Tips and tricks to open a bigger window on arts and humanities research

Erzsébet Tóth-Czifra
Open Science Officer at DARIAH
@etothczifra

03.11.2022, Mendrisio

DARIAH-EU
Digital Research Infrastructure
for the Arts and Humanities

Why Open Science really matters to me...

Opening up access to scholarly knowledge is a key value of the academy in the 21st century.

Still, sometimes we still need to cross actual, physical borders to cross virtual ones (such as paywall or proprietary licenses).
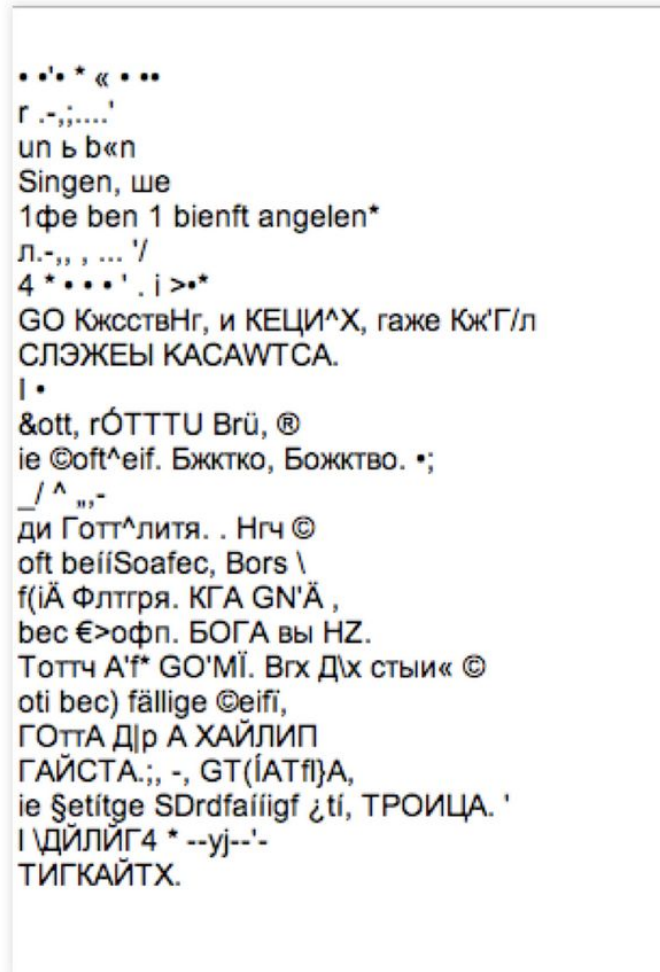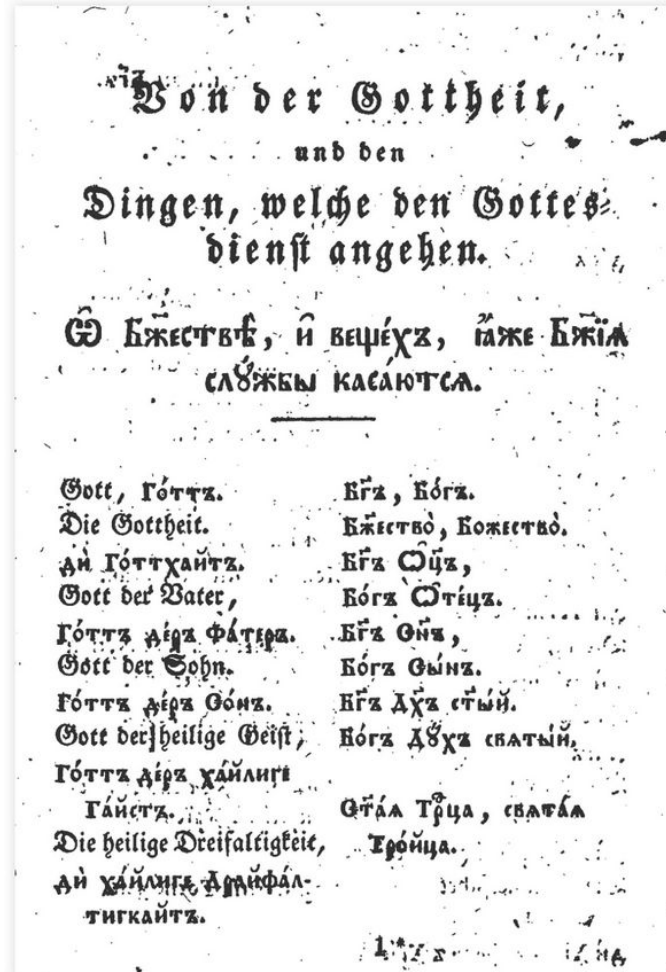
Luckily, things are changing...

What do you see?

• •'• * « • ••

r .-,;....'
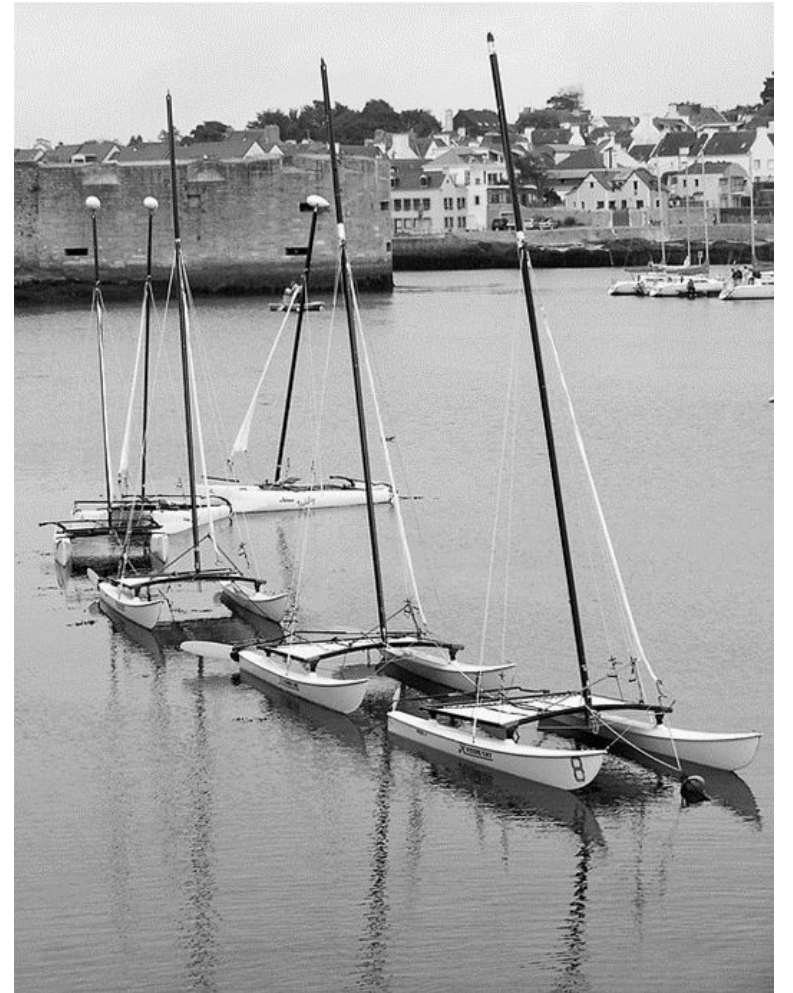
un ь b«n

Singen, ше

1фе ben 1 bienft angelen*

л.-,, , ... '/

4 * • • • ' . ¡ >•*

GO КжсствНг, и КЕЦИ^Х, гаже Кж'Г/л

СЛЭЖЕЫ КАСАWTCА.

I •

&ott, rÓTTTU Brü, ®

ie ©oft^eif. Бжктко, Божктво. •;

_/ ^ „,-

ди Готт^литя. . Нгч ©

oft beííSoafec, Bors \

f(iÄ Флтгря. КГА GN'Ä ,

bec €>офп. БОГА вы HZ.

Тоттч A'f* GO'MЇ. Вгх Д\х стыи« ©

oti bec) fällige ©eifï,

ГОттА Д|р А ХАЙЛИП

ГАЙСТА.;, -, GT(ÍATfl}A,

ie §etítge SDrdfaííigf ¿tí, ТРОИЦА. '

I \ДЙЛЙГ4 * --yj--'-

ТИГКАЙТХ.

What do you see?

Рѣчникъ ма́лый Daß ist Kleines Wörterbuch, 1837 edition. Source: Google Books.

A major challenge in Arts and Humanities data sharing workflows: the cross-sector continuum of data curation

Cultural Heritage Institutions

- Selecting, hosting, curating, preserving cultural artefacts
- Digitizing them and making them available online
- Rights management: opening or closing away collections

Digital Humanities Labs

- Cleaning, standardizing and enriching these materials
- Opening up existing shared resources for large-scale, computational analysis in ways that align with researcher requirements

Researcher Communities

- Analyzing, enriching, reusing, interpreting collections
- Both large-scale and small-scale analyses
- New computational methods change the ways in which we interact with our heritage, enable us to ask new questions
- New insights
- Digital storytelling to a broader audience

Research Infrastructures, Publishers

- Hosting, preserving, publishing, sustaining content generated by scholars
- Interlinking different content types
- Discovery environments

What is in the term...

+ rich provenance, capturing research steps

Source: https://www.nosci.mk/?page_id=26

Once building on each others' work is truly possible…

Real life examples

Polifonia: playing the soundtrack of our history

https://polifonia-project.eu/

Watch a demo here: https://aimusicfestival.eu/en/programs/2021/areas/concerts-amp-demo-research/demo-research-2-polifonia

IIIF – sustainably linking, embedding, sharing digital resources

From mauscripts to innovative and unexpected ways to acces history

Welcome to Transcribe Bentham!

By Louise Seaward, on 6 December 2017

'*Many hands make light work. Many hands together make merry work*', wrote the philosopher and reformer, Jeremy Bentham (1748 – 1832) in 1793.

In this spirit, we cordially welcome you to *Transcribe Bentham*, a double award-winning collaborative initiative which is crowdsourcing the transcription of Bentham's previously unpublished manuscripts.

Anyone can start transcribing at our Transcription Desk. Your transcripts will contribute to the production of Bentham's *Collected Works* and preserve Bentham's writings into the future.

UCL Home » / Transcribe Bentham » /

Home

Transcription Desk

Blog

About Us

People

Jeremy Bentham

Recent Posts

- Transcription Update – 22 November 2019
- Transcription Update – 31 October 2019
- Transcription Update – 30 September 2019
- Preliminary download of Jeremy Bentham, Writings on Political Economy, Volume IV: Circulating Annuities and other writings on National Debt
- Transcription Update – 30 August 2019
- Transcription Update – 31 July 2019

Enter your email address

- Philisopher, social and legal reformer

- Digitizing his 40.000 untranscribed folios (in 8 years!)

- https://blogs.ucl.ac.uk/transcribe-bentham/

- WikiMedia instance to deliver images to volunteers to transcribe the texts in a machine-readable format (TEI-XML)

- Huge success: more than 22.000 manuscripts transcribed, 96% quality checked.

From manuscripts to new ways to access history

https://transkribus.eu/Transkribus/

- Automatic handwritten text recognition and transcription

- The more people using it the more useful it gets

- Starting out from a digitization project → creating a mechanisms by which others can auto transcribe their texts on the large → changing how people do history, how people can access history, the questions we might ask from these resources etc.

...but sometimes it is the small steps that makes the biggest diference. Everyone can be an open scholar!

Data Basics With Databases – The Wonders of Data Material

BY STEFAN KARCHER · PUBLISHED 2018-10-05 · UPDATED 2018-12-06

Vacation time is over and – which is a bit surprising for PhD students – I had much time to think about my dissertation and my academic future. I did some researches and picked up many new interesting fields someone should deal with. However, from time to time, I got angry about an issue which came across a lot. Have you ever noticed how stupid some subject-related databases are? I mean, having them helped us find sources and texts, which enriched our own work, was a wonderful opportunity 10 years ago; but today, 10 years later the relation between the whole bunch of data sets became the focus of some researchers' attention.

The need (or wish) to work with the full material of a database leads to the wide discussion about Open Access in science. I don't want to deepen the aspect here in this post, but to make my opinion clear: Open Access is a great convenience for an open and free science. I support this position and I think my contribution to the scientific society is to share research results and raw data in an easy, accessible way. Yet, I understand the need of legal restrictions, payment, and license policies. Publishers, universities, and scholars invest so much time and money to develop their systems; and, of course, charges and copyright restrictions are necessary for financial profitability. I'm happy to pay for access if that's the price to support scientific progress. One can't be so naive to think open science is for free. Someone must pay for it and even my time/work as a PhD costs resources and money. To make a long story short: If a database is hidden behind payment or license restrictions, there should be good reasons for it. And I hope the reason is not profit, but the necessity to keep the system running.

In this context, it is very important to differ between two types of access. The

InFoDiTex was nominated for a DH Award 2018 in the category "Best Use of DH For Public Engagement." We are very happy about our placement as "2nd Runner Up" and we would like to thank all those who voted for us!

Congratulations to **the winners and all the other great DH projects!**

UPCOMING EVENTS

ABOUT & CONTACT

The *Interdisciplinary Forum of Digital Textual Sciences* at the University of Heidelberg is an open meeting for (junior) researchers in all fields of Digital Humanities. It was founded by doctoral students who meet every month during the semester turn for

Blogging

Licensing

Open Access sharing resources beyond the research paper

Public engagement

publications

Article
Grades of Openness: Open and Closed Articles in Norway

Susanne Mikki *, Øyvind L. Gjesdal and Tormod E. Strømme

University of Bergen Library, University of Bergen, 5007 Bergen, Norway; oyvind.gjesdal@uib.no (Ø.L.G.); tormod.stromme@uib.no (T.E.S.)
* Correspondence: susanne.mikki@uib.no; Tel.: +47-909-082-66

DARIAH-EU
Digital Research Infrastructure for the Arts and Humanities

CC BY

PUBLIC HUMANITIES

How to open a bigger window on your research?

Step-by-step practices

Traveling through the research data life cycle
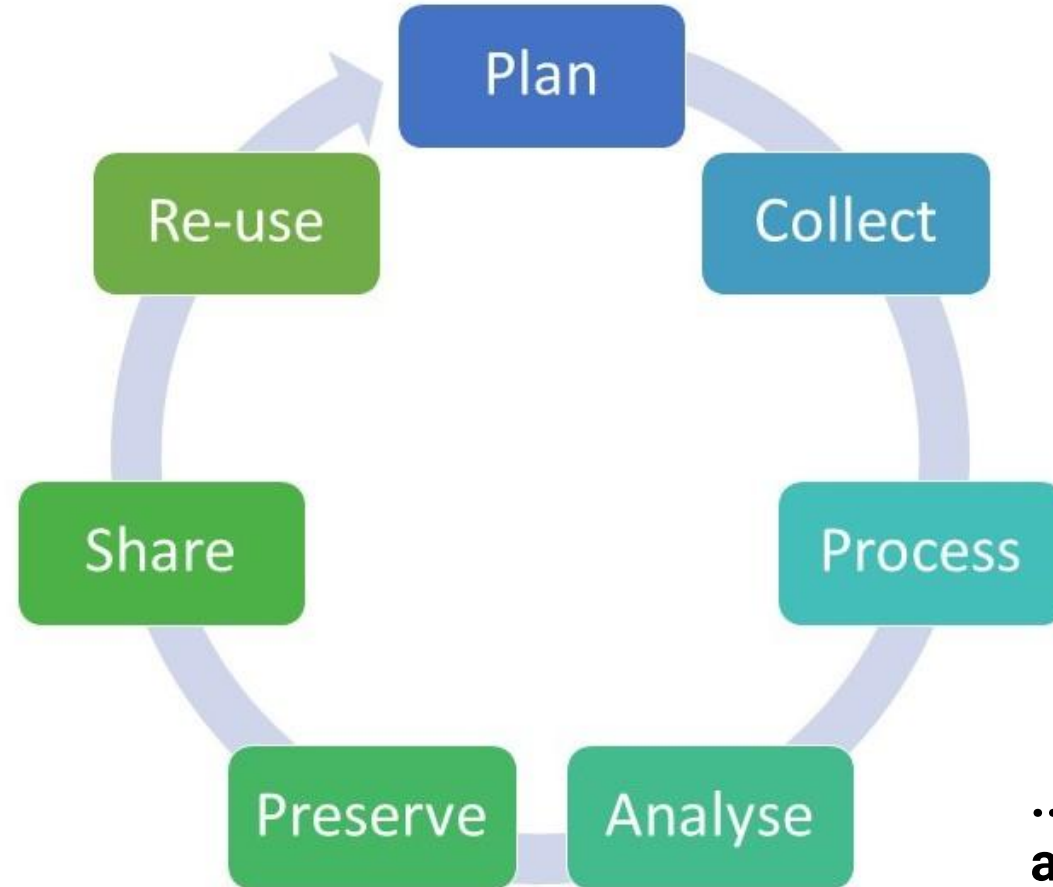
...which is never as linear as put forward here ;-)

A model of research workflow

- Never as linear as one would expect

- Data sharing should be kept in mind from the beginning

- " Your primary collaborator is yourself from 6 months now and your past self doesn't anser emails" (Rachel Ainsworth)

Plan

Collect

Process

Analyse

Preserve

Share
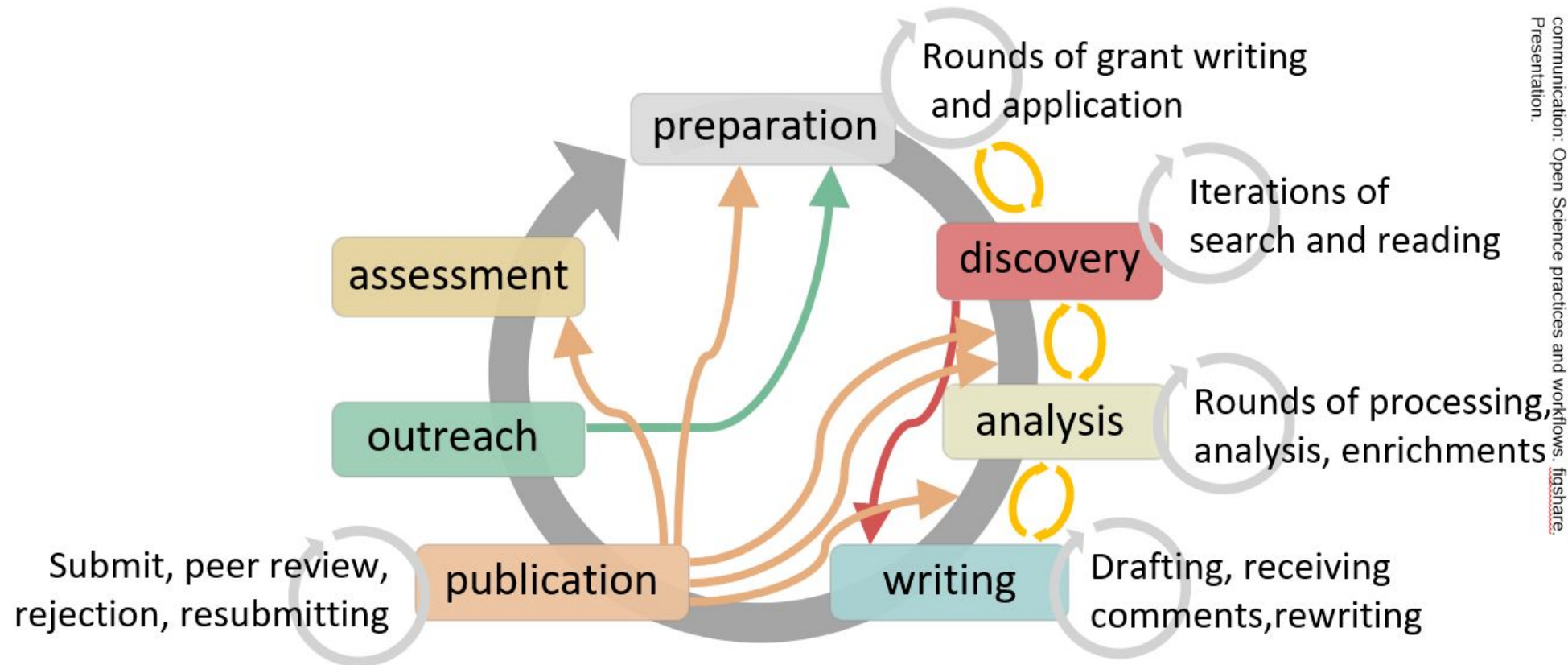
Re-use

1. Collection and reuse

Checklist to keep in your pocket during your first visit to the (digital) archive

December 10, 2018   Other   Open Access

Open Data for Humanists, A Pragmatic Guide

Edmond, Jennifer; Tóth-Czifra, Erzsébet

In the arts and humanities, digital data production is still expensive, challenging and time-consuming. We all know this, and yet the results of these processes often in the end can't be reused by other researchers, meaning that we reinvent (or redigitise) the wheel far too often. This resource is aimed at giving practical advice for arts and humanities scholars who are willing to take their first steps in research data management but don't know where to begin. Our approach to data management views it as a reflective process that exposes and tweaks existing behaviours, rather than one that introduces specific tools. It is intended to encourage awareness of one's own processes and mindfulness about how they could be more open and how and how small changes across three points in your research workflow can make big differences.

Preview

Page: 3 of 4   —   +   Automatic Zoom

With Source Data: *because the more you can share of your data, the more likely is that someone can use or will reuse it.*

1. Establish your sharing rights while you are at the archive, or otherwise accessing sources. Remember that librarians and archivists are your best allies in this, and will generally be very willing to help you understand reuse conditions for their material (or, in the case of your local, institutional research librarians, your own)
   a. Have the documents I need been digitised??
      i.   If so, how can I obtain digital copies? Can I access the documents remotely? (E.g. through your institutional search interface? an open API? via a downloadable or emailed file?)
      ii. If so, do digital copies or their description exist in multiple versions, how should I keep track of them?
      iii. If not, may I photograph documents myself or order photographs? Is there a charge for this?
      iv. May I share any photographs I take away? Under what conditions? Under what license(s) (such as the Creative Commons CC-BY)?
   b. Who is in charge of the documentation and curation of the material I am planning to work on? How can I learn about the documented ownership and history of curation of the document?
   c. Are there any sensitivities (eg. personal data) in the data I should be aware of,

Files (106.7 kB)

Open Data for Humanists, A Pragmatic Guide

With Source Data: *because the more you can share of your data, the more likely is that someone can use or will reuse it.*

1. Establish your sharing rights while you are at the archive, or otherwise accessing sources. Remember that librarians and archivists are your best allies in this, and will generally be very willing to help you understand reuse conditions for their material (or, in the case of your local, institutional research librarians, your own)
   a. Have the documents I need been digitised??
      i. If so, how can I obtain digital copies? Can I access the documents remotely? (E.g. through your institutional search interface? an open API? via a downloadable or emailed file?)
      ii. If so, do digital copies or their description exist in multiple versions, how should I keep track of them?
      iii. If not, may I photograph documents myself or order photographs? Is there a charge for this?
      iv. May I share any photographs I take away? Under what conditions? Under what license(s) (such as the **Creative Commons CC-BY**)?
   b. Who is in charge of the documentation and curation of the material I am planning to work on? How can I learn about the documented ownership and history of curation of the document?
   c. Are there any sensitivities (eg. personal data) in the data I should be aware of, and what are the best practices for using this data for research purposes?
   d. Are there any specific references or identifiers you recommend I capture in my personal metadata to facilitate reuse?
   e. Can I deposit any data surrogates I create with you or link mine with somehow with your collection? Would you like me to make you aware if I deposit them elsewhere (e.g. in my institutional repository or other data repositories)? Whom should I contact with this information?
   f. How would you like your institution to be cited/acknowledged in any publications in which I use them?

How do we find data for reuse? Discussion:

- When was the last time you used data collected, curated or generated by someone else?
- Where/how did you find it and **which factors helped its discoverability?**
- How could you **access** the material in question?
- How did you trust in the creators?
- **How the limitations of the data set (incompleteness, uncertainties) were indicated?**
- How your research methods affected the collection of your data?
- How the collection of your data affected your research methods?

2. Processing and analyzing

How will I know how the dataset I'm interested in had been 'cooked'?

- A major twist in FAIR research culture: the separation of data from its context of creation.

- Explaining how the data had been 'cooked': rich provenance metadata (incl. the description of the software environment) is of crucial importance for both cultural heritage professionals.

- Standards: bridges between repositories, enable to bring together isolated data and to give them a richer context, improving their readability.
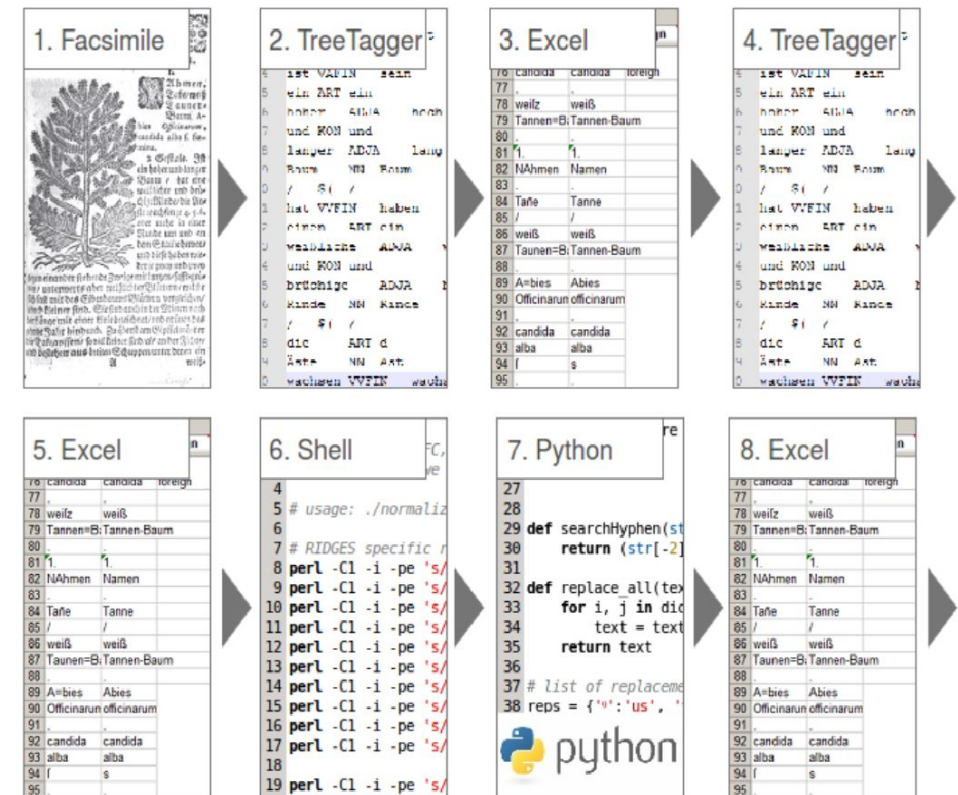
Data documentation and metadata

- **How can you minimize the hassle for other people to find the materials you used and created?**

- Your documentation should indicate finding aids and other resources used

- **Not everything has to be kept!**

- Once you have developed a suitable data model, you are also advised to develop a data dictionary which documents the model.

- This document may contain the following information:
  - a list of all the column names used in the data spreadsheet
  - description of the purpose and the contents of these different columns, explaining abbreviations etc.

Image source: Pixabay, CC0.

How others can make sense of your data?

An example: interview data

• The audio file of the interview

• The interview transcript in the form of a digital text file

• The discussion guide or questionnaire which explains the methodological approach and is necessary for the comprehensibility of the results of the study.

• The project explanation as well as the declaration of consent of the interviewee, which documents compliance with the legal provisions of the GDPR.

• The codebook which e.g. documents the development categories and variables used

• The documentation of the procedure for anonymization and pseudonymization

• The indexing information (metadata), which guarantees the citaability of the interview and its findability.
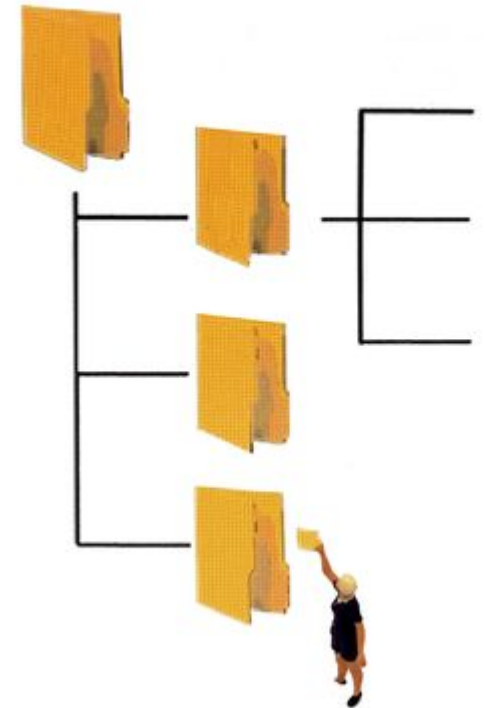
Image source: CESSDA , CC-BY 4.0

Leveraging on the open, on the digital

...and making it work for her career advancement!

- Background in German studies

- Building, encoding, and annotating a corpus of parliamentary debates in XML-TEI – a guided process, reaching out for mentorship

- Data sharing ☐ gaining recognition, new collaborations and post-doc position(s)

- Inernmediate thougs ☐ blog

- Opportunities to openly share papers even if publishing in closed access

Naomi Truan
Wissenschaftliche Mitarbeiterin
at Uni Leipzig

@Berlinguistin

All publications available online in HAL and ORTOLANG :
https://cv.archives-ouvertes.fr/naomi-truan

Blogging & tweeting research experience:
https://icietla.hypotheses.org

FAIR data workflows made concrete: from researchers to researchers

Research Data Management Workflows
Volume 1: Creating and analyzing multilingual parliamentary corpora

Document initiated by Erzsébet Tóth-Czifra, Open Science Officer, DARIAH Coordination Office Berlin, Germany, and conceived and prepared in collaboration with Naomi Truan, University of Leipzig, Germany

Creating and analyzing multilingual parliamentary corpora

Harmonization of the sub-corpora	Quality assurance	Reuse
- Source materials are converted into a TEI format with the help of the GROBID suite	- Manual checks, iteratively, fine grains where needed	- Tóth reuse history already available - Fellow scholars working on parliamentary debates from various disciplines

Step 1	Step 2	Step 3	Step 4	Step 5	Step 6

Selection of the data relevant to the research question		Creation of the TEI annotation framework		Data sharing and publication	
- Public, open data - English, French, German and parliamentary debates revolving around the European Council meetings		- TEI annotation framework in two levels: TEI Header (metadata) and on the level of the text (part-of-speech) - Automated part-of-speech tagging with the help of the TXM software		- Corpora and annotation framework on ORTOLANG - Papers: on HAL - Reflections: on a Hypotheses blog	

Tóth-Czifra & Truan 2021 | CC BY 4.0

Overview of the data workflow 'Creating and analyzing multilingual parliamentary corpora'
[Download the schema as a PDF on HAL-SHS to make it machine-readable]

In this resource, you can follow a step-by-step description of a research data workflow involving the annotation of multilingual parliamentary corpora (French, German, British) according to the guidelines of the Text Encoding Initiative (TEI). Read further if you are interested in working with the TEI, analyzing parliamentary corpora, or simply would like to see a validated example of how FAIR and open data is implemented in the context of a PhD dissertation in Corpus Linguistics.

Openness barometer

What	Ideal case	If open sharing is not possible, you can still:
Publications (papers, books, book chapters)	Immediate Open Access with pre-prints shared in public repositories	✔ Legally share pre-prints or post-prints (also known as green Open Access) ✔ Specify and communicate its embargo period and openly share your work once it's over ✔ If the publication cannot be Open Access due to copyrighted third-party material, you can still open share a version without that material, only referring to them.
Data or other underlying material (multimedia, visualizations, code, project documentation etc.)	Immediate Open Access, CC-BY, shared and archived in a data repository and if preferred, also linked back to institutional or project websites.	✔ Publish your resources in repositories with restricted access conditions, closed access or embargo – this way, it can still be FAIR ✔ If the data itself sensitive or cannot be legally, ethically shared, you can still share its anonymized version or its documentation (how the methodology had been set up, sharing the empty questionnaire, standards used etc.) ✔ Share contact information to data creators, curators or controllers
Source material (coming from memory institutions or other third parties)	Public domain or free to reuse under Creative Commons licenses	✔ Clarify reuse conditions, location and contact info ✔ Share everything else (e.g. The enrichments you made, the resulting publications without including copyrighted 3rd party materials) ✔ Arrive at compromises with the rights holder like making these resources available in institutional environments, websites, this way setting up virtual reading rooms, or publish them in environments (websites or repositories) with institutional authentication control.

File naming conventions
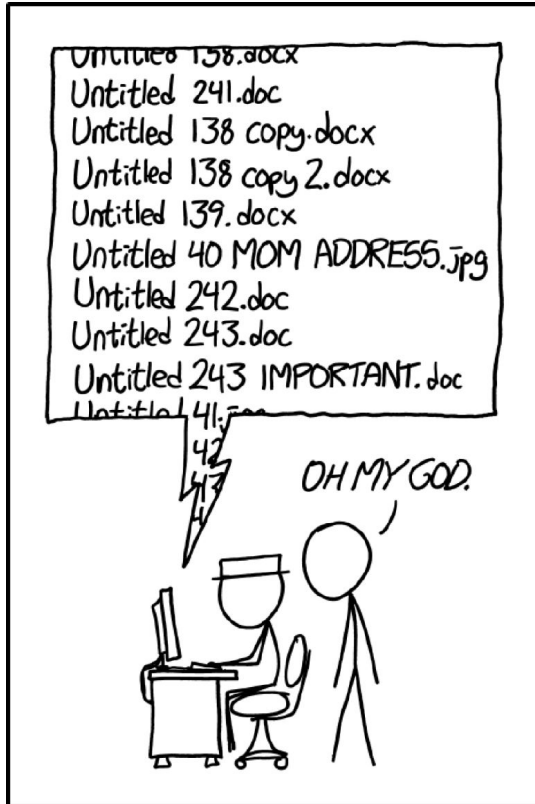
PROTIP: NEVER LOOK IN SOMEONE ELSE'S DOCUMENTS FOLDER.

Looks familiar?

File naming conventions
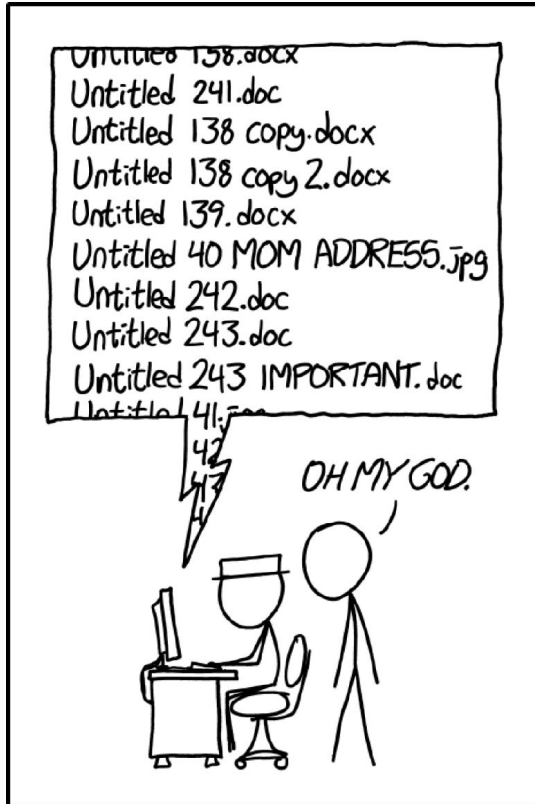
VS.
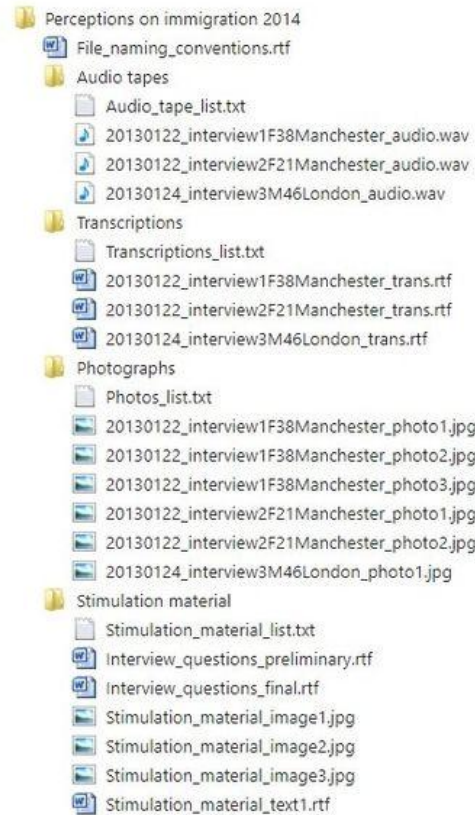
The specifics usually matter less than just having some.
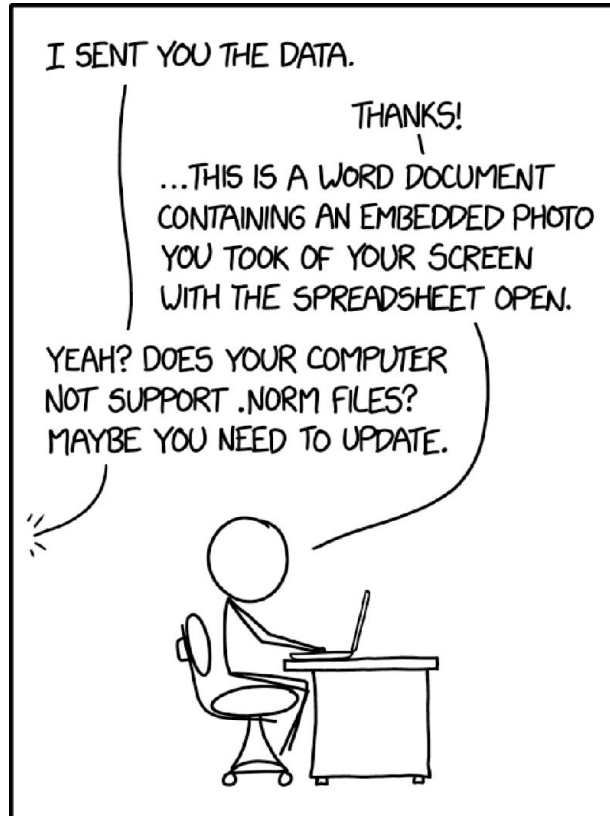
Common elements (UK Data Service):
- Version number
- Date of creation (date format should be YYYY-MM-DD);
- Name of creator;
- Description of content;
- Name of research team/department associated with the data;
- Publication date;
- Project number.

Can I run your data on my tools?

Image source: Stanford Library
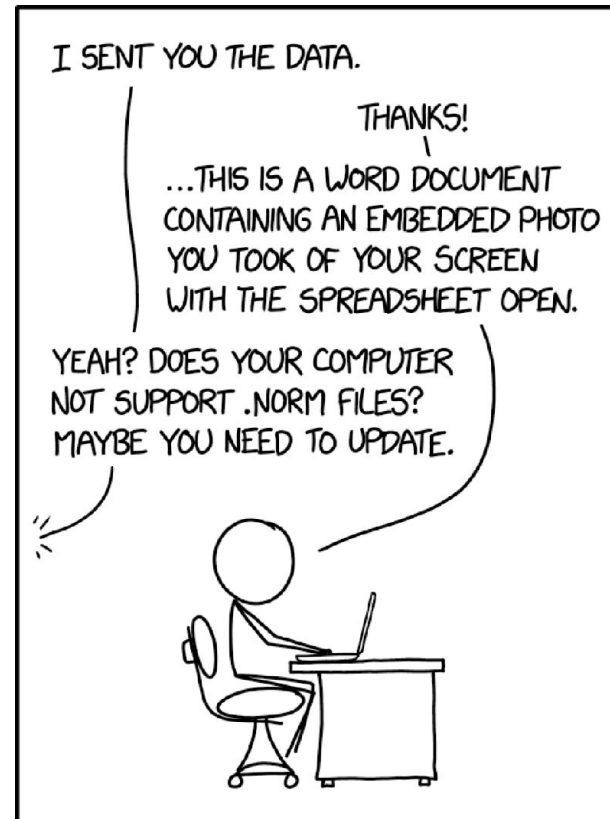
Open formats

SINCE EVERYONE SENDS STUFF THIS WAY ANYWAY, WE SHOULD JUST FORMALIZE IT AS A STANDARD.

Image source: Stanford Library

	Type	• Preferred format(s)	• Non-preferred format(s)
Statistical data			
Programming languages	Text documents	• PDF/A (.pdf) • ODT (.odt)	• Microsoft Word (.doc) • Office Open XML (.docx) • Rich Text File (.rtf) • PDF other than PDF/A (.pdf)
Images (raster)			
Images (vector)			
Audio	Plain text	• Unicode text (.txt)	• Non-Unicode text (.txt)
Video	Markup language	• XML (.xml) • HTML (.html) • Related files: .css, .xslt, .js, .es	• SGML (.sgml) • Markdown (.md)
Computer Aided Design (CAD)			
Geographical information (GIS)	Programming languages	• MATLAB • NetCDF • TextFabric	
Images (georeference)			
Raster grid	Spreadsheets	• ODS (.ods) • CSV (.csv)	• Microsoft Excel (.xls) • Office Open XML Workbook (.xlsx) • PDF/A (.pdf)
3D			
RDF	Databases	• SQL (.sql) • SIARD (.siard) • CSV (.csv)	• Microsoft Access (.mdb, .accdb) • dBase (.dbf) • HDF5 (.hdf5, .he5, .h5)
Computer Assisted Qualitative Data Analysis (CAQDAS)			
	Statistical data	• SPSS Portable (.por) • STATA (.dta) • DDI (.xml)	• SPSS (.sav) • SAS (.7dat; .sd2; .tpt)
Abbreviations and acronyms			

Formats preferred by the DANS repository. See the full list here:
https://dans.knaw.nl/en/about/services/easy/information-about-depositing-data/before-depositing/file-formats

5 star development scheme for Open Data

Source: https://5stardata.info/en/

Ontologies, vocabularies

- Put structure on your messy data so that it opens up to others (people, machine, applications, related databases).
- Check whether some of the general topics and terms (persons, locations, concepts) that you focus on have already been assigned persistent identifiers or URIs in one of the ontologies that are relevant for your field.
- Ontologies are one of the ways in which we can make datasets interoperable,

The Basel Register of Thesauri, Ontologies & Classifications (*BARTOC*)
The CLARIN Concept Registry and the DARIAH/ACDH collection of vocabularies.

PARTHENOS YouTube Videos

'An Ontologist and a Data Scientist walk into a bar: Data in Research Projects'

An Ontologist and a Data Scientist walk into a bar: Data in Research Projects
Unlisted

https://www.youtube.com/watch?v=WNG1iLB4KtA&index=1&list=PLKq1g7snsFGc7f1_Aidypmz62d7i6Uh4x

Metadata standards in the Arts and Humanities (teaser)

- **TEI** (Text Encoding Initiative): www.tei-c.org
- **CEI** (Charter Encoding Initiative): http://www.cei.lmu.de/index.php
- **MEI** (Music Encoding Initiative): https://music-encoding.org/
- **CMDI** (Language Resources, CLARIN):
- **IIIF** (International Image Interoperability Framework): https://iiif.io/
- **EAD** (Encoded Archival Description, for finding aids): https://www.loc.gov/ead/
- **Dublin Core** (description of digital documents): http://dublincore.org/

The choice of appropriate standards is more of a community issue than a technical one.
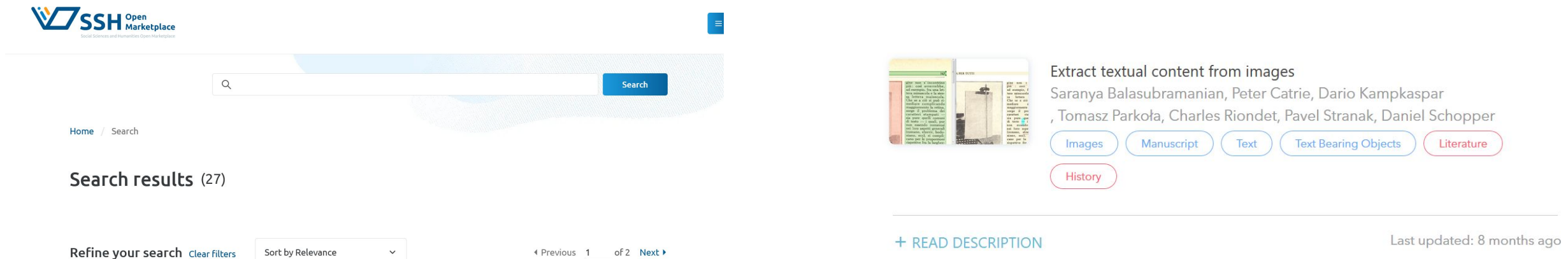
Adapted from: 'Future Proof and FAIR Research Data: Open Data Management Best Practices and First Steps', Ulrike Wuttke: https://www.fosteropenscience.eu/node/2603

The Standardization Survival Kit

An overlay platform dedicated to promoting a wider use of standards within the Arts and Humanities:

- Documenting existing standards by providing reference materials.
- Fostering the adoption of standards.
- 18 scenarios: Heritage science scenarios + "traditional" DH ones → Living memory of best practices
- Developed within the framework of the EU project PARTHENOS:

  http://ssk.huma-num.fr/#/

Give a passport with your data − prepare a readme file

- A readme file provides information about a data file and is intended to help ensure that the data can be correctly interpreted, by yourself at a later date or by others when sharing or publishing data.
- <u>Standards-based metadata</u> is generally preferable, but where no appropriate standard exists, for internal use, writing "readme" style metadata is an appropriate strategy.

Exercise:

1. Go to <u>https://cornell.app.box.com/v/ReadmeTemplate</u> where you will find a readme template

2. What are the main components of the document?

3. Are these well-aligned with your research processes and data? How would you adapt it for your own research?

Source: Cornell University, Research Data Management Service Group.
https://data.research.cornell.edu/content/readme

Image source:
https://www.feeldesain.com/travel-tag-texts.html

Linking datasets with publications Exercise:

Work in groups.

Consider the following three articles.

- To what extent can the data sets that are mentioned in the articles be accessed?
- Are the data sets also in preferred formats?
- Which kinds of additional documentation would further increase their accessibility and reusability for other disciplinary communities?

https://doi.org/10.1080/0969594X.2016.1194257 (https://tinyurl.com/datainarticles1)

http://doi.org/10.1111/lang.12172 (https://tinyurl.com/datainarticles3)

https://doi.org/10.1515/JDH-2021-1004?locatt=label:JDHFULL (https://tinyurl.com/datainarticles5)
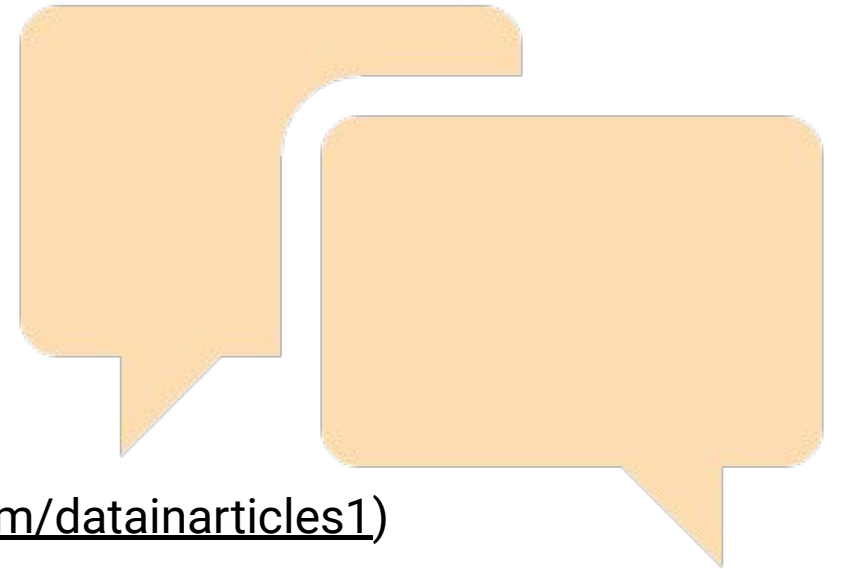
3. Discovery and sharing

Data repositories for Arts and Humanities

How about this?

What is an API?

- It stands for "Application Programming Interface."

- Some projects provide interfaces that allow for multiple ways of access to data

- Machines to machines; including all kinds of environments; machine-readable formats; data can be accessed directly from common programing languages (R, Python etc.)

- API queries can be executed live and directly and that the output can be controlled more precisely

- APIs are customizable to specific research questions and needs: researchers can develop their own analysis tools and connect them with data via APIs.

Using APIs for more flexible data discovery and reuse

vs.

https://pro.europeana.eu/post/using-jquery-with-the-europeana-api

Data discovery via discovery portals

Digital Humanities and Cultural Heritage

Research outcomes

Search by title, author, abstract, DOI, orcid...

SEARCH

Advanced Search

| SUMMARY | PUBLICATIONS 3,696,053 | RESEARCH DATA 262,743 | SOFTWARE 752 | OTHER RESEARCH 141,846 |

This community gathers research results, data, scientific publications and projects related to the domain of Digital Humanities. This broad definition includes Humanities, Cultural Heritage, History, Archaeology and related fields.

Curated by: 👤 Achille Felicetti

Created: 01-Mar-2018   Members: 76

Subjects

modern art, monuments, europe
field walking, frescoes, LIDO meta
history, excavation, Arts and Hum
coins, temples, numismatics, lithi
archaeology, digital cultural herit

Exercise 2:

- Go to: https://dh-ch.openaire.eu/
- Add a search term
- What are the search variables?
- Are you happy with the results?

How to select a repository that best fits your research?

SO...
WHERE WOULD YOU STORE THIS?

LIBRARIAN

AWKE

LEARNING HOW TO ARCHIVE DATA

1. Use an external data archive or **repository already established for your research domain** to preserve the data according to recognised standards in your discipline.

2. If available, use **an institutional research data repository**, or your research group's established data management facilities.

3. Use a cost-free data repository such as **Zenodo**.

4. Search for other data repositories here: **re3data.org**.

Source: https://www.openaire.eu/opendatapilot-repository-guide

Pick a position and discuss !

- The subsequent use of data requires more knowledge than the collection of new data.
- I often feel unsure about the reuse conditions of Cultural heritage data dat are relevant for me.
- It is not easy to apply standards to my work.
- The publication of research data does not contribute to building a reputation.
- The management and publication of research data causes costs, which I can't carry.
- If I publish my research data, somebody might scoop me and publish findings based on my data.
- When I publish my research data, my research becomes completely transparent and even the smallest errors become apparent.
- My research belongs to me!
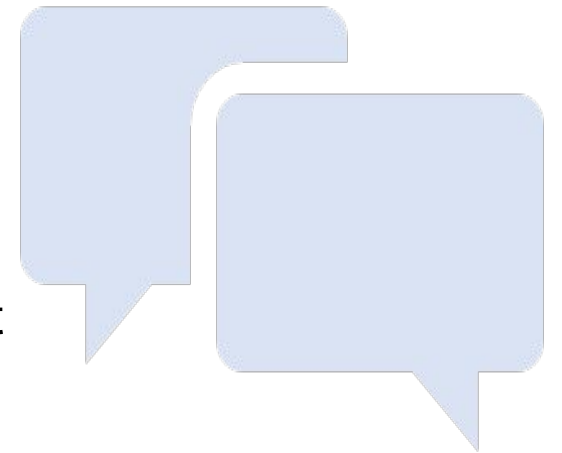
Adapted from: 'Future Proof and FAIR Research Data: Open Data Management Best Practices and First Steps', Ulrike Wuttke: https://www.fosteropenscience.eu/node/2603

Special flavours of Open Access in the Arts and Humanities?

Line up!

"*I have the impression that openly sharing my publication is not that easy in my discipline*"

Disagree

Agree

1   10

The ongoing COVID-19 pandemic provides yet more evidence of the importance of unrestricted access to scientific and scholarly outcomes

How to put all this to practice?

Source: Pixabay, CC0.

Early-career researchers

What do we want?

Get published

...and build academic profile

Finish that dissertation

Find a permanent job
(within or outside academia)

With thanks to **Lisa Matthias** for the following slides.

Publish where you want,
Open Access is still
possible.

YAY!

50

Open Access

⊘ **GOLD OPEN ACCESS**
Publishing your scholarly work in an OA journal or as an OA monograph.

⊘ **GREEN OPEN ACCESS**
Providing OA to a version of a work published in a closed-access journal or with a closed-access publisher.

51
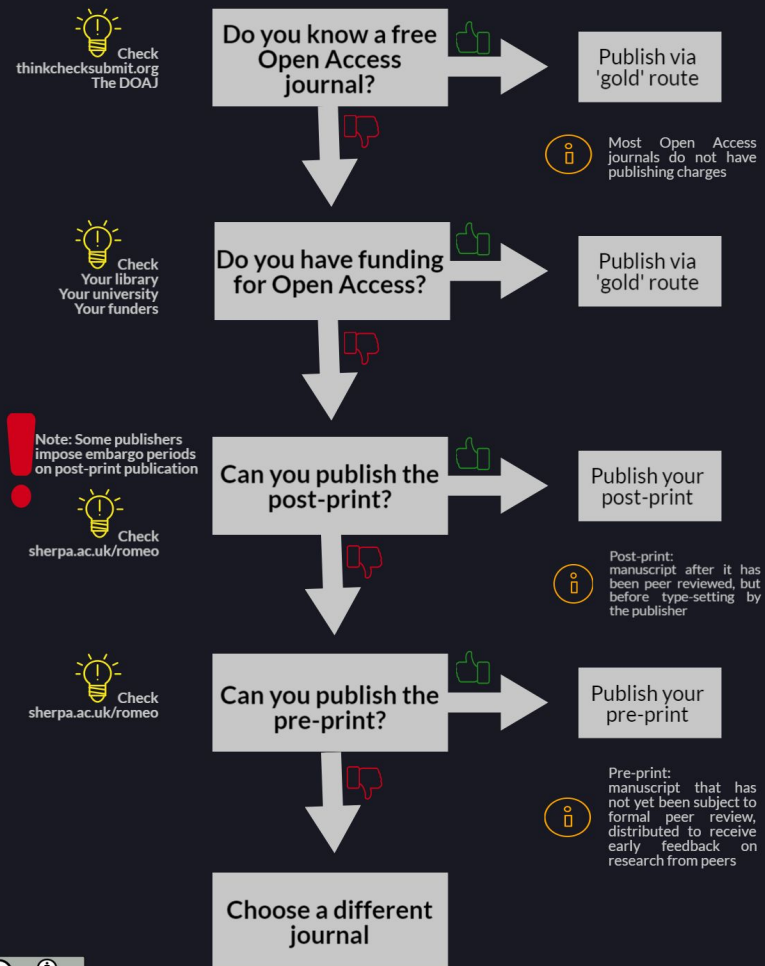
Why self-archiving?

Scott B. Weingart
@scott_bot

A colleague and I have a chapter on literary network analysis forthcoming that, on account of a long delay in publication, has a footnote with the most controversy I've ever stuffed into a single sentence. Meep. Wish we had another few months and few thousand words to unpack it.

> [2] Between this chapter's completion in early 2016 and its push to press in 2019, much has changed in the Literary Network Analysis landscape. Although a full treatment is beyond the scope of this chapter, "Distributed Character: Quantitative Models of the English Stage, 1550–1900" (Algee-Hewitt 2017), "Fictional Systems: Mass-Digitization, Network Analysis, and Nineteenth-Century Australian Newspapers" (Bode 2017), and "The Computational Case against Computational Literary Studies" (Da 2019) are worth special mention, as is the recent decentering of Moretti from the long narrative of Literary Network Analysis.

1:38 PM · Aug 5, 2019 · Twitter Web App

10 Retweets   **46** Likes

Self archiving? Easy to say so...

Will I be plagiarized?

What exactly can/should I put online?

Will anybody find my preprints anyway?

All my mistakes and uncertainties will be visible?

Is it good for my career or am I just giving away my resources?

How to find a safe home to my data?

Why self-archiving?

- Early visibility, putting a time-stamp on your
- work and ideas (vs. the fear of being plagiarized)

- Opening up the scholarly dialogue

- Ensuring proper citability

- Defining the conditions of re-use, versioning

- Taking (back) control ;-): if you put your publication into
- a publicly maintained platform
-
- Enabling text and data mining on your corpus of publications

Ten myths around open scholarly publishing

Literature review   Science and Medical Education   Science Policy

Jonathan P Tennant[1], Harry Crane[2], Tom Crick[3], Jacinto Davila[4], Asura Enkhbayar[5], Johanna Havemann[6], Bianca Kramer[7], Ryan Martin[8], Paola Masuzzo[9], Andy Nobes[10], Curt Rice[11], Bárbara S Rivera-López[12], Tony Ross-Hellauer[13], Susanne Sattler[14], Paul Thacker[15], Marc Vanholsbeeck[16]

March 11, 2019

https://peerj.com/preprints/27580/

How do I know which version should I put online?

- From an early draft to the final publishers' versions

- Do I have a right to put the publisher's version?

https://v2.sherpa.ac.uk/romeo/

https://v2.sherpa.ac.uk/fact/

Source: DARIAHOpen

The life cycle of a publication

- **Author version:** corresponds to any text with content that is directly produced by its authors. This includes the initial manuscript and its subsequent versions, the manuscript submitted for review to a conference or a journal.
  - **Initial manuscript:** first form of the author version disseminated in open access.
  - **Submitted manuscript:** first form of the author version sent to a conference or a journal for peer review.
    - **Revised manuscript after review:** last author version as transmitted to the conference or journal after peer review.
- **Publisher's version:** document possibly formatted by the conference or journal publisher and distributed by the latter. If the copyright is owned by the publisher, authors are not allowed to share this version.

Source: University of Cambridge Open Access Service

Which license should I attach the paper?

- Which what?
- Indicating the conditions of re-use
- Anyone has not heard of Creative Commons?
- **The baseline: being attributed: CC-BY**
- With CC-0 as an option if you just want your content be available (e.g. meta-data)
- How to deal with 3rd party material?
- Am I afraid of commercial re-use?
– The –NC (non commercial) extension
– Fear of being translated and cited without an authorisation?
– Example of possible difficulties: Online advertisement, private universities,
- Do I want to impose open access?
– The –SA (share alike) extension
– Example of possible difficulties: reuse in text and data mining contexts

How the Creative Commons licenses are different from each other?

	CC Symbol	Description
1	(cc) (i) BY	**[BY] By Attribution** Permits all uses of the original work, as long as it is attributed to the original author (Note: Attribution is in all six licences)
2	(cc) (i) (o) BY SA	**[BY-SA] By Attribution – Share Alike** As above, but any derivative work must also use a similar license, hence "Share Alike"
3	(cc) (i) (=) BY ND	**[BY-ND] By Attribution – No Derivatives** Licensed works are free to use / share with attribution, but does not permit derivative works from the original
4	(cc) (i) (s) BY NC	**[BY-NC] By Attribution – Non-Commercial** Licensed works are free to use / share / remix with attribution, but does not permit commercial use of the original work
5	(cc) (i) (s) (o) BY NC SA	**[BY-NC-SA] By Attribution – Non-commercial – Share Alike** Does not permit commercial use of the original work, and any derivatives from it must use a similar licence
6	(cc) (i) (s) (=) BY NC ND	**[BY-NC-ND] By Attribution – Non-Commercial – No Derivatives** Does not permit any commercial use or derivatives of the original work. *Note: this is the most restrictive of CC licenses, and is often regarded as a "free advertising" license*

"A license is a formalized promise not to sue!" (Pawel Kamoczki)

Green Open Access:
Common Practices

WHERE

Institutional or disciplinary repositories, sometimes author's personal website.

WHEN & WHAT

Check the publisher's self-archiving policies to know when & which version you are allowed to deposit.

Some examples from the SSH

HumanitiesCommons (all humanities disciplines):https://hcommons.org/
BodoArXiv (medieval studies):https://bodoarxiv.org
Cogprints (cognitive sciences): http://cogprints.org/
LingBuzz (linguistics): https://ling.auf.net/lingbuzz
MediArXiv (media, film and communication studies): https://mediarxiv.com
PhilArchive (philosophy): https://philarchive.org
PsyArXiv (psychological sciences): https://psyarxiv.com/
PhilSci-Archive (philosophy of science):http://philsci-archive.pitt.edu
SocArXiv (social sciences): https://socopen.org
SSH communities in Zenodo www.zenodo.org

What about these?

RESEARCHGATE

Sued by Elsevier.

ACADEMIA.EDU

Received several takedown notices from Elsevier.

SSRN

Bought by Elsevier.

What about these?

	R^G	A	HAL
SUPPORTS EXPORT & HARVESTING	No	No	Yes
LONG-TERM PRESERVATION	No	No	Yes
BUSINESS MODEL	Commercial	Commercial	Non-profit
FULFILLS FUNDER OA POLICIES	No	No	Yes

62

What about these?

SUPPORTS EXPORT &
HARVESTING

LONG-TERM PRESERVATION

BUSINESS MODEL

FULFILLS FUNDER OA
POLICIES

63

Adapted from: https://openreflections.wordpress.com/2017/05/09/academia-edu-and-self-branding

The aim of this scholarly blog is to provide a single space where information around the emerging landscape of Open Humanities can be disseminated, discussed and evaluated.

Come and join us to share your views on or experience with open research practices or suggest new topics to cover.

ARCHIVES

BILLETS

One more word about ResearchGate/Academia.edu and why using these platforms will never be equal to proper self-archiving

BY **ERZSÉBET TÓTH-CZIFRA** · PUBLISHED 31/05/2020 · UPDATED 05/06/2020

In the past couple of years, we saw a massive advancement of platforms, tools and networks that enable us to share our research legally and freely. Even so, the use of ResearchGate and Academia.edu, two of the most popular but also heavily criticized social networks that compete

TAGS

Arts and humanities Cultural heritage DARIAH DARIAH communities Data discovery Data management Data reuse Data skills Didactics Digital critical editions Digital cultural Heritage Digital Humanities Digitization EOSC Gold Open Access Green open Access Humanities Licensing Open Access Open Access Books Open Access Week 2019 OpenAIRE Open data Open humanities Open peer review Open scholarly infrastructure Open

What are the important criteria for you when choosing a publication venue for your work?

Source: Pixabay, CC0.

What comes to your mind about the term 'Open Access publication?'

Source: Pixabay, CC0.

Gold Open Access: common concerns

PAYING TO PUBLISH
Majority of OA journals does not charge APCs.

AUTHOR RIGHTS
Author rights still protected, granting greater reuse right to readers.

AVAILABILITY
Considerable differences regarding the availability of OA journals.

QUALITY
OA articles and monographs subject to the same editorial & peer review standards as closed-access counterparts.
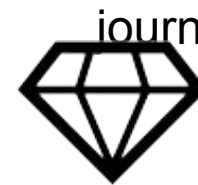
ROYALTIES
OA does not influence print sales. Consult publisher.

Should I pay or should I go?

"Article Processing Charge" (APC) „Book Processing Charge" (BPC)

- When the journal/publisher asks you to pay to be "open access"

- Native OA journals, hybrid journals, APC-free journals

- In many cases, there is a possibility to reduce/waive the charges
-

- Hount for special issues!

- Very good read:
http://fossilsandshit.com/the-term-article-processing-chargeis-misleading/

6 innovations from the humanities that make Open Access publishing a reality to everyone

BY ERZSÉBET TÓTH-CZIFRA · PUBLISHED 28/03/2019 · UPDATED 28/10/2019

https://dariahopen.hypotheses.org/529

Free to read, free to publish publication venues (sample)

https://dariahopen.hypotheses.org/630

Emerging community publishing models

PubPub

Pricing

PubPub's goal is to provide a high-quality, open-source, institution- offer a Free tier with no user or publishing limits, as well as an Orga want to support our mission).

Free

For publishing communities of all kinds.
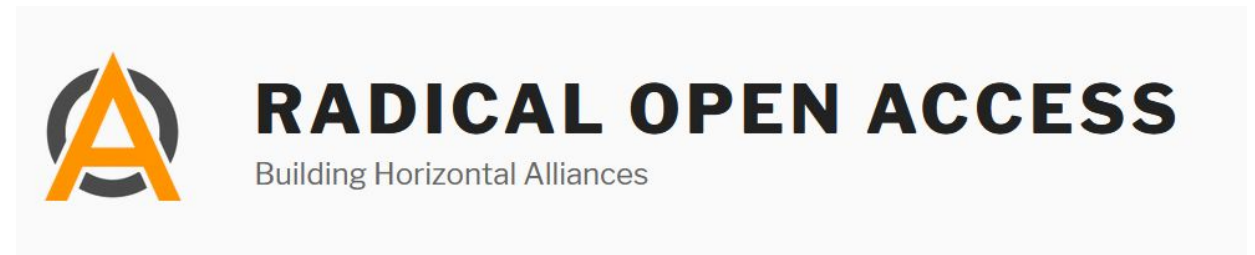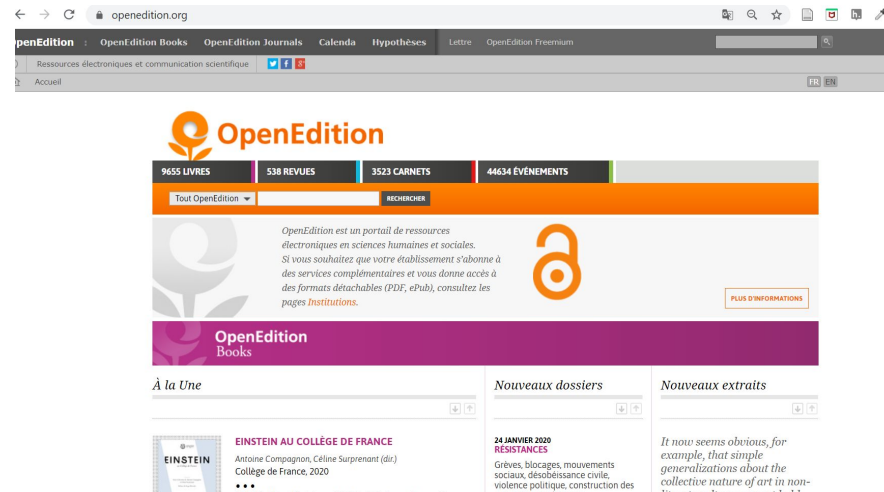
Free, Forever

- All of PubPub's core drafting, review, publishing, and discussion features
- Your own pubpub.org subdomain
- Unlimited publications
- Unlimited users

Quote from Open a GLAM Lab

The book's front cover

Cultural heritage organisations have historically provided access to and preserve cultural heritage. The shift towards the digital has presented new opportunities for experimentation and innovation. The fast pace of technological developments impacts society and culture worldwide. Some institutions may not be ready for this. This is the world of GLAM Labs. Labs and Lab-style work challenge the traditional approach and use new, existing and emerging technologies to make their collections available in innovative, engaging and unexpected ways. Labs experiment, collaborate, take risks, sometimes fail, and always push boundaries.

Open a GLAM Lab, p. 32

Read the Book

How to find an Open Access publication venue for my work?

1. Word of mouth. What are the venues preferred (read) in your community?

2. Check what is available locally. Does your institution have publication funds, agreements with publishers/publishing platforms, own university press?

3. Browse DOAJ, DOAB, OAPEN or OpenEdition.

SO YOU'RE TELLING ME

THERES A CHANCE

memegenerator.net

Source: memegenerator.com

Once you made your choices...

## What should I ask a publisher about Open Access?

- Will I have to pay a fee? How much is it? What does it pay for?

- Is the publisher for-profit or not-for-profit?

- What peer-review systems do they have in place?

- Do they create Open Access editions?

- Do they insist on an embargo period?

- Is their Open Access edition just a downloadable PDF?

- Are the Open Access editions easily discoverable? How is the work distributed?

- Do they let you keep your copyright?

Think. Check. Submit. helps researchers identify trusted journals for their research. Through a range of tools and practical resources, this international, cross-sector initiative aims to educate researchers, promote integrity, and build trust in credible research and publications.
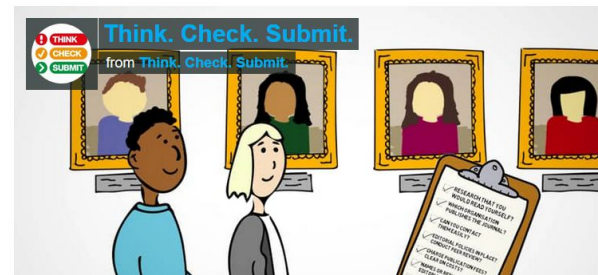
Latest news

Think. Check. Submit. looks forward to collaboration on new research programme
17th October 2019
A new research programme at Texas Tech University in the US, funded by the US National Science Foundation, aims...Read more...

OAPEN joins Think. Check. Submit.
7th October 2019
3rd October 2019   OAPEN (Open Access Publishing in European Networks) has joined the group of organisations endorsing...

https://thinkchecksubmit.org/

https://blogs.openbookpublishers.com/what-should-i-ask-a-publisher-about-open-access/

How to recognize predatory journals/publishers?

6 ways to spot a Predatory Journal:

1. Always check the website thoroughly

2. Check if the journal is a member of DOAJ, COPE, OASPA or STM

3. Check the journal's contact information

4. Research the editorial board

5. Take a look at their peer review process and publication timelines

6. Read through past issues of the journal

Source: rxcommons.org

https://predatoryjournals.com/journals/

Érdeklődés a ELTE oldalán található Tóth-Czifra Erzsébet munkájával kapcsolat

Aniko Varga <a.varga@globeedit.com>
címzett: én

Tisztelt Tóth-Czifra Erzsébet,

Szeretnék érdeklődni, hogy érdekelné-e ajánlatom, miszerint publikálnánk munkáját nyomtatott könyv formájában. Esetleg elküldhetem önnek brosúránkat Pdf formátumban, kiadónk szolgáltatásainak bemutatásával.
Örömömre szolgálna, ha hallanék önről kedves Tóth-Czifra Erzsébet.

Sincere regards,

Aniko Varga

Acquisition Editor

e: a.varga@globeedit.com
w: www.globeedit.com

Redeem the code xmas_17 and get 10€ off on each purchased book at:
www.morebooks.de

GlobeEdit
is an imprint of:
SIA OmniScriptum Publishing
Brivibas gatve 197, LV-1039, Riga
Latvia, European Union

Management:
Dr. Wolfgang Philipp Müller

Visibility, discoverability, usage
...through technology!

Adding permanent identifiers to books improves crosslinking: DOIs

Visibility, discoverability, usage ...through technology!

Permanent identifiers improve attribution to scholars: ORCID

Open Access made concrete: lessons learned from the NeDiMAH project

Start thinking about OA publishing at the beginning of the project!

Agree your plans with the team members!

Easier than one might think!

Search

Register   Log In

News Feed

Members

Groups

Sites

CORE Repository

Help & Support

HC Organizations

About

Roadmap

Team Blog

James L. Smith

Dr

University College Cork

@scrivenersmith

scrivenersmith

0000-0002-1370-3923

scrivenersmith.com

Following 28 members

Academic Interests

Cultural geography

Digital humanites

Ecocriticism

Ecomaterialism

About

Dr James L. Smith is Postdoctoral Research Fellow in the School of English and Digital Humanities at University College Cork, working on the 2019-23 Ports, Past and Present project. His work is at the intersection of the blue, environmental, spatial and digital humanities. His first monograph is *Water in Medieval Intellectual Culture: Case-Studies from Twelfth-Century Monasticism* (Brepols, 2018). James is the editor of *The Passenger: Medieval Texts and Transits* (punctum books, 2017), and co-editor of the Open Library of the

https://scrivenersmith.com/open-scholarship/

Open Scholar Stars interview series: Interview with Dr. James L. Smith

In the first episode of the DARIAH Open Scholar Stars interview series, we talk to Dr. James L. Smith, Irish Research Council postdoctoral fellow in geography at Trinity College Dublin, editor and Open Access advocate about the many ways in which creativity and openness fuels his scholarship.

> One thing I have said in the past that I will say again here is that early career researchers should not be afraid to publish openly, but should also not be ashamed if they encounter obstacles. This is especially apparent in group projects. It is always possible and desirable to push for open access and open scholarship from within any team, organisation or committee that you are a part of, but sometimes you do not get to make the decision. This is fine, and not something to be ashamed of. That said, it is very important for older scholars to listen when their junior colleagues are advocating for ambitious open solutions, and to do their best to assist them. Everyone will win.

Sharing publications in a broad sense

Do you have publications/digital scholarly objects that cannot be placed on a bookshelf?
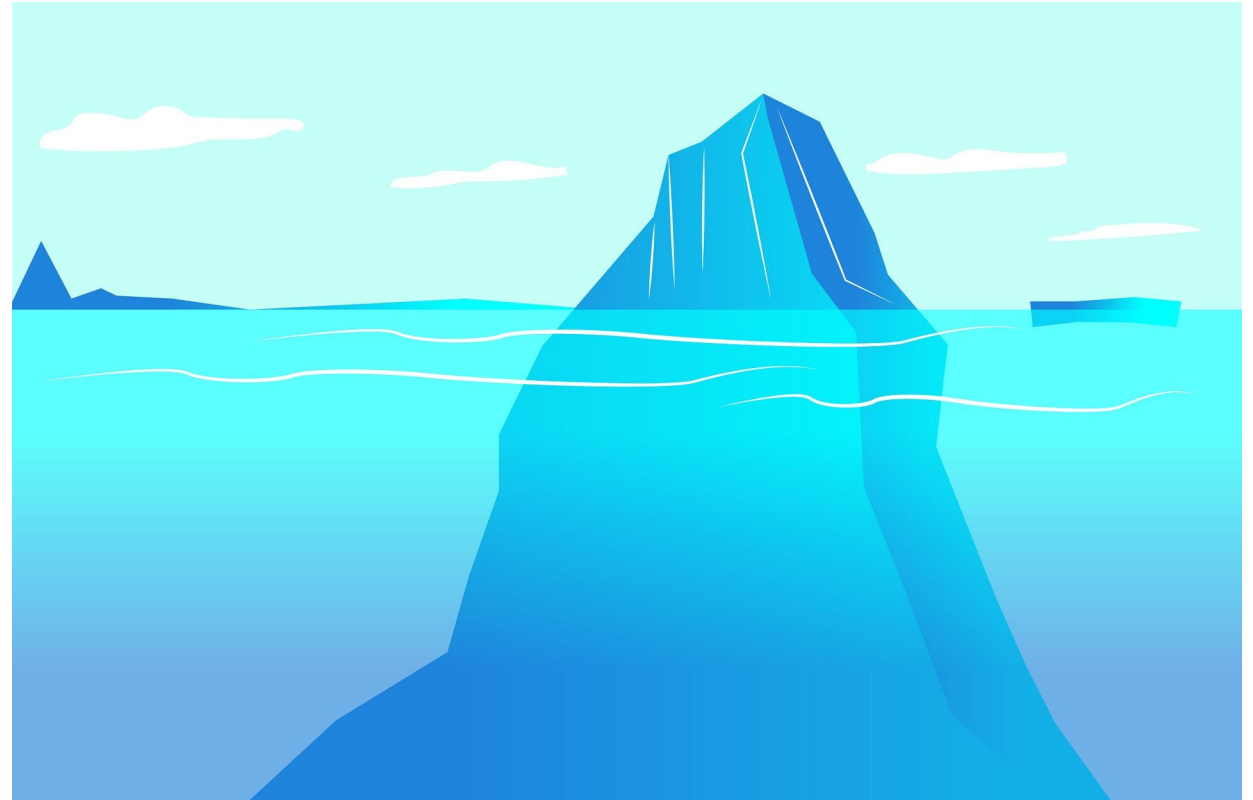
Source: Pixabay, CC0.

Ready, set, explore, reuse!
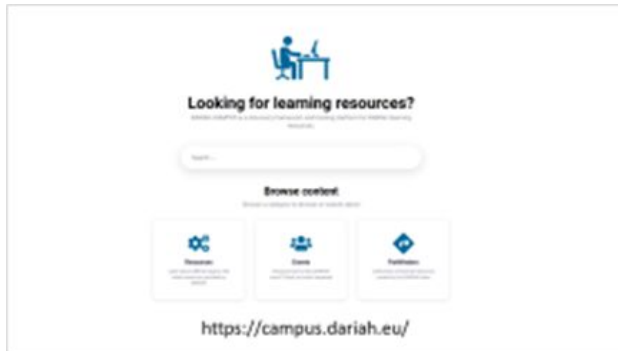
Training materials on [DARIAH Campus](https://campus.dariah.eu/) (contains Open Science!)
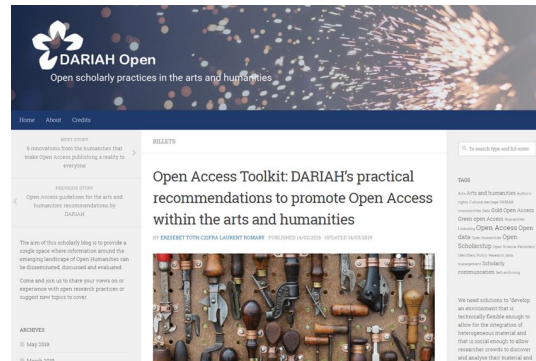
[DARIAH Open blog](https://campus.dariah.eu/)

[OpenMethods](#) platform
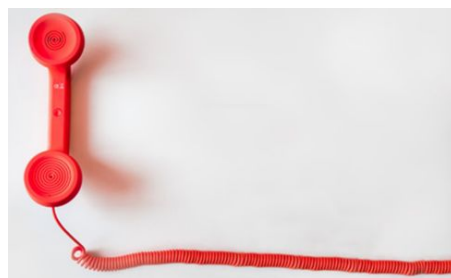
Open Science [Helpdesk](#)

Annual [bursary](#) for first Open Access monographs in Digital Humanities

DARIAH Research Data Management [Working Group](#)

Further useful resources

Epistemic cultures matter!

See also: the FOSTER Open Science taxonomy:

Running order

1. Special flavours of Open Science in the Arts and Humanities disciplines

2. What are we talking about when talking about humanities data?
    Exercise 1: What are your research data?
    Exercise 2: What are others' research data?

3. Traveling through the research data life cycle
    Exercise 1: What makes a data set well documented?
    Exercise 2: Where can I find relevant data sets?
           (test drive with OpenAIRE)

You are more than welcome to raise your hand or use the chat throughout the presentation.

Epistemic cultures matter!

Open Access books?

Small disciplines

Qualitative research

Access to cultural heritage data?

Multilingualism?

What all this tells us about the nature of research data in the humanities?

- Multilingualism.

- Humanities research lives from enrichment of data (**layers of interpretation**) □ data curation happens in a cotinuum: the way cultural heritage resources are made available form a continuum with layers of analysis based on them.

- Problematic to **distinguish between primary data** (raw data) **and secondary data.**

- **Access to machine readable artifacts and digital collections** is crucial

- **Shared ownership** between data creators, data curators and the human subjects, researchers, Cultural Heritage Institutions and opublishers.

- Layers of analysis are separated by institutional and infrastructural silos and only in the rarest cases can they stay connected with each other.

- Humanities are a **very broad research discipline**, many specific research contexts, but also increasingly interdisciplinary research.

Gaining access to Cultural Heritage resources can get pretty complex though…

„At noon, we arrived in Speyer, where the chapter had already allowed us access to the archives to compare our copies of documents with the originals. We had also been promised that we would receive further material. However, it took us eight days to find out what we wanted to know. Because here it is like everywhere else, nothing happens without a multitude of difficulties. [...] The chairman forced us to dine with him every evening. Only once did we have our peace. We also had to spend every evening with the archivist,who was awarded with a gold medal worth 25 ducats."

Andreas Lamey, 1769, quoted after Voss 2002: Schöpflin, p. 604.

Source: Henrike Hoffmann, Moritz Mähr and Daniela Zetti: How the search enginge came to the historians? Presentation slide from the haring the experience: Workflows for the Digital Humanities workshop, Neuchatel, 06.12.2019.

A data (re)use case study that highlights many of the challenges DH research is facing:

Data Basics With Databases – The Wonders of Data Material

BY **STEFAN KARCHER** · PUBLISHED 2018-10-05 · UPDATED 2018-12-06

Vacation time is over and – which is a bit surprising for PhD students – I had much time to think about my dissertation and my academic future. I did some researches and picked up many new interesting fields someone should deal with. However, from time to time, I got angry about an issue which came across a lot. Have you ever noticed how stupid some subject-related databases are? I mean, having them helped us find sources and texts, which enriched our own work, was a wonderful opportunity 10 years ago; but today, 10 years later the relation between the whole bunch of data sets became the focus of some researchers' attention.

The need (or wish) to work with the full material of a database leads to the wide discussion about Open Access in science. I don't want to deepen the aspect here in this post, but to make my opinion clear: Open Access is a great convenience for an open and free science. I support this position and I think my contribution to the scientific society is to share research results and raw data in an easy, accessible way. Yet, I understand the need of legal restrictions, payment, and license policies. Publishers, universities, and scholars invest so much time and money to develop their systems; and, of course, charges and copyright restrictions are necessary for financial profitability. I'm happy to pay for access if that's the price to support scientific progress. One can't be so naive to think open science is for free. Someone must pay for it and even my time/work as a PhD costs resources and money. To make a long story short: If a database is hidden behind payment or license restrictions, there should be good reasons for it. And I hope the reason is not profit, but the necessity to keep the system running.

In this context, it is very important to differ between two types of access. The

InFoDiTex was nominated for a DH Award 2018 in the category "Best Use of DH For Public Engagement." We are very happy about our placement as "2nd Runner Up" and we would like to thank all those who voted for us!

Congratulations to **the winners and all the other great DH projects!**

UPCOMING EVENTS

InFoDiTex
INTERDISZIPLINÄRES FORUM DIGITALER TEXTWISSENSCHAFTEN
JUNIOR RESEARCH INFRASTRUCTURE

ABOUT & CONTACT

The *Interdisciplinary Forum of Digital Textual Sciences* at the University of Heidelberg is an open meeting for (junior) researchers in all fields of Digital Humanities. It was founded by doctoral students who meet every month during the semester turn for

Stefan Karcher, "Data Basics With Databases – The Wonders of Data Material," in *INFODITEX -BLOG*, 2018-10-05, https://infoditex.hypotheses.org/245.

" During vacation, I had an idea for a future project for which data from a database is needed. Some key features and search functions in the text sources of the database are available online for free, but it's not possible (and maybe illegal, too) to parse it with a script. That is why I made something silly: I told my idea to a responsible person and asked for raw data, plain texts, and license policies. (I will describe the project idea in another post if everything works as expected). **Within some days, I received an answer: they will not confer about IF they grant me access, but about HOW they can do it! Let's dig up the treasure.***"