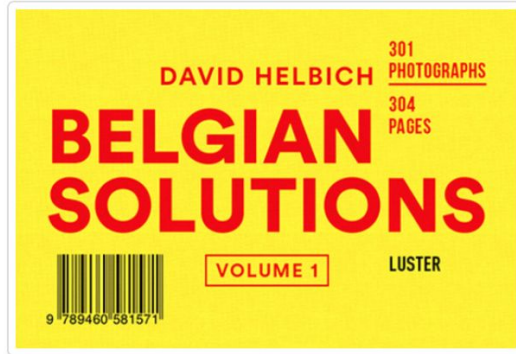


# Interdisciplinary collaboration between CLARIN and DARIAH: a Belgian Solution

Sally Chambers, DARIAH-BE National Coordinator, CLARIAH-VL Project Coordinator  
Ghent Centre for Digital Humanities Ghent University, Belgium



## Belgian Solutions

**Not every solution is an answer to a problem.**

*Belgian Solutions* gathers over 300 images of witty, absurd and at times hilarious hands-on solutions for our everyday environment. The project started in 2006 as an ongoing series of photos by Brussels-based artist David Helbich. Since Helbich posted his pictures on social network platforms in 2009, the project went viral, growing rapidly with contributions posted by Belgian solutions-spotters from around the globe.

This book is the sixth, updated and improved edition of the original, best-selling publication, and now there's also a second volume available: <http://www.lusterweb.com/items/346-belgian-solutions---volume-2?locale=en>.

<https://www.belgian-solutions.be>

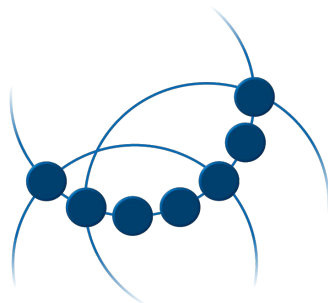
# (Re)introducing CLARIN and DARIAH

CLARIN, a European Research Infrastructure for Language Resources and Technology, makes digital language resources available to scholars, researchers, students and citizen-scientists from all disciplines, especially in the Humanities and Social Sciences.

DARIAH is a pan-European infrastructure that enables excellent research in the Arts and Humanities by exchanging and sustaining tools, services, data and knowledge from its member countries and facilitating the wide uptake of digital methods.

**CLARIN**

Common Language Resources and  
Technology Infrastructure



**DARIAH-EU**

Digital Research Infrastructure  
for the Arts and Humanities

**Concerns:** Request for the setting-up of DARIAH as a European Research Infrastructure Consortium (DARIAH-ERIC)



Dear Mr Genet,

**KBR**

Belgium hereby requests, through the host country of France, the European Commission to set-up the DARIAH research infrastructure as a European Research Infrastructure Consortium with the statutes attached hereto.



The initial commitment as a member will be for the full five years as of the establishment of DARIAH-ERIC.



The representing entity in the DARIAH-ERIC governing bodies will be the *Service public fédéral Politique scientifique* (in English: Belgian Science Policy Office / Acronym: BELSPO).



The National Representatives in the DARIAH-ERIC General Assembly will be *Mr Chris De Loof, attaché "ICT Transversal" at the Service public fédéral Politique scientifique.*



The National Coordinating Institution will be the *Ghent Centre for Digital Humanities (University of Ghent).*



I have read the attached statutes of DARIAH-ERIC and I confirm my approval on them.



Sincerely yours,



Philippe Courard  
Secretary of State for Science Policy




**Belgium, Founding Member of  
DARIAH-EU**





## Members and Partners

[Home](#) > [Network](#)

-  Member Countries
-  Observer Countries
-  Cooperating Partners

### Belgium

Members



NATIONAL  
COORDINATION

PARTNER  
INSTITUTIONS



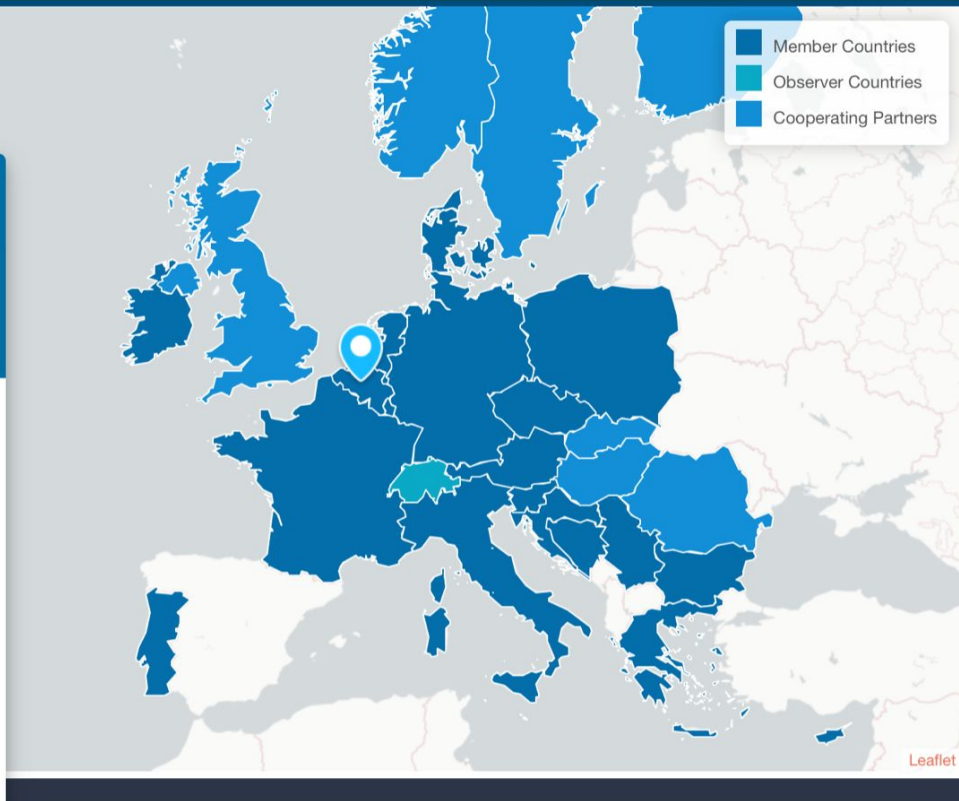
[Visit the DARIAH-BE website](#)

Representing Entity:  
**Belgian Science Policy Office**

National Representative:  
**Chris De Loof**

National Coordinating Institution:  
 [Ghent Centre for Digital Humanities, Ghent University](#)

National Coordinator:  
 **Sally Chambers**



## Ghent Centre for Digital Humanities

### Activities:

- **advice and guidance** throughout the research project lifecycle where digital tools, methods or collections are used with a specific focus on:
  - **collaborative databases**
  - **digital text analysis**
  - **geospatial analysis**
- **research data management and training**



GhentCDH

### Research groups:

- **Inter-faculty Research Centre**
- **Hosted by the Faculty of Arts and Philosophy**
- **Faculty of Arts and Philosophy Library**
- **Ghent University's HPC Infrastructure**



<http://www.ghentcdh.ugent.be>

## University of Antwerp's Platform for Digital Humanities

### Activities:

- **Introductory BA course**
- **MA Track in Digital Text Analysis**
- **Summer School in Digital Humanities**
- **platform{DH} Lecture Series**
- **Antwerp Time Machine**

platform{DH}

### Research Groups:

- **ACDC** (Antwerp Centre for Digital humanities and literary Criticism)
- **CLIPS** (Computational Linguistics and Psycholinguistics)
- **CSG** (Centre for Urban History)
- **University Library**



<http://uahost.uantwerpen.be/platformdh/>

## Instituut voor de Nederlandse Taal

### Activities:

- **Language Technology Services**
- **Depositing service for textual material**
- **Disclosure of linguistic resources** (mainly for Dutch)

### Dutch Language Institute:

- **Bi-national institute** funded by Flemish and Dutch governments
- **CLARIN-B technical centre**
  - for Flanders
  - for the Netherlands

<http://www.ivdnt.org/>



/instituut voor de Nederlandse taal/

## VUB Platform for Digital Humanities (DIGI)

### Activities:

- **Support** digitally-enabled research across the Arts and Humanities
- **focus on**
  - **digital collections**
  - **historical sociolinguistics**
  - **socio-economic history**

### Research Groups:

- **History**
- **Linguistics**
- **Literature**
- **Philosophy**
- **DINF** (University Library Artificial Intelligence Lab)

<https://digi.research.vub.be/home/>



## Leuven Digital Humanities Researchers

### Activities:

- **sharing of knowledge and digital tools**
- **co-organising** DH workshops, lectures, summers schools and conferences
- **long-term preservation strategy** for research data from DH projects at and beyond the university
- **Advanced Master in Digital Humanities**

### Research Groups:

- **Ancient History & Archaeology**
- **Centre for Computational Linguistics**
- **Artes Library (DH Commons initiative)**

<https://www.arts.kuleuven.be/digitalhumanities>




## LARGE SCALE RESEARCH INFRASTRUCTURES IN FLANDERS

FLEMISH PARTICIPATION IN INTERNATIONAL RESEARCH INFRASTRUCTURES 2020




DARIAH: unlocking the scientific potential of digital research in the humanities.

**Social Science and Humanities Research Infrastructures Concertation  
Group (CIS-INFRA-SSH)**

Representing Entity: BELSPO  
National Representative: Chris De Loof  
General Assembly (DARIAH-EU)

**Federal**

**Flanders**

**Federation  
Wallonia-Brussels**

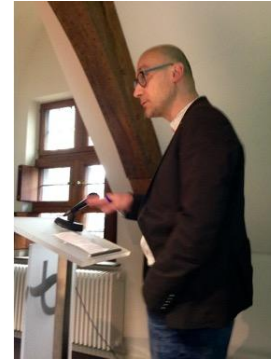


National Coordinating Institution: Ghent University  
National Coordinator: Sally Chambers  
National Coordinator Committee (DARIAH-EU)

**Federal**

**Flanders**

**Federation  
Wallonia-Brussels**





# Common Language Resources Infrastructure in Flanders

New prospects

Vincent Vandeghinste  
Els Lefever  
Walter Daelemans  
Tim Van de Cruys  
Sally Chambers

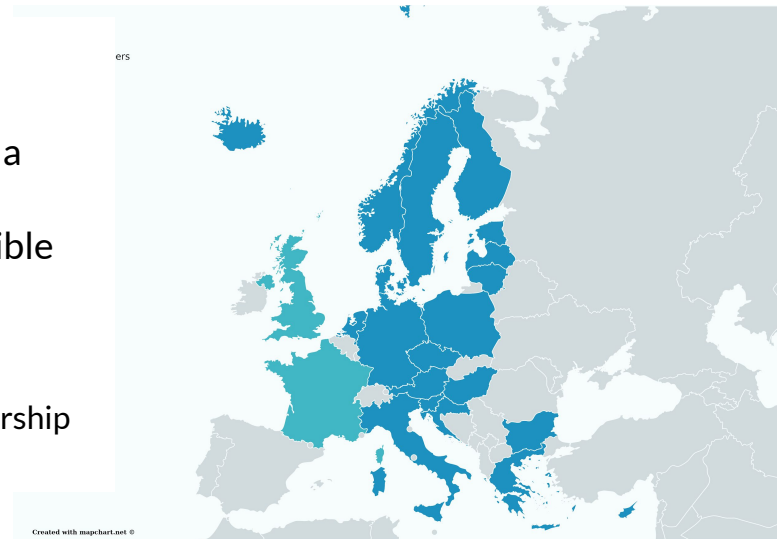
/instituut voor  
de Nederlandse  
taal/



**KU LEUVEN**

# Renewed membership of Flanders through CLARIN-Belgium in CLARIN-ERIC

- Only countries and international organisations can become members of ERICs
- Before 2018: The Dutch Language Union (DLU) was a member of CLARIN-ERIC in lieu of Flanders
- 2018: **Flexit** - construction with DLU no longer possible
- Solution:
  - Apply for political support from Flanders (approved October 2020)
  - Flanders supports founding of CLARIN-BE for membership (ongoing)





# Combining the strengths of CLARIN and DARIAH

- **CLARIN and DARIAH are the two most important European-level research infrastructures in the Humanities.**
- **Together, they provide European researchers with sustainable tools, services, and data to perform research in Computational Linguistics (CLARIN) and the Arts and Humanities (DARIAH).**
- **CLARIAH-VL combines these strengths by embedding high-quality, user-friendly tools and resources into the workflows of humanities researchers in Flanders and beyond, paving the way for Flemish participation in the European Open Science Cloud.**



Update Coronavirus

Regulations

PhD fellowships

Postdoctoral fellowships

Sabbatical bench fee

Research projects

Infrastructure

> [Medium-scale research infrastructure](#)

## International research infrastructure

### Objectives

Research infrastructure comprises all facilities and sources that promote the performance of cross-border and strategic basic research across all scientific disciplines. Besides scientific infrastructure, this includes collections, natural habitats, corpora and databases (including digital opening up).

The International Research Infrastructure (IRI) programme governing the participation in and/or funding of international research infrastructure aims to support the Flemish participation in and/or funding of international investment initiatives that are carried out at large-scale international or supranational facilities to which the Flemish Government contributes and/or whose strategic importance for Flanders can be demonstrated.

### Contact:

+32 2-550 15 42

[infrastructuur@fwo.be](mailto:infrastructuur@fwo.be)

Call closed

<https://www.fwo.be/en/fellowships-funding/research-infrastructure/international-research-infrastructure/>

# Challenge 1: Data-Level Access to Digital Resources

## Service Infrastructure Component (SIC) 2: DH Research Dataset Aggregation

- **Goals:**
  - Provide data-level access to key Belgian resources
  - Develop sustainable data curation workflows
  - Engage researchers with the data
- **DATA-KBR-BE**



# Challenge 2: LOD Ecosystem for the Humanities

## Service Infrastructure Component (SIC) 3: Linked (Open) Humanities Data Tool Suite

- **Goals:**
  - Support extraction, contextualization & modelling of entities
  - Model semantic relations
  - Link to LOD Reference Services
- SIC 3.1: Entity Referencing Toolkit
- SIC 3.2: LOD Creation and Publication Toolkit

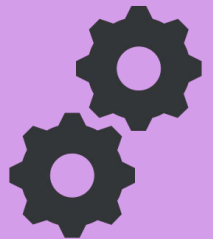


# Challenge 3: Humanities-Centred Artificial Intelligence

Service Infrastructure Component (SIC) 5:

## Digital Text Analysis Dashboard and Pipeline

- **Goals:**
  - Provide access to benchmarked NLP tools & resources
  - Web-based interface for training advanced models
  - Automate digitisation pipeline
- Flemish AI Academy





# Embedding services into the *Humanities Data Lifecycle*

 **Coordinate**

Project management,  
policy, consultation,  
user engagement

**Research Needs**

 **Aggregate**

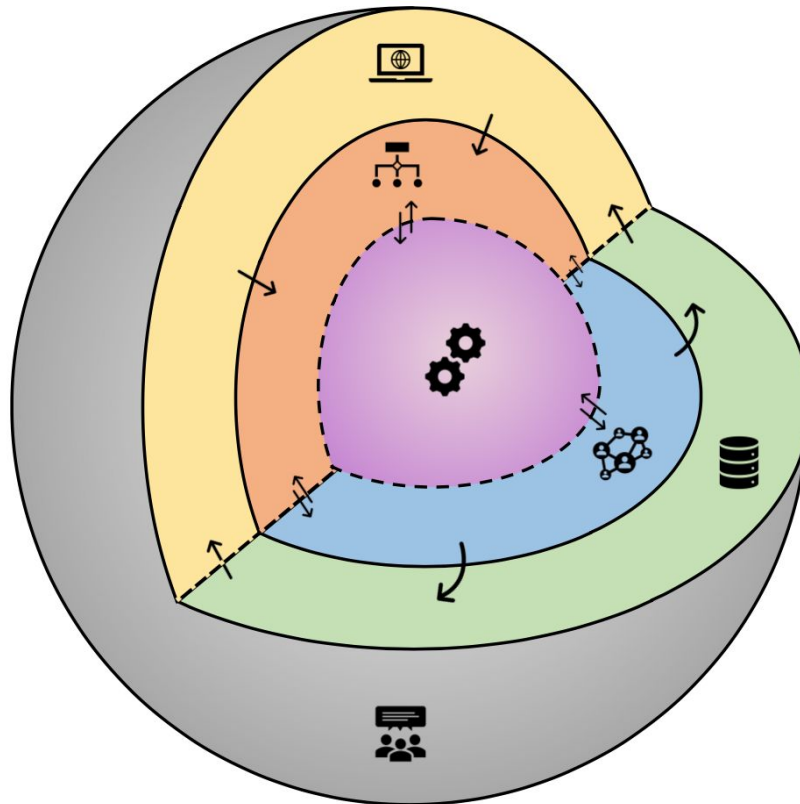
Capture, creation,  
scraping, digitisation,  
automatic transcription


**Raw Data**

 **Enrich**

Pre-processing,  
clean-up, manual  
annotation, editing


**Structured Data**



 **Contextualise**

Ontologise, linking,  
modelling, visualisation

**Linked Open Data**

 **Analyse**

Content, network,  
relational, spatial,  
visual, and stylistic  
analyses

**Processed Data**

 **Sustain**

Prepare data for  
long-term storage in  
trusted (open) data  
repositories, reusability

**Reusable Data**

# Collaboration with Cultural Heritage Institutions



## **KBR**

KBR is the national scientific library and collects all Belgian publications. It preserves, manages and studies more than 8 million documents, a rich cultural and historical heritage. KBR provides access to all information in its collections, facilitates research and offers a broad cultural experience.



## **State Archives (RA-AE)**

The tasks of the State Archives are numerous: acquiring archives, making them accessible, valorising, conducting scientific research, preserving and managing more than 315 kilometres of archives. CEGESOMA is part of the State Archives.

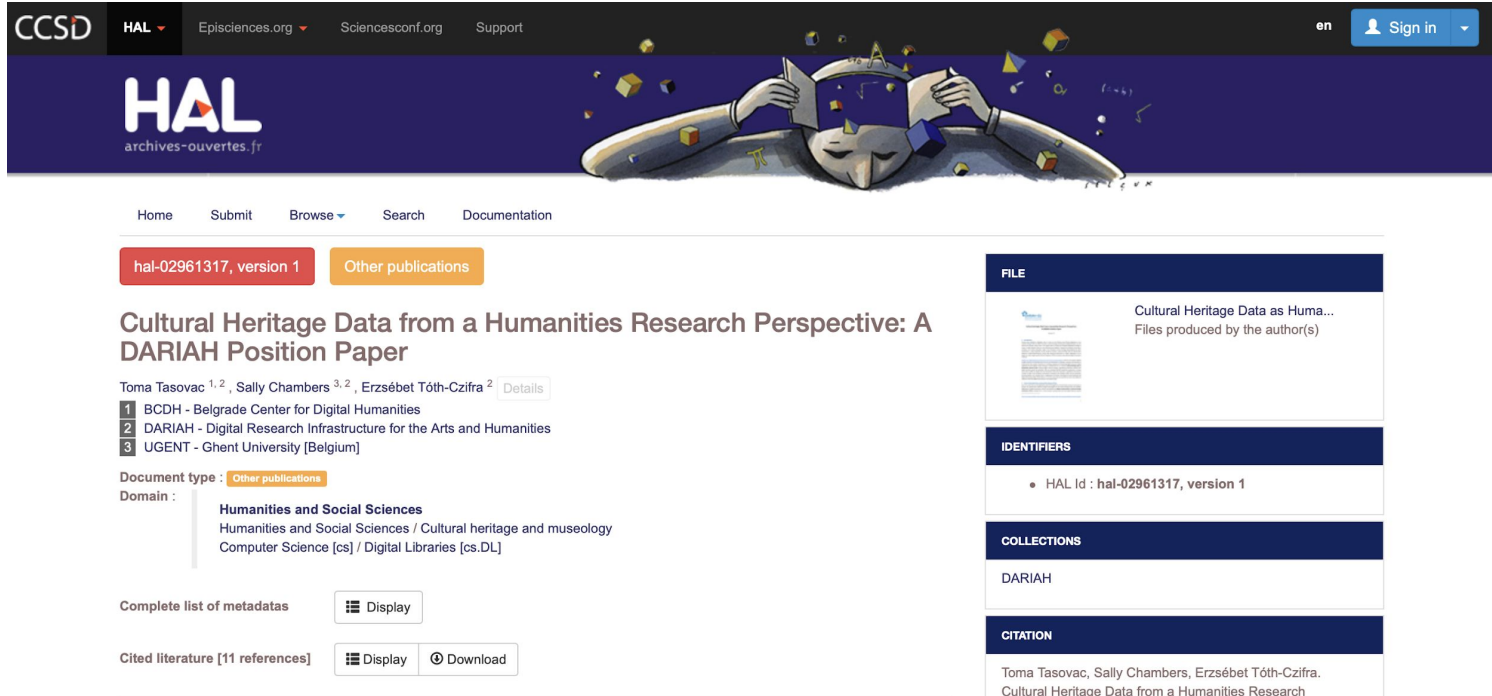


## **Royal Museums for Art and History (KMKG-MRAH)**

The KMKG-MRAH encompass various renowned museums in Brussels: the Art & History Museum, the Museums of the Far East, the Halle Gate and the Musical Instruments Museum (MIM).

[https://www.belspo.be/belspo/fsi/index\\_en.stm](https://www.belspo.be/belspo/fsi/index_en.stm)

# Cultural Heritage Data as Humanities Research Data?



The screenshot shows the HAL website interface. At the top, there is a navigation bar with the CCSD logo, a language dropdown set to 'en', and a 'Sign in' button. Below this is a banner featuring the HAL logo and an illustration of a bird-like creature reading a book with floating geometric shapes. The main content area includes a navigation menu (Home, Submit, Browse, Search, Documentation) and two buttons: 'hal-02961317, version 1' and 'Other publications'. The title of the record is 'Cultural Heritage Data from a Humanities Research Perspective: A DARIAH Position Paper' by Toma Tasovac, Sally Chambers, and Erzsébet Tóth-Czifra. A list of affiliations follows: 1 BCDH - Belgrade Center for Digital Humanities, 2 DARIAH - Digital Research Infrastructure for the Arts and Humanities, and 3 UGENT - Ghent University [Belgium]. The document type is 'Other publications' and the domain is 'Humanities and Social Sciences'. There are buttons for 'Complete list of metadatas' (Display) and 'Cited literature [11 references]' (Display, Download). On the right, a sidebar contains sections for 'FILE', 'IDENTIFIERS' (HAL id: hal-02961317, version 1), 'COLLECTIONS' (DARIAH), and 'CITATION'.

CCSD HAL Episciences.org Sciencesconf.org Support en Sign in

HAL archives-ouvertes.fr

Home Submit Browse Search Documentation

hal-02961317, version 1 Other publications

## Cultural Heritage Data from a Humanities Research Perspective: A DARIAH Position Paper

Toma Tasovac<sup>1, 2</sup>, Sally Chambers<sup>3, 2</sup>, Erzsébet Tóth-Czifra<sup>2</sup> [Details](#)

- 1 BCDH - Belgrade Center for Digital Humanities
- 2 DARIAH - Digital Research Infrastructure for the Arts and Humanities
- 3 UGENT - Ghent University [Belgium]

Document type : Other publications

Domain : Humanities and Social Sciences  
Humanities and Social Sciences / Cultural heritage and museology  
Computer Science [cs] / Digital Libraries [cs.DL]

Complete list of metadatas

Cited literature [11 references]

**FILE**

Cultural Heritage Data as Huma...  
Files produced by the author(s)

**IDENTIFIERS**

- HAL id : hal-02961317, version 1

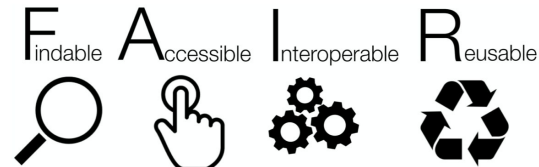
**COLLECTIONS**

DARIAH

**CITATION**

Toma Tasovac, Sally Chambers, Erzsébet Tóth-Czifra.  
Cultural Heritage Data from a Humanities Research  
Perspective: A DARIAH Position Paper. 2020. (hal-02961317)

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



## Resource families

### Introduction

The aim of the CLARIN Resource Families initiative is to provide a comprehensive overview of available language resources in the CLARIN infrastructure across all disciplines of the humanities, social sciences and human language technology. The listings are sorted by language family and resource type.

The listings for each family include the most important metadata such as corpus size, text sources, time periods, annotations and licences as well as concordancers, whenever available. In addition to the resource listings, an overview is provided of other existing valuable language resource families in the CLARIN infrastructure.

#### Corpora

- Computer-mediated communication corpora
- Corpora of academic texts
- Historical corpora
- L2 learner corpora
- Literary corpora
- Manually annotated corpora
- Multimodal corpora
- Newspaper corpora
- Parallel corpora
- Parliamentary corpora
- Reference corpora
- Spoken corpora

#### Lexical Resources

- Lexica
- Dictionaries
- Conceptual Resources
- Glossaries
- Wordlists

#### Tools

- Normalization
- Named entity recognition
- Part-of-speech tagging and lemmatization
- Tools for sentiment analysis

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

# Database of Southern Dutch Dialects (DSDD)

The screenshot displays the DSDD search interface. On the left, a map of Southern Dutch dialects is shown with various colored dots representing different dialectal forms. The search results for the word 'kikker' are displayed on the right. The results include 85 dialect words, a definition, and a list of original definitions. Below the search results, a legend shows the symbols used for different dialectal forms.

**kikker** Concept

85 dialectwoorden — aakpui · aakpuit · aaktepui · aaktepuit · aaptepuit · akepui · akepuit · akerpuit · bospuit · meer... (76)

*Definie*  
Een algemeen voorkomende amfibie zonder staart. Hij is vaak te vinden in vochtige omgevingen.

**Oorspronkelijke definities**

Aantal	Dialectwoord	Opgetekend in	Toon op kaart
28	aakpui	Aalter · Bassevelde · Bellem · Boekhoute · meer... (14)	

Legend:

- puid, puit
- vors
- kivors
- pui
- kwavors
- kwakkerd
- kuit
- kwakvos
- kikker
- aakpui
- kwakker
- aaktepui
- kinkvors
- aaktepuit
- kwekkerd
- paddenmoek
- kinkpuit
- vros
- kikbil
- kikmauw
- kikvorsje
- kruts
- kwekker
- akepuit
- kwak
- puiden, puiten
- puitje
- kikkerd
- kinkeluut
- kinkenduut
- naakte puid, naakte puit
- kinkeluter
- naaktepui
- aaptepuit
- groenjager
- kikpuit
- kikvors
- lookvors
- pad
- vorsje
- aakpuit
- akepui
- broeknachttegaal
- keven
- kikbil
- kikkenbil
- kink
- kweekvors
- puider
- weerkwakkerd
- akerpuit
- bospuit
- bruinling
- duiker
- fret
- getpuit
- groenduiker
- groene puit
- groenvors
- haakpuit
- hakpui
- kienkvuit
- kienpuit
- kikbeul
- kikbilletje
- kikkebil
- kikkenduut
- kikkertje
- kinkerd
- klabotskop
- koelekop
- koelkop
- kokkerd
- kokvors
- kroddel
- kwaker
- kwek
- loker
- meun
- naakpui
- naakpuit
- puiodder
- sprinkhaan
- vek
- vrosje

<https://dsdd.ivdnt.org/DSDD/search>



# DATA-KBR-BE

Facilitating data-level access to KBR's digitised and born-digital collections for digital humanities research

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



**KBR**  Where time  
is treasured

Belgian Science Policy Office  
  
**belspo**

# Collections as Data

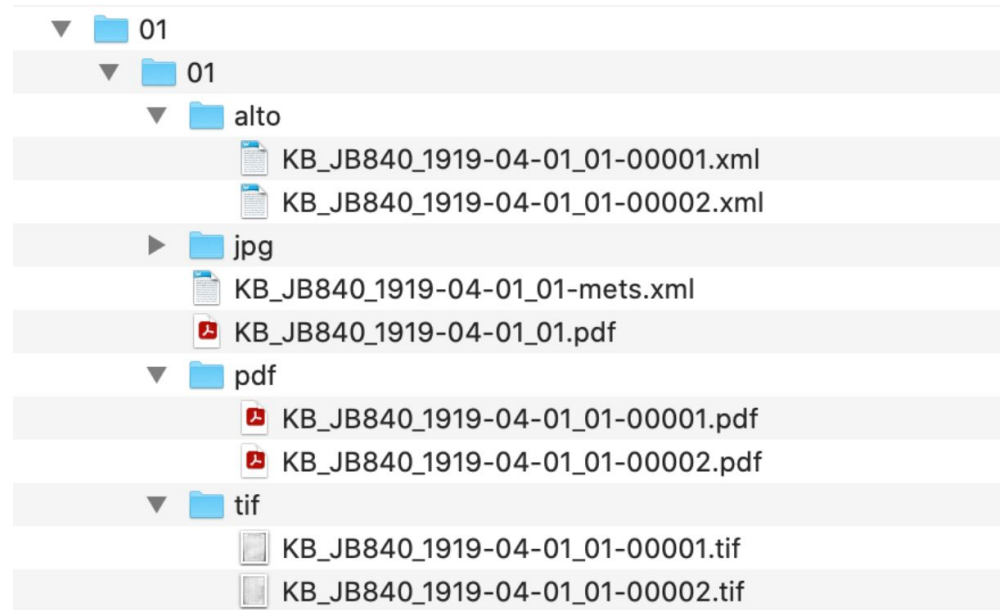
Providing data-level access to digital collections is a primary challenge for undertaking digital humanities research. In the United States, the flagship initiatives, ‘[Always Already Computational: Collections as Data](#)’ and ‘[Collections as Data: Part to Whole](#)’, define ‘*Collections as Data*’ as a “conceptual orientation to collections that renders them as ordered information, stored digitally, so that they are inherently amenable to computation”. The initiative was established to document, exchange experience and share knowledge to encourage cultural heritage institutions to implement ‘collections as data’ in their own institutions. **DATA-KBR-BE will kick-start the implementation of ‘Collections As Data’ in Belgium.**



**<https://collectionsasdata.github.io>**

# Data-level access

**Providing access to the underlying files of digitised cultural heritage resources to facilitate data analysis by means of tools and methods developed in the field of digital humanities**



# From Collections to Data



## KBR BelgicaPress

KBR Digitised Newspapers	Language	First	Last	H
Vlaamsche Post (De)	NL	1915	1916	
Ons Vaderland	NL	1917	1922	
Libre Belgique (La)	FR	1918	1950	H
Nation Belge (La)	FR	1918	1950	H
Peuple Wallon (Le)	FR	1918	1918	H
Standaard (De)	NL	1918	1950	H
Vers L'Avenir	FR	1918	1950	H
Schelde (De)	NL	1919	1936	H
Drapeau Rouge (Le) / Voix du peuple (La)	FR	1921	1950	H
Rof(o)de Vaan (De)	NL	1921	1950	H
Volksgazet (De)	NL	1921	1950	H
Wallonie (La)	FR	1921	1950	H
Vlaanderen	NL	1922	1932	H
Gentenaar (De)	NL	1923	1950	H
Grenz Echo	DE	1932	1945	H
Nieuwsblad (Het)	NL	1932	1950	H
Vlaanderen-Jong-Dietschland	NL	1934	1934	H
Pays réel (Le)	FR	1936	1944	H
Volk en Staat	NL	1937	1944	H
Cité Nouvelle (La)	FR	1938	1947	H

- 01
  - 01
    - alto
      - KB\_JB840\_1919-04-01\_01-00001.xml
      - KB\_JB840\_1919-04-01\_01-00002.xml
    - jpg
      - KB\_JB840\_1919-04-01\_01-mets.xml
      - KB\_JB840\_1919-04-01\_01.pdf
    - pdf
      - KB\_JB840\_1919-04-01\_01-00001.pdf
      - KB\_JB840\_1919-04-01\_01-00002.pdf
    - tif
      - KB\_JB840\_1919-04-01\_01-00001.tif
      - KB\_JB840\_1919-04-01\_01-00002.tif

# Interdisciplinary Research Scenarios

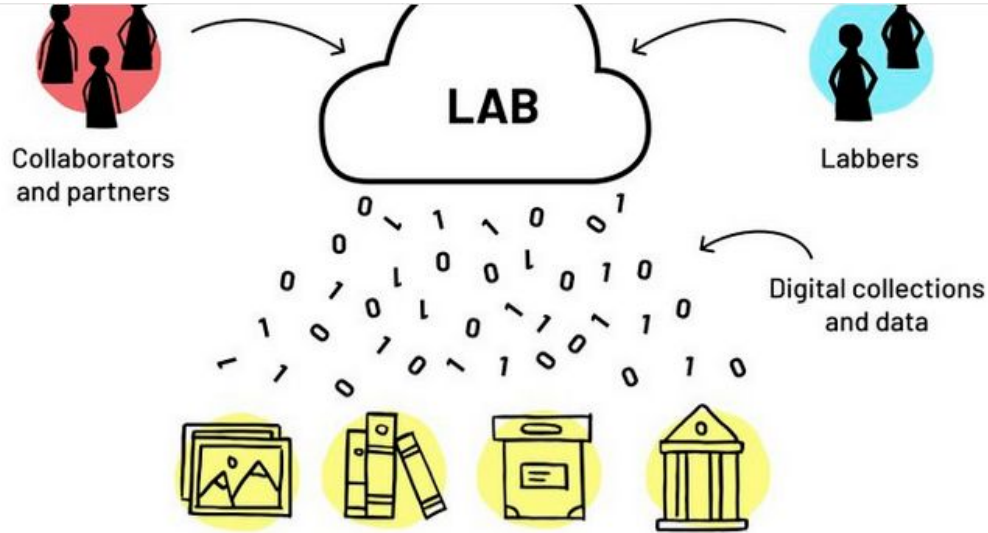
- **Collective Action Belgium** led by [GhentCDH](#)<sup>2</sup>, focuses on social history in the Interbellum and World War Two period and aims to trace the dynamics of contention, strikes, demonstrations and other forms of collective action in Belgium as reported in Belgian newspapers;
- **The feuilleton in Belgium**, led by [ACDC](#)<sup>2</sup>, focuses on literary studies in the period 1830–1930 and aims to map the publication of literature in Belgian newspapers across the first century of the Belgian nation state;
- **History of Belgian Journalism**, led by [ULB](#)<sup>2</sup> and KBR, focuses on media history from 1886 until now and aims to trace the history of Belgian journalism through the lens of critical discourses about journalism as in Belgian newspapers.



A	B	C	D	E	F	G	H
ID	Text	Nom (entier original)	nom	prénom	renvoi	Sources	Sexe
5	Abbeïos Th : (mo) Collabore à forger de la Centrale du personnel enseignant (socialiste), L'Émancipe (créé en 1919).	Abbeïos Th	Abbeïos	Th	NA	no	?
8	Abessart H : Rédacteur à La Meuse après 1940-1945 (CRHC 98).	Abessart H	Abessart	H	NA	CRHC 98	H
9	Abressart Adolphe: critique de théâtre et musical à La Métropole vers 1937-1938.	Abressart Adolphe	Abressart	Adolphe	NA	NA	H
11	Abs Fernand: dit Abel Fernand, J.D.D.Le Soir, 14 avil 1998. (Bruxelles, 1911 - 1999) Comédien. Avant la Seconde Guerre mondiale, est élève de Jules Delaire, animateur du Théâtre du Marais. Également comédien de radio puisqu'il prend part aux premières émissions dramatiques de TRF à laquelle il reste attaché jusqu'au milieu des années 1970' en tant que comédien, metteur en ondes puis chef du service dramatique. Eve du théâtre en 1961, acteur Miroche du Théâtre de Poche (il cite en Belgique "Le Roi se meurt" d'Éugène Ionesco), il se produit aussi au Théâtre national, au Théâtre de Bruxelles 11 et aux Galeries (NDB).	Abs Fernand	Abs	Fernand	NA	J.D.D.Le Soir,	H
14	Ackemans Jan : Collaborateur de De Goede Boodschap (1902-1913), collaborateur à De Aankondiger van Lee (1906-1913) et à Koningsch Handelsblad (1907-1912), collaborateur à 14 Gazet van Belfaar (1906-1913).	Ackemans Jan	Ackemans	Jan	NA	NA	H



# International GLAM Labs Community



<https://glamlabs.io/books/>



## GLAM Workbench

Home

About >

Help >

Data sources >

GLAM Labs >

Trove >

DigitalNZ >

Archives >

Libraries >

Web Archives >

Museums >

Government >

Suggest a topic

Ask a question

## Welcome to the GLAM Workbench

Here you'll find a collection of tools, tutorials, examples, and hacks to help you work with data from galleries, libraries, archives, and museums (the GLAM sector). The primary focus is Australia and New Zealand, but new collections are being added all the time. Let me know if there's some GLAM data you'd like me to explore – [suggestions](#) are always welcome!

### Quick start

The resources in the GLAM Workbench are created and shared as Jupyter notebooks. Jupyter lets you combine narrative text and live code in an environment that encourages you to learn and explore. Jupyter notebooks run in your browser, and you can get started without installing any software – just click on a few links!

Here's how to find your way around the GLAM Workbench and open up a notebook.



### Table of contents

Quick start

What is GLAM data?

What can I do with GLAM data?

Do I need to be able to code?



Image: State Library of Victoria, <http://handle.slvvic.gov.au/10381/342096>

Sherratt, Tim, Wilms, Lotte, & Lingstadt, Kirsty. (2020, April). LIBER Webinar: Setting Up A GLAM Workbench In Your Library. Zenodo. <http://doi.org/10.5281/zenodo.3743193>

About KBR

Legal Deposit

Venue Hire

Plan Your Visit

**KBR** 

Research & Collections ▾

Cultural Offer ▾

Services ▾

[HOME](#) > [RESEARCH](#) > [DIGITAL RESEARCH LAB](#)

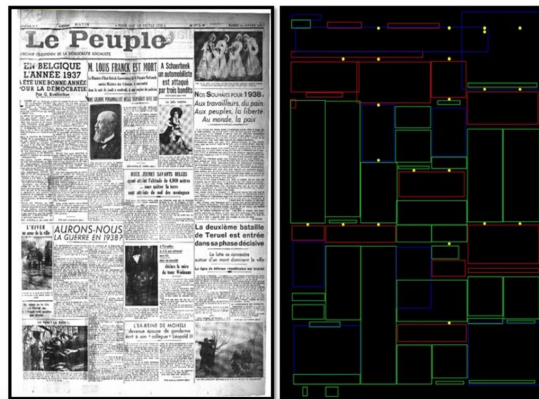
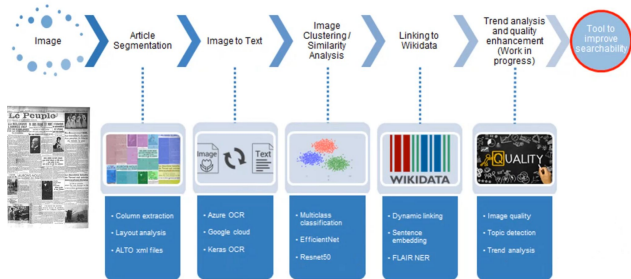
# Digital Research Lab

**A lab for text and data mining research on KBR's digitised and born-digital collections.**

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

# Data Science expertise

## Document Layout Analysis using Belgica Press



Document layout analysis using  
BelgicaPress



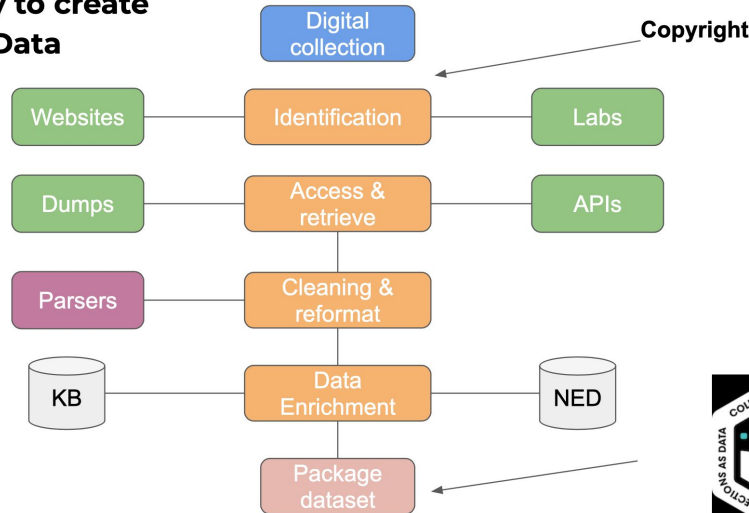
# Data Science expertise



<https://tw06v072.ugent.be/kbr/>

# Collections as Data methodology

## A methodology to create Collections as Data



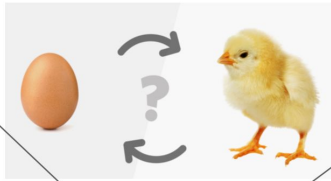
Candela, G., Sáez, M. D., Escobar Esteban, Mp., & Marco-Such, M. (2020). *Reusing digital collections from GLAM institutions*. *Journal of Information Science*: <https://doi.org/10.1177/0165551520950246> & <http://rua.ua.es/dspace/handle/10045/109460>



# Collections as Data @ KBR

Research Scenarios

Digital collection



Search in KBR's digitised newspapers

The portal BelgicaPress gives you access to 169 Belgian newspapers referred to as 'major newspapers', most of which appeared daily (1814-1970).

Copyright?

Identification

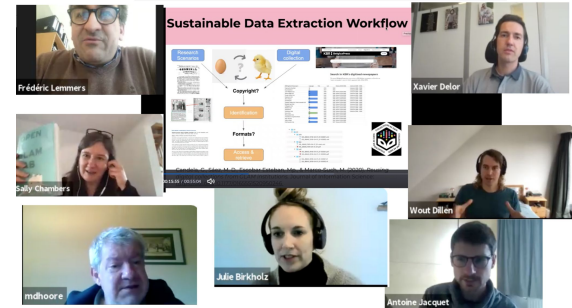
Formats?

Access & retrieve

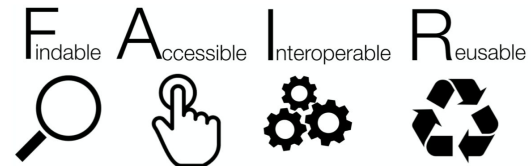


KBR Digitized Newspapers	Language	First	Last	History (1918-1950)	Literature (1830-1930)
Vlaamische Post (Dn)	NL	1915	1918		Literature (1830-1930)
Ons Vaderland	NL	1917	1922		Literature (1830-1930)
Libre Belgique (La)	Fr	1918	1950	History (1918-1950)	Literature (1830-1930)
Nation (Engl (La)	En	1918	1950	History (1918-1950)	Literature (1830-1930)
Peuple (Walon (La)	Fr	1918	1950	History (1918-1950)	Literature (1830-1930)
Standaard (Dn)	NL	1918	1950	History (1918-1950)	Literature (1830-1930)
Verscheider	NL	1918	1950	History (1918-1950)	Literature (1830-1930)
Schiede (Dn)	NL	1919	1950	History (1918-1950)	Literature (1830-1930)
Drapere (Engl (La) / With the people (La)	NL	1921	1950	History (1918-1950)	Literature (1830-1930)
Hilfsagent (Dn)	NL	1921	1950	History (1918-1950)	Literature (1830-1930)
Wolkens (La)	Fr	1921	1950	History (1918-1950)	Literature (1830-1930)
Vlaanderen	NL	1922	1950	History (1918-1950)	Literature (1830-1930)
Genève (Dn)	NL	1922	1950	History (1918-1950)	Literature (1830-1930)
Genex Echo	NL	1932	1945	History (1918-1950)	Literature (1830-1930)
Neuzeitlich (Dn)	NL	1932	1950	History (1918-1950)	Literature (1830-1930)
Vlaanderen Jong-Duitsland	NL	1934	1944	History (1918-1950)	Literature (1830-1930)
Payk (Dn)	NL	1936	1944	History (1918-1950)	Literature (1830-1930)
Nik in Saks	NL	1937	1944	History (1918-1950)	Literature (1830-1930)
OH Nouvelle (La)	Fr	1938	1947	History (1918-1950)	Literature (1830-1930)

- 01
  - 01
    - 01 KB\_IBB840\_1919-04-01\_01-00001.xml
    - 01 KB\_IBB840\_1919-04-01\_01-00002.xml
  - 02
    - 01 KB\_IBB840\_1919-04-01\_01-mets.xml
    - 01 KB\_IBB840\_1919-04-01\_01.pdf
  - 03
    - 01 KB\_IBB840\_1919-04-01\_01-00001.pdf
    - 01 KB\_IBB840\_1919-04-01\_01-00002.pdf
  - 04
    - 01 KB\_IBB840\_1919-04-01\_01-00001.tif
    - 01 KB\_IBB840\_1919-04-01\_01-00002.tif



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>



# Collections

[www.kbr.be/en/collections/](http://www.kbr.be/en/collections/)



## Newspapers

The collection In terms of diversity and volumes, KBR's collection of newspapers is by far the [...]

*[Continue reading](#)*



## Belgian Bibliography

The Belgian Bibliography makes inventories of catalogue records of all non-periodical publications (monographs, brochures, offline electronic [...])

*[Continue reading](#)*



## Colloquium Saving the Web

30 October 2019

On 18 October 2019, the colloquium 'Saving the Web: the Promise of a Belgian Web Archive' [...]

*[Continue reading](#)*

# Web and Social Media Archiving at KBR



2017 – 2019  
Focus on  
archiving websites

[PROMISE website](#)



2020-2022  
Focus on  
social media archiving

[BE-Social Website](#)

# Born-Digital Collections as Data @ KBR

KBR



## Save the social media of Belgium

Capture a social media account as that appears from national/social/historical and or cultural importance for Belgium.

\*Vereist

What is the name of the organization/event?

Jouw antwoord

What is the twitter url? (e.g @Destandaard) \*

Jouw antwoord

besocial Home Collections

THIS COLLECTION

Description

Timespan

Content

Seeds

Access

Metadata

<< Collapse



## mini pilot twitter accounts

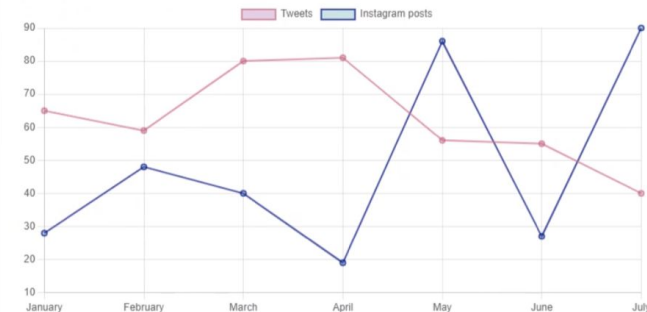
### Description

This is a social media collection, harvested as part of the BESOCIAL project at the Royal Library of Belgium (KBR). The folk

These are all Twitter accounts of our selection which are harvested once per day.

### Timespan

This section lists when the content of this collection was harvested and which timespan the content covers.



### Content

By using Natural Language Processing (NLP), we extracted named entities such as locations or persons.

# Born-Digital Collections as Data @ KBR



Home Collections

COLLECTIONS MENU

Collections List

Access

Collapse

## Collections

This page contains a list of social media collections, harvested for the BESOCIAL project.

### Collection List

#### mini pilot twitter accounts

5213 social media posts

[HTML](#) [EAD \(XML\)](#) [MARC \(XML\)](#)

#### new test

5213 social media posts

[HTML](#) [EAD \(XML\)](#) [MARC \(XML\)](#)

#### Twitter institutions

5213 social media posts

[HTML](#) [EAD \(XML\)](#) [MARC \(XML\)](#)

#### mini pilot twitter hashtags

5213 social media posts

[HTML](#) [EAD \(XML\)](#) [MARC \(XML\)](#)

#### mini pilot twitter hashtags small

5213 social media posts

#### test end date sven

5213 social media posts

[HTML](#) [EAD \(XML\)](#) [MARC \(XML\)](#)

#### Twitter search test

5213 social media posts

[HTML](#) [EAD \(XML\)](#) [MARC \(XML\)](#)

#### RabbitMQ Twitter Search test

5213 social media posts

[HTML](#) [EAD \(XML\)](#) [MARC \(XML\)](#)

#### Twitter mixed language corona hashtags

5213 social media posts

[HTML](#) [EAD \(XML\)](#) [MARC \(XML\)](#)

#### new test (shared)

5213 social media posts

[HTML](#) [EAD \(XML\)](#) [MARC \(XML\)](#)

#### flickr test

5213 social media posts

[HTML](#) [EAD \(XML\)](#) [MARC \(XML\)](#)

#### IDLab

5213 social media posts

[HTML](#) [EAD \(XML\)](#) [MARC \(XML\)](#)

#### sven test

5213 social media posts

[HTML](#) [EAD \(XML\)](#) [MARC \(XML\)](#)

#### sven

5213 social media posts

[HTML](#) [EAD \(XML\)](#) [MARC \(XML\)](#)

#### public news

5213 social media posts

[HTML](#) [EAD \(XML\)](#) [MARC \(XML\)](#)

#### Twitter French corona hashtags

5213 social media posts

[HTML](#) [EAD \(XML\)](#) [MARC \(XML\)](#)

#### Twitter Dutch corona hashtags

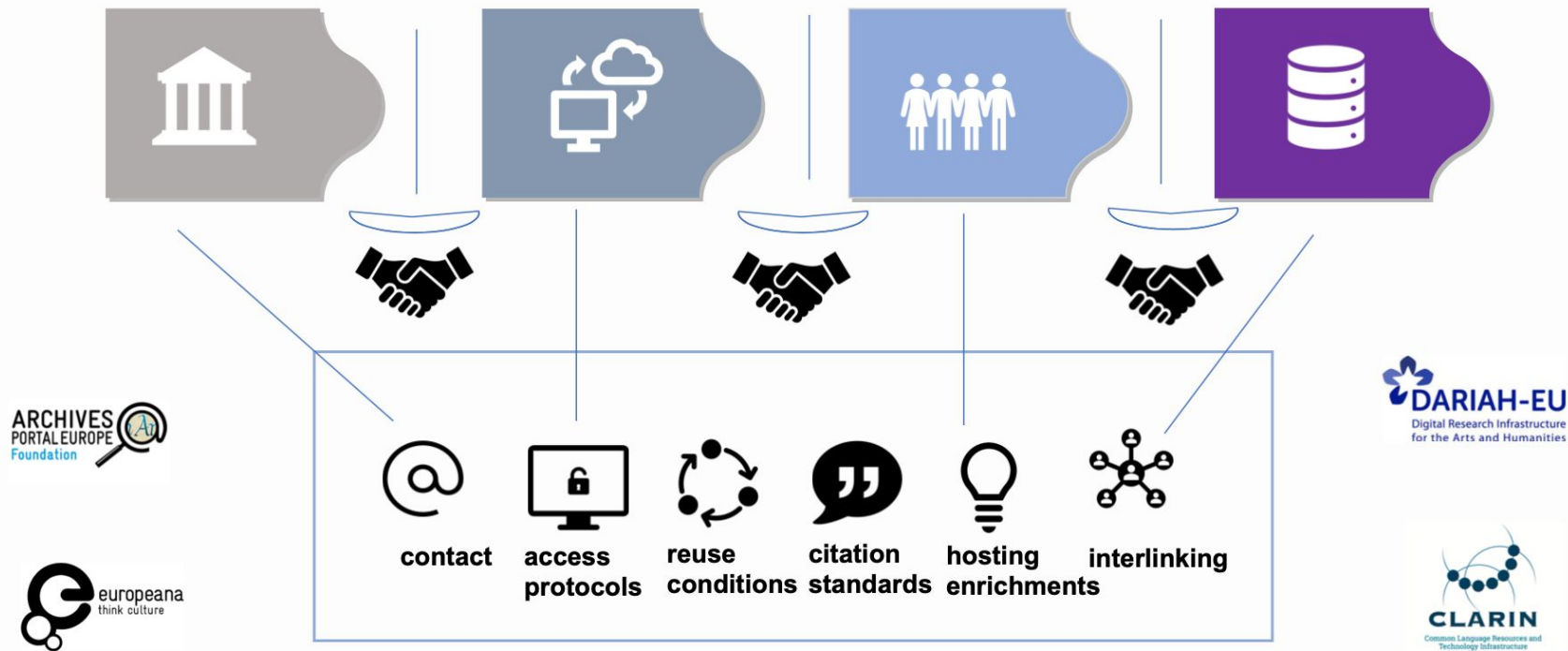
5213 social media posts

[HTML](#) [EAD \(XML\)](#) [MARC \(XML\)](#)



With thanks to: Dr. Erzsébet Tóth-Czifra, DARIAH-EU

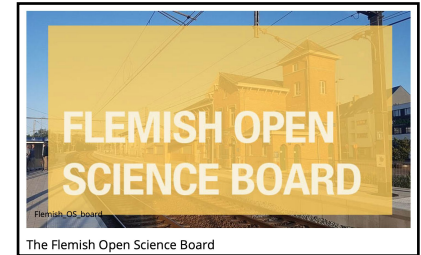
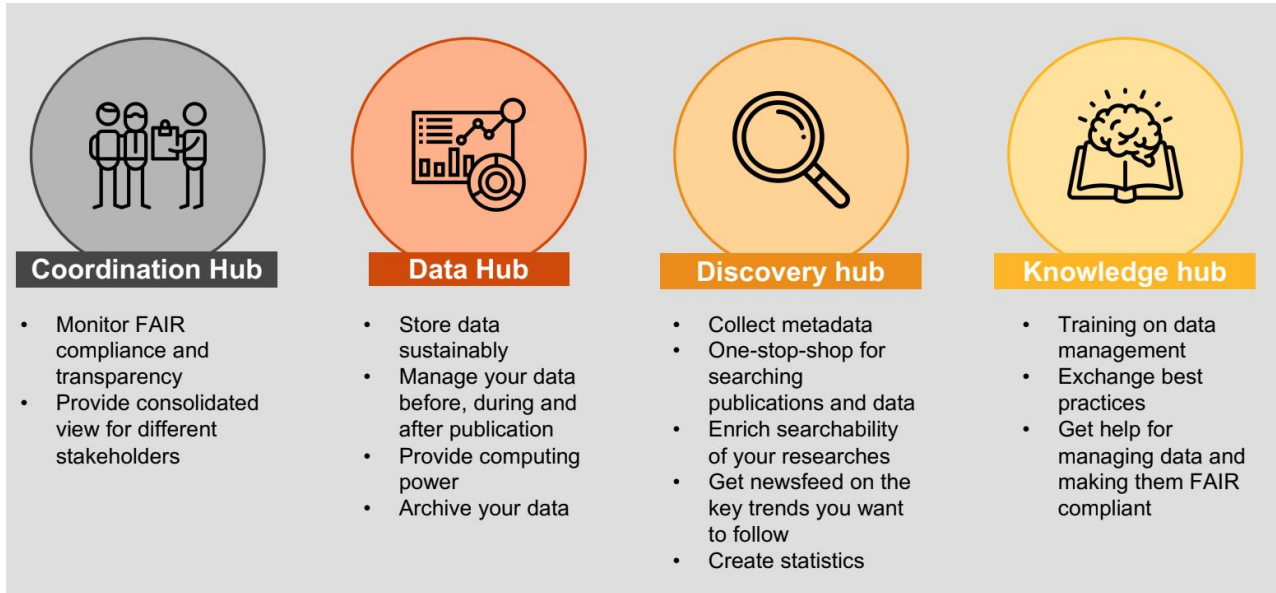
# The Heritage Data Reuse Charter: facilitating the handshake between all parties involved in data sharing



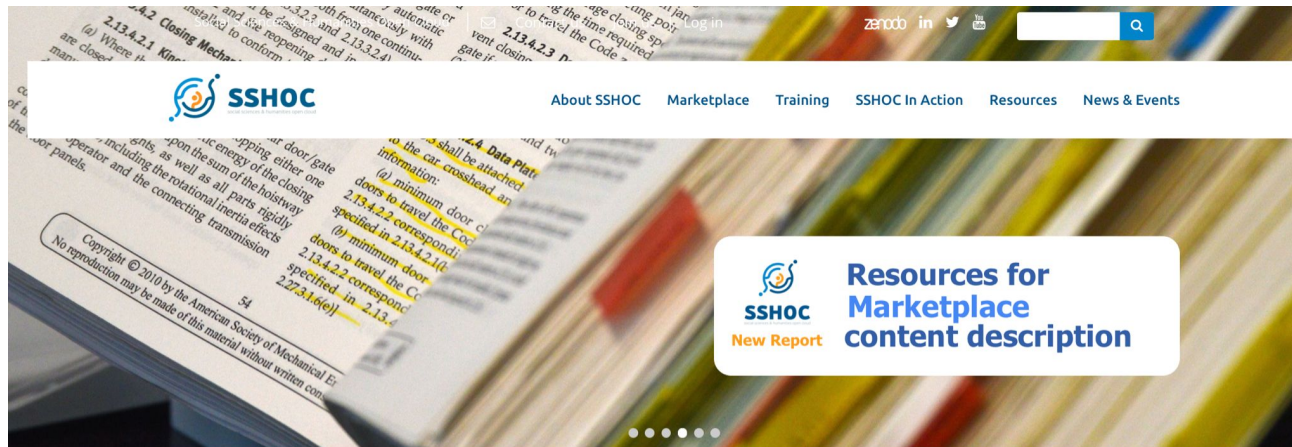
How to put all this to practice? <https://datacharter.hypotheses.org/>



# CLARIAH-VL: a vital role in preparing Humanities Researchers for the European Open Science Cloud



# Social Sciences and Humanities Open Cloud (SSHOC)



## SSHOC Principal Goals



SSHOC will create the social sciences and humanities area of the **European Open Science Cloud (EOSC)** thereby facilitating access to flexible, scalable research data and related services streamlined to the precise needs of the SSH community.



SSHOC will leverage and interconnect existing and new infrastructures from the **SSH ERICs** to foster synergies across disciplines and expedite interdisciplinary research and collaboration.



To maximise the efficiency and effectiveness of data re-use **SSHOC** will apply Open Science practices and **FAIR** principles to data management.



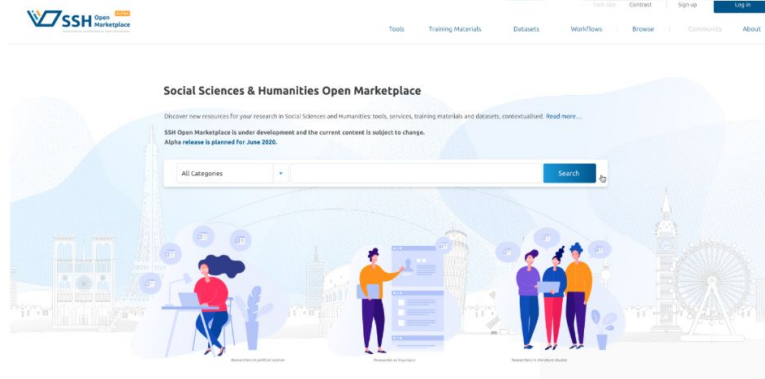
**SSHOC** will set up an appropriate governance model for the social sciences and humanities area of the **EOSC**, taking into account the specificities of different sub-domains within the two disciplines.

<https://sshopencloud.eu>

# Social Sciences and Humanities Open Cloud (SSHOC)

## SSH Open Marketplace as one of the SSHOC services

- Discovery portal for SSH resources
  - Tools & services
  - Training materials
  - Workflows
  - Datasets
  - Publications
- 3 guiding principles
  - Contextualisation
  - Curation
  - Community



Beta version: [marketplace.sshopencloud.eu](https://marketplace.sshopencloud.eu)



<https://sshopencloud.eu>

# Social Sciences and Humanities Open Cloud (SSHOC)



Report

Tools & Services

Training Materials

Publications

Datasets

Workflows

Browse

Home / Workflows / Collaborative Digital Edition of a Musical Corpus



## Collaborative Digital Edition of a Musical Corpus

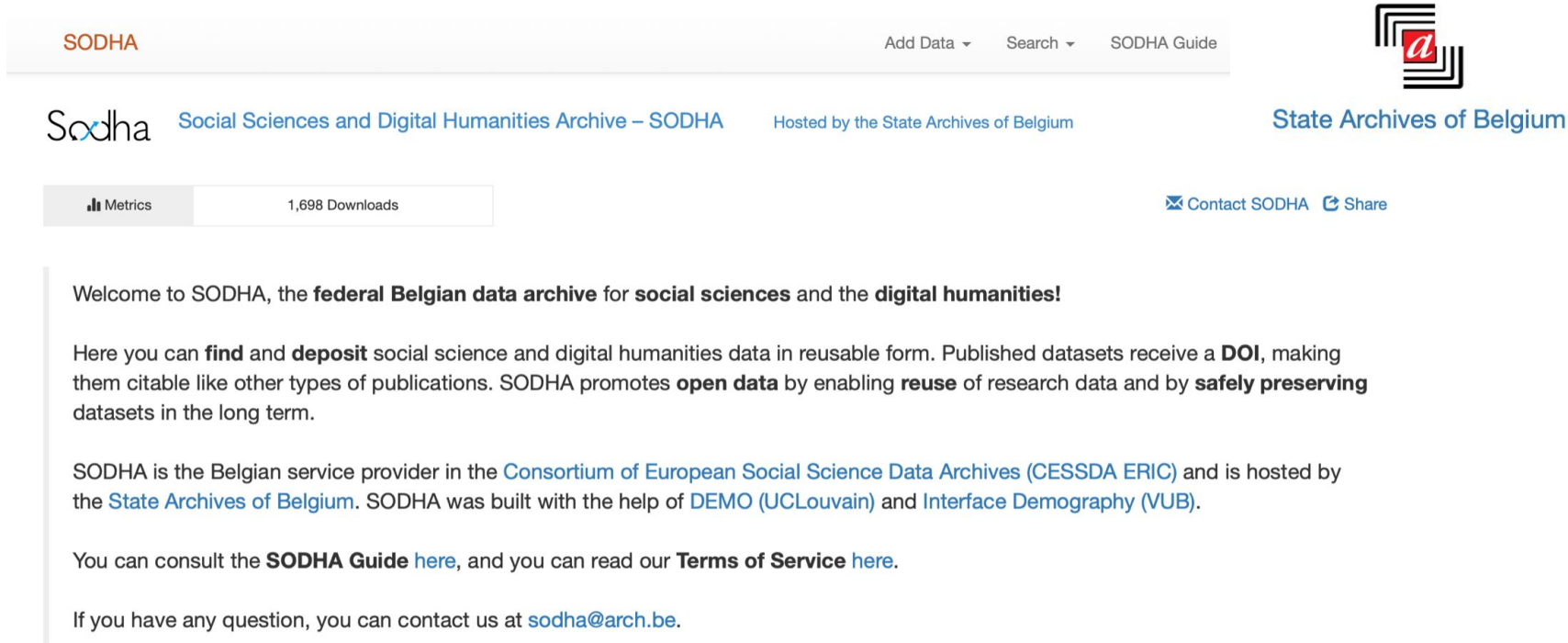
A project aims to do a digital edition of a musical corpus. The researchers need to be able to encode a broad range of musical documents in a machine-readable structure. The data to be encoded may include the musical content as provided by the composer (notes, pitches, durations, dynamics, etc.), information on the score (incipit, lyrics writer, etc.), information added by a performer when interpreting the content (timing, phrasing, various annotations, etc.), information on the visual appearance of the score (page layout, musical font, etc.) and analyses of the content in any of the other domains. The edition will be structured around a database in order to allow the users to explore it more easily. Furthermore, the project intends to be collaborative, which means it will offer anyone interested the possibility to contribute.

Steps (9)

- 1 Create a digital corpus of musical compositions.
- 2 Change format into MEI files if necessary.
- 3 Collect and organize metadata
- 4 Enrich the metadata of MEI files
- 5 Enrich the musical edition with critical apparatus metadata
- 6 Enrichment of MEI files
- 7 Present in an open source digital CMS the result of the musical edition.
- 8 Offer anyone interested the possibility to contribute
- 9 Have a reflexion on the quality of the produced data.

<https://marketplace.sshopencloud.eu/search?categories=workflow&order=label>

# Social Sciences and Humanities Data Archive (SODHA.be)



The screenshot shows the SODHA website interface. At the top, there is a navigation bar with the SODHA logo on the left, and 'Add Data', 'Search', and 'SODHA Guide' on the right. Below the navigation bar, the main header includes the 'Sodha' logo, the text 'Social Sciences and Digital Humanities Archive – SODHA', and 'Hosted by the State Archives of Belgium'. On the right side of the header is the State Archives of Belgium logo. A metrics box shows '1,698 Downloads'. There are links for 'Contact SODHA' and 'Share'. The main content area contains a welcome message, a description of the archive's mission, information about its affiliation with CESSDA ERIC and the State Archives of Belgium, and links to the SODHA Guide and Terms of Service. A contact email address is also provided.

SODHA

Add Data ▾ Search ▾ SODHA Guide

Sodha Social Sciences and Digital Humanities Archive – SODHA Hosted by the State Archives of Belgium

State Archives of Belgium

Metrics 1,698 Downloads

Contact SODHA Share

Welcome to SODHA, the **federal Belgian data archive** for **social sciences** and the **digital humanities**!

Here you can **find** and **deposit** social science and digital humanities data in reusable form. Published datasets receive a **DOI**, making them citable like other types of publications. SODHA promotes **open data** by enabling **reuse** of research data and by **safely preserving** datasets in the long term.

SODHA is the Belgian service provider in the [Consortium of European Social Science Data Archives \(CESSDA ERIC\)](#) and is hosted by the [State Archives of Belgium](#). SODHA was built with the help of [DEMO \(UCLouvain\)](#) and [Interface Demography \(VUB\)](#).

You can consult the **SODHA Guide** [here](#), and you can read our **Terms of Service** [here](#).

If you have any question, you can contact us at [sodha@arch.be](mailto:sodha@arch.be).

<https://www.sodha.be/dataverse/sodha>

# Social Sciences and Humanities Data Archive (SODHA.be)

SODHA

Add Data ▾

Search ▾

SODHA Guide

Sign Up

Log In



Social Sciences and Digital Humanities Archive – SODHA

Hosted by the State Archives of Belgium



Social Sciences and Digital Humanities Archive – SODHA > 16th Century Low Countries (shapefiles)

[✉ Contact Depositor](#) [🔄 Share](#)



## 16th Century Low Countries (shapefiles)

Version 1.0

Jongepier, Iason, 2021, "16th Century Low Countries (shapefiles)", <https://doi.org/10.34934/DVN/Y6DOAD>, Social Sciences and Digital Humanities Archive – SODHA, V1

[☰ Cite Dataset ▾](#)

Learn about [Data Citation Standards](#).

### Dataset Metrics [?](#)

3 Downloads [?](#)

### Description [?](#)

This dataset (ZIP-file) contains two shapefiles (provinces and major cities) of the 16th Century Low Countries) to be used in GIS and one example output map. (2021-01-19)

### Subject [?](#)

Arts and Humanities; Earth and Environmental Sciences

### Keyword [?](#)

Low Countries, Borders, Provinces, 16th Century, Historical geography

<https://doi.org/10.34934/DVN/Y6DOAD>



# Embedding services into the *Humanities Data Lifecycle*

 **Coordinate**

Project management,  
policy, consultation,  
user engagement

**Research Needs**

 **Aggregate**

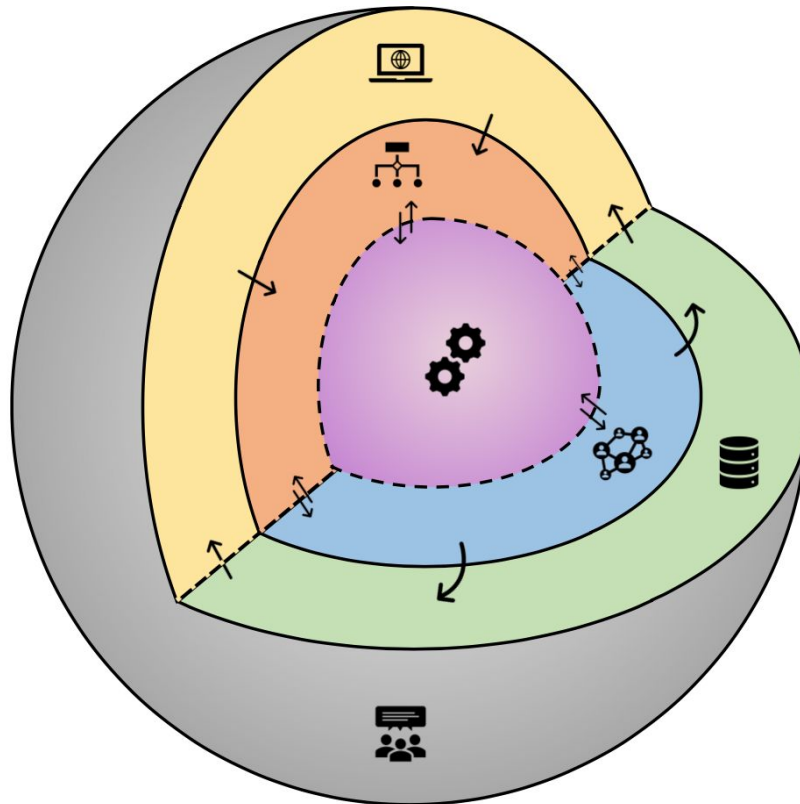
Capture, creation,  
scraping, digitisation,  
automatic transcription


**Raw Data**

 **Enrich**

Pre-processing,  
clean-up, manual  
annotation, editing


**Structured Data**



 **Contextualise**

Ontologise, linking,  
modelling, visualisation

**Linked Open Data**

 **Analyse**

Content, network,  
relational, spatial,  
visual, and stylistic  
analyses

**Processed Data**

 **Sustain**

Prepare data for  
long-term storage in  
trusted (open) data  
repositories, reusability

**Reusable Data**

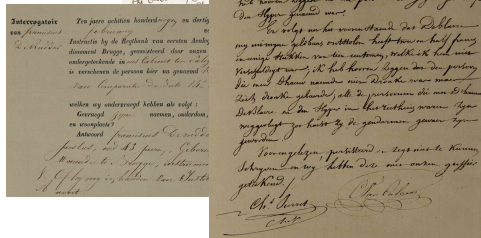
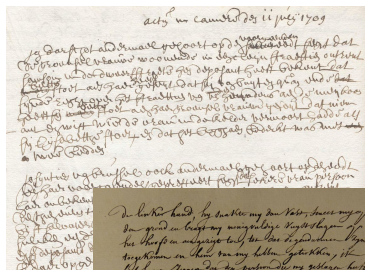
# CLARIAH: enabling data-driven research in the humanities



IIF Corpus  
Management Platform



International  
Image  
Interoperability  
Framework



### Crowd-sourcing

The screenshot shows a web interface for crowd-sourcing. On the left, there's a 'Documents' list with columns for ID, name, and count. Below it is a 'Code System' tree with categories like 'Plaats', 'Taal', and 'Tijdsperiode'. On the right, there's a 'Document Browser' showing a list of documents and a 'Text' view of a document snippet. The text is in Dutch and mentions 'PRO-JUSTITIA' and 'Alfred LeFranois Onderzoek-Rechter'.

### Handwritten Text Recognition

The screenshot shows a handwritten text recognition interface. It displays a document image with a red box highlighting a section of text. Below the image, the transcribed text is shown in a digital font. The interface includes a toolbar with various tools and a keyboard layout at the bottom.

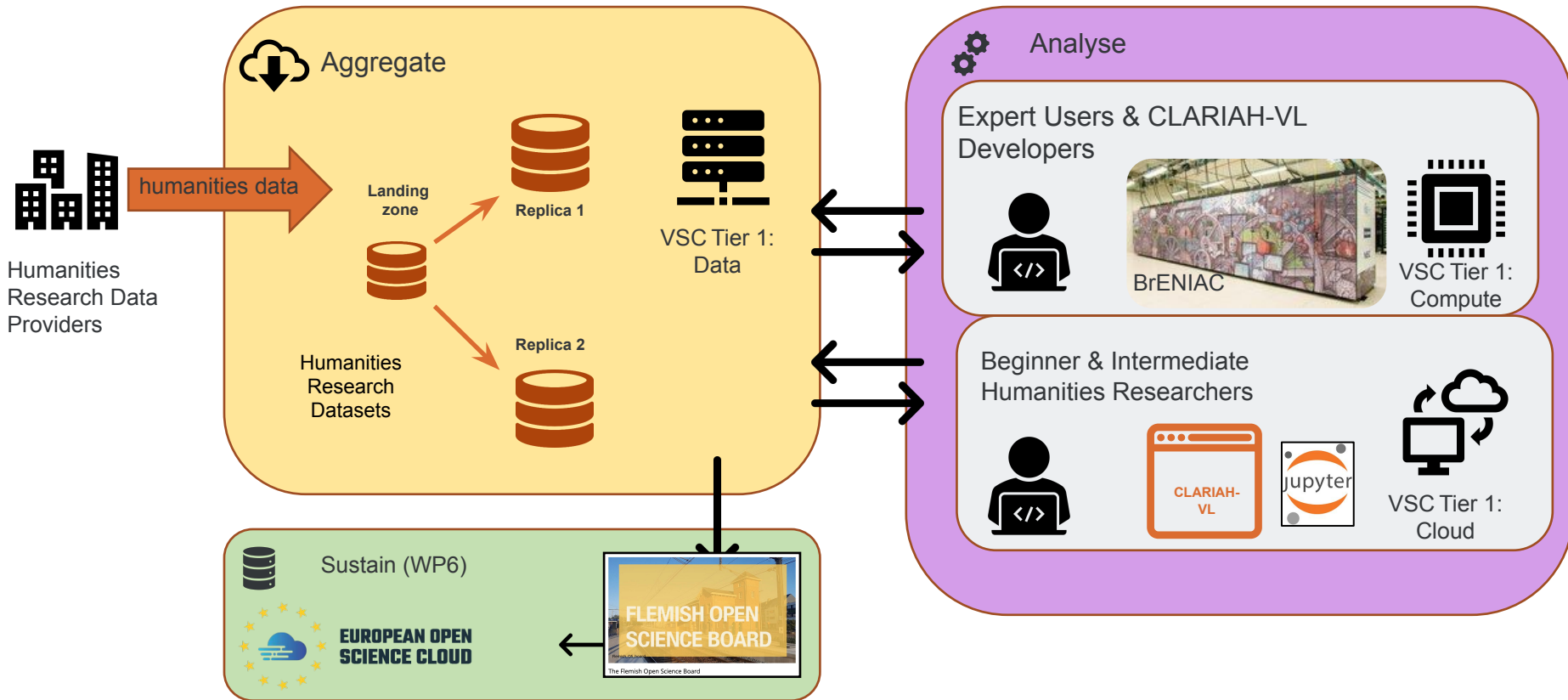


## Digital Text Analysis Pipeline

- Named Entity Extraction
- Social Network Analysis
- Sentiment & Emotion Mining



# Data-driven research in the arts and humanities



**¡Muchas gracias!**

**Sally Chambers | [sally.chambers@ugent.be](mailto:sally.chambers@ugent.be)  
Ghent Centre for Digital Humanities  
Ghent University, Belgium**