Digital Humanities Research Initiative

Building Digital Humanities

Libraries as hubs? The role of libraries in building DH, *10th November 2022*

Libraries as next generation data infrastructures: conceptualising ‘Collections as Data’ and ‘Data Spaces’

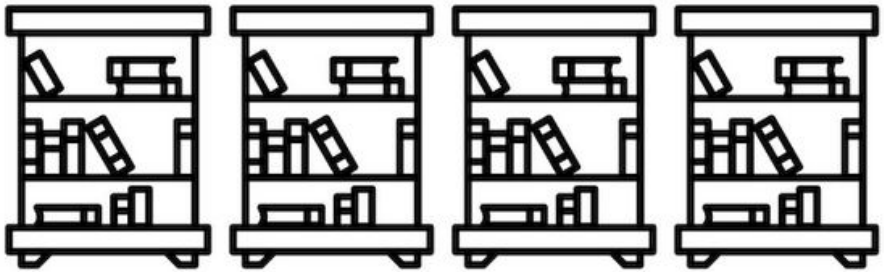
Sally Chambers, KBR, Royal Library of Belgium,
Ghent Centre for Digital Humanities and DARIAH-EU

GHENT UNIVERSITY

KBR Where time is treasured

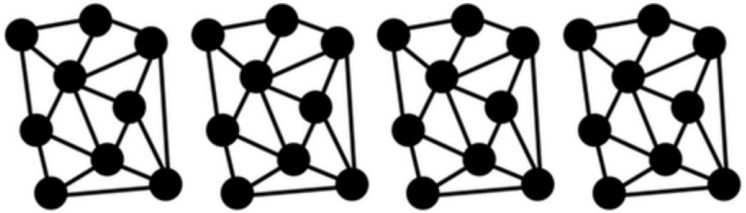
DARIAH-EU
Digital Research Infrastructure for the Arts and Humanities

Collections as Data: “Always Already Computational”



… and “Part to Whole”

Padilla, Thomas, Allen, Laurie, Frost, Hannah, Potvin, Sarah, Russey Roke, Elizabeth, & Varner, Stewart. (2019, May 22). Final Report --- Always Already Computational: Collections as Data (Version 1). Zenodo. http://doi.org/10.5281/zenodo.3152935



Always Already Computational: Collections as Data final report and project deliverables: https://osf.io/mx6uk/wiki/home/

https://collectionsasdata.github.io/part2whole/

Newspapers as Data

Bibliothèque nationale du Luxembourg
Open Data



https://data.bnl.lu/data/historical-newspapers/

STARTER PACK

250MB

of digitised newspapers

- ✓5 days of news
- ✓5 newspaper issues
- ✓22 pages
- ✓D'Wäschfra (1868)
- ✓Public Domain, CC0 (See copyright notice)
- ✓Best for getting started & developing

DOWNLOAD (ZIP)

TEXT ANALYSIS PACK

2GB

of processed newspapers data

- ✦38 years of news (1841-1878)
- ✦23663 processed newspaper issues
- ✦97285 processed pages
- ✦510505 extracted articles
- ✦Public Domain, CC0 (See copyright notice)
- ✦Best for getting started with text analysis

DOWNLOAD (ZIP)

GROUND TRUTH PACK

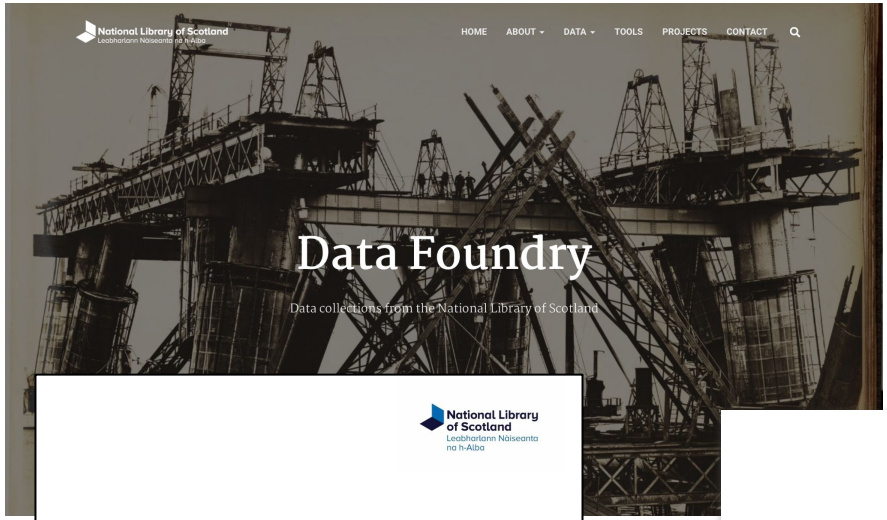
33.000

transcribed text lines

- ✦Text line based OCR
- ✦19.000 text lines in Antiqua
- ✦14.000 text lines in Fraktur
- ✦Transcribed using double-keying (99.95% accuracy)
- ✦Public Domain, CC0 (See copyright notice)
- ✦Best for training an OCR engine

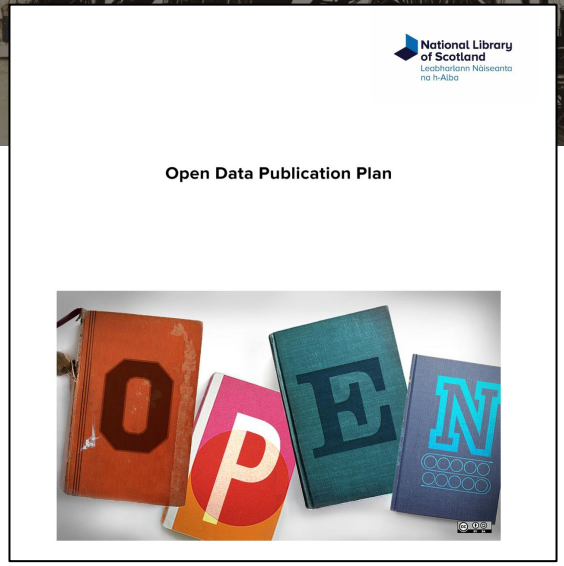
DOWNLOAD (ZIP)

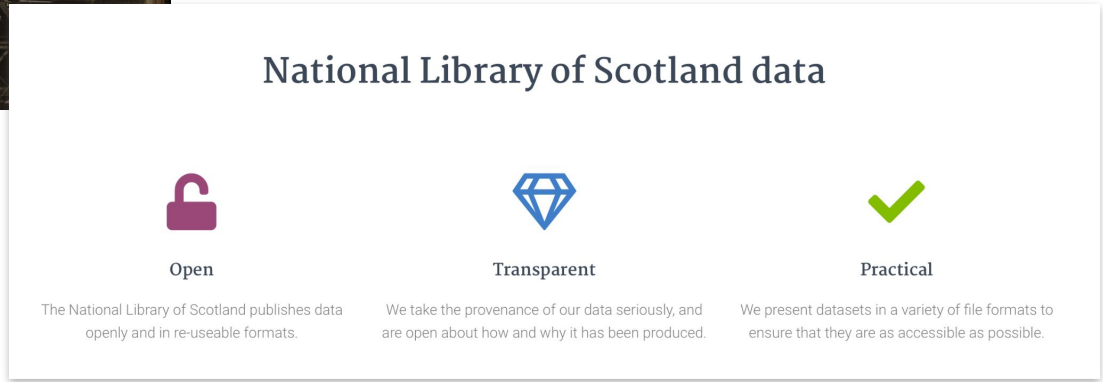
National Library of Scotland: Data Foundry



Ames, S. (2021) "Transparency, Provenance and Collections as Data: The National Library of Scotland's Data Foundry." LIBER Quarterly 31 (1): 1–13. https://doi.org/10.18352/lq.10371.

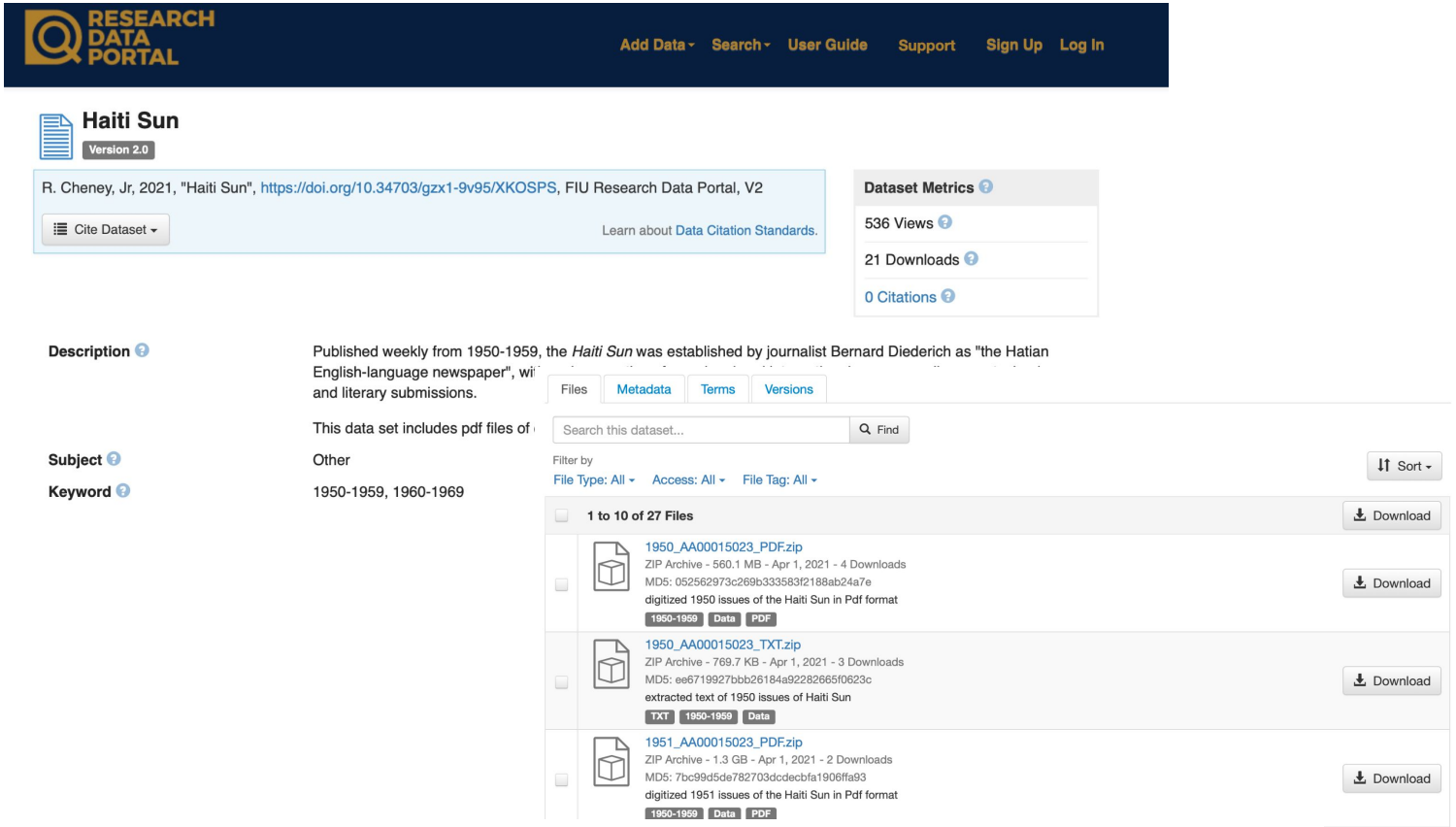
https://data.nls.uk





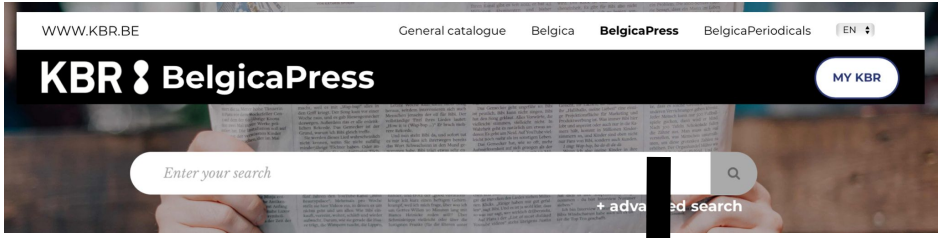
https://data.nls.uk/download/national-library-of-scotland-open-data-publication-plan.pdf

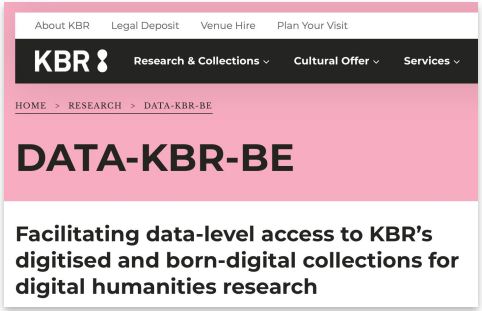
Digital Library of the Caribbean as Data

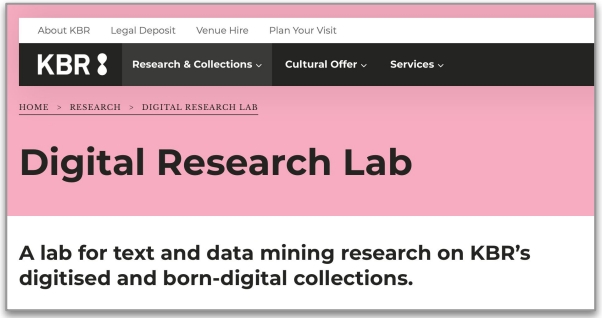


https://dataverse.fiu.edu/dataverse/dloc-as-data

Collections as Data @ KBR, Royal Library of Belgium



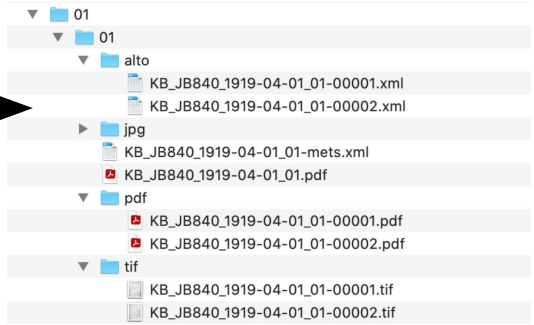




https://www.kbr.be/en/projects/digital-research-lab/

https://www.kbr.be/en/projects/data-kbr-be/





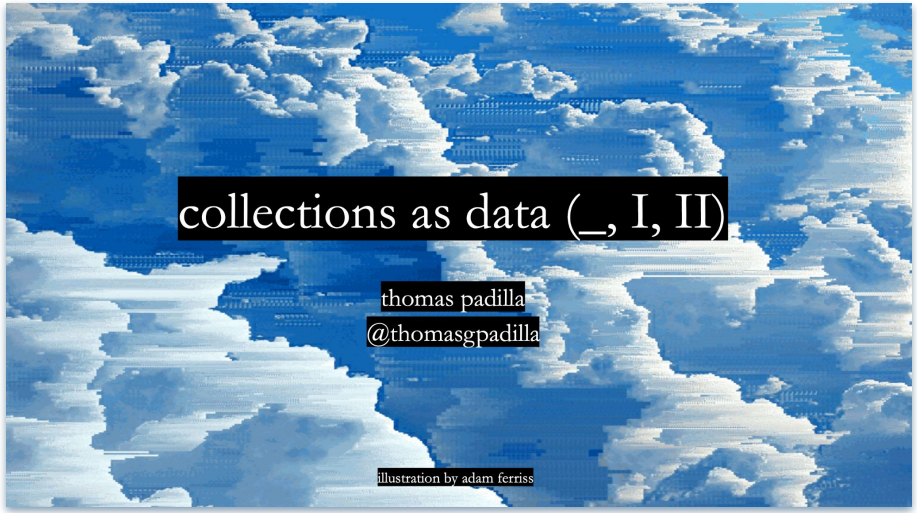
GHENT UNIVERSITY

KBR

Findable Accessible Interoperable Reusable

Europeana Research Community Café - Collections as Data





https://pro.europeana.eu/event/europeana-research-community-cafe-collections-as-data

Thomas Padilla's presentation

https://docs.google.com/presentation/d/1XM71isRXEjKesaWPZjSrZiKytq2XC2QJ-gqKWZF

GHENT UNIVERSITY

International GLAM Labs Community

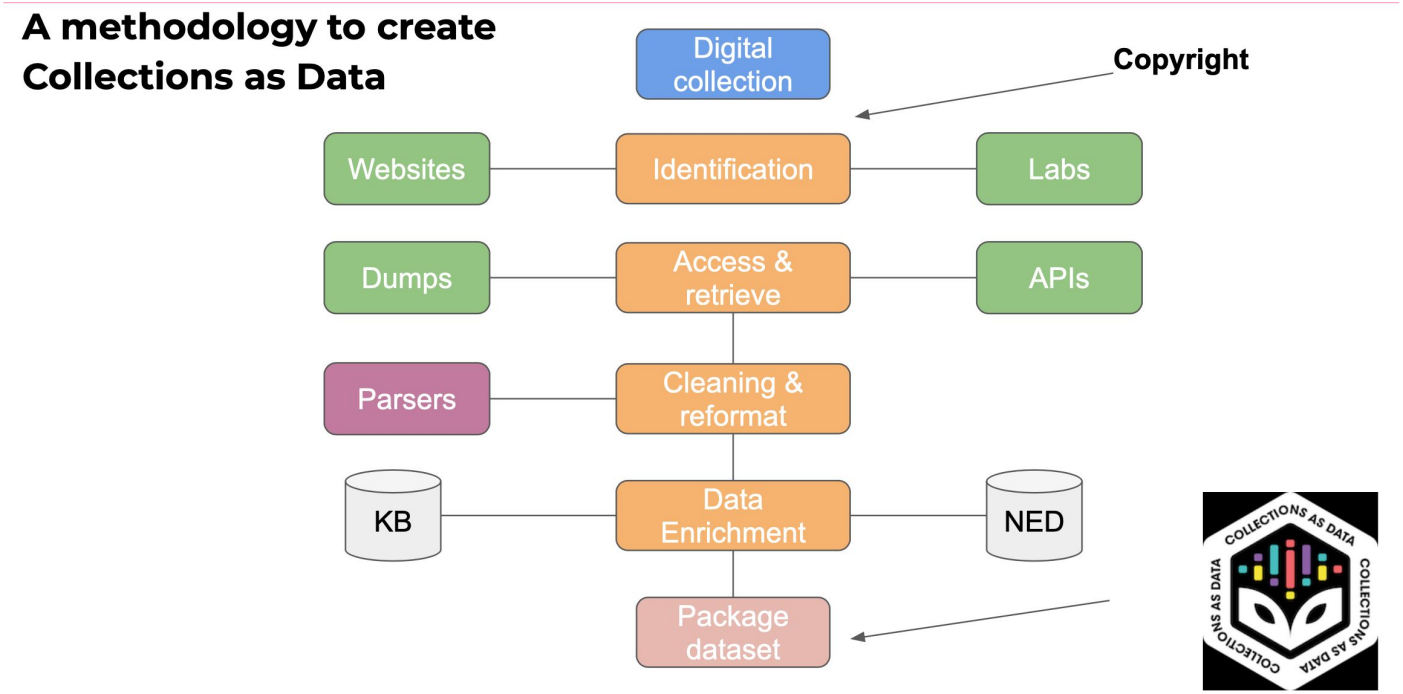




https://glamlabs.io/books/

Chapter: Rethinking Collections as Data:
https://glamlabs.pubpub.org/pub/urhm68yr/release/1

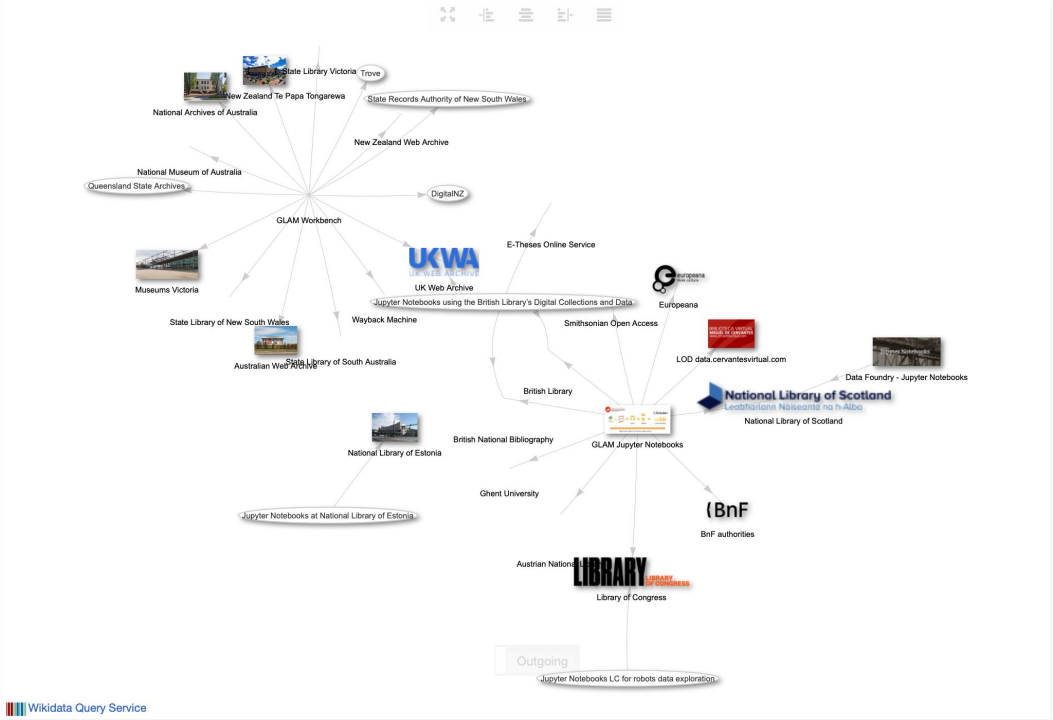
Collections as Data Methodology



Candela, G., Sáez, M. D., Escobar Esteban, Mp., & Marco-Such, M. (2020). *Reusing digital collections from GLAM institutions.* Journal of Information Science: https://doi.org/10.1177/0165551520950246 & http://rua.ua.es/dspace/handle/10045/109460

Computational access to digital collections

International GLAM Labs Community



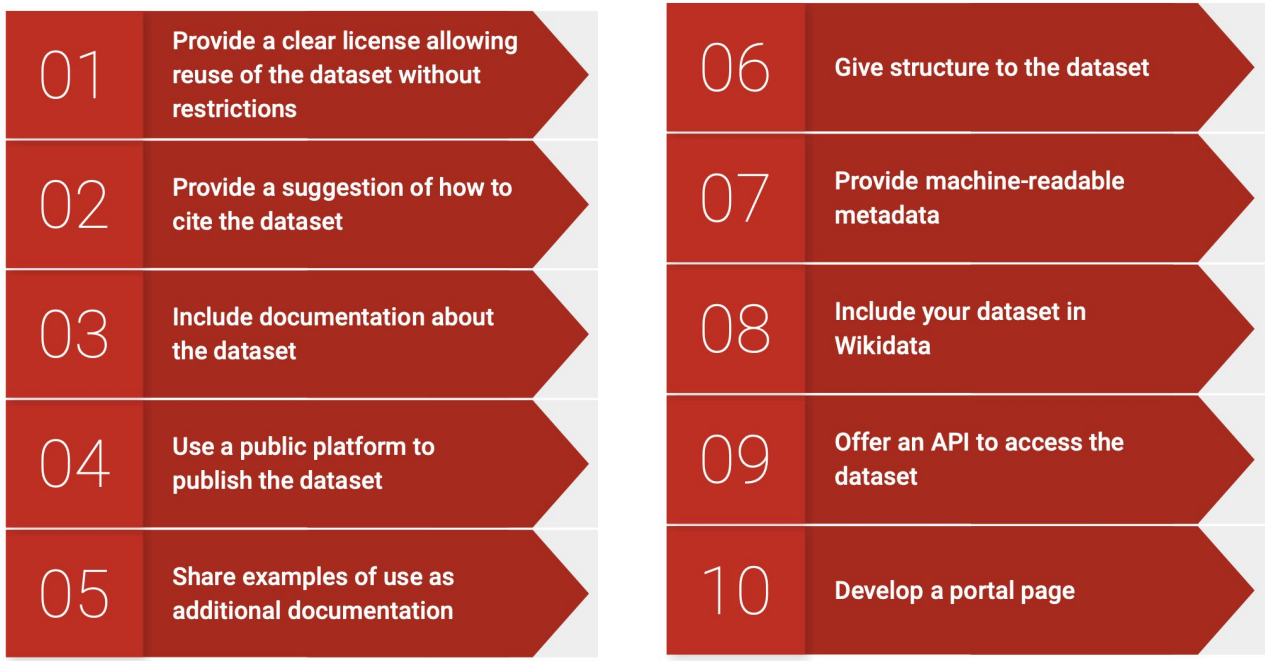
International GLAM Labs Community: Computational Access to Digital Collections

https://glamlabs.io/computational-access-to-digital-collections/

Towards a Collections as Data Checklist

International GLAM Labs Community

This approach provides an easy to apply method to encourage small and medium-size org to publish their digital collections as Collections as data.

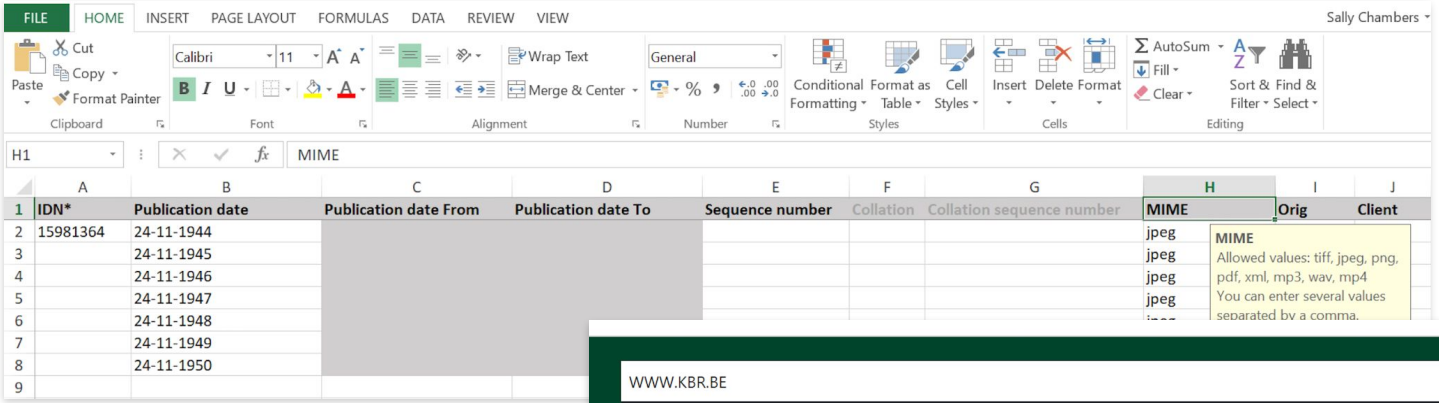


With special thanks to Gustavo Candela (Alicante, Spain) and Nele Gabriëls (Leuven, Belgium)

Methods for Data Extraction

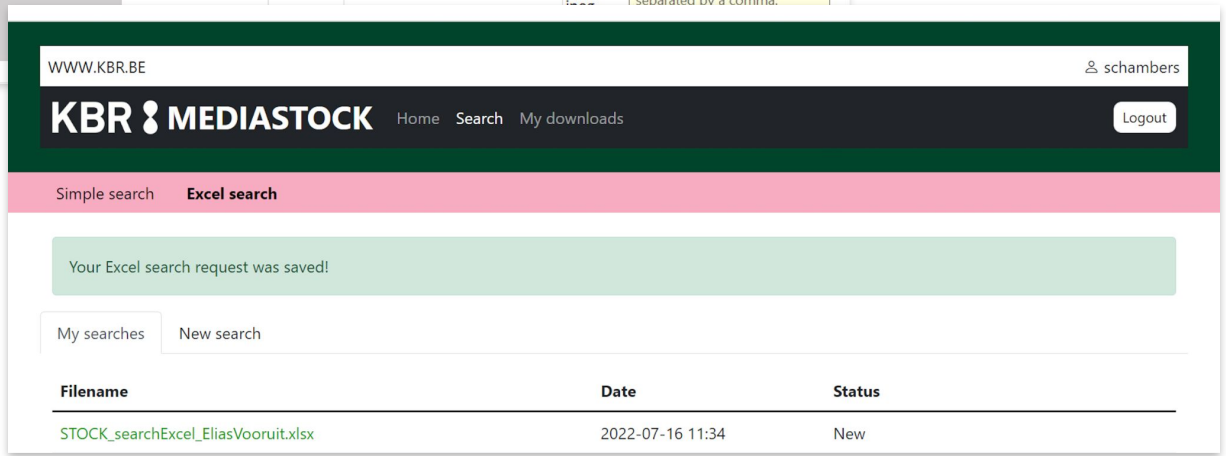
KBR Where time is treasured

Vooruit, 24 November, 1944 - 1950



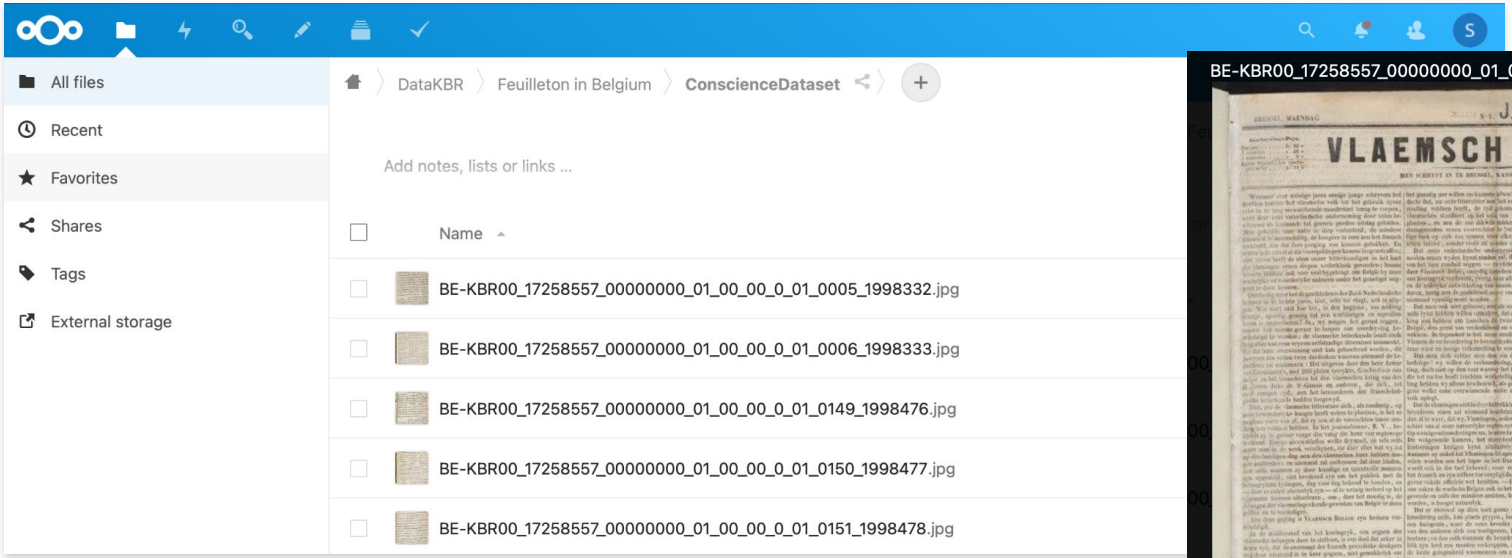
	A	B	C	D	E	F	G	H	I	J
1	IDN*	Publication date	Publication date From	Publication date To	Sequence number	Collation	Collation sequence number	MIME	Orig	Client
2	15981364	24-11-1944						jpeg		
3		24-11-1945						jpeg		
4		24-11-1946						jpeg		
5		24-11-1947						jpeg		
6		24-11-1948								
7		24-11-1949								
8		24-11-1950								
9										

MIME
Allowed values: tiff, jpeg, png, pdf, xml, mp3, wav, mp4
You can enter several values separated by a comma.



Methods for Data Sharing: Next Cloud

KBR Where time is treasured

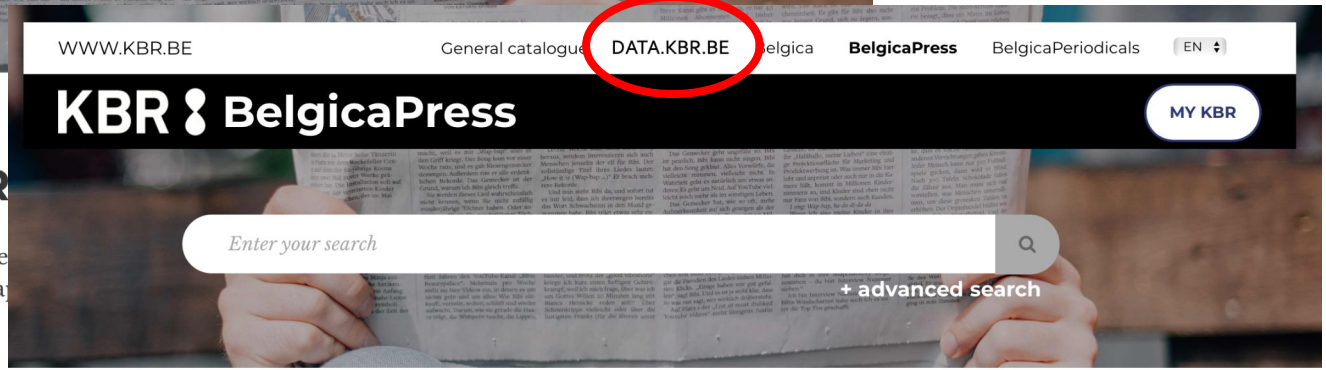




GHENT UNIVERSITY

Towards the ‘DATA-KBR-BE’ platform





Search in KBR

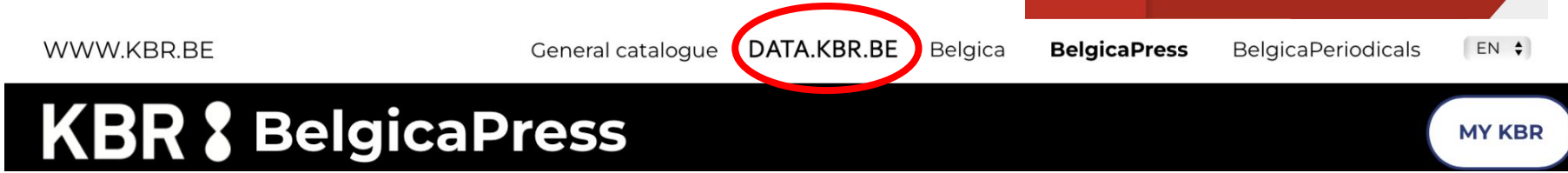
The portal **BelgicaPress** give
newspapers", most of which a

Search in KBR's digitised newspapers

The portal **BelgicaPress** gives you access to **121** Belgian newspapers referred to as "major newspapers", most of which appeared daily (**1814-1970**).

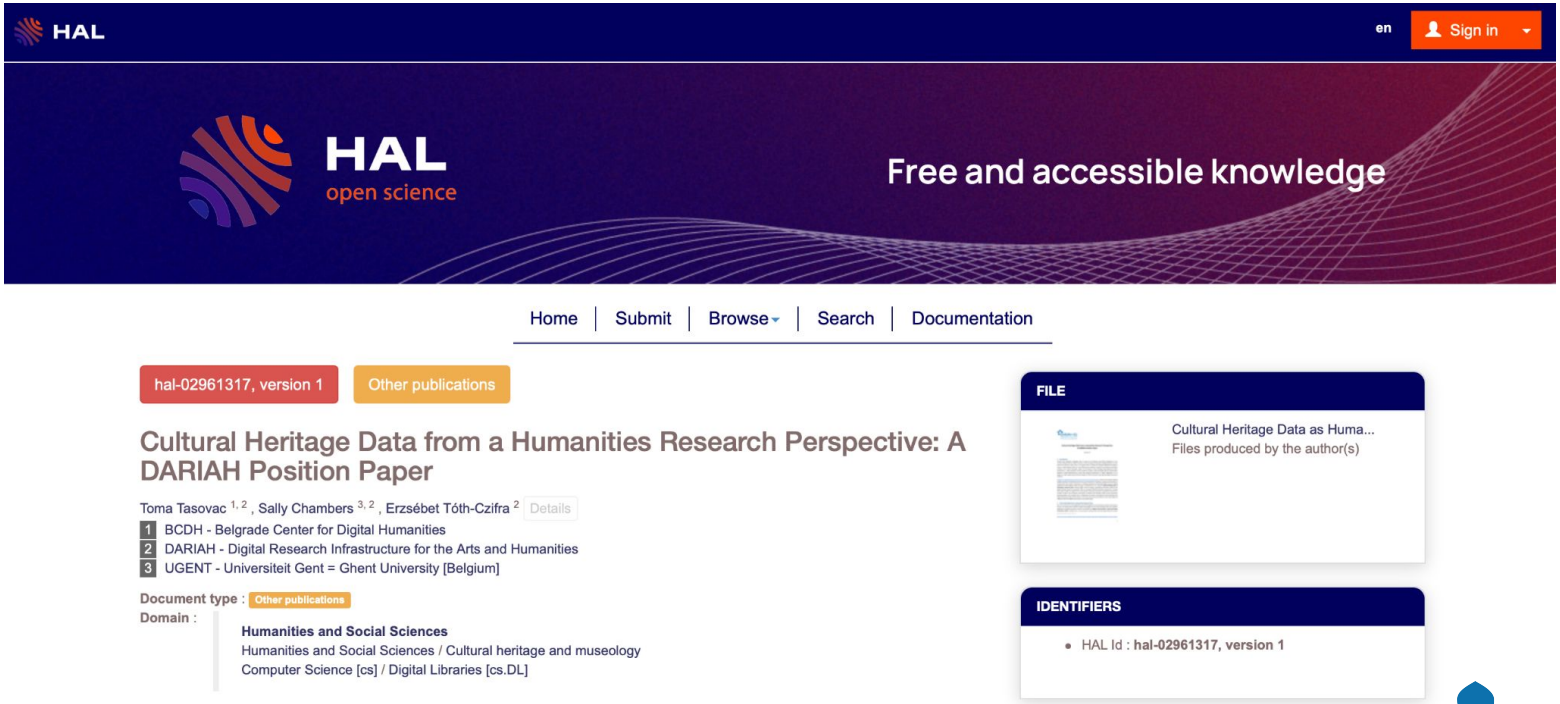
Towards the ‘DATA-KBR-BE’ platform

10 Develop a portal page



- **Sustainably publish curated cultural heritage datasets**
  - Facilitating and documenting the dataset building process (**also: KBR Virtual Lab**)
  - Research Data Management (FAIR) Best Practices
- **Data Formats => Checklist # 6**
  - For the ‘objects’ in the dataset (e.g. ALTO.xml; METS.xml, TIFF, .jpg, .pdf)
  - For the dataset itself (e.g. RO-Crate?)
- **Data Licensing** (ideally not public domain only) **=> Checklist # 1**
- **Data Citation** (easily citable) **=> Checklist # 2**
- **Publication:** Webpage? Searchable? Repository? Versioning? **=> Checklist # 4**
- **Data sharing:** KBR Data Send + Researcher’s cloud-based system (e.g. GhentCDH NextCloud)

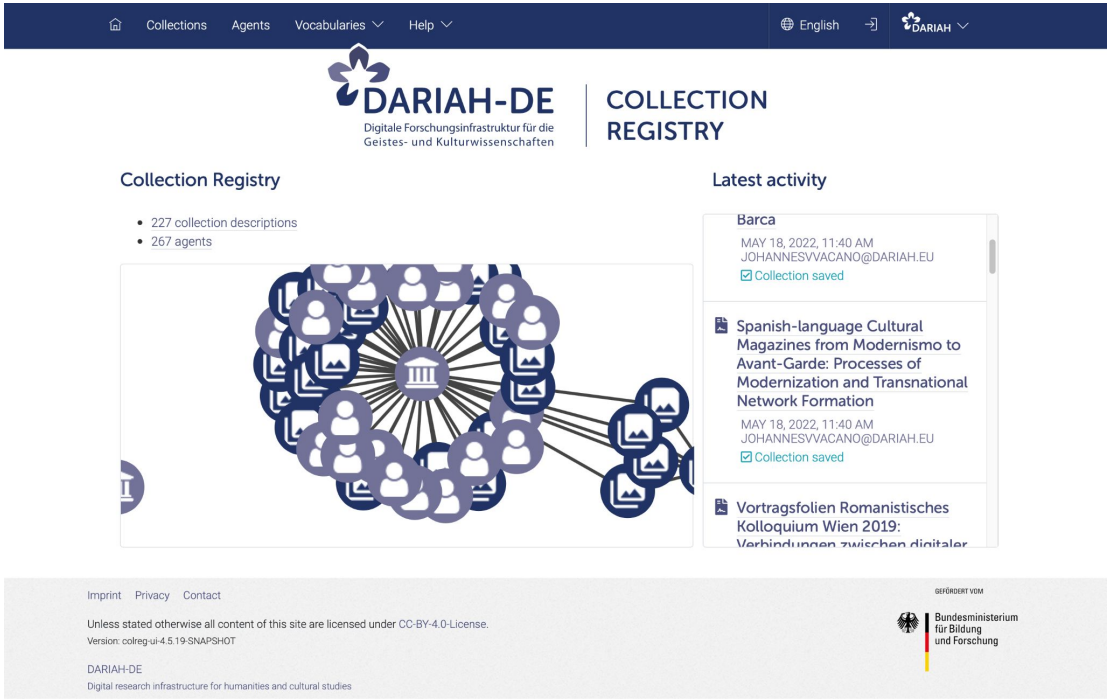
Cultural Heritage Data is Humanities Research Data



Tasovac, T., Chambers, S. and Tóth-Czifra, E. (2020) Cultural Heritage Data from a Humanities Research Perspective: A DARIAH Position Paper. https://hal.archives-ouvertes.fr/hal-02961317

DARIAH-EU
Digital Research Infrastructure for the Arts and Humanities

DARIAH ‘Collections as Data’ ?



DARIAH-EU
Digital Research Infrastructure for the Arts and Humanities

Findable Accessible Interoperable Reusable

COLLECTIONS AS DATA

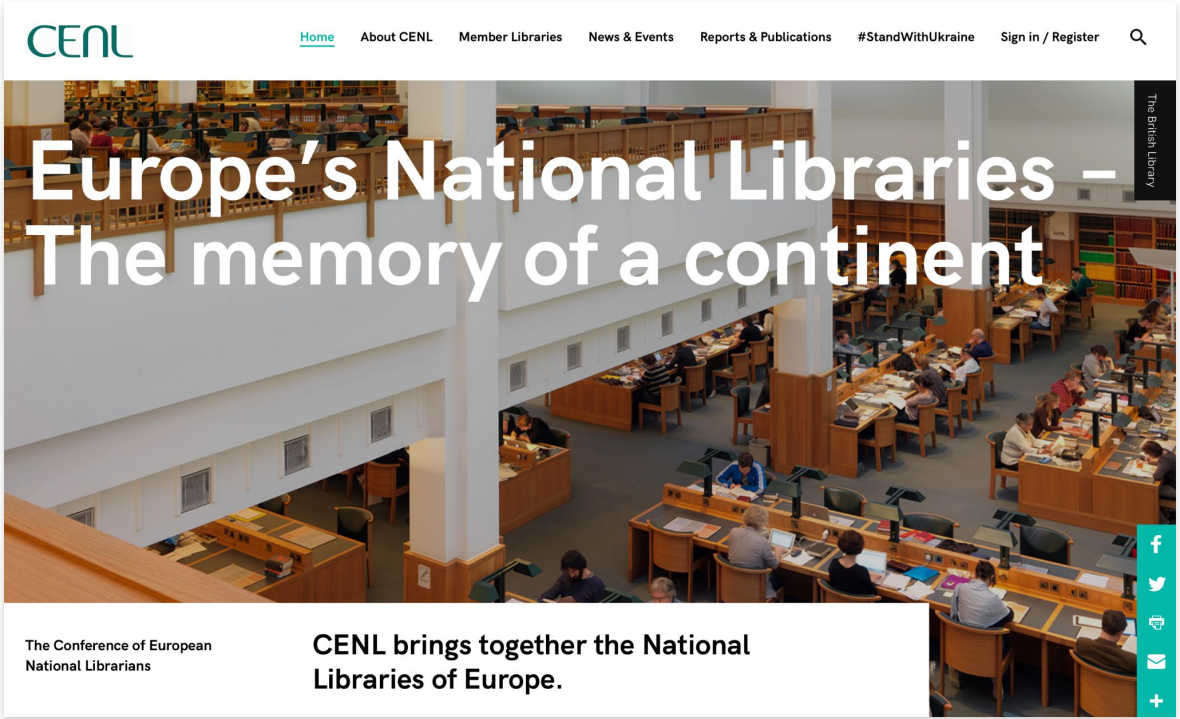
https://colreg.de.dariah.eu/colreg/?lang=en

CLARIN 'Collection as Data'?



https://www.clarin.eu/resource-families

Towards a CENL Dialogue Forum: Libraries as Data Infrastructure?

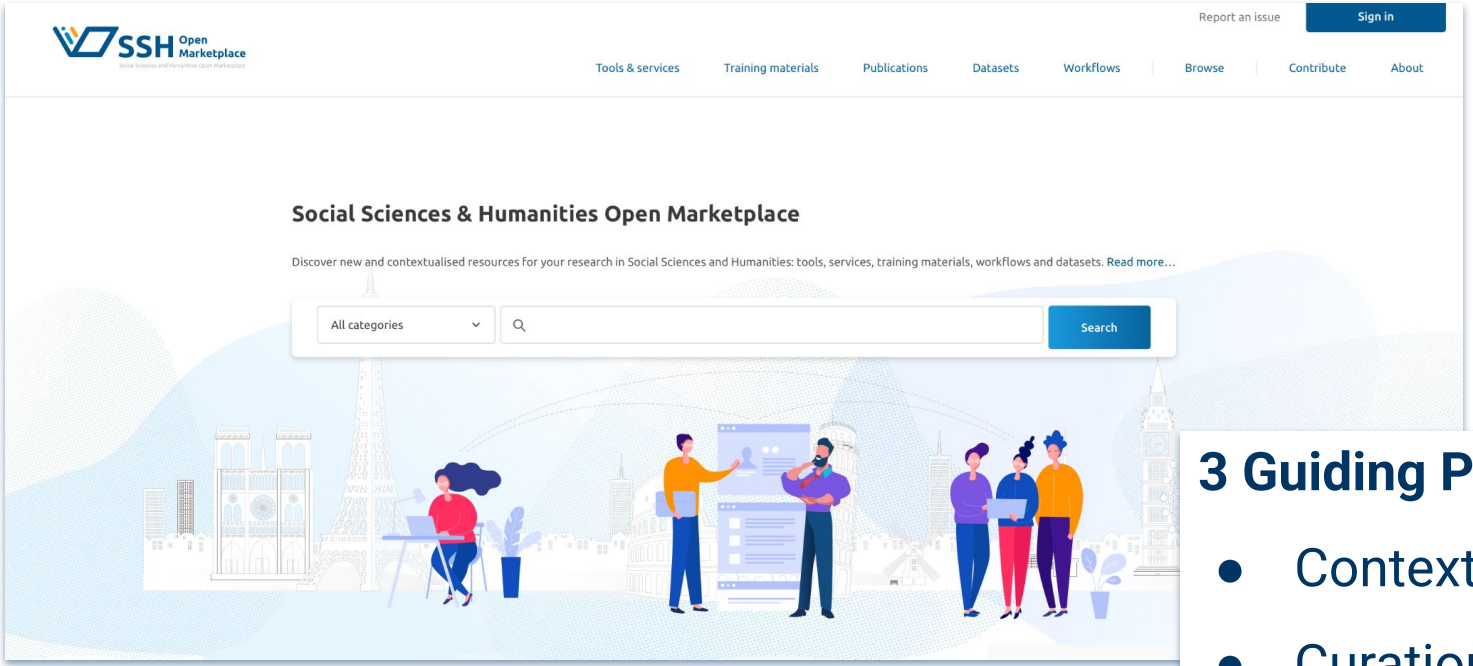


https://www.cenl.org/

DARIAH-EU
Digital Research Infrastructure for the Arts and Humanities

CLARIN

Social Sciences and Humanities Open Marketplace



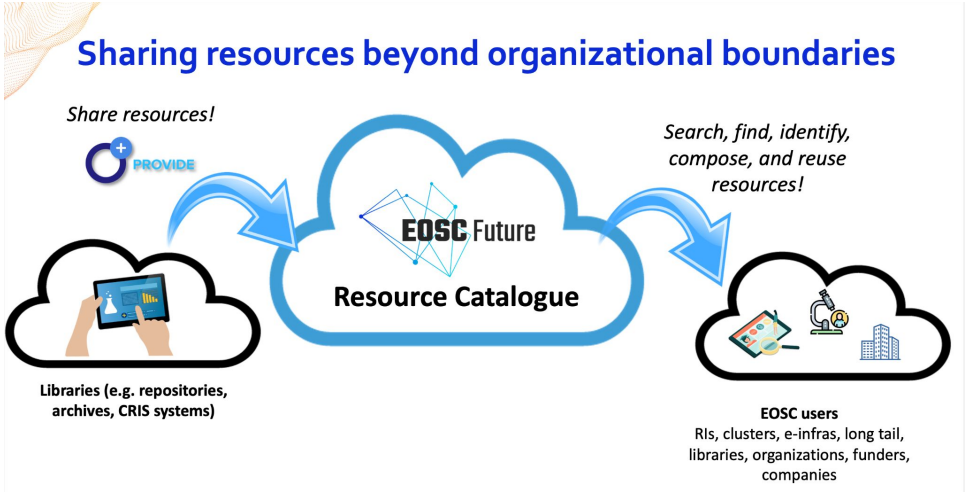
https://marketplace.sshopencloud.eu

3 Guiding Principles:

- Contextualisation
- Curation
- Community

European Open Science Cloud (EOSC) and Libraries

EUROPEAN OPEN SCIENCE CLOUD



Libraries as EOSC providers

How and why libraries should become EOSC providers – Inge Van Nieuwerburgh, University of Gent

Current and upcoming functionalities for libraries – Paolo Manghi, OpenAIRE

A community presentation from Susanne Blumesberger, University of Vienna Library & Lisa Hönegger, AUSSDA

Libraries as intermediaries – Sara Garavelli

Presentation – recording

Presentation:
https://eoscfuture.eu/wp-content/uploads/2022/04/ProviderDays_Libraries_Presentation2Inge_clean.pdf

Recording: https://www.youtube.com/watch?v=_vxHJNU78Pw

Towards a European Data Space for Cultural Heritage

European Commission

Search

Shaping Europe's digital future

Home Policies Activities News Library Funding Calendar Consultations

Home > News & Views > Commission proposes a common European data space for cultural heritage

DIGIBYTE | Publication 10 November 2021

Commission proposes a common European data space for cultural heritage

Today, the European Commission publishes a recommendation on a common European data space for cultural heritage. The aim is to accelerate the digitisation of cultural heritage assets.

EUROPEANA FOUNDAT

The objectives

The Commission recommends Member States to accelerate the digitisation of all cultural heritage monuments and sites, objects and artefacts for future generations, to protect and preserve those at risk, and boost their reuse in domains such as education, sustainable tourism and cultural creative sectors.

Related topics

Supporting media and digital culture

Digital Cultural heritage

europeana pro ABOUT US JOIN THE NETWORK THE EUROPEANA WEBSITE SHARE YOUR DATA Search for ...

Projects data-space

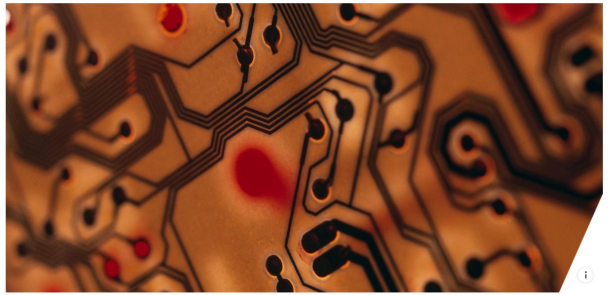


Harry Verwayen

General Director , Europeana Foundation

Europeana wins tender to deploy the common European data space for cultural heritage

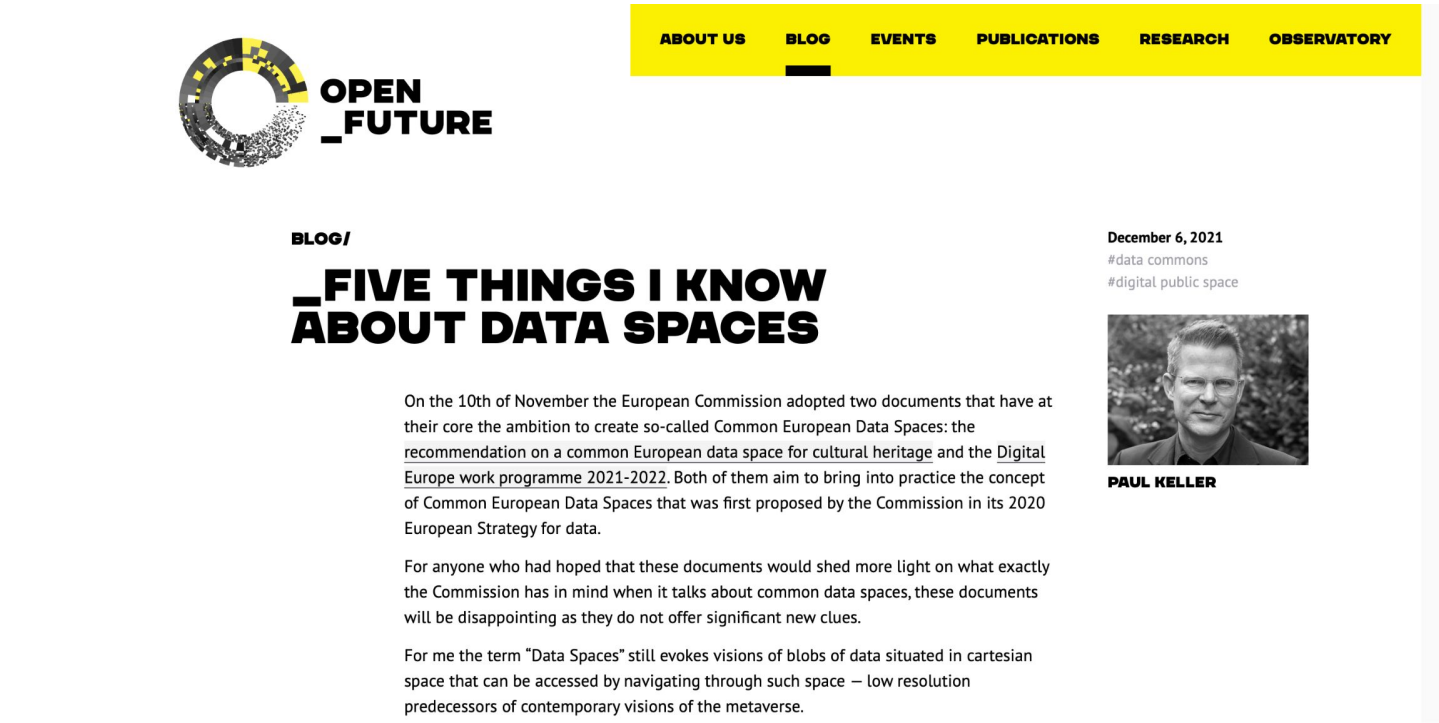
We are delighted that in the coming years, Europeana will be at the heart of the European Commission's data space for cultural heritage, continuing our work together with 18 partners and a great network, to empower the cultural heritage sector in its digital transformation.



https://digital-strategy.ec.europa.eu/en/news/commission-proposes-common-european-data-space-cultural-heritage

https://pro.europeana.eu/post/europeana-wins-tender-to-deploy-the-common-european-data-space-for-cultural-heritage

“5 things I know about Data Spaces …”



OPEN _FUTURE

ABOUT US | BLOG | EVENTS | PUBLICATIONS | RESEARCH | OBSERVATORY

BLOG/

_FIVE THINGS I KNOW ABOUT DATA SPACES

December 6, 2021

#data commons

#digital public space

PAUL KELLER

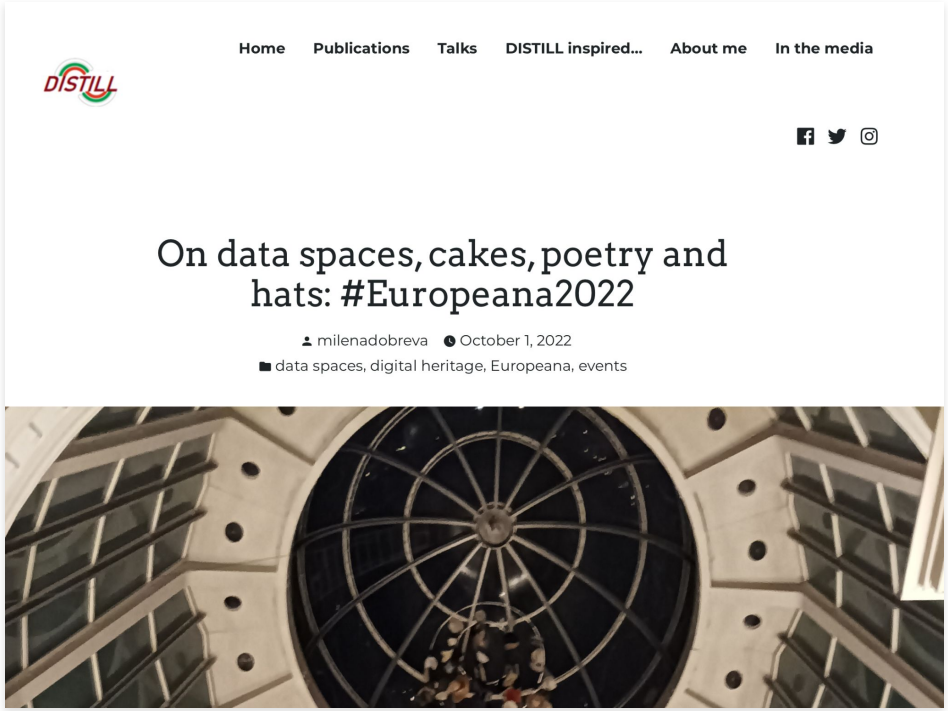
On the 10th of November the European Commission adopted two documents that have at their core the ambition to create so-called Common European Data Spaces: the recommendation on a common European data space for cultural heritage and the Digital Europe work programme 2021-2022. Both of them aim to bring into practice the concept of Common European Data Spaces that was first proposed by the Commission in its 2020 European Strategy for data.

For anyone who had hoped that these documents would shed more light on what exactly the Commission has in mind when it talks about common data spaces, these documents will be disappointing as they do not offer significant new clues.

For me the term “Data Spaces” still evokes visions of blobs of data situated in cartesian space that can be accessed by navigating through such space — low resolution predecessors of contemporary visions of the metaverse.

https://openfuture.eu/blog/five-things-i-know-about-data-spaces/

The bold new world of data spaces …



- **Where is actually this data space?** How are users 'entering' there – do they go anywhere specific, or will it underpin our web searches on any platforms we might be using (as far as we access the web the way we do today)?
- **What is expected the user to know for using the dataspace?** If the space allows the use of intelligent tools, does this mean that the successful users will be some sort of geeks or data scientists?
- **How the current content providers will contribute to this dataspace –** should they change something in their infrastructures? Are any new providers going to emerge? What would be their roles?
- **Which institutions will be able to act as data owners?** What conditions will guarantee an institution will be successful as a data owner contributing to the data space? Are the small institutions at a disadvantage here?
- **How many intermediaries – data providers – will be needed** and what exactly are they going to do? Is Europeana of today a particularly strong data provider which can grow further into becoming the data space enabler?
- **What technologies can be orchestrated within the data spaces** to support the mixing, matching, reuse and packaging of results for the users?
- **What will be the right implications of mixing objects from different data owners and data providers?** How will rights be inherited? How the data space will identify cases where the current rights' statements are problematic?
- **What platforms and tools** are already there to help develop the data space? What will be the underpinning architecture?

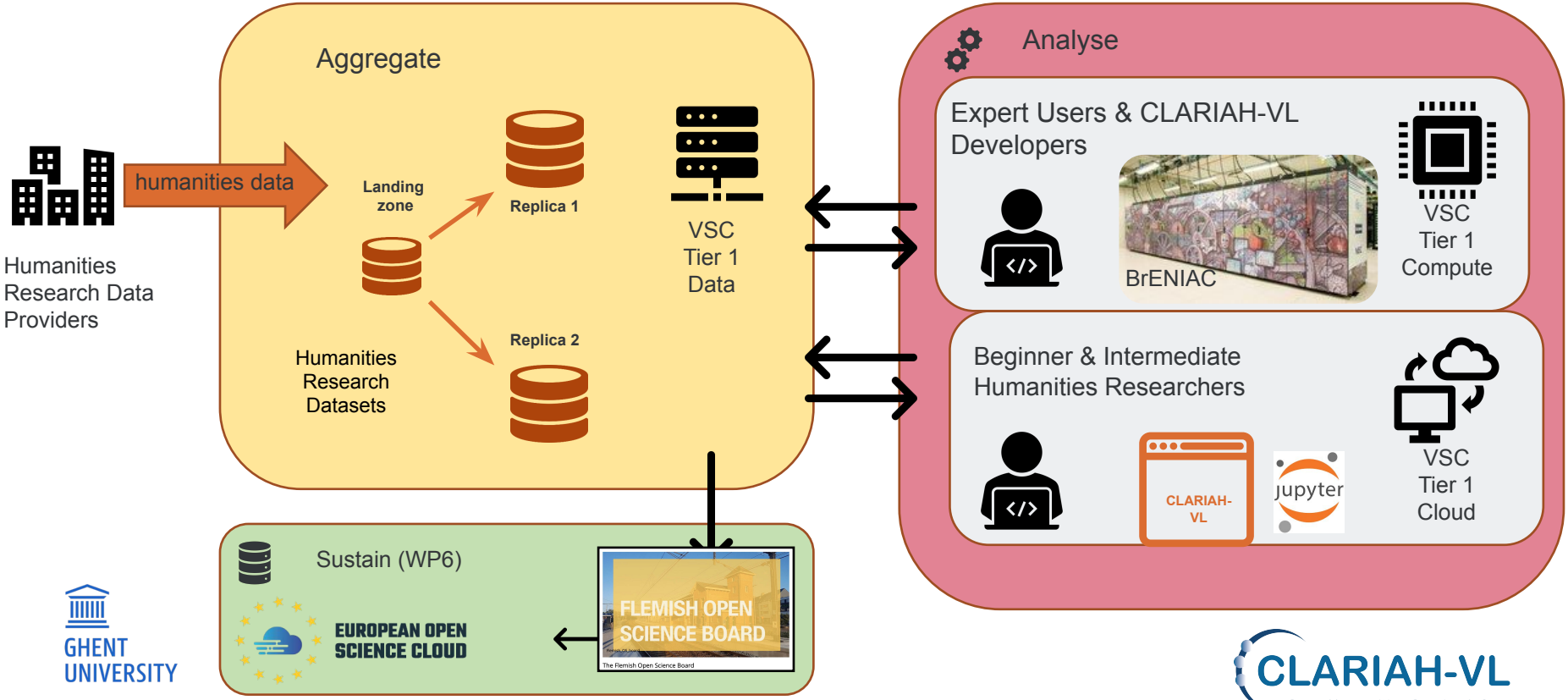
https://distill.page/2022/10/01/on-data-spaces-cakes-poetry-and-hats-europeana2022/

Are there Data Engines in Data Spaces?



https://www.acd-engine.org/

Data-driven research in the arts and humanities



GHENT UNIVERSITY

CLARIAH-VL
Open Humanities Service Infrastructure

DARIAH-EU
Digital Research Infrastructure for the Arts and Humanities

Cultural Heritage Data as Humanities Research Data?



Save the date for the DARIAH Annual Event 2023

By Eliza Papaki | DARIAH News Slides, Home Page Events, News | October 26, 2022

We are excited to share the dates of the DARIAH Annual Event 2023, planned to take place on **June 6th to June 9th** in Budapest, Hungary. The venue for this year's event is Eötvös Loránd University, Faculty of Humanities, Trefort Garden Campus. June 6th will be a day for DARIAH internal meetings, followed by the conference on June 7th to June 9th.

Topic of this year's event is **Cultural Heritage Data as Humanities Research Data?**. Please save the date and stay tuned, we will be publishing more information on the conference and the call for papers in the coming weeks!

Conference themes:

- Sustainable data curation workflows for corpus building
- Experimental Data Spaces for Cultural Heritage
- Advancing digital methods for the analysis of cultural heritage

https://www.dariah.eu/2022/10/26/save-the-date-for-the-dariah-annual-event-2023/

Libraries as hubs? The role of libraries in building DH, *10th November 2022*

Thanks for listening!

Sally Chambers
sally.chambers@kbr.be

GHENT UNIVERSITY

KBR Where time is treasured

DARIAH-EU
Digital Research Infrastructure for the Arts and Humanities