

Title CLARIN ERIC Strategy 2024-2026
Version 1.0
Author(s) CLARIN ERIC Strategy Taskforce
Date 4 December 2023
Status Approved
Distribution GA, SAB, NCF, BoD
ID CE-2023-2333



Decision

The General Assembly approves the CLARIN ERIC Strategy 2024-2026 (CE-2023-2333).

Language as Social and Cultural Data

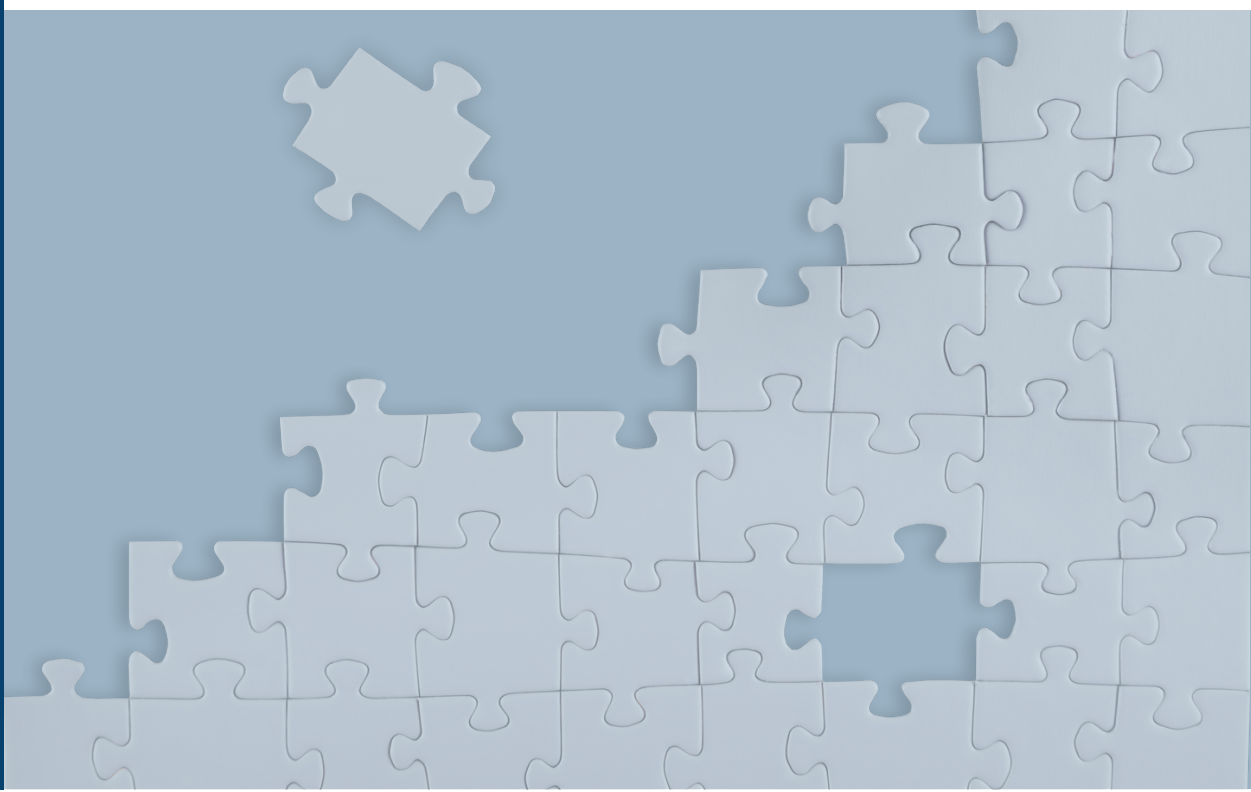
*Enhanced Infrastructural Support
for Research Using
Language Materials
in the Era of AI*

CLARIN ERIC
Strategy 2024-2026



TABLE OF CONTENTS

BACKGROUND	2
INTRODUCTION	3
HOW CLARIN WORKS	4
CLARIN'S STRATEGIC ORIENTATION	6
Supporting Scientific Excellence and Innovation	6
Fostering Strategic Partnerships	7
Forging Impact Pathways	8
CLARIN'S STRATEGIC AGENDA	9
Pillar 1: Our Users	10
Pillar 2: Our Offer	12
Pillar 3: Our Landscape	15



Background

This document presents the CLARIN ERIC Strategy for the period 2024-2026 and will form the basis for the work plan for this phase. The new strategy builds on those aspects from the CLARIN ERIC Strategy 2021-2023 that remain relevant, and also expands the previous strategy in areas that are developing and/or gaining significance for the infrastructure.

In order to ensure that the main strategic objectives encompass an appropriate scope and balance the interests of all the relevant stakeholders, the strategy was developed by the strategy task force: Darja Fišer (representing the Board of Directors), Krister Lindén (representing the National Coordinators' Forum), Bente Maegard (representing CLARIN's thematic committees), and Twan Goosen, Alexander König, Elisa Gorgaini and Karina Berger (Central Hub; representing both technical and coordinating functions). Version 0.1 was discussed at the first CLARIN Strategy Days from 29 to 30 June 2023 in Ljubljana. The meeting was attended by the Board of Directors (BoD), the National Coordinators' Forum (NCF), chairs of CLARIN's thematic committees and representatives of the Central Hub.

Version 0.2 was discussed with the Scientific Advisory Board (SAB), the task force, BoD and committee chairs. Version 1.0 has been submitted to the General Assembly (GA) for approval. A condensed version of the strategy will be published on the CLARIN website.



Introduction

CLARIN's Vision: All digital language resources and tools from all over Europe and beyond are accessible through a single sign-on online environment for the support of researchers in the humanities and social sciences.

CLARIN's Mission: Create and maintain an infrastructure to support the sharing, use and sustainability of language data and tools for research in the humanities and social sciences.

CLARIN's Strategic Orientation: 'Language as Social and Cultural Data: Enhanced Infrastructural Support for Research Using Language Materials in the Era of AI'. Language is a carrier of cultural content and information. Language also plays a role as the reflection of scientific and societal knowledge, as an instrument for human communication and persuasion, as one of the central aspects of the identity of individuals, groups, cultures and nations, as an instrument for human cognition and creative expression, and as a formal system. Moreover, language materials form a considerable part of the historical records which are considered cultural heritage. By giving access to language material, CLARIN facilitates the comparative research perspective for complex phenomena, and enables the development of data-driven analytics and computational modelling of these phenomena. The resources CLARIN makes available also fuel the training and development of large language models, a key technology of Artificial Intelligence (AI), as well as the development of methodological frameworks for the analysis of heterogeneous data.

CLARIN's Key Scholarly Domains: Linguistics, Language Studies, Language Technology, Artificial Intelligence, Literary Studies, History, Journalism and Media Studies, Communication Studies, Ethnography and Anthropology, Migration Studies, Political Studies, Culture Studies, Sociology and Psychology.

This document outlines CLARIN's strategy to fulfil its commitment to **enabling high-quality research and innovation**, as well as the measures aimed at strengthening its accessibility, sustainability and resilience for the period 2024-2026. CLARIN's strategic planning is aligned with the [ERA Policy Agenda 2022-2024](#) and will contribute not only to the way science in Europe is performed, with an emphasis on **collaboration, inclusiveness** and **open access**, but also to goals that reach beyond advancing scientific knowledge, such as those outlined in the [United Nations Sustainable Development Goals](#) and the [Paris Agreement](#). Taking on board the recommendations from the [3rd Report on the Implementation of the ERIC Regulation](#), CLARIN's strategic implementation agenda has been designed to improve open access, support exploitation, foster connectivity and stimulate integration. It also aims to capitalise on CLARIN's potential to **provide solutions to global challenges** with strengthened efforts towards EU sovereignty in critical and emerging technologies, including those relevant to the fair green and digital transition. In order to reach these objectives, CLARIN will provide the foundations for ground-breaking research and innovation addressing global challenges, such as climate change and global health issues, and for helping Europe to collaborate and compete globally, including by exploiting technologies such as Artificial Intelligence (AI).

CLARIN ERIC, as one of the infrastructures that has been established under the umbrella of the European Strategy Forum on Research Infrastructures (ESFRI), was established in 2012 with the vision to provide first-class resources for researchers that critically rely on the availability of language resources, as well as services that give access to relevant digital language data and tools for processing language materials from all over Europe.

CLARIN's mission is rooted in the wide acknowledgement of the role of **language as societal and cultural data**, and the increased potential for comparative research of cultural and societal phenomena across different languages. With its richly faceted nature, range of potential uses and inherent connection to questions of identity and origin, language is an integral part of the humanities and social sciences research, as well as an interesting phenomenon from the perspective of information science, data science, language technology and Artificial Intelligence. This makes CLARIN a crucial pillar for the support of researchers from a very broad spectrum of disciplines, with different needs and skill sets.

In its 11 years of existence, CLARIN has reached a mature operational stage. According to the [ESFRI Landmark Monitoring Review](#), CLARIN's performance is fully satisfactory and meets its objectives, provides valuable services to the academic community on a truly **pan-European, multilingual and multidisciplinary** level, has contributed significantly to the development of natural language processing tools and techniques that have been widely adopted in the research community, and is located in the right intellectual space to ensure that European scholars can make a positive and important contribution to the development of AI. However, further attention is needed to ensure CLARIN's long-term technical, financial, and organisational sustainability, including risk assessment and management. To ensure a continuous, leading role in the landscape, an ambitious plan for change and development that addresses the very dynamic landscape in which CLARIN operates and takes into account other infrastructure initiatives, the GLAM sector (Galleries, Libraries, Archives, and Museums), as well as industry, is essential.

How CLARIN Works

CLARIN is a distributed digital research infrastructure. CLARIN currently has 26 member and observer countries (23+3) and consists of a federation of more than 200 nodes that offer sustainable access to resources, technology, and knowledge in-kind (status on 7 December 2023, when this document was finalised).

The highest decision-taking body in CLARIN ERIC is the **General Assembly (GA)**, which consists of representatives of the ministries or funding agencies of the member countries. It is assisted by an international **Scientific Advisory Board (SAB)**, which conducts annual assessments of CLARIN's progress and advises the GA on strategic issues such as vision, new initiatives, work plans and quality assurance.

Day-to-day management is in the hands of the **Board of Directors (BoD)**, which receives its guidelines from the GA, as well as input from other committees. The BoD is in charge of preparing the multi-annual strategy, annual work plan and corresponding budget, as well as the annual report, which need to be approved by the GA.

Member and observer countries create their own **national consortia** that contribute to the construction and operation of the CLARIN infrastructure, as well as to the outreach to communities of use. By combining the diversified nature of a distributed infrastructure with a cooperative governance model, CLARIN can work towards the set objectives in a truly collaborative manner.

This is primarily achieved through two essential channels. First, national coordinators from each member and observer country come together in the **National Coordinators' Forum (NCF)**. As members of the NCF, national coordinators act as the main liaisons between CLARIN ERIC and the national consortia, coordinate national activities, exchange ideas and experiences, and advise CLARIN ERIC's Board of Directors. The second channel is the CLARIN centre network, which offers sustainable access to resources, services, and knowledge. The activities of the technical centres are coordinated through the **Standing Committee for CLARIN Technical Centres (SCCTC)**, which also has one representative from each member or observer country. The **Knowledge Infrastructure Committee (KIC)** coordinates the activities of the Knowledge Centres.

The BoD can create **Thematic Committees** that help to implement the strategy. Currently, CLARIN ERIC has the **Knowledge Infrastructure Committee, CLARIN Legal and Ethical Issues Committee, CLARIN Committee for Standards and Interoperability, User Involvement Committee, and Centre Assessment Committee.**

The **Central Hub**, which is responsible for financial management, communication, organisation of events and supporting cross-country collaboration, is essential for the realisation of CLARIN's mission and the implementation of the strategic agenda. The employment of central personnel is arranged through an agreement with Utrecht University.

CLARIN's Strategic Orientation

Supporting Scientific Excellence and Innovation

Language is the carrier of cultural content, and is the lens through which social, cultural and scientific dynamics can be studied, both in the present time and historically. By giving access to language material, CLARIN facilitates the comparative research perspective for complex phenomena, such as human communication and cognition, the identity of individuals, groups, cultures and nations, and the expression of ideas and theories. It also provides digital resources for the development of data-driven analytics and computational modelling of these phenomena. The resources CLARIN makes available fuel the training and development of large language models, a key technology of **Artificial Intelligence**, as well as the development of methodological frameworks for the analysis of heterogeneous data.



Promoting data registries and data management services that comply with the **FAIR principles** (Findable, Accessible, Interoperable, Reusable) underpins all aspects of CLARIN's strategy, and the interoperability paradigm of what is now known as the **Open Science agenda** has been one of CLARIN's distinguishing features from the outset. Their importance will only increase in the coming period as the [2023 Lund Declaration on Maximising the Benefits of Research Data](#) calls for reinforcing, accelerating, and maximising the benefits of FAIR and open research data in Europe, within scientific communities and through research infrastructures (RIs), in order to increase the overall research and innovation performance of the ERA and strengthen the outreach to and impact on industry and society. The declaration further points out that access to reusable high-quality research data is crucial for strengthening and advancing knowledge within and across disciplines, as well as enabling comparative research agendas. What is more, it determines how efficiently new challenges and emerging crises, such as the COVID-19 pandemic and the Russian aggression in Ukraine, can be tackled by the research community, further underlining the importance of Open Science, and its ability to support effective and rapid responses to future crises.

In addition, FAIR and open research data can also bring increased societal benefit during non-crisis times. CLARIN is committed to actively promoting open data, open source code, and open standards in order to support **reproducible and replicable research**. With this, CLARIN contributes to **sustainable, cross-disciplinary and responsible (re)use of research data**, thereby facilitating new knowledge discovery and maximising the benefit of investment by academia, industry, and society.

In combination with the inherent **multilinguality** of Europe and the growing attention to **language equality** and **digital inclusion**, it is CLARIN's ambition to consolidate its role in supporting the emerging research agendas for the Social Sciences and Humanities (SSH) domain and to contribute to the innovation potential of advanced models for interaction between people, data and technology for data processing. This is facilitated by the strong embedding of the developers of tools and data collections in their local, culturally specific context, and the interoperability paradigm for the model of collaboration between the centres involved. With this, CLARIN is well positioned both in terms of computing resources and availability of qualified experts to meet the current and future needs of the research and innovation community, including Artificial Intelligence, as called for in the [2023 Tenerife Declaration on the Global Dimension and Sustainability of RIs](#).

Fostering Strategic Partnerships

CLARIN's core community consists of **academic researchers, developers and lecturers** from a range of disciplines who work with language data and who have made crucial contributions to the construction and operation of the CLARIN infrastructure in the form of resources, technology, and knowledge. Over the next years, the collaboration with academic parties will be reinforced and broadened. Among other initiatives, outreach to existing and new user categories will be pursued through the CLARIN Ambassador Programme and the CLARIN Trainer Network.

Even though CLARIN's main focus is to enable curiosity-driven fundamental research and scientific excellence, CLARIN also has significant innovation potential. It cooperates with a variety of stakeholders from outside of academia, including **industry, governmental organisations**, and the **GLAM sector** (Galleries, Libraries, Archives, and Museums) in the role of contributors as well as users of data, tools and know-how. Collaboration with non-academic parties is forged both at the national level and at the level of CLARIN ERIC. Some commercial companies and GLAM organisations have even formalised their collaboration as members of a national CLARIN consortium. Currently, 10 out of the 26 consortia have the national library as a node. Commercial parties are involved in 3 out of the 26 consortia. Given the important role of language data in the cultural heritage domain, CLARIN's collaboration with the GLAM sector will be strengthened through European projects that aim to open up cultural heritage sources with digital methods, such as Advancing Frontier Research In the Arts and Humanities (ATRIUM) and the European Collaborative Cloud for Cultural Heritage (ECCCH). The rising importance of language and speech technology in member states' digital transformation offers CLARIN exciting opportunities for cooperation with industry. Taking into account the recommendations of the [Report on cooperation of ESFRI RIs with industry](#), a Working Group of CLARIN National Coordinators has been formed to prepare an [Industry-Liaison Strategy and Innovation Action Plan for 2023-2024](#) in collaboration with non-academic parties.

In the broader RI ecosystem, CLARIN is well positioned in the [ESFRI cluster of Social and Cultural Innovation](#) and the [ERIC Forum](#). It also has a steering role in the [Science Clusters](#), especially in the **SSHOC cluster** (with particularly strong partnerships with **DARIAH**, **CESSDA** and **EHRI**), and is actively contributing to the [European Open Science Cloud \(EOSC\)](#). This has opened up interesting opportunities for the rearticulation of CLARIN's position in an ecosystem that is shaped through service federation and whose activities are aimed at cross-disciplinary exploitation and the exchange of knowledge and best practices among research organisations.

Forging Impact Pathways

CLARIN's primary goal is to contribute to scholarly excellence and scientific progress by offering access to state-of-the-art resources and tools, as well as to best practices, training and expertise. However, CLARIN's embedding in the very centre of the **data science** and **Artificial Intelligence** communities, whose most influential innovations are entering most academic fields as well as our everyday lives at an unprecedented speed, also helps to position CLARIN as an enabler of research to address societal challenges. These include broad and pressing issues such as climate change and inequality, for instance by using improved processing of language and speech signals in order to develop assistive technologies for people with a speech disorder. But while big data and algorithms based on machine learning are of significant value for research, the economy, and society in general, they also pose a threat for social justice. Overcoming these difficulties requires researchers and industry alike to adopt guidelines for **responsible data science practices** and **ethical AI**. CLARIN aims to take a role in supporting the development of use cases that can help furthering the understanding of the pitfalls of data-driven methods.

CLARIN's potential for societal impact is also reinforced by the growing attention to support measures for language equality, and by use cases and proofs of concept for CLARIN tools in non-academic contexts. In the recently started Horizon Europe (HEU) framework, **multidisciplinary agendas aimed at cultural inclusivity** are expected to play an important role. In addition, aligning with consortia that plan to submit HEU proposals may bring additional potential for impact from support for methodological frameworks driven by language data.

The functionality offered by CLARIN's resource discovery service, the Virtual Language Observatory (VLO), forms a critical component for (the realisation of) the [European Language Data Space](#) (LDS). This concept is envisaged as a vehicle for language data provision, based on a model that allows monetising the available service offer and addressing the needs of the industry sector. The LDS is part of the European strategy for data, aiming at a single market for data in order to ensure Europe's **global competitiveness** and **data sovereignty**.

Due to CLARIN's distributed nature and open-access policy, the collection of quantitative data to track the socio-economic impact (SEI) of resources and CLARIN at large is far from straightforward. While a quantitative causal chain from (access to) resources to an end result (e.g. a publication or policy) is not obvious, in qualitative terms CLARIN's 10-year history provides ample evidence of its socio-economic impact. For example, the rapidly developed machine translation tools between [Ukrainian and Finnish](#) and [Ukrainian and Czech](#) to support Ukrainians demonstrate how a speedy response to a societal need was possible due to the readily available CLARIN resources.

In this strategy period, CLARIN will develop quantitative and qualitative indicators for scientific and socio-economic impact that align with its core value of open access, including the formulation of impact-specific Key Performance Indicators (KPIs)/Key Impact Indicators (KIIs). Following the ESFRI Policy Brief on Impact Assessment of RIs, CLARIN will take a customised, mixed-method approach (including bibliometric analysis, case studies, document analysis and stakeholder survey), taking into account the longitudinal nature of impact assessment.

CLARIN's Strategic Agenda

The implementation agenda of the CLARIN strategy 2024-2026 has been organised into three pillars that are key to CLARIN's operations:

Pillar 1: Our Users

Existing and potential new users with academic and non-academic interests.

Pillar 2: Our Offer

Distributed technical and knowledge infrastructure with interoperable data centres and collaborative knowledge centres across Europe and beyond.

Pillar 3: Our Landscape

Programme for organisational development, coordination of action lines and collaboration with stakeholders and embedding in the broader context.

Each pillar is organised into **focus areas** (*What is important to us?*). Focus areas are broken down into **goals** (*What do we want to achieve?*), which are then further divided into **individual strategies** (*How will we achieve it?*). They will form the basis for a detailed work plan and budget allocation for each of the coming years in order to ensure that the descriptions of the envisaged actions are sufficiently precise and match the available human and financial resources. All of the strategies presented below share three common objectives: increasing the potential for uptake, **increasing the potential for impact and increasing the sustainability** of CLARIN's role and service offer in the RI landscape.

Pillar 1: Our Users

Focus Area 1: Academic User Base

Goal 1.1: Improve outreach to CLARIN's existing academic user base.

- **Strategy 1.1.1:** Continue to identify CLARIN's key user groups and their needs in terms of research data, (large) language models, tools, documentation, expertise and training.
- **Strategy 1.1.2:** Improve the website and other communication channels to better reflect CLARIN's mission, network, offer, as well as impact, and to better cater for the diverse communities served by CLARIN.
- **Strategy 1.1.3:** Develop a plan for partnerships and systematic visibility at national and international scientific and training events that are key for the communities served by CLARIN to showcase the RI services and stimulate engagement and collaboration.
- **Strategy 1.1.4:** Develop a CLARIN Trainers' Network and training materials to serve not only end users, but also trainers and data stewardship communities such as the EOSC Data Stewardship Task Force, RDA interest groups, and national data stewardship programmes.
- **Strategy 1.1.5:** Improve central, national and distributed support services, as well as offer of expertise, via a newly created user forum, helpdesks, training, workshops, and online tutorials, with special attention to multilingual, multidisciplinary and transnational research agendas.

Goal 1.2: Broaden CLARIN's academic user base portfolio.

- **Strategy 1.2.1:** Extend and improve the use of web analytics to better understand how new users find and interact with CLARIN's offer and develop plans to optimise their experience.
- **Strategy 1.2.2:** Survey the academic landscape and identify priority new and emerging user communities for this strategy period and understand their needs via surveys, focus groups, and interviews.
- **Strategy 1.2.3:** Review and adjust the CLARIN Ambassadorship model to ensure the user communities targeted in the previous strategy period are adequately absorbed into the existing user base, as well as to target the identified new priority user communities, including those beyond academia.
- **Strategy 1.2.4:** Review and adjust plans to reach out to new users via CLARIN communication instruments.



Focus Area 2: Non-Academic User Base

Goal 2.1: Develop systematic engagement with non-academic communities that have an interest in CLARIN's outputs and can fuel technical or societal innovations.

- **Strategy 2.1.1:** Identify and extend existing relationships and collaborations with non-academic R&D communities from industry, the GLAM sector, the public sector, NGOs, educators, science journalists, policy makers and the wider audience interested in topics such as language diversity and multilinguality, digital revolution and digital language equality, language technology and Artificial Intelligence via national nodes and centres.
- **Strategy 2.1.2:** Develop a plan for systematic visibility at key academic-industry, public sector and policy events, join academic-industry associations and open up participation at CLARIN events, such as hackathons and the CLARIN Annual Conference, to researchers working in tech companies, GLAM institutions, and NGOs and projects in the public sector.
- **Strategy 2.1.3:** Develop a pilot program to brand CLARIN also as an expert reference authority and advisor supporting European industry, public sector and decision makers on the scientific, sociological and technological aspects of AI/NLP digital language innovation in key industry sectors by mapping socio-economic industry needs and matching them with CLARIN expertise, and by proactively promoting CLARIN ERIC to relevant non-academic stakeholders.

Pillar 2: Our Offer

Focus Area 3: Quality and Interoperability of Data, Tools, Models and Metadata

Goal 3.1: Improve the quality of metadata served by CLARIN and thereby increase the quality of discoverability and reusability of resources and tools.

- **Strategy 3.1.1:** Develop and implement policies and workflows for quality assurance and monitoring for aggregated and curated metadata through a combination of automatic procedures (e.g. link checking) and manual curation (e.g. Metadata Curation Taskforce), and by maximising synergies with existing curation efforts and discovery portals (e.g. VLO, Switchboard, CLARIN Resource Families, SSH Open Marketplace). Integrated monitoring and issue handling (for VLO, VCR, CRF) will be instrumental to making this scalable and sustainable.
- **Strategy 3.1.2:** Analyse current practices with respect to issue handling and develop approaches to a more formalised way of aggregating and routing of issues reported or identified in other ways, including logging and monitoring.
- **Strategy 3.1.3:** Pilot quality tiers for metadata.

Goal 3.2: Improve the coverage and quality of data, models and tools offered through the CLARIN infrastructure.

- **Strategy 3.2.1:** Perform a coverage and gap analysis as well as a technical check on availability, single-sign-on support etc. on resources and tools that will serve as a basis for defining priorities for future development as well as training, outreach and communication initiatives, especially considering possible new application areas (e.g. those that focus on the UN's SDGs or on the EU's HE missions) and fields/topics that are strategically important to CLARIN.
- **Strategy 3.2.2:** Develop procedures for collecting, creating and documenting test suites for tools, models and data sets, as well as procedures for their quality assurance and quality control that are fully aligned with the Open Science agenda and FAIR data principles.
- **Strategy 3.2.3:** Develop a more coordinated approach to attract more deposits of existing and new resources, including sensitive data (e.g. from the health domain, oral history). Improve the existing depositing guidelines and actively approach researchers with interesting and relevant datasets that are not yet available in the CLARIN infrastructure.

- **Strategy 3.2.4:** Position CLARIN as an infrastructure focusing on high-quality language data sets. Share definitions of quality (in terms of format and content), best practices and experiences between the centres with regards to achieving such elevated quality standards. Source data quality success stories and practices from the national consortium activities.
- **Strategy 3.2.5:** Highlight quality and coverage of the CLARIN NLP tools by establishing an interactive 'BLARK meets Test Automation' dashboard that shows the tools that are available in combination with a recent test run on some standard test files, e.g. sourced from the ParlaMint corpus.
- **Strategy 3.2.6:** Improve the existing citation guidelines and practices by actively approaching authors, editors and conference chairs to improve citations of CLARIN datasets and tools in scientific publications.

Goal 3.3: Improve interoperability of metadata, data, models and tools.

- **Strategy 3.3.1:** Integrate the interoperability of metadata at several levels: between CLARIN centres (e.g. via core metadata recommendations, facilitating the use of common vocabularies, and a general CLARIN gateway service for FAIR digital objects), but also with partners operating in the GLAM sector (e.g. ECCCH, EHRI, EURHISFIRM), and other data providers and users that can benefit from CLARIN's processing services (e.g. through support of PubMed identifiers in the Language Resource Switchboard). Explore interoperability with commonly popular repositories like Zenodo, Open Science Framework (OSF), Kaggle, HuggingFace or Github.
- **Strategy 3.3.2:** Explore possibilities to provide a more unified access to NLP tools and models provided by CLARIN centres, e.g. by providing standard programming interfaces. Highlight the added value in comparison with existing, widely used NLP frameworks such as SpaCy.
- **Strategy 3.3.3:** Explore possibilities to provide a more unified access to the key corpora provided by CLARIN centres, e.g. by offering them through a common concordancer that would importantly flatten the learning curve for users and facilitate comparable analyses.



Focus Area 4: Accessibility and Usability of Data, Models, Tools and Services

Goal 4.1: Achieve a high level of accessibility and usability for the central technical services.

- **Strategy 4.1.1:** Actively monitor current and emerging technologies, standards and legal requirements with respect to accessibility and ensure they are supported.
- **Strategy 4.1.2:** Define and implement accessibility and usability requirements and policies for central infrastructure services.

Goal 4.2: Foster a high level of accessibility, usability and interoperability between technical services offered by national nodes and centres.

- **Strategy 4.2.1:** Raise awareness by communicating the findings from Goal 4.1 to the national nodes and centres.
- **Strategy 4.2.2:** Enable easy adoption of the policies, guidelines and designs developed for national nodes and centres.
- **Strategy 4.2.3:** Facilitate the use of processing services for sensitive data through federated national centres.

Focus Area 5: Collaboration and Knowledge Exchange

Goal 5.1: Consolidate the CLARIN community for sharing knowledge and experience.

- **Strategy 5.1.1:** Identify systematic opportunities and develop practices to improve interaction and minimise fragmentation of work between all CLARIN centres and committees.
- **Strategy 5.1.2:** Review and adjust the CLARIN Annual Conference model so that it remains a rich and inspiring source of knowledge exchange for the community.
- **Strategy 5.1.3:** Improve and disseminate the collection of Impact Stories, Tour de CLARIN contributions, CLARIN Cafés, the CLARIN best practice papers and the CLARIN Zotero Library collections to better reflect the knowledge and experience of the CLARIN network.
- **Strategy 5.1.4:** Develop best practices and templates for Data Management Plans (DMP) for researchers collecting language data as a new means of knowledge sharing.

Goal 5.2: Integrate K-centres more into the ‘bigger picture’.

- **Strategy 5.2.1:** Extend impact, data curation, DMP templates, CLARIN Resource Families, and training initiatives with experts from K-centres.
- **Strategy 5.2.2:** Promote increased use of K-centres’ expertise by researchers via a variety of means, such as Tour de CLARIN and Impact Stories.
- **Strategy 5.2.3:** Monitor K-centres, including certification, annual reporting, certification, mid-term meetings, looking at possible new K-centres, etc.

Goal 5.3: Enhance collaboration across ‘borders’, i.e. across countries, across types of entities, across fields of expertise.

- **Strategy 5.3.1:** Identify systematic opportunities and develop practices to better support collaboration ‘across borders’.
- **Strategy 5.3.2:** Continue to contribute to cross-national and cross-disciplinary events, such as the Digital Humanities Helsinki Hackathon.

Pillar 3: Our Landscape

Focus Area 6: Sustainability and Impact

Goal 6.1: Extend and reinforce membership.

- **Strategy 6.1.1:** Develop a growth strategy to attract new members, with the final goal of including all EU and associated countries as well as interested Third Countries in and outside of Europe.
- **Strategy 6.1.2:** Develop plans to illustrate the added value of membership and agreements to existing members.

Goal 6.2: Monitor sustainability and impact.

- **Strategy 6.2.1:** Develop mechanisms to maximise synergies of strategies, action plans and projects at national and EU level.
- **Strategy 6.2.2:** As a follow-up to the ESFRI Landmark Monitoring exercise, develop a methodology for socio-economic impact assessment. In this context, the KPI framework and implementation will be revisited and adjusted as needed.
- **Strategy 6.2.3:** As a follow-up to the ESFRI Landmark Monitoring exercise, develop an annual risk assessment and management plan taking into account a variety of scenarios on financial, technological and personnel levels.
- **Strategy 6.2.4:** Identify and address current and foreseen scalability issues and risks as well as a continuous evaluation loop of infrastructure components of the technical infrastructure, such as AAI (e.g. growing number of SPs which need to be supported, impact on architecture) and metadata curation (e.g. support, issue handling, link checking capacity).

Goal 6.3: Strengthen governance.

- **Strategy 6.3.1:** Review and adjust the role of Chief Operations Officer vis-à-vis the BoD and SAMBA vis-à-vis NCF, as well as the composition of these committees.
- **Strategy 6.3.2:** Review and adjust committee membership cycle as well as the annual planning, monitoring and reporting of their work plan.
- **Strategy 6.3.3:** Review and adjust the format (in-person, hybrid, virtual) and cycle of internal meetings (GA, SAB, BoD, NCF, SAMBA, committees, K-centre representatives), stimulate joint meetings (annual CLARIN Strategy Summit) and encourage environmentally friendly travel options in order to contribute to the European 'green agenda'.
- **Strategy 6.3.4:** Review and adjust the interaction with GA and SAB in order to maximise two-way information exchange using optimal communication instruments.

Goal 6.4: Reinforce CLARIN as a key enabler of language-based state-of-the-art research and innovation.

- **Strategy 6.4.1:** Identify opportunities and improve CLARIN's offer with regard to emerging technology, such as Large Language Models (LLMs) and AI.
- **Strategy 6.4.2:** Contribute to Europe's 'green agenda' in view of the energy consumption of LLMs by identifying ways to avoid the generation of (similar) models twice or more.

Focus Area 7: Central Hub

Goal 7.1: Review the Central Hub's support role.

- **Strategy 7.1.1:** Analyse what helps national consortia and centres the most in building up (guidance vs. independence). Investigate and evaluate an alternative role of the central technical team towards centres. Some possibilities include: regular open hours, improved and consolidated documentation on the CLARIN website (e.g. consolidated FAQs, guides for typical tasks/problems), in-person guidance and prefab packages for new centres. Keep in mind lessons learned from previous approaches (e.g. CLARIN DSpace).
- **Strategy 7.1.2:** Evaluate and adjust existing support for CLARIN's committees and encourage active contribution of Central Hub members to committee work.

Goal 7.2: Further professional capacity of the Central Hub.

- **Strategy 7.2.1:** Include impact development and monitoring, events coordination and quality assurance in Central Hub capacity.
- **Strategy 7.2.2:** Facilitate mobility, training, knowledge sharing and team development for Central Hub staff.

Focus Area 8: National Nodes

Goal 8.1: Increase focus on national contributions to the CLARIN infrastructure.

- **Strategy 8.1.1:** Strengthen contributions from national nodes and centres to CLARIN's efforts to increase the quality, coverage and interoperability of data, tools and metadata, i.a. assisting CLARIN's users in finding, accessing, reusing, developing, and depositing LLMs for specific research problems.
- **Strategy 8.1.2:** Improve contributions from national nodes and centres to CLARIN's outreach and training initiatives, such as events calendar, Trainers' Network, Tour de CLARIN and Impact Stories.

Goal 8.2: Share, collaborate and align national agendas.

- **Strategy 8.2.1:** Reinforce the sharing and alignment of existing and future national agendas at the level of the National Coordinators' Forum (NCF).
- **Strategy 8.2.2:** Analyse and reinforce the exchange of experience and cross-country collaboration, e.g. forming clusters of national nodes with similar development goals or similar expertise using the K-centre framework.
- **Strategy 8.2.3:** Promote expertise within committees and encourage cross-national exchange and collaboration.

Goal 8.3: Encourage good communication practices between the Central Hub and national nodes to increase consistency across the network and improve branding and visibility.

- **Strategy 8.3.1:** Contribute to CLARIN's dissemination initiatives, such as Newsflash, Tour de CLARIN and Impact Stories.
- **Strategy 8.3.2:** Promote the adoption of common citation practices and templates that allow for the citation of language resources by means of a Persistent Identifier.
- **Strategy 8.3.3:** Promote the adoption of a common Style Guide/Information Sheet (including guides on content and design for a variety of outputs (e.g. online content, internal documents, templates)).
- **Strategy 8.3.4:** Develop benefits and incentives for CLARIN nodes and centres to present themselves as part of CLARIN (a larger, stable, long-term body) when showing outputs or achievements (in addition to their individual institutions).

Focus Area 9: Collaboration beyond CLARIN

Goal 9.1: Reinforce alignment and collaboration with other European infrastructures.

- **Strategy 9.1.1:** Continue close collaboration with fellow RIs in SSH Open Cluster and EU Science Clusters at management level to ensure optimal visibility, representation and participation with reference to the EC, EOSC and ESFRI, as well as in relevant projects and working bodies.
- **Strategy 9.1.2:** Reinforce CLARIN's position as a building block/intermediary node in the EOSC ecosystem. Among other benefits, this could importantly expand CLARIN's user base and widen its disciplinary engagement.
- **Strategy 9.1.3:** Analyse collaboration models across RIs and develop a plan to adopt the identified best practices.
- **Strategy 9.1.4:** Promote collaboration with fellow RIs at operational level, fostering regular interactions (on knowledge/best practices exchange) between CLARIN and other relevant infrastructures.
- **Strategy 9.1.5:** Monitor the development of the Alliance for Language Technologies (ALT-EDIC) and prepare relevant action plans.

Goal 9.2: Strengthen collaboration at global level.

- **Strategy 9.2.1:** Explore opportunities for collaboration with the Linguistic Data Consortium (LDC) and similar sister organisations outside of Europe.
- **Strategy 9.2.2:** Explore collaboration models with relevant partners or networks outside of Europe, such as (but not limited to) Canada, Australia, Japan, Brazil and Mexico.
- **Strategy 9.2.3:** Explore opportunities for collaboration with non-academic partners outside of Europe.



CLARIN ERIC
c/o Utrecht University
Drift 10
3512 BS Utrecht
The Netherlands

CE-2023-2333
November 2023

