Value proposition

The *Common Language Resources and Technology Infrastructure* CLARIN and its Swiss node CLARIN-CH

Table of Contents

CLARIN as pan-European distributed research infrastructure

CLARIN ERIC[1], short for *Common Language Resources and Technology Infrastructure,* is a pan-European research infrastructure which provides easy and sustainable access to a broad range of language data and tools to support research in social sciences and humanities, and beyond.

Currently, through its member countries, CLARIN connects hundreds of researchers as well as research and data infrastructures in 24 European and 2 non-European countries. It enables cross-country collaboration among academics, but also with industry, policymakers, cultural institutions, and the public. CLARIN supports scholars who want to engage in pioneering data-driven research, contributing to a multilingual European Research Area, by offering resources in Open Access and promoting the curation and depositing of data in alignment with the requirements for the interoperability of data and services, CLARIN paved the way for large-scale data sharing and increased reuse of resources.

CLARIN is a networked federation of centres, which may be language data repositories (C-centers), service and data sharing centres (B-centers), or knowledge/expertise centres (K-centers). CLARIN centres share with the community their expertise, their language resources, tools, and provide a large array of services. Thanks to its central hub and its centers, CLARIN supports the sharing, use and sustainability of digital language data, offers advanced tools to explore, exploit and analyse data, and gives access to top level expertise in language sciences.

CLARIN in Switzerland: CLARIN-CH

To enable the participation of Switzerland in CLARIN ERIC, several Swiss higher education institutions (HEIs) and the Academy for Humanities and Social Sciences founded the **CLARIN-CH Consortium** in 2020 with the mission is to join the European CLARIN community and to build an active and impactful sustainable national network that is dedicated to Open Science and FAIR principles. By 2022, all Swiss HEIs joined the consortium and jointly co-fund its activities.

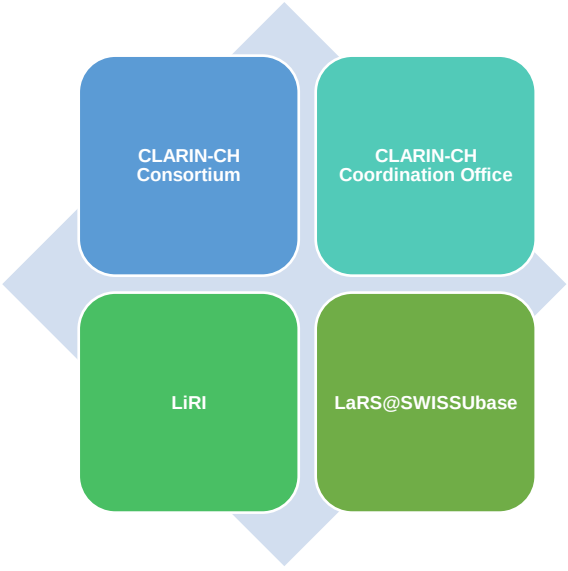
[1] The ERIC legal form (*European Research Infrastructure Consortium*) simplifies the founding and operating of internationally coordinated research infrastructure networks.

Parallelly, CLARIN and its Swiss node received an A-rating by the SNSF and are, thus, included in SERI's 2023 Swiss Roadmap for Research Infrastructures. This confirms its strategic importance for the Swiss scientific community working with language data in the context of Open Science. As targeted, Switzerland joined CLARIN ERIC as observer on January 1, 2023, and plans to become full member early 2025. CLARIN-CH offers thus to the Swiss language related research community the unique possibility to be integrated in the largest, constantly evolving, and very active international infrastructure in the language sciences, but also other language related disciplines such as deep learning, automatic speech recognition, machine translation, artificial intelligence, social sciences, digital humanities.

To run its activities at the national level, the CLARIN-CH Consortium has a solid and fruitful partnership with the **Linguistic Research Infrastructure LiRI**[2] and the **Language Repository of Switzerland** LaRS@SWISSUbase[3]. LiRI is a national technology platform to serve the needs of quantitative research in Switzerland and beyond. LiRI provides support services, facilities, and equipment to the scientific community, it enables collaborative research and deals with big data. LiRI services and expertise deal with deal with the entire data life cycle, support Open Science and Open Research Data. LaRS@SWISSUbase is a FAIR-compliant national platform for the publication and long-term archiving of linguistic research data on Swiss servers (SWITCH).

CLARIN-CH, as the Swiss branch of the European CLARIN, consists of:

- CLARIN-CH Consortium of Swiss HEIs – the kernel of the national network
- CLARIN-CH Coordination Office – in charge with the operation of CLARIN-CH and its driving force
- LiRI – the technical data and service providing center
- LaRS@SWISSUbase – the FAIR-compliant national repository



The strong partnership between these four entities enables the participation of Switzerland in CLARIN and the existence of a national ecosystem of research infrastructure and network to better support Swiss scholars in their research and in managing their language data in the spirit of Open Science and FAIR principles.

[2] https://www.liri.uzh.ch/en.html
[3] https://www.lars.uzh.ch/en.html

Benefits from being a member of CLARIN-CH

The CLARIN-CH Consortium makes possible that Switzerland is a member of CLARIN. By getting actively involved in CLARIN-CH, each Swiss HEI brings its formal and financial support to the national consortium. At the same time, being member of the CLARIN-CH consortium comes with a series of benefits both at the individual and at the institutional level.

At the individual level: European CLARIN

1.1 Full access to CLARIN digital language resources

CLARIN offers monolingual and multilingual digital language resources in written, spoken, or multimodal form available for numerous European languages and families of languages, as well as to advanced tools to explore and exploit such data sets wherever they are located, including several technical webservices.

- More than 5500 corpora, organised in 15 families[4]
    - 23 Computer-mediated communication corpora
    - 24 corpora of academic texts
    - 76 historical corpora
    - 75 L2 learner corpora
    - 33 legal corpora
    - 45 literary corpora
    - 70 manually annotated training corpora and corpus collections
    - 17 multimodal corpora
    - 31 newspaper corpora
    - 8 oral history corpora
    - 82 parallel corpora
    - 35 parliamentary corpora
    - 30 reference corpora
    - 81 sign language resources
    - 148 spoken corpora

[4] https://www.clarin.eu/resource-families

- More than 390 lexical resources:
    - 98 language models
    - 83 lexica
    - 95 dictionaries
    - 29 conceptual resources
    - 33 glossaries
    - 58 wordlists
- More than 60 NLP tools:
    - 95 corpus query tools
    - text normalization
    - named entity recognition
    - part-of-speech tagging and lemmatization
    - tools for sentiment analysis

- CLARIN Catalogue for automatic metadata harvesting: the Virtual Language Observatory (VLO). The VLO is a means of exploring language resources and tools. It provides an easy-to-use interface, allowing for a uniform search and discovery process for a large number of resources from a wide variety of domains. Facets make it easy to explore and access available resources. A powerful query syntax makes it possible to carry out more targeted searches as well[5].
- CLARIN search engine in language resources hosted at a variety of institutions: the CLARIN Federated Content Search is a technology to access and search resources that are available at different locations. Using a common search page, someone can search simultaneously over the resources available at various places. Technically this means that each location has a search service for the resources provided by them and the FCS client accesses all of them, sending a query. In CLARIN, FCS is used to search within language resources at different CLARIN centres. Someone using FCS searches simultaneously at various centres and receive the answer to a query on a common website[6].
- CLARIN execution environment for automatic annotation of text corpora:
    - *Language Resource Switchboard*: a tool that helps to find a matching language processing web application for a set of data. After uploading a file or entering a URL, one can select which task to perform. The Switchboard will then provide the user with a list of available CLARIN tools to analyze the input[7].
    - *Weblicht*: an execution environment for automatic annotation of text corpora offered by the German node of CLARIN, CLARIN-D. Linguistic tools such as tokenizers, part of speech taggers, and parsers are encapsulated as

[5] https://www.clarin.eu/content/virtual-language-observatory-vlo
[6] https://www.clarin.eu/content/content-search
[7] https://www.clarin.eu/content/language-resource-switchboard

web services, which can be combined by the user into custom processing chains. The resulting annotations can then be visualized in an appropriate way, such as in a table or tree format[8].

1.2 Full access to CLARIN Knowledge Infrastructure

CLARIN facilitates a secure a continuous transfer of knowledge and expertise between all members through the services provided by CLARIN Knowledge Centres (K-centres). These centres are a cornerstone of the CLARIN Knowledge Infrastructure, one of the main components ensuring a continuous transfer of knowledge between all players involved in the construction, operation, and use of the infrastructure.

The mission of CLARIN Knowledge infrastructure is to ensure that the available knowledge and expertise does not exist as a fragmented collection of unconnected bits and pieces but is made accessible in an organised way to both the CLARIN community and the social sciences and humanities research community more widely. The focus of CLARIN is on language resources (in all modalities, from all regions and with any topical orientation) and K-centres serve researchers and educators from any discipline where language plays one of its many roles, ranging from object of study, a means of communication or expression, a means to store and extract information, object of learning or teaching activities, to training source for data-driven analytics, and many others. K-centres share their knowledge and expertise on one or more aspects of the domain covered by the CLARIN infrastructure and can be mostly found in CLARIN countries, but also beyond[9]. Examples of services provided by K-centers (online courses, training materials, best-practice documents, guidance in getting access to and using data and tools).

1.3 Learning Hub

Access to a large array of services provided by the CLARIN community: The CLARIN Learning Hub gives access to open educational resources on various topics, including full online training modules to learn new skills and materials to design new university courses, training, and workshops. Additionally, the hub contains best practices and guidelines

[8] https://weblicht.sfs.uni-tuebingen.de/weblichtwiki/index.php/Main_Page
[9] https://www.clarin.eu/content/knowledge-centres

developed in educational projects, such as UPSKILLS, or created in collaboration with other research infrastructures[10].

1.4 Funding Hub

Access to numerous European individual funding opportunities, such as Teaching with CLARIN, Trainer Network Programme, Mobility Grants for training and scientific events for sharing technical expertise and know-how in building the CLARIN infrastructure and to reinforce international collaborations. In addition, the central CLARIN office and the national consortia can provide **support for preparing EU-funded projects**, such as help for networking and documentation, assistance in writing the proposal, small grants that can be used to cover certain costs coming with the preparation of a project proposal, facilitation services regarding the Open Science and FAIR Data requirements[11].

1.5 Participation in European collaborative projects

Increased involvement in European and national collaborative projects:

Since 2015, CLARIN ERIC has been participating as partner in European projects funded by the European Commission and addressing topics related to research infrastructures, such as in the Horizon 2020 programme or as part of Horizon Europe[12]. Another example is the European initiative to create the ParlaMint corpus[13], which contributes to the creation of comparable and uniformly annotated multilingual corpora (more than 17 languages) of parliamentary sessions. Scholars from all CLARIN countries are called to participate for the development of the corpus[14].

CLARIN-CH

1.6 Access to services proposed by LiRI, LaRS@SWISSUbase and other CLARIN-CH members

1.6.1 Free access to the Linguistic Corpus Platform (LCP@LiRI), including VIAN-DH for text and multi-modal data annotation

LCP@LiRI is a platform for the hosting of complex linguistic annotated data allowing refined search and quantitative analyses. Infrastructure which allows hosting of corpora not for archiving, but for online searching and querying the data, is a key element of ORD.

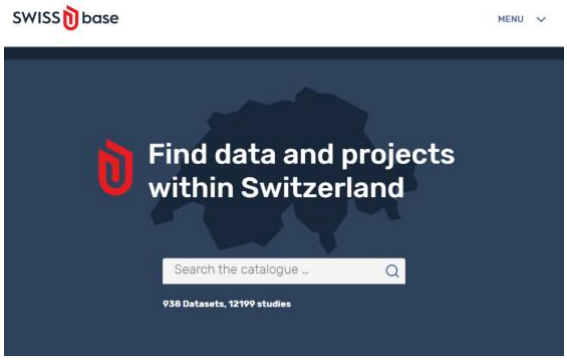
[10] https://www.clarin.eu/content/learning-hub
[11] https://www.clarin.eu/funding
[12] https://www.clarin.eu/content/eu-projects-and-international-initiatives
[13] https://www.clarin.eu/parlamint
[14] https://www.clarin.eu/parlamint-project-information

The LCP not only simplifies work with corpora by offering even complex queries via an easy-to-use interface, but also facilitates the reproducibility and reusability of the analyses by other researchers. Import and export functions accept and offer standard data formats and all conceivable forms of annotation can be added as any number of layers. This fosters the reusability of these corpus data, e.g. by adding new annotation layers.

Within LCP@LiRI, ORD principles are strengthened by explicitly encouraging data sharing and reuse free of charge. VIAN-DH is a web application tool for automated annotation and transcription tasks for multi-modal data collections that is integrated in the LCP. With VIAN-DH, classical methods of interaction linguistics are combined with modern methods such as automatic speech and gesture recognition to combine qualitative with quantitative methods. The software is open source, enabling data transparency, sharing and reproducibility according to ORD principles.

1.6.2 FAIR repository with tailored metadata which is harvested by the CLARIN catalogue

As a national platform for the publication of linguistic research data, LaRS@SWISSUbase provides Swiss HEIs with a reliable data infrastructure. LaRS facilitates access to research data and projects across the linguistic domain. Furthermore, a team of data curators contribute towards a high level of data and metadata quality.



The interest of scholars in publishing their data with LaRS@SWISSUbase is in increase, and this is due, on the one hand, to the key features of the repository, as well as to the proactive approach by the CLARIN-CH Coordination Office to encourage and support data owners for increasing the FAIR-ness of their data and corpora.

1.6.3 Swissdox@LiRI: discount on subscription

Swissdox@LiRI, the largest database of Swiss media texts, is hosted at LiRI. It al-lows the distribution and the use for academic purposes of journalistic content despite its regular

restrictive copyright conditions. The database contains a daily growing collection of at the moment 24 million media articles of more than 250 media titles.

Swissdox@LiRI

Until today, researchers of more than 80 projects at nine academic institutions have down-loaded data. The data is provided in CSV format which allows further automatic processing, e.g. natural language processing. The contract with SMD Swiss Media Database allows free use of the data within an academic project, but only derivates of the data may be redistributed and archived. To allow reproducibility of analyses based on Swissdox@LiRI data and as much compatibility with the FAIR principles as possible, the query the user defined to get the data set is provided in JSON format and can be published and archived. This allows researchers whose institutions have access to Swissdox@LiRI to reproduce the data basis.

Swissdox@LiRI, the LCP and the web application tool VIAN-DH for automated annotation and transcription tasks for multi-modal data collections, offer **state-of-the-art language technology solutions** for big data to researchers not only in the linguistic community, but also for researchers in many other disciplines like psychology, political science, communication & media, business administration, finance, medicine, and film studies.

1.6.4 Vouchers for LiRI services

In collaboration with LiRI, CLARIN-CH offers to researchers from its member institutions vouchers for LiRI services such as access to swissdox@LiRI database and statistical consulting.

I am very grateful for the CLARIN-CH voucher that gives me the opportunity to have a free statistical consulting for my PhD thesis in linguistics. The voucher is especially valuable to me since it will for sure increase the quality of my work.

(Franziska Keller, PhD student, University of Fribourg)

1.6.5 Access to Swiss-AL platform (ZHAW)

Swiss-AL is a language data platform for Applied Sciences (Digital Discourse Lab, ZHAW). The platform hosts large multilingual written text corpora, mostly obtained by crawling Swiss web domains. The platform includes an analysis workbench specifically addressing researchers from applied, non-linguistic disciplines, e.g., researchers from social sciences, law, psychology, and architecture, and will contain dedicated open educational resources that are being developed within the swissuniversities ORD program. The workbench allows for quantitative linguistics analysis, including recent machine learning based methods.

1.7 Access to CLARIN-CH knowledge hub, training, and support

1.7.1 National working groups, on relevant topics as ORD, sensitive data, legal issues, how to "open" "closed" research data

CLARIN-CH working groups (WG) are groups of researchers (i.e. CLARIN-CH members, other national and European scholars) that are interested in language-, resource- and infrastructure-related topics, and who work together with peers in a formalized and sustainable environment. The purpose is to bring together expertise on a specific topic, prepare joint research projects and to serve the community, which can be scholarly and technical. The long-term goal is to build and to extend the CLARIN and CLARIN-CH infrastructure.

Working Group: ORD projects for linguistic data

This WG[15] was founded in spring 2023 by members of three ORD projects funded by the swissuniversities' 2022-2024 programme for ORD:

(i) Upgrading the linguistic ORD-ecosystem UpLORD[16]
(ii) Swiss-AL: Linguistic ORD Practices for Applied Sciences[17]
(iii) CHORD Data-sharing skills in corpus-based research on talk-in-interaction[18]

The aim of this WG is to **join their forces** to:

- better address the ORD requirements for linguistic data

[15] https://clarin-ch.ch/working-groups/linguistic-ord-practices
[16] https://www.liri.uzh.ch/en/projects/UpLORD.html
[17] https://www.zhaw.ch/en/linguistics/research/swiss-al-linguistic-open-research-data-practices-for-applied-sciences/
[18] https://www.chord-talk-in-interaction.usi.ch/

- upgrade workflows and interoperability of existing infrastructure services (LiRI, Swiss-AI, LaRS@SWISSUbase) and adapt them to answer the needs of linguistic disciplines (i.e. interactional linguistics)
- document and promote best practices by type of data (e.g. textual corpora, multimodal corpora)
- raise awareness and training about ORD practices in the context of teaching, research, publishing, and applied linguistics
- build a robust practice of data curation in general and by discipline

Working Group: Management of Sensitive and Personal data, Ethical and Legal issues for linguistic data

This WG[19] addresses questions regarding the management of sensitive and personal data, as well as ethical and legal issues for linguistic data in Switzerland and aims to disseminate information and advice about these topics in the Swiss scientific community using language data.

The WG started in September 2023 with a kick-off event, which focused on data collection, protection and preservation and their associated procedures, with respect to different types of linguistic data (e.g., multimodal, historical, experimental, sociolinguistic, data from social media, data from different age groups). The slides and the recording of the three-keynote talked are shared on the CLARIN-CH Documentation Platform.

During the kick-off event, we collected questions from the community (discipline-specific) about managing sensitive and personal data, as well as copyright issues especially when it comes to social media and internet data. To answer these questions, we invite researchers and professionals (e.g. from grant and research support offices, from the legal services of higher education institutions, from the cantonal legal services) to share from their experience and to answer participants' questions in the form of a webinar.

Working Group: Swiss Learner Corpora and SLA

This WG[20] aims at increasing the accessibility and reusability of Swiss Learner Corpora for Learner Corpora Research and Language teaching. In Switzerland, there are numerous multimodal corpora which cannot be accessed and reused for various reasons. In this context, this WG has the following objectives:

- Make an inventory of existing Swiss learner corpora
- Make an analysis of the reasons for which these resources are not accessible
- Make recommendations about how to render these resources accessible and reusable

[19] https://clarin-ch.ch/working-groups/sensitive-personal-data
[20] https://clarin-ch.ch/working-groups/swiss-learner-corpora

- Produce best practice documentation about how to avoid having resources that cannot be openly shared and reused, such as transcription conventions, anonymisation, consent, etc.

The WG started in December 2023 with a kick-off meeting, whose purpose was to launch the inventory of existing Swiss learner corpora and datasets. As a second step, a survey to collected more detailed information about each corpus and dataset identified is carried out to understand the reasons for which these resources are not accessible and how to increase their FAIR-ness.

1.7.2 Training opportunities: webinars, workshops, hands-on sessions, doctoral courses

Monthly webinars to address essential topics related to the management of open- and FAIR-research data organised as part of the CLARIN-CH working group on *Management of Sensitive and Personal data, Ethical and Legal issues for linguistic data:*

- Protection of personal and sensitive linguistic data: Legal aspects
- Overview of anonymisation tools
- Protection of personal and sensitive linguistic data: Technical aspects
- Case study on anonymisation of textual data
- Case study on anonymisation of video and audio data
- Intellectual property rights in linguistic data
- Legal aspects of collecting and sharing social media data (copyright)
- Cross-border projects: which laws apply?
- Legal aspects of sharing and reusing linguistic data (licenses)

Workshops, training sessions and hands-on sessions offered via the **partnership with LiRI and LaRS@SWISSUbase**:

- linguistic data management according to the ORD practices
- tools and infrastructure developed by LiRI, such as the Linguistic Corpus Platform and the swissdox@LiRI media database
- statistics for linguistics data proposed by LiRI

Doctoral schools, workshops and courses offered via the **CLARIN-CH co-branding programme[21]:**

- Managing Languages, Arguments and Narratives in the Datafied Society (LAND) Spring School: Questioning Large Language Models, understanding (multimodal) argumentation in the age of AI and social media.
- Doctoral workshop: Digging into Data – Methods for the Collection and Mining of Various Types of Linguistic Data

[21] https://clarin-ch.ch/co-branding/start

- Winter School on Corpus Data for the Analysis of Discourse, Interactions and Arguments
- Linguists in tech: closing the skills gap
- International Pragmatics Association (IPrA) Mentoring Program

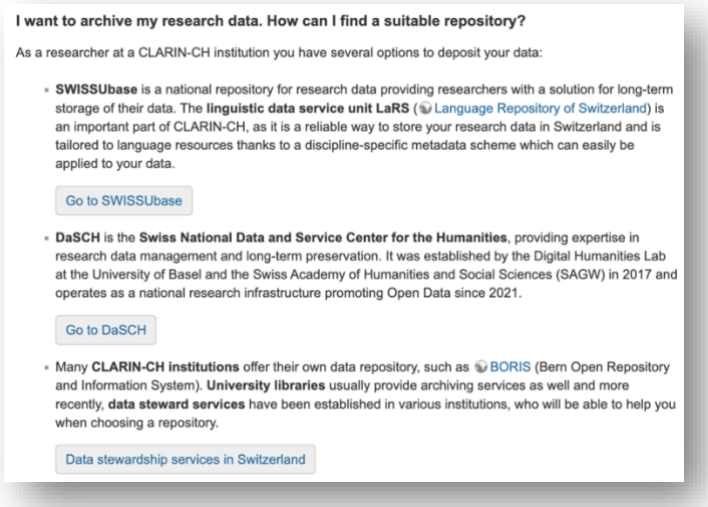
1.7.3 Documentation platform and FAQ

The CLARIN-CH Documentation Platform[22] offers useful information relevant at the different steps of your data life cycle, which are usually covered by the Data Management Plan. The Platform offers best practices and resources that may be helpful for researchers to engage in FAIR-compliant data management in the context of CLARIN-CH.

Overview of topics

Research data lifecycle
1) Data Management Planning
2) Collecting data
3) Processing data
4) Sharing data
5) Archiving data
6) Reusing data

FAIR data
Copyright
Licenses
Data protection
Data access and security
Metadata standards
Standard data formats



1.7.4 Helpdesk: support for data management, including DMP, data protection, legal aspects, FAIR data

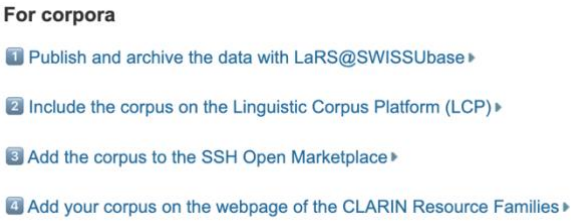
The CLARIN-CH Helpdesk, consisting of the members of the CLARIN-CH Coordination Office, provide support for data management, including making a DMP in the context of the CLARIN-CH ecosystem of infrastructures, data protection, legal aspects, FAIR data.

[22] https://clarin-ch.ch/documentation-platform/start

This can be via email, or online meetings. When necessary, more specialized support is searched within the larger CLARIN and CLARIN-CH network.

1.7.5 Increase the FAIR-ness of your corpora campaign

This campaign consists in reaching out to owners of corpora created in Swiss HEIs to explain the existing options to increase the FAIR-ness of the corpora and provide support during the process.



1.8 Participation in CLARIN-CH coordinated projects

1.8.1 Coordination and synergy-identification for better and sustainable cooperation

Since the founding of the national consortium in 2020, the CLARIN-CH Coordination Office plays an important role of coordination and synergy-identification within the Swiss scientific and infrastructural community using language data.
Some notable examples are:

- the implementation of a solid and fruitful partnership with LiRI (UZH), the LaRS@SWISSUbase repository and the Swiss-AL platform (ZHAW) to increase the access of Swiss researchers to specialized services (e.g. increasing the degree of FAIR-ness of Swiss data and corpora, statistical consulting, the creation of a national linguistic corpus platform, discounts on access to the biggest Swiss media database).
- the creation of national and cross-institutional WGs that address crucial issues raised by dealing with language data in the context of Open Science.
- the participation in two joint funding applications, among which a project that aims at building an unprecedented national FAIR-compliant ecosystem of federated infrastructure for language data.

1.8.2 Funding opportunity by supporting and writing joint funding applications

The UpLORD project: Upgrading the linguistic ORD-ecosystem

UpLORD is a swissuniversities ORD-funded 2-year project (2023-2024) hosted by the University of Zurich, with the support of the Zurich University Library and CLARIN-CH. Since 2018, a consortium of partners has been working on building a national ecosystem

of infrastructures, which covers the whole linguistic data lifecycle according to ORD requirements (FAIR principles: Findable, Accessible, Interoperable, Reusable) from data generating, processing, and analyzing to data sharing and archiving. This ecosystem includes the national technology platform LiRI and the national repository for publishing and archiving linguistic data (SWISSUbase) as service providers, a database of Swiss media texts and a platform for hosting of and searching in large text and audio/video corpora.

The project focuses on upgrading workflows and interoperability of existing infrastructure services, establishing working groups on the national level, documenting and promoting best practices, raising awareness and training about ORD practices in the context of teaching, research, and publishing, and building a robust practice of data curation. In the long-term, this project will significantly contribute to a strong foundation for a sustainable ORD strategy for linguistic data in Switzerland.

The FAIR-FI-LD project: Moving towards a national FAIR-compliant ecosystem of Federated Infrastructure for Language Data

FAIR-FI-LD is a joint project proposal submitted to the swissuniversities ORD programme (2024-2025) by the University of Zurich, the Zurich University of Applied Sciences, Università della Svizzera italiana and CLARIN-CH. This project addresses the question of reduced interoperability between the existing national services for language data, which include, up to now, the Linguistic Research Infrastructure (LiRI), the Swiss-AL Platform for Applied Sciences (ZHAW), the national repository for the publication and long-term preservation of language data LaRS@SWISSUbase, and various smaller tools and services. This reduces the potential for collaboration and data reuse.

With the foundation of the CLARIN-CH consortium in 2020, Swiss higher education institutions started to work together to build a FAIR-compliant, sustainable, and expandable CLARIN-CH ecosystem of federated infrastructure to answer the needs of researchers and professionals using language data in Switzerland and beyond.
The FAIR-FI-LD project describes the vision of moving towards a national FAIR-compliant ecosystem of federated infrastructure for language data, and proposes concrete steps towards this mid- and long-term goal, in compliance with the Swiss ORD strategy,

by prototyping:

- interoperable underlying software using NLP techniques and exploratory AI techniques
- harmonized metadata between the existing Swiss infrastructure components and the European CLARIN infrastructure
- CLARIN federated content search (FCS) to query each component of the infrastructure

- a FCS multilingual landing page hosted on the CLARIN-CH website
- a frontend of the VIAN-DH@LiRI environment to visualize, query and analyze multimodal talk-in-interaction data

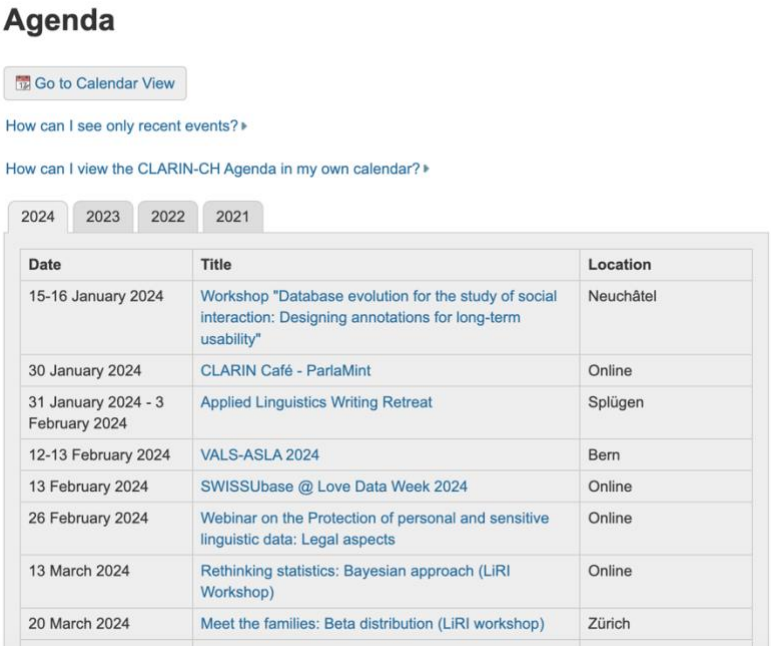
by producing:

- documentation and training to support the use of the infrastructure and inform about legal and ethical issues related to language data in the context of Open Science,

… and by planning the future collaboration with further stakeholders and aggregation of further tools and services.

1.8.3 Community building

To foster the exchange of information within members of the CLARIN-CH community and to foster exchanges, we created the CLARIN-CH Agenda.



Agenda

Go to Calendar View

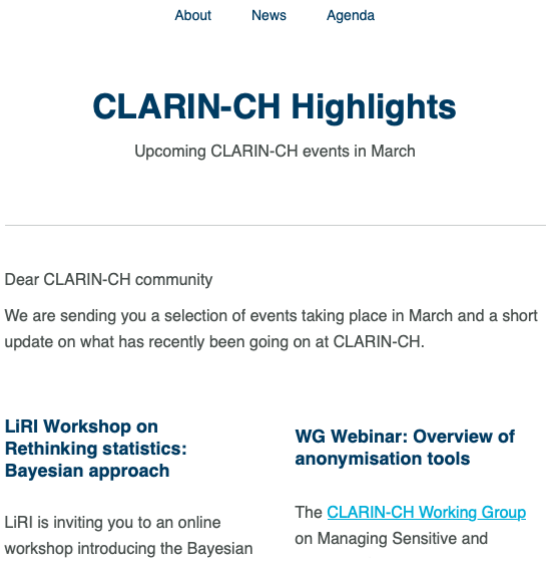
How can I see only recent events?

How can I view the CLARIN-CH Agenda in my own calendar?

2024 2023 2022 2021

Date	Title	Location
15-16 January 2024	Workshop "Database evolution for the study of social interaction: Designing annotations for long-term usability"	Neuchâtel
30 January 2024	CLARIN Café - ParlaMint	Online
31 January 2024 - 3 February 2024	Applied Linguistics Writing Retreat	Splügen
12-13 February 2024	VALS-ASLA 2024	Bern
13 February 2024	SWISSUbase @ Love Data Week 2024	Online
26 February 2024	Webinar on the Protection of personal and sensitive linguistic data: Legal aspects	Online
13 March 2024	Rethinking statistics: Bayesian approach (LiRI Workshop)	Online
20 March 2024	Meet the families: Beta distribution (LiRI workshop)	Zürich

The CLARIN-CH Newsletter and the monthly CLARIN-CH News highlights are means to keep up to date the members of the CLARIN-CH with information about what has been happening in CLARIN and CLARIN-CH, such as activities, events, training opportunities, services, calls for participations, and other relevant information.



About News Agenda

CLARIN-CH Highlights

Upcoming CLARIN-CH events in March

Dear CLARIN-CH community

We are sending you a selection of events taking place in March and a short update on what has recently been going on at CLARIN-CH.

LiRI Workshop on Rethinking statistics: Bayesian approach

LiRI is inviting you to an online workshop introducing the Bayesian

WG Webinar: Overview of anonymisation tools

The CLARIN-CH Working Group on Managing Sensitive and

CLARIN-CH Day 2024: Open Research Data – Challenges and Opportunities

The event is the first of a series of annual meetings of the CLARIN-CH community. It is organized by the CLARIN-CH consortium in cooperation with its member institutions and aims to support the scientific community in their challenges when it comes to Open Research data. It seeks to foster exchange and to enable the encounter between researchers and data management experts.

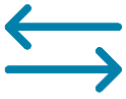
The 2024 edition aims to bring together experts and researchers to discuss challenges and opportunities, and to open the dialogue on standards and practices of open research data as well as the legal and ethical aspects of processing and sharing linguistic data. The event builds on the work done by two CLARIN-CH Working Groups, which address essential topics related to Open Research Data.

At the institutional level:

1.9 Increased national and international visibility

for Swiss assets (research projects, data, tools, methodologies, expertise, etc.) through (1) automatic integration into the largest and most visible online catalogue of resources and (2) through its Tour de CLARIN, which periodically highlights prominent activities and accomplishments of a particular national network. As such, national resources (corpora, databases, language models, dictionaries) and the tools created and developed in Switzerland, as well as the expertise available at the universities and research institutes members of CLARIN-CH, will be made available to the CLARIN community.

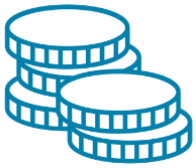
1.10 Adoption of international standards to ensure interoperability



Similarly to CLARIN, CLARIN-CH aims to connect researchers across international borders by offering access to data and services in line with the FAIR Principles. For this, CLARIN-CH and its partners work for the adoption of international standards to ensure interoperability in the construction of national databases and infrastructure. The two main pillars of the CLARIN-CH ecosystem, LiRI and LaRS@SWISSUbase, already comply

with the general norms of Open Science and the corresponding requirements of the SNSF. In terms of metadata, LiRI practices comply with contributes to the dissemination of unified standards and best practices in linguistics. In this context, at the national level, all data and publications issued from CLARIN-related projects and use of resources and/or tools are in line with the FAIR Principles.

1.11 Cost reductions and a maximization of the profitability



By sharing language resources nationally and internationally, the effective costs invested for creating new resources (corpora, database, tools are diminished. Furthermore, CLARIN-CH represents a framework in national actions and investments can be united at the national level so that their efficiency, their effectiveness, their impact, as well as the dissemination of their activities, can be increased.

1.12 Cooperation within and beyond the SSH field

Since its founding in 2020, CLARIN-CH has rapidly become an important stakeholder and a driving force for the Swiss research infrastructure landscape, as highlighted by the fruitful cooperation opportunities with other national infrastructures, such as FORS and DaSCH, other nodes of European ERICs with Swiss participation, such as DARIAH-CH, the SSHOC-CH cluster, as well as the future Swiss EOSC node.

The SSHOC-CH cluster[23] has the mission to cluster SSH research infrastructures in Switzerland (national infrastructures and national nodes of international infrastructures) to ensure an exchange and cooperation to support research projects and researchers, to identify and create synergies and, where possible, to develop joint platforms and services or make existing ones interoperable.

Impact of CLARIN and its Swiss node

As a well-established research infrastructure, CLARIN has the experience and infrastructure necessary for **serving the scientific community**, notably disciplines that profit from existing language resources (e.g. archelogy, history, psychology, sociology, to name but a few), and the non-scientific community (e.g. libraries, archives, and museums, governmental organisms, and the industry sector). As such, the CLARIN-CH network serves not only scientific communities working on language data, but also the communities from

[23] https://sshoc.ch/start

other disciplines, which are not interested in language per se by providing skills, expertise, and tools for the automatic treatment of all sorts of language related data sets.

CLARIN provides the necessary infrastructure to allow the **creation of relevant technologies** in the fields of deep learning, automatic speech recognition, machine translation and artificial intelligence, such as NLP pipelines, speech processing systems, machine translation systems, environments for manual annotation and evaluation. Within CLARIN-CH, LiRI contributes to providing infrastructure and services in the areas of software development, NLP, Human Language Technology and offer services, such as collecting and processing language data, development, and maintenance of purpose-built applications, long-term archival of research data, including tailored workshops, consulting, and coaching.

Finally, by joining CLARIN, the CLARIN-CH community will have a role to play in the transformational change that targets the implementation of a knowledge-based economy with potential for secure and sustainable economic growth both at the national and the European level. CLARIN and its Swiss node are non-profit organizations, so their contacts to industry are primarily focused on the exchange of information. Nevertheless, they are open to cooperation with the private sector and can thus encourage start-ups, which in turn can positively impact the national and the European industry and thus also research and society.