CLARIN2022 Panel: CLARIN and Libraries Infrastructures Working Together

Moderator: Martin Wynne
Panelists: Sally Chambers, Andreas Witt, Peter Leinen
CLARIN and Libraries: What are the Next Steps?

Sally Chambers, DARIAH-EU, KBR, Royal Library of Belgium and Ghent Centre for Digital Humanities 🐦 @schambers3
Cultural Heritage Data as Humanities Research Data

Collections as Data?

https://collectionsasdata.github.io

Resource Families

Introduction
The aim of the CLARIN Resource Families project is to provide a comprehensive overview of available language resources in the sciences and humanities. The listings are sorted by language families and can be filtered by type.

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

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.


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