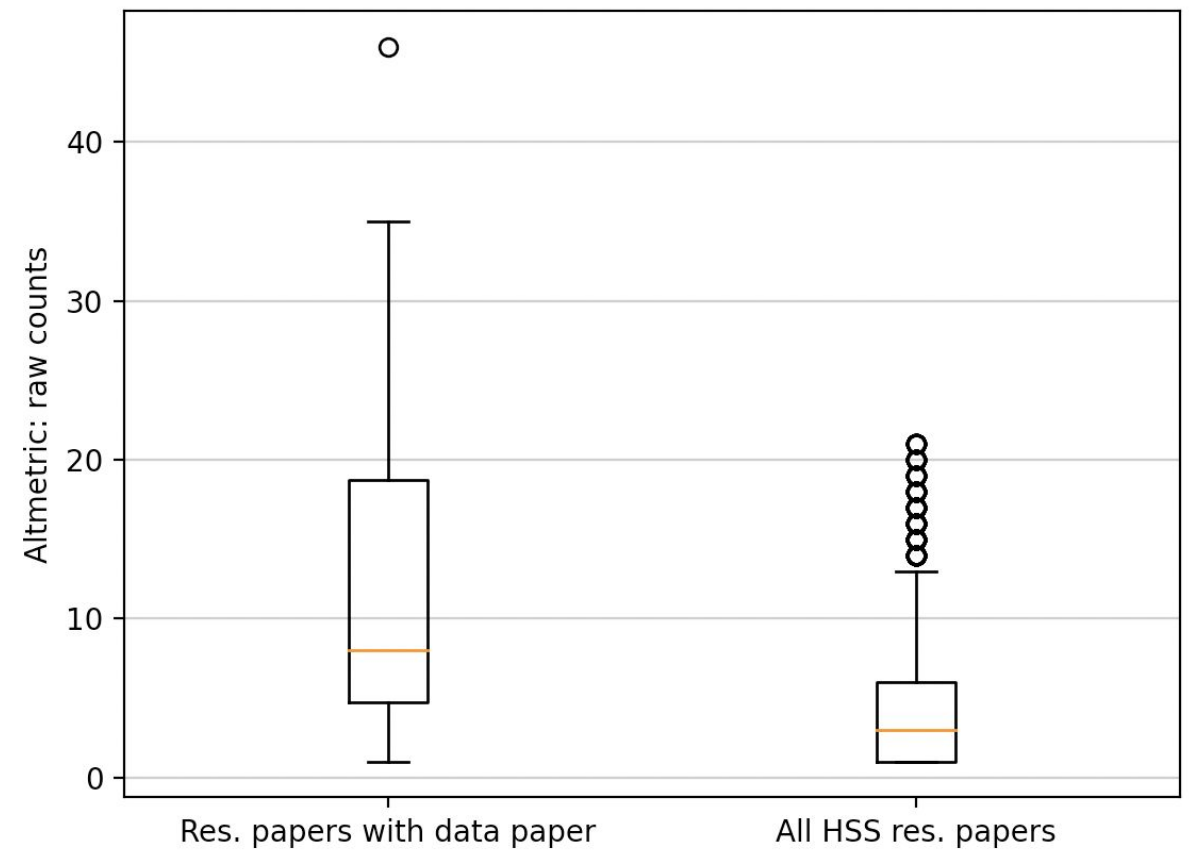
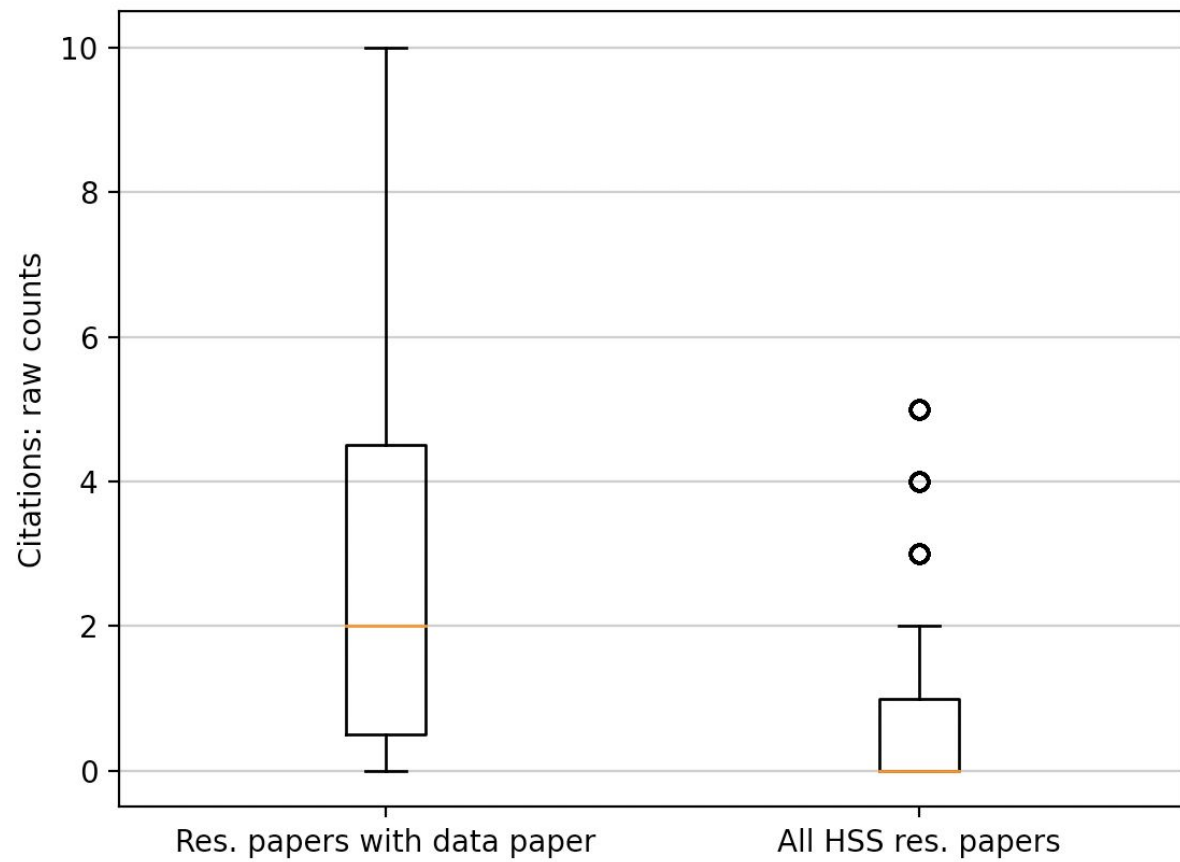
# Data papers and impact



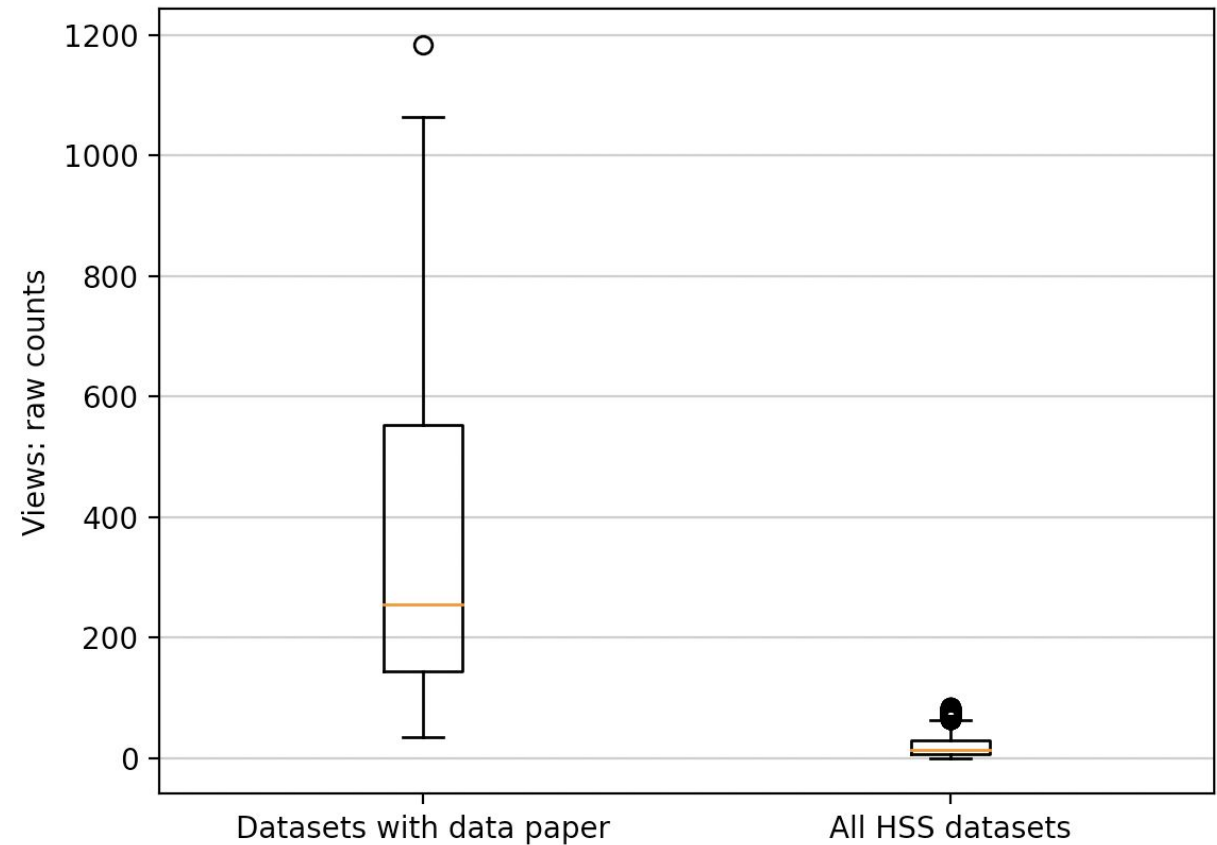
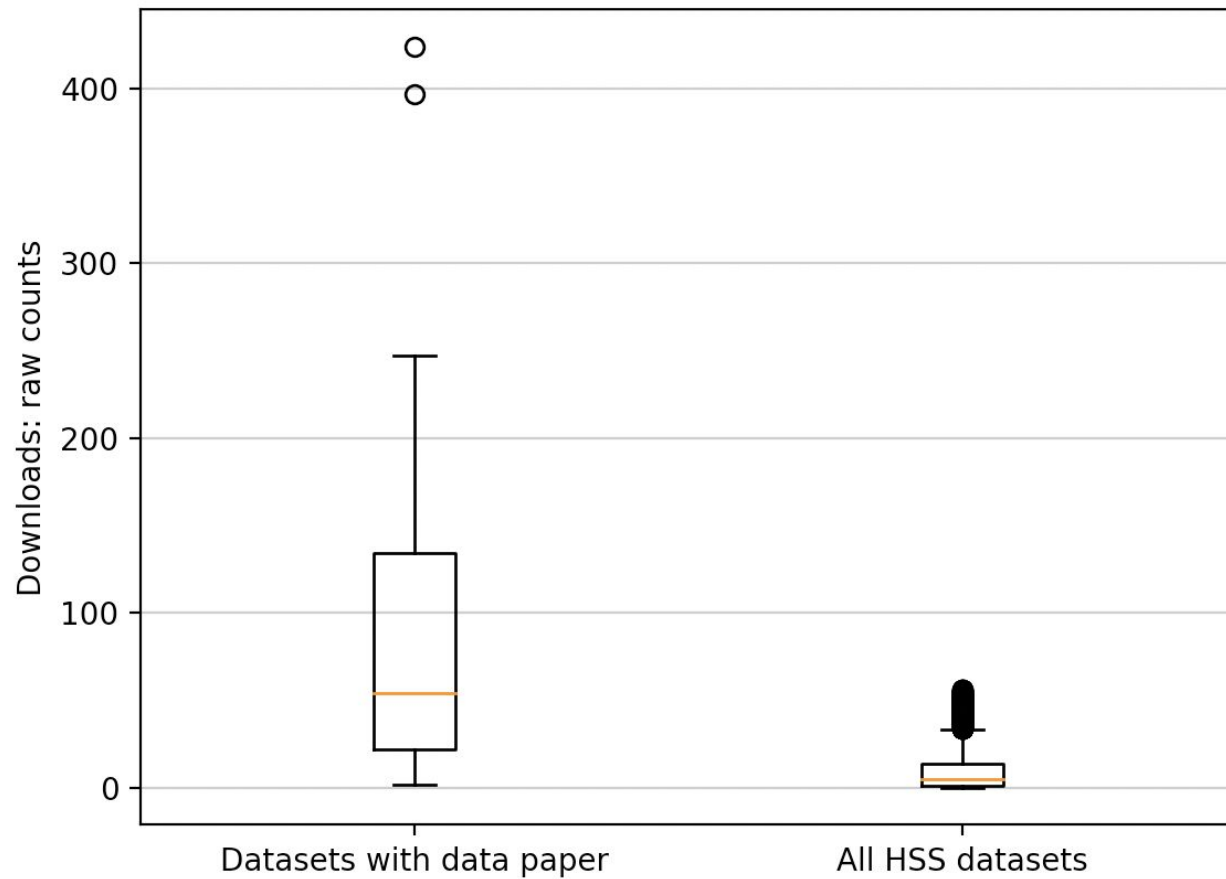
# Data papers and impact

variable1	variable2	rho	p-value	strength	significant?
altmetric_research	tot_citations_datapaper	<u>0.68</u>	<u>&lt; 0.001</u>	<u>moderate/strong</u>	<u>yes</u>
altmetric_research	views_datapaper	0.47	0.06	moderate	borderline
altmetric_research	downloads_datapaper	<u>0.59</u>	<u>0.01</u>	<u>moderate</u>	<u>yes</u>
tot_citations_research	altmetric_datapaper	0.39	0.06	weak	borderline
tot_citations_research	tweets_datapaper	0.38	0.06	weak	borderline

# Data papers and impact



# Data papers and datasets



Credits: Nilo Pedrazzini and Marton Ribary

# Keep in touch...



**@up\_johd**

**#johdpapers and #johdCfP: newly published papers and call for papers**

**#johdnews: news about JOHD (activities and events)**

**#johdagenda: external events about open data, data sharing and related matters**

**#showmeyourdata: share an image of your data and describe it**



**Newsletter:**

**Every 6 months, an update on all our activities, publications, new members of our team**

**Recordings of the events organised by JOHD (e.g. *Open Humanities Data Forums*)**

**Recordings of presentations to conferences and events**

**Short videos on key topics for JOHD e.g. how to define humanities data?**





# Thank you!

[Barbara.mcgillivray@kcl.ac.uk](mailto:Barbara.mcgillivray@kcl.ac.uk)