

Language as a Crucial Domain for EOSC

Andreas Witt

EOSC Symposium
Berlin, Germany
23 October 2024

CLARIN



Language as data

- Language is everywhere: dealing with language data is not limited to the Humanities disciplines such as linguistics
- Language is used in all sciences as a means to communicate scientific work
- Since the rise of Large Language Models (LLMs), language resources have become the foundation of the digital landscape
- Large scale infrastructures dealing with language data are already in place, e.g. in Europe the ERIC CLARIN

CLARIN in a nutshell

- Common Language Resources and Technology Infrastructure
- **ESFRI** roadmap 2006, **ESFRI** ERIC status 2012, Landmark 2016
- Easy and sustainable access for schools
 - **digital language data** (written, spoken, video, or multimodal)
 - **tools** to discover, analyse, and combine data wherever they are located
 - **single sign-on** environment
- Ecosystem for **knowledge exchange**
- Some services **integrated in EOSC**



- ERIC members
- Observers
- Countries with participating centres
- Centre Providing Data
- Centre Providing Metadata
- Knowledge Centre



CLARIN as an EOSC node

- CLARIN ERIC, as a member of the SSHOC initiative, aims to create an EOSC Node, offering training, knowledge exchange, and federated language resources from its 73 centres. As a node, CLARIN could enhance interoperability, expand resource access across disciplines, and provide specialized services for large language models.
- CLARIN's experience with service federation could add value to developing the EOSC node concept.

