Common Language Resources and Technology Infrastructure

Strategy 2021-2023 at a glance

The Common Language Resources and Technology Infrastructure (CLARIN) is a virtual platform for everyone interested in language. **CLARIN offers access to language resources, technology, and knowledge, and enables cross-country collaboration among academia, industry, policy-makers, cultural institutions, and the general public.**

Researchers, students, and citizens are offered access to digital language resources and technology services to deploy, connect, analyse and sustain such resources. In line with the Open Science agenda, CLARIN enables scholars from the Social Sciences and Humanities (SSH) and beyond to engage in and contribute to cutting-edge, data-driven research. Furthermore, CLARIN facilitates researchers' participation in the European Research Area and the mission-driven programmes of Horizon Europe and the wider global context. The GLAM sector (Galleries, Libraries, Archives, and Museums) can use CLARIN as a hub for services to access and enrich digital cultural heritage objects and to establish collaborations with academic partners. For governmental institutions and the private sector, CLARIN offers tools and services to maximize the informational value of the growing amount of digital data. CLARIN is at the forefront of technological innovations, such as deep learning, automatic speech recognition, machine translation, and artificial intelligence. By organising the training of new generations of skilled data professionals, CLARIN contributes to reinforcing digital literacy and reducing the shortage of critical skills in the field of data analytics.

CLARIN is a distributed data infrastructure, governed and coordinated by CLARIN ERIC, a European Research Infrastructure Consortium, currently joined by 24 member and observer countries. CLARIN ERIC is recognised as a Landmark by the European Strategy Forum on Research Infrastructures (ESFRI). As a distributed research infrastructure, CLARIN consists of a federation of centres that offer resources, technology, and knowledge.
To promote the interoperability of resources and technology, CLARIN encourages the use of common metadata standards.

Impact

Language technology has become one of the most influential technological innovations of the data science era. Whether it is researchers using resources and tools to address new research questions, governments and industry applying text-mining algorithms to find valuable patterns in large amounts of language data and discriminating valid information from misinformation, or citizens using applications such as automatic speech recognition, machine translation or autocomplete, language technology is omnipresent. Unsurprisingly, experts have high hopes of applying smart language technologies to develop appropriate measures for today's societal challenges, such as climate change, inequality, pandemics, and illnesses. Each of these leave trails of language data, allowing researchers to study their causes and effects as well as the social and cultural dynamics stirred by them.

Priority areas 2021-2023

In the strategy period 2021-2023, CLARIN will focus on four priority areas to increase the potential for uptake and impact:

Sustainability

The stage of maturity now reached by CLARIN calls for greater attention to be paid to technical, financial, and organisational sustainability. CLARIN has had a considerable impact on a variety of fields, including digital history, media studies, language variation, election studies, social signal processing, translation studies, speech pathology, and migration studies. In addition, CLARIN helps in shaping the development of the European research infrastructure landscape, by actively contributing to advancing the models for collaboration among different research infrastructures. In the strategy period 2021-2023, further collaboration will be sought with non-European parties. To ensure CLARIN's future, the membership base will be consolidated and extended. Policies will be developed to protect valuable resources that are in danger of becoming inaccessible. Finally, the financial portfolio will be diversified and viable models for collaboration with industry will be articulated and promoted.

Technical infrastructure

The technical infrastructure is the foundational layer of CLARIN. Through a single sign-on environment, people from all over the world can access the resources and technology offered by CLARIN. Over the next three years, CLARIN will further invest in the three components that make up the architecture of the technical infrastructure: robustness, interoperability, and innovation. The visibility and interoperability of resources and tools will be enhanced by synchronising various sources of information and preparing for further integration into the European Open Science Cloud, the SSH Open Marketplace and other relevant platforms. Adherence to the FAIR Data Principles (i.e., making digital objects Findable, Accessible, Interoperable, and Re-usable) will be advocated both within the CLARIN community and to third parties. Relevant technological trends are closely watched and examined for their potential to be incorporated in the CLARIN infrastructure.

Knowledge infrastructure

In addition to resources and technology, CLARIN offers a broad range of expertise related to language data and tools. CLARIN's knowledge infrastructure is not only supported by thematic knowledge centres and committees, but also by large numbers of lecturers and teachers who have integrated CLARIN in their courses. Numerous initiatives, such as an annual conference, various workshops and events, an ambassadors programme, and Tour de CLARIN – highlighting contributions from national CLARIN consortia and CLARIN centres – help the exchange of knowledge both within the existing CLARIN community and with potential new communities of use and stakeholders. **By deploying this knowledge infrastructure and maintaining partnerships, it is ensured that CLARIN is aligned with the research agendas of the digital humanities, computational social sciences, and data science at large.** In the next strategy period, efforts will be made to reach out to policy-makers, library networks, science journalists, and the wider audience interested in the dynamics of data science. The visibility of services offered by CLARIN will be increased by improving the information structure of the CLARIN ERIC website, by offering materials like video tutorials and showcases, and setting up a sustainable network of CLARIN trainers.

Organisational development

In recent years, CLARIN ERIC has gradually evolved from a relatively small project organisation into a permanent organisation with a professional governance model and competent support office. To optimise the use of human resources, efforts will be made to evaluate the current organisation and explore opportunities for reinforcement of collaboration among various CLARIN bodies and staying connected with existing and new networks. In alignment with other research infrastructures, arrangements will be made to ensure that the right instruments are in place to develop the capacities of the central staff. These endeavours will increase the expertise and skills of the central staff needed to build, maintain, and innovate the CLARIN infrastructure, while simultaneously reinforcing the career perspectives of the central staff. ERIC Forum, an alliance of European research infrastructures, has proven to be a particular fruitful consultative body in this respect.

A more elaborate version of the strategy 2021-2023 can be found on the CLARIN ERIC website.