

Open Research Data Challenges



Interoperability:

Incompatible data formats and non-standardized metadata make it difficult to find and re-use existing datasets.



Personal and sensitive data, copyright and intellectual property rights

Being familiar with the legal framework and handling data appropriately is important for every researcher.



Lack of know-how and training

There is a need for documentation and training in ORD practices in order to implement Open Science standards successfully.

Proposed solution

Building a FAIR-compliant, sustainable and expandable CLARIN-CH ecosystem of federated research infrastructure to address the needs of researchers and professionals using language data in Switzerland

swissuniversities' ORD projects

1) The UpLORD project (2023-2024)

Host institution: UZH

Project partners: LiRI, Zürich UB, CLARIN-CH Coordination Office

The UPLORD project aims to enhance linguistic ORD in Switzerland by improving **workflows, interoperability, and data curation**.

It also focuses on automating repository harvesting, promoting best practices, and providing ORD training.

2) Swiss-AL project (2023-2024)

Host institution: ZHAW

Swiss-AL (Applied Linguistics) is an extensive collection of text data in all Swiss national languages, a linguistic processing pipeline, and a browser-based analysis workbench, enabling researchers to explore copyright-protected data on public language use. It makes data accessible and findable on a dedicated workbench, by processing language data (e.g., websites, social media) in an interoperable, standardized fashion and by fostering reusability for different purposes.

3) CHORD-talk-in-interaction project (2023-2024)

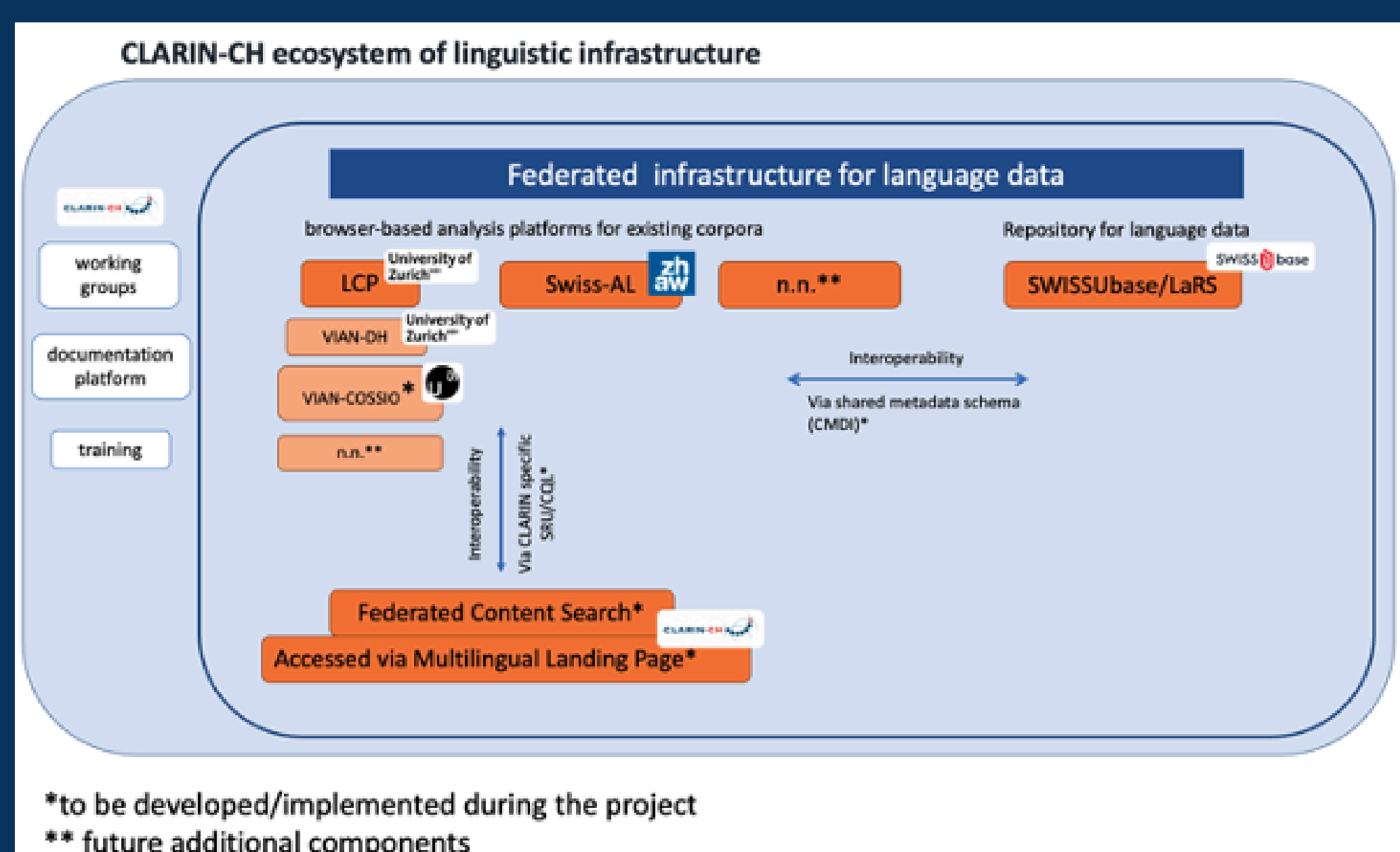
Host institution: USI

Project partners: UNINE, UNIBAS, UNIL

Audio-video recorded and transcribed corpora of spoken language in interaction are particularly complex and offer several challenges when it comes to practices that generate Open Research Data. The project explores possibilities of improvement with a particular focus on a) researchers' data-sharing skills related to corpora of spoken language in interaction and b) the Swiss situation.

4) The FAIR-FI-LD project:

As a follow-up project starting in July 2024, FAIR-FI-LD focuses on prototyping interoperable software, harmonizing metadata, implementing federated content search, and prototyping a platform to visualize, query and analyze multimodal data, supported by comprehensive documentation and training.



Other CLARIN-CH Actions

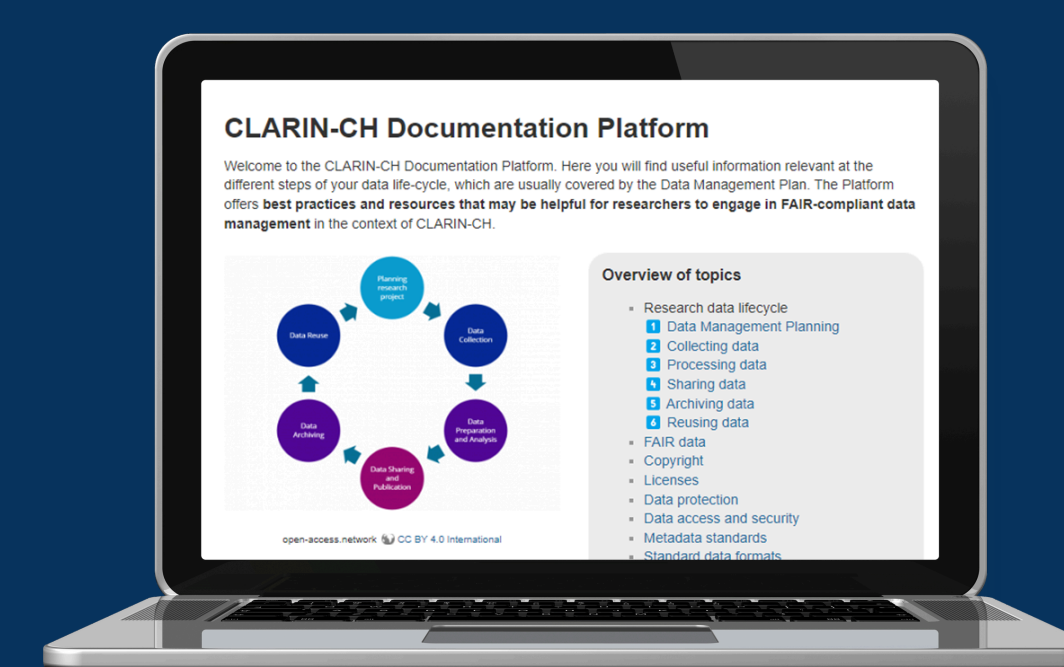
6) National Working Groups

CLARIN-CH WGs are groups of researchers working together on language, resources, and infrastructure topics, to enhance expertise, submit joint research projects, and ultimately build and extend the CLARIN infrastructure, for example:

- ORD projects for linguistic data
- Management of Sensitive and Personal data, Ethical and Legal issues for linguistic data
- (Swiss) Learner Corpora and SLA



7) Documentation Platform



... provides answers to questions regarding FAIR data and open research data management

... contains specific information adapted to the Swiss context

... is regularly updated with insights gained from our working group webinars

8) Support for increasing data FAIR-ness

- Publish and archive the data with LaRS@SWISSubase
- Include the corpus on the Linguistic Corpus Platform (LCP)
- Add the corpus to the SSH Open Marketplace
- Add your corpus on the webpage of the CLARIN Resource Families

9) Vouchers for LiRI services

CLARIN-CH is offering vouchers to researchers from its member institutions for LiRI services such as:

- Statistical consulting
- NLP support
- Swissdox@LiRI



10) CLARIN-CH Helpdesk

The CLARIN-CH Helpdesk, consisting of the members of the CLARIN-CH Coordination Office, provides support for data management, including making a DMP in the context of the CLARIN-CH ecosystem of infrastructures, data protection, legal aspects, FAIR data.

Participate in our survey on data formats

We are currently administering a survey to determine which data formats are most widely accepted among the research community.

Please help us facilitate collaboration among researchers and promote interoperability by filling in the survey for your linguistic subfield!



Scan this code to get to our linktree!



5) Future projects

Aims:

- Extend the national infrastructure by adding new components, e.g. platform for learner corpora
- Consolidate and ensure the sustainability of the components implemented during FAIR-FI-LD

LINKEDIN

<https://www.linkedin.com/company/clarin-ch>

EMAIL

contact@clarin-ch.ch

MAILING LIST

<https://clarin-ch.ch/maillinglist>

WEB

<https://clarin-ch.ch>