



CLARIN Annual Conference 2022

Prague, Czechia

CLARIN2022 Panel: CLARIN and Libraries Infrastructures Working Together

Moderator: Martin Wynne

Panelists: Sally Chambers, Andreas Witt, Peter Leinen



CLARIN Annual Conference 2022

Prague, Czechia

CLARIN and Libraries: What are the Next Steps?

Sally Chambers, DARIAH-EU, KBR, Royal Library of Belgium and
Ghent Centre for Digital Humanities  @schambers3





CLARIN Annual Conference 2022

Prague, Czechia

Cultural Heritage Data as Humanities Research Data



HAL open science Free and accessible knowledge

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^{1,2}, Sally Chambers^{3,2}, Erzsébet Tóth-Czifra² Details

- 1 BCDH - Belgrade Center for Digital Humanities
- 2 DARIAH - Digital Research Infrastructure for the Arts and Humanities
- 3 UGENT - Universiteit Gent = 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]

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

IDENTIFIERS
• HAL id : hal-02961317, version 1



European Commission

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

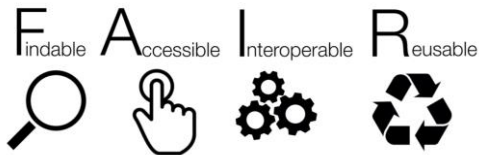
Commission proposes a common European data space for cultural heritage

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.

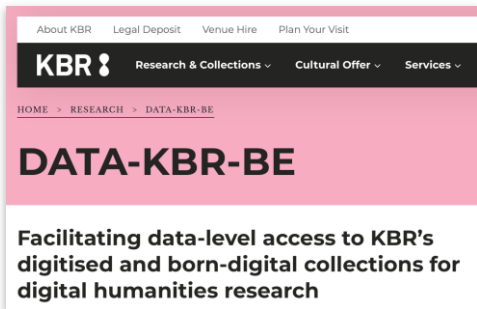
EUROPEANA FOUNDATION

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>

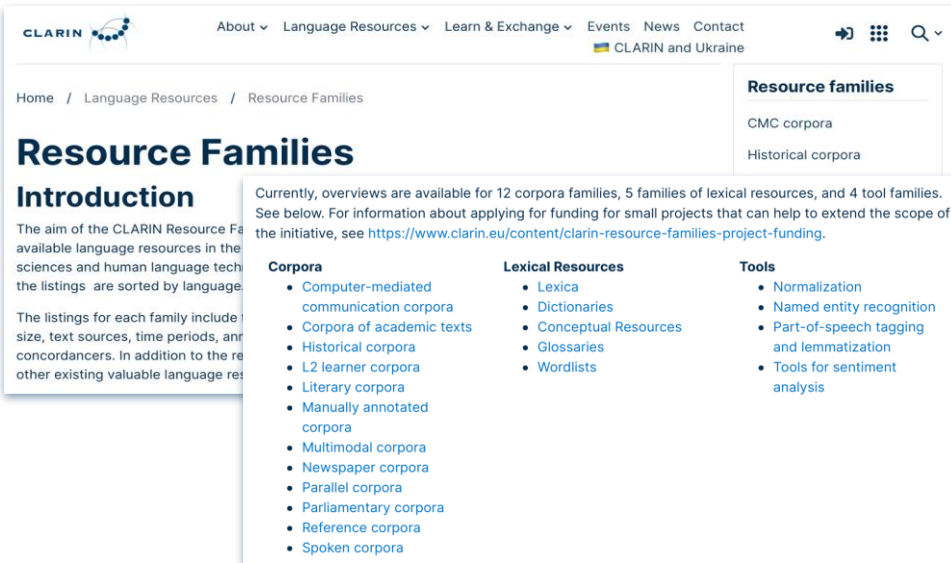
Collections as Data?



<https://collectionsasdata.github.io>



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



CLARIN

About Language Resources Learn & Exchange Events News Contact

CLARIN and Ukraine

Home / Language Resources / Resource Families

Resource Families

Resource families

- CMC corpora
- Historical corpora

Introduction

The aim of the CLARIN Resource Families project is to provide an overview of available language resources in the field of digital linguistics and the sciences and human language technology. The listings are sorted by language.

The listings for each family include information on the size, text sources, time periods, and concordancers. In addition to the existing valuable language resources, the project also aims to create new resources.

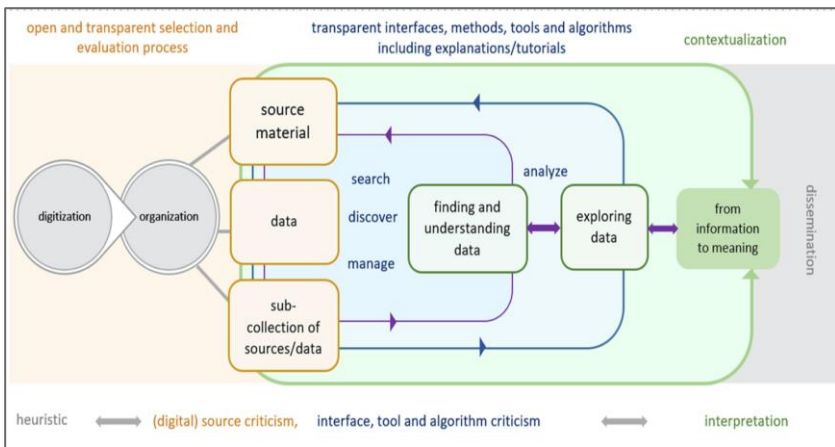
Currently, overviews are available for 12 corpora families, 5 families of lexical resources, and 4 tool families. See below. For information about applying for funding for small projects that can help to extend the scope of the initiative, see <https://www.clarin.eu/content/clarin-resource-families-project-funding>.

Corpora	Lexical Resources	Tools
<ul style="list-style-type: none"> 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 	<ul style="list-style-type: none"> Lexica Dictionaries Conceptual Resources Glossaries Wordlists 	<ul style="list-style-type: none"> Normalization Named entity recognition Part-of-speech tagging and lemmatization Tools for sentiment analysis

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

From Collections to Corpora

Oberbichler, S. et al. (2021) Integrated interdisciplinary workflows for research on historical newspapers: Perspectives from humanities scholars, computer scientists, and librarians. *JASIST*, August 2021. <https://doi.org/10.1002/asi.24565>




digital humanities quarterly

submissions | about dhq | dhq people | contact

2020
Volume 14 Number 2

2020.14.2 | XML | PDF | Print

Digital Humanities and Natural Language Processing: “Je t’aime... Moi non plus”

Barbara McGillivray <bm517_at_cam_dot_ac_dot_uk>, The Alan Turing Institute and University of Cambridge
Thierry Poibeau <thierry_dot_poibeau_at_ens_dot_fr>, CNRS / Ecole normale supérieure — PSL/ Université Sorbonne nouvelle
Pablo Ruiz Fabo <ruizfabo_at_unistra_dot_fr>, Laboratoire LILPa, Université de Strasbourg

Abstract

In spite of the increasingly large textual datasets humanities researchers are confronted with, and the need for automatic tools to extract information from them, we observe a lack of communication and diverging goals between the communities of Natural Language Processing (NLP) and Digital Humanities (DH). This contrasts with the wealth of potential opportunities that could arise from closer collaborations. We argue that more efforts are needed to make NLP tools work for DH datasets so that that NLP research applied to humanities data receives more attention, leading to the development of evaluation approaches tailored towards relevant research questions. This has the potential to bring methodological advances to NLP, while at the same time confronting DH datasets with powerful state-of-the-art techniques.

McGillivray, B., Poibeau, T., & Ruiz Fabo, P. (2020). Digital Humanities and Natural Language Processing: “Je t’aime... Moi non plus”. *DHQ, Digital Humanities Quarterly*. 2020, 14.2

<http://www.digitalhumanities.org/dhq/vol/14/2/000454/000454.html>