Title	Progress Report 2022
Version	1.0
Author(s)	BoD, CLARIN Office
Date	24 November 2022
Status	Approved
Distribution	SAB, NCF, BoD, GA
ID	CE-2022-2080

Resolution
The General Assembly approves the CLARIN ERIC Progress Report 2022 (CE-2022-2080).

Table of Contents

1   Introduction

This document presents a report on the progress achieved within CLARIN ERIC by October 2022. It builds on the CLARIN ERIC Strategy 2021-2023 (CE-2020-1709)[1] and, more specifically, on the Work Plan for 2022 (CE-2021-1909), which was approved by the General Assembly (GA) in November 2021. For each of the four strategy areas and for each of the goals identified in these documents, this report describes the progress made since the autumn of 2021. The structure adopted in the above-mentioned documents will serve as a point of reference, that is, this report focuses on activities aimed at the construction, operation, and uptake of the research infrastructure in the Social Sciences and Humanities (SSH) and beyond, for each of the following four priority areas:

- Knowledge Infrastructure
- Technical Infrastructure
- Sustainability
- Organisational development.

1.1   The Impact of COVID-19: Alternative Activities and Instruments

The COVID-19 pandemic continued to considerably affect the activities of the central office staff, even though face-to-face meetings gradually started again in the second half of 2021. Overall, working from home remained the predominant modus operandi, with all the implications of virtual meetings and online models of collaboration. Similar as for many other organisations in the field, this brought with it some challenges, but also the opportunity for new initiatives.
A dedicated web page (https://www.clarin.eu/covid-19) presents an overview of the initiatives related to COVID-19 that were supported by CLARIN. As in 2021, the main aim of these initiatives was to offer language resources and natural language processing tools that can be of relevance for the study of phenomena related to COVID-19.

Similar to 2021, in 2022 the planning of most activities that require face-to-face meetings was suspended and the available budget for activities involving travel could not be put to use as originally envisaged. Part of the available resources were re-allocated in order to facilitate and/or boost the strategic action lines:
- Project ParlaMint

- Populating the CLARIN web section devoted to impact stories with new items
- Increasing the volume of funding for the CLARIN Resource Families initiative

---

[1] The CLARIN Strategy 2021-2023 is available in online version as well. See also the two-page summary of the strategy.

- Building up expertise in organising and facilitating hybrid events
- Consolidation of the development of materials for '*Training with CLARIN*' through:
  - A new call for submissions for a dedicated session during the CLARIN conference
  - The development of a series of training materials aimed at presenting the core CLARIN services and encouraging their use in the classroom, carried out within the framework of the UPSKILLS project
  - A new collaboration with the University of Bologna (Department of Translation Studies) for setting up student internships at Master level, aimed at translating and adapting CLARIN training materials
- Continuation of the collaboration with a senior advisor aimed at the further development of a strategy for innovation and collaboration with non-academic partners.

This report includes more detailed accounts of the additional or adjusted activities as part of the reporting for the relevant priority areas.

ParlaMint

ParlaMint is a project aimed at compiling a collection of comparable parliamentary datasets with curated corpora of recent parliamentary debates in multiple languages about the COVID-19 pandemic. It also includes corpora from before the pandemic that serve as the basis for comparison. ParlaMint I started in Q3 of 2020. After the first successful first data release (available through the CLARIN.SI repository), it was decided in 2021 to also fund a second stage of the project and to promote the cross-country collaboration on a multilingual dataset as CLARIN's flagship project. For follow-up project ParlaMint II, funding has been made available that will allow a continuation of the activities from Q4 2021 until Q1 2023. The use of the dataset has been stimulated in several ways. For instance, the use of the ParlaMint data was reported in several papers submitted for the ParlaCLARIN III workshop held at LREC2022. Moreover, in the Horizon2020 project EOSC Future, the ParlaMint data release is one of the data pillars for a study on how the metadata for life science data on COVID-19 and parliamentary data can be aligned through harmonised labeling.

1.2   Role of the KPI Framework

The progress achieved in the past 12 months is reported in terms of a combination of quantitative and qualitative categories. For the quantitative progress measures, the framework for Key Performance Indicators (KPIs), which was adopted in 2018 and formally endorsed by the GA, has been used as a basis (CE-2018-1266). The purpose of this framework is to help the CLARIN community describe the progress achieved in developing and operating the CLARIN research infrastructure. The framework is likely to undergo adjustment as a consequence of the launch of the monitoring framework for landmark RIs that is planned for 2022-2023. The landmark monitoring will be based on the ESFRI report

published in 2019. From 2019 onwards, for a subset of the KPIs suggested in the ESFRI report CLARIN has been collecting performance figures. The KPIs presented in this report will again be based on this subset. In addition some notes and figures are provided explaining the style of reporting the publication output, which is based on a mixture of methods.

1.3   Overview of Sections

In Section 2, for each of the four strategic priority areas, the progress with regard to the following elements will be described:
- The plans as included in the Strategy 2021-2023 ([CE-2021-1709](#)), labelled below as *Where We Stand* and *Where We Want to Go*, which by default have 2020 as their point of reference (these passages are highlighted light green)
- The actions planned in the Work Plan for 2022 ([CE-2021-1909](#)) (these passages are highlighted light blue)
- The achievements in relation to the Work Plan for 2022 ([CE-2021-1909](#)), extended with more detailed reports on the actions undertaken in response to the COVID-19 pandemic (the subsections describing the progress are labelled below as *Achievements in 2022*).

In Section 3, three additional topics will be addressed. Section 3.1 reports on the implementation of the agreed model for governance, management and organisation. In addition, an overview of some coordination and organisational activities will be presented. Section 3.2 identifies the strategic priorities that have been identified as the ones that require continued attention in the work plan for 2023. Section 3.3 provides some details on salient publications and web statistics.

Throughout the sections, the report will highlight the role of workshops and virtual events organised since Q4 2021. An interesting side effect of organising virtual events is that they are typically accessible to a wider audience. The various CLARIN events were mostly attended by participants from multiple continents. The availability and maturity of the demonstrated expertise that the CLARIN staff has developed in 2020 and 2021 in setting up events for the community and in offering nodes in the network a chain of support tools for virtual events is widely appreciated. Indeed, the offer has been further extended, and now includes the capacity and expertise to support an even more challenging format: hybrid events. Section 4 presents some concluding remarks.

2　CLARIN's Priority Action Lines Aimed at Enhancing Uptake

The action lines listed in the Work Plan for 2022 (CE-2021-1909) were devised to contribute to the longer-term objectives described in the CLARIN ERIC Strategy 2021-2023. The overall aim of the action lines is to enhance the uptake of the CLARIN Infrastructure. Following the structure of the CLARIN ERIC Strategy, these action lines are categorised into four priority areas:

- Knowledge Infrastructure
- Technical Infrastructure
- Sustainability
- Organisational development.

The strategy is set up in alignment with the emerging Open Science policies and promotes adherence to the FAIR data principles: open and easy access to data, paradigms for standardised data citation, support for data re-use and data repurposing, as well as to the related paradigm of responsible and explainable data science, that are specifically relevant for the uptake the CLARIN services. In parallel, the aim is to initiate and support the application of novel, data-driven methods for the SSH domain that has become feasible by the advance of the digital humanities and computational social sciences and by treating language resources as a rich source for fields such as data science, language technology, and artificial intelligence (AI).

As the overarching priority, CLARIN ERIC's strategy is strongly focused on increasing the uptake of the CLARIN services available, taking into account the different types of users of the CLARIN infrastructure: providers (who produce resources and tools and want to store and disseminate their results), SSH scholars (who use CLARIN to do better, faster and more innovative research in one or more stages in the research workflow), and educators (who teach new generations of scholars and professionals with a thorough understanding of the added value of digital scholarship and digital technologies). Understanding the needs of these user categories in general, and the specific needs of the different research communities, is seen as a precondition for meaningful demonstrations of the vital importance of language data for discovering new ways for machines to interact with humans, and for humans to interact with machines. And more generally, insight in user needs is seen as the foundation for generating lasting impact and for 'making things happen' based on language resources.

2.1 Knowledge Infrastructure

Scientific breakthroughs are rarely the work of a single person. They often involve many different partners from various sectors, disciplines and countries. Within Europe's vast innovation landscape, research infrastructures act as brokers of resources, technology, and knowledge. They provide a platform for scientists, industry, governments, and citizens to exchange information and foster collaborations among these stakeholders.

CLARIN's primary mission is to support and accelerate research excellence. The CLARIN Knowledge Infrastructure (KI) has been developed as the 'glue' for the various communities engaged with CLARIN, and as the structure that secures a continuous transfer of knowledge between the diverse parties involved in the construction, operation and promotion of the infrastructure in all relevant communities of use. The deployment and further development of this knowledge infrastructure is a crucial pillar for optimising the exchange of expertise and the creation of a rich offer of online materials that can be used for instruction, explanation and teaching regarding the use of CLARIN.

The implementation of the KI strategy requires close coordination with activities aimed at maintaining and enhancing the Technical Infrastructure (see more details in section 2.2) and with national consortia, which play a key role in providing use cases, training materials, and capacity for knowledge centres (K-centres).

The implementation of the Knowledge Infrastructure strategy is partly supported by dedicated committees. Their mission and annual plans are coordinated with the Board of Directors (BoD):
- Knowledge Infrastructure Committee (KIC)
- User Involvement Committee (UIC)
- CLARIN Legal Issues and Ethics Committee (CLIC)
- CLARIN Standards Committee (CSC).

For some of the larger initiatives, ad hoc Task Forces and/or Steering Committees have been appointed, such as the CLARIN Resource Families Task Force.

Where We Stand

1. A network of CLARIN K-centres that bring together expertise, e.g. for a specific language, topical domain, data modality, etc.
2. Tour de CLARIN: This initiative gives an impression of the richness of the network and the ongoing activities in different countries. The series features national consortia and centres, and highlights their prominent tools, resources and, most importantly, the rich community of researchers developing and using the infrastructure.
3. Funding schemes for:
   a. Workshops, webinars and training sessions
   b. Events aimed at stimulating the uptake of CLARIN in specific disciplines and regions
   c. Mobility grants to enable the exchange of expertise among individual researchers, educators and technical experts.
4. A programme for reaching out to new communities, supported via:
   a. The Ambassadors Programme
   b. A series of CLARIN Café events, introducing CLARIN's potential for diverse audiences
   c. Information platforms about training opportunities, such as the Digital Humanities Course Registry.

Where We Want to Go

1. A website that is better suited to cater for the diverse community served by CLARIN and that offers a more contemporary user experience.
   a. More diversified communication about the Technical Infrastructure and its use, aimed at both non-technical and more technically sophisticated audiences
   b. Better visibility of the training and educational initiatives and the services offered by K-centres
   c. Contextualisation of the CLARIN Resource Families by providing links to publications, guides, and tutorials
2. A wider range of topics covered by the K-centres
3. Strengthening of the CLARIN networks:
   a. Incentives for closer interaction among the K-centres
   b. A sustainable network of CLARIN trainers to better enable the education of end-users of the infrastructure
4. Agile support:
   a. Adjustment of the mobility grants model and other funding instruments to better support cross-country technical development, research, teaching, and documentation
   b. Central expertise and support for the CLARIN consortia to address the increased need for virtual events and meetings.

Actions Planned in Work Plan 2022

The following specific activities were planned for the strategic priority area Knowledge Infrastructure (see CE-2021-1909, section 3.1):

- To be able to increase the effectiveness of the online communication instruments for the various target audiences, some dedicated monitoring activities will be set up:
  - The user experience of the new website design will be periodically evaluated.
  - The interest in new information formats such as podcasts will be investigated.
- Organisation of the 2022 CLARIN Annual Conference
- Interaction among the K-centres to be stimulated by events for and/or exchange between K-centres
- In the context of the project UPSKILLS:
  - New learning content related to data acquisition skills (text and speech processing) and data handling skills (data standards and repositories) will be created.
  - The best practice guidelines for integrating research and research infrastructures into teaching and learning activities will be finalised by September 2022.
- Other activities foreseen to support education and training under the campaign labelled 'Teaching with CLARIN':
  - Publish a set of guidelines for lecturers to help them make their training materials FAIR
  - Further development of the training and education section of the CLARIN website
  - Continue the development of the CLARIN Training Suite and the Trainers' network
  - Optimise the conditions for using CLARIN resources in the context of academic 'teaching abroad' tracks
- Further extension of Tour de CLARIN and the Impact Stories portfolio
- The CLARIN Ambassadorship Programme will be extended to more disciplines
- Enhancement of the Zotero platform for sharing publication output describing the design or use of CLARIN
- Preparing a book about CLARIN in view of the celebration of the 10th anniversary in 2022, to be published in open access

- Specific communication plans:
  - Clarifying under what conditions and to whom services and resources are available (with reference to the FAIR Data Principles), and ensuring that this information is visible and clear to users
  - Providing information about best practices of use and collaboration opportunities
  - Inclusion of a section that highlights CLARIN's value proposition for non-academic parties, such as the public sector and industry
  - Providing online access to teaching and training materials, such as presentations, videos, case studies, posters, slides, and templates for presentations
- Accelerating the integration of key strategic datasets, in order to attract new communities of use.

General KI instruments (ongoing action lines)
- The network of CLARIN K-centres that bring together expertise in a certain domain, topic, data modality, etc.
- Mobility grants (travel support for meetings between technical experts, researchers and educators, or writing retreats for researchers to prepare publications using the CLARIN infrastructure or for teachers to develop training materials based on the CLARIN infrastructure)
- Training and education modules promoted via dedicated channels
- The online CLARIN-DARIAH DH Course Registry
- Dissemination of what CLARIN has to offer via blog posts and social media.

Online Events planned for 2022
Since face-to-face events were expected to be hindered by the pandemic-related travel limitations, the following virtual outreach events were planned:
- CLARIN Café (ca. 6 per year)
- Webinars (ca. 1-2 per year)
- Online tutorials (ca. 1-2 per year).

For the activities of the UIC and KIC, committee-specific plans for 2022 have been prepared.

Achievements in 2022

In the last year, the CLARIN Knowledge Infrastructure has grown and a stable set of activities are now on the agenda of both the central and the local nodes. The pandemic, and the concomitant limits to mobility, have not proven to be an impediment to further development; indeed, there has been an ever-increasing number of collaborations for the exchange and production of shared knowledge within CLARIN. In what follows the main achievements are presented, referring to more detailed reports when needed.

CLARIN Website

The CLARIN website is the first point of access to information about CLARIN and the gateway to its expertise and wealth of resources. In 2021, in order to address the growing demands of CLARIN's increasingly diverse community and audience, the website was relaunched featuring a new style and improved functionality. In the course of 2022, the content has been adjusted to the new structure on a large number of pages. A new content item that will be published soon is a promotional video clip that has been produced with the aim of introducing CLARIN's value proposition to new audiences. In the near future other clips zooming in on specific aspects of CLARIN or on the activities in the national consortia will be produced, using the same design style.

As an example of the new style and improved functionality, the section on 'Language Resources' now offers a quick access route to datasets, tools and services, and also contains hand-picked illustrative examples for the data and tool entries.

A new section called 'Learn and Exchange' was added, which includes all content dedicated to knowledge exchange and teaching, such as training materials, videos, case studies, posters, slides and presentations. Also located in 'Learn and Exchange' is a new initiative featuring a series of 'Impact Stories', which was launched in September 2021. The series aims to showcase high-quality and innovative research that benefits from CLARIN tools and resources. The impact stories highlight the excellent research enabled by the CLARIN infrastructure and demonstrate the wider impact that CLARIN and the social sciences and humanities have on broader societal issues. The stories can be found here.

Impact Stories

In this series we showcase high-quality and innovative research that uses CLARIN tools and resources. These impact stories illustrate the huge variety of disciplines that use the CLARIN infrastructure, highlight the excellent research linked to it, and demonstrate the wider impact that CLARIN and the social sciences and humanities have on broader societal issues.

Read more →

Fig 1. The entrance page for the section featuring impact stories.

The funding and support section has also been reviewed and improved, and now features a one-stop Funding Hub, which provides information on and access to all funding calls and event support options.

The work on the website revision is still ongoing, with a review of its structure and content currently underway. New navigation pages are being piloted with the aim of making the website more user-friendly. Improved and/or additional sections for teaching materials and industry collaboration case studies are in their planning stages.

Knowledge Centres

The network of CLARIN K-centres is a crucial instrument for the exchange of knowledge and expertise on the various aspects of the domain covered by the CLARIN infrastructure. The number of K-centres has grown to 26 and several applications and re-certifications are underway. The centres are listed on the K-centre section of the CLARIN website that describes the role and topical coverage of the K-centres.

A first workshop for all K-centre managers was held in December 2020. The purpose of the workshop was to make the K-centres better aware of each other's profile and activities in order to facilitate collaboration. In 2021-2022, the organisation of a follow-up face-to-face workshop was hampered by the travel limitations. Instead, the focus shifted to producing a collection of best practices and a common information base (CE-2022-2037). Initiatives to increase the K-centres' visibility include coverage of K-centres as part of the Tour de CLARIN series (see section below), the organisation of dedicated CLARIN Cafés, and the publication

of a best practices papers collection as part of the Zotero library, with links to K-centres and their expertise.

KPI – Number of Certified K-centres

September 2018: 10 certified CLARIN K-centres
September 2019: 19 certified CLARIN K-centres
September 2020: 23 certified CLARIN K-centres
September 2021: 25 certified CLARIN K-centres
September 2022: **25 certified CLARIN K-centres**

Training and Education

Training-related information and materials are of utmost importance to ensure the uptake of CLARIN. In 2022, the resources for training and education, and the action lines facilitated by them received a boost through various initiatives and projects. A summary is presented below.

The online Digital Humanities Course Registry (DHCR), a joint initiative of CLARIN ERIC and DARIAH-EU, provides information about courses and training activities in the Digital Humanities from Europe and beyond. The DHCR database contains information about 505 courses from 166 institutions, of which 122 courses are currently displayed in the public view and actively maintained by the national moderators. The DHCR is promoted among lecturers and students, by means of presentations at scientific conferences, and social media channels, such as Instagram and YouTube. Between September 2021 and September 2022, the platform had 9863 unique page views from all over the world. (In the previous year there were 8179 page views.)

Teaching with CLARIN is the umbrella for the activities aimed at the increased awareness of the added value of the CLARIN resources for academic teaching. Through this initiative 12 examples have been collected of online tutorials and teaching materials that use CLARIN language resources and tools to support skills development in the areas of linguistics, language technologies, digital humanities, data science, research infrastructures and research data management. The metadata of each collection have been added to the SSH Training Discovery Toolkit, that provides an inventory of training materials relevant for the social sciences and humanities.

Based on a recent analysis, a set of recommendations will be produced and a use case will be included in an upcoming publication, *Towards FAIRification of training resources and catalogues – lessons learnt from research communities,* to be submitted to the journal Scientific Data. The publication is a joint effort of training coordinators that are affiliated to a Community of Practice (CoP) task force: OpenAire, CNR Italy, DANS, SSHOC, ELIXIR, EIFL, JISC/EOSC Synergy, EMBL, and CLARIN.

Events specifically addressing topics related to training and education

- On 3 March 2022, CLARIN and DARIAH jointly organised an online industry outreach workshop as part of a workshop series initiated by H2020 project ENRIITC. The aim of the event was to identify common challenges regarding the skills needed to reach out beyond academia, as well as a set of common SSH training topics that are relevant for non-academic entities when considering the research infrastructure service offer. The event was attended by 32 representatives from SSH research infrastructures and related networks. The main conclusion was that the training strategy proposed ENRIITC for the training of Industry Contact Officers and outreach towards the industry generally suits the needs of the SSH, but that more bespoke training is desirable.
- During the SSHOC final conference (6-7 April 2022), CLARIN participated in the SSHOC'n Tell Challenge to demonstrated how the CLARIN Virtual Collection Registry and Switchboard can be integrated not only into research but also into teaching to help students enhance their research data discovery, processing skills and citation practices.
- On 13 July 2022, CLARIN organised a workshop at TaLC 2022 with three tutorials: (1) UPSKILLS tutorial, demonstrating how the external language datasets can be processed via Switchboard and then annotated with WebLicht; (2) ParlaMint tutorial, showing how key corpus linguistic techniques can be applied to parliamentary discourse to investigate women's representation in the parliament; and (3) a CLiC tutorial that gave the participants a hands-on experience with the basic concordance, tagger and annotation tools suited to address research questions around properties of narrative fiction.

In the context of the UPSKILLS project a survey among 93 lecturers in language and linguistics programmes revealed that more than half of them used research data repositories, catalogues and mailing lists to locate existing language resources suitable for teaching, including CLARIN national repositories, the CLARIN research infrastructure, the Resource Families and the CLARIN VLO. The respondents also highlighted a number of practical challenges when trying to access and use existing language resources. The survey results will inform the further development of the services and material offered to support training and education.

As of October 2022, the learning content and guidelines developed by the UPSKILLS consortium partners will be available for piloting in language-related programmes. In UPSKILLS CLARIN has developed teaching materials on the topic of *Introduction to Language Data: Standards and Repositories* that can be integrated into any course or programme dealing with language research and teaching.

The department of interpreting and translation of the University of Bologna approached CLARIN with a request for *student internships*. As a result, we have designed an internship programme that will be coordinated by the Training and Education Officer together with an experienced researcher or teacher from a CLARIN member country. The main objective of the first round of internships will be to test existing training materials on core CLARIN services and to localise them to the interns' preferred target languages.

CLARIN Annual Conference 2022

The tenth CLARIN Annual Conference will be held in Prague and can be attended in-person as well as virtually. The programme will be mostly hybrid, and is thus an opportunity for the further development of the capacity and skills needed to support the increasingly wide range of format variants that are likely to become commonplace in the field in the years to come. All CLARIN committees will be able to meet again in person (with the hybrid option) in the morning before the start of the conference programme.

The call for abstracts was launched in January 2022 and followed the successful format of previous editions. The NCF acted as the Programme Committee (chaired by Prof. Tomaž Erjavec, Slovenia). A subcommittee of the PC was formed to prepare the background information for all major decisions regarding the conference programme, the format and all other content matters. This subcommittee, chaired by the PC chair, prepared a set of proposals for the selection of the keynotes, acceptance/rejection of the paper submissions, and the event format, taking into account the lessons learnt from the experience with the virtual format of CLARIN2021.

Overall, the call for abstracts attracted 21 submissions, of which 16 were accepted. The accepted papers have been grouped into five categories: Language resources and CLARIN centres, tools and workflows, legal questions, curation of language resources, and research cases. In order to maximise the potential for interaction, all authors of selected papers have been allocated a plenary presentation slot of 15 minutes in one of the topical sessions.

The three keynotes will cover the following three topics: Collaboration with libraries, human labelling for AI, and speech processing in the industry context. A special slot has been reserved for the outcome of the H2020 project SSHOC (2019-2022) and the envisaged consolidation of the collaboration between the RIs from the SSH domain under the name of *SSH Open Cluster* (see also sections 2.3 and 4). The tradition of the CLARIN Bazaar session has been maintained. In addition, two panel sessions have been set up. The first focuses on the collaboration between CLARIN and libraries, while the second informs the audience about the complementarity of the range of discovery platforms that have recently become available for the wider SSH field, as well as the non-academic communities that are served. The popular Student Session has attracted 13 PhD student stories, which will be presented as both short oral presentations during the plenary session and as posters at the Bazaar. Finally, a special session entitled *Teaching with CLARIN* has been organised for academic teachers, which will showcase the efforts made by teachers, lecturers and trainers in the CLARIN network to meet the needs for training and education.

User Involvement Initiatives

As in previous years, CLARIN User Involvement (UI) activities are organised at two levels: the central level and the local level. Some central UI activities are carried out within the framework of EU-funded projects. Local activities are carried out by national consortia and centres, but coordination of the cross-country efforts is provided by the User Involvement Committee. In 2022, with the easing of the COVID-19 situation, face-to-face events have become possible again. However, various forms of virtual outreach and online events stayed on the agenda and are likely to remain an important element of the CLARIN outreach strategy.

In the period between October 2021 and September 2022, 15 workshops and training events were organised by the CLARIN Office (including some organised within the framework of funded projects): 9 virtual events and 6 face-to-face events. A complete list of CLARIN-related events can be found on the events page. Among the centrally organised events, the below are particularly noteworthy:

- The continuation of the CLARIN Café series, with five cafés organised in collaboration with various communities. This type of short event, with short thematic presentations followed by discussion, has proven to be a successful model. The topics were: legal issues, the integration of CLARIN in teaching, new important initiatives at the national level, corpora and tools for their exploration. CLARIN Cafés continue to meet the interest of the public. The number of registered participants varied from 50 to over 100 for each café
- Other noteworthy events organised by CLARIN are:
  - 'ENRIITC Your Industry Outreach: Workshop for Social Sciences and Humanities' (online); co-organised with DARIAH ERIC, 11 March 2022 (link)

- CLARIN and Libraries workshop; in-person event, 9-10 May 2022, The Hague (link)
- Two UPSKILLS multiplier events (hybrid and online):
    - 'Linguists in Tech - a Swiss Perspective', 8 June 2022, Lugano (link)
    - 'Fit for the Job Market? Teaching Transferable Skills to Future Language Experts', 5 July 2022, Graz (link)
- ParlaCLARIN workshop@LREC 2022, June 2022, Marseille (link)
- Two training events:
    - 'Full-Text Resource Processing' (online, in the context of the Europeana DSI-4 project)
    - 'Teaching with CLARIN', co-located with TALC 2022 in Limerick
- CLARIN also supported two major DH events for young researchers: the Helsinki Digital Humanities Hackathon 2022 and the European Summer University in Digital Humanities 2022.

Attendance Figures for Centrally Organised Events

In the period from September 2021 to September 2022, a total of 15 events were centrally organised. Three of these were face-2-face events with two of them accommodating also remote participation. The face-2-face events were attended by a total number of 86 unique participants (based on names indicated during registration) from 25 countries. While the majority of participants came from Europe, the events also attracted participants from further afield. See Figure 1 below for an indication of the spread of countries. Note that some participants may have participated in several events.
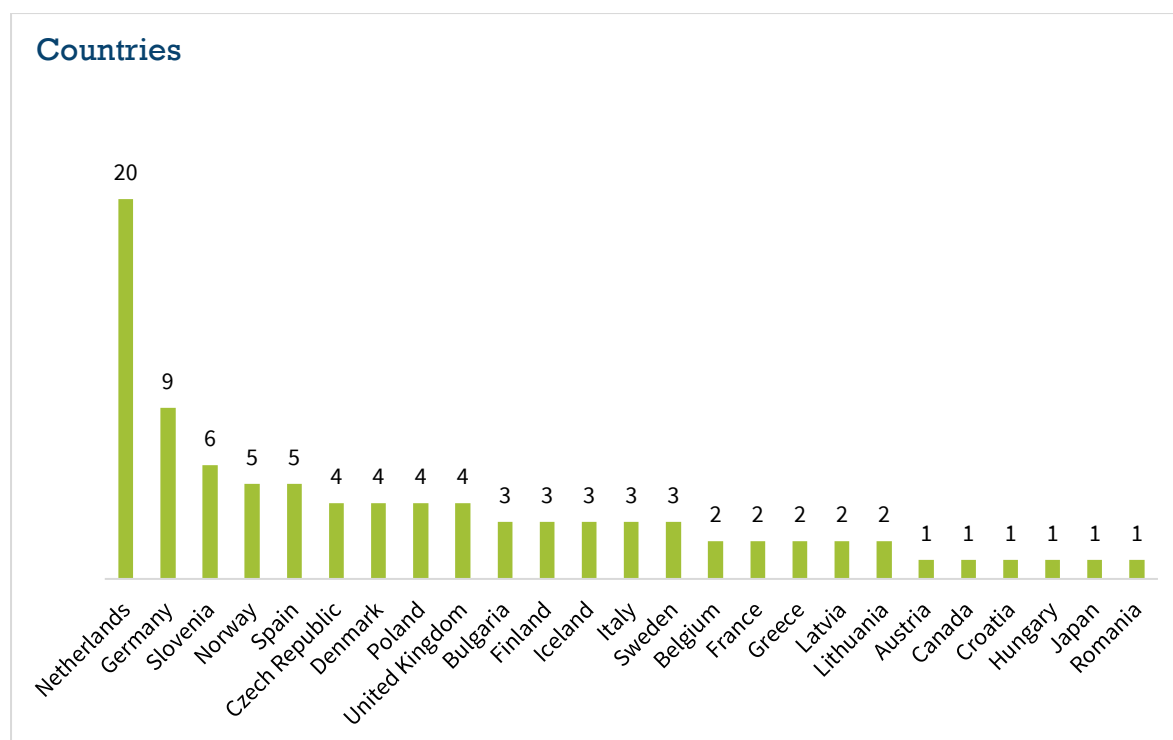
Countries

Fig 2. Number of participants of centrally organised in-person events per country of origin

Of the centrally organised events 12 were held in virtual format only. These events were attended by a total number of 888 unique participants (based on names indicated during registration) from 74 countries. As with the in-person events, the virtual events attracted mostly but not only participants from Europe. See Figure 3 below for an indication of the spread of countries. Again, some participants may have participated in several events.

Countries

Attendance Figures for Locally Organised Events

At the local level, 189 UI activities and events were (co-)organised by the various national consortia (both virtual and face-to-face). See Figure 4 for an overview of the types of events organised. Overall, around 5000 participants attended such events.
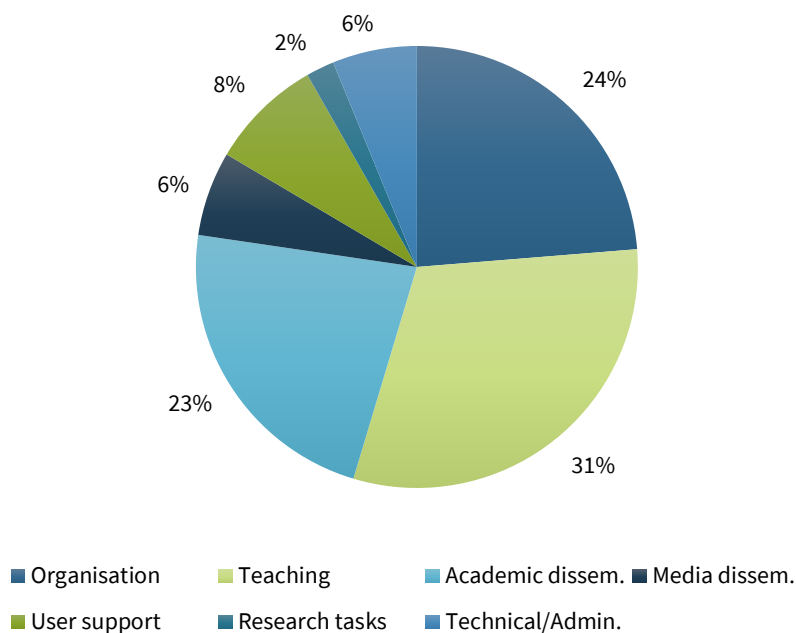
Types of Activities

Fig 4. A breakdown of locally organised UI events per type.

The UI events and activities reported present a broad variety of types, domains and disciplines (AI, SSH, libraries, legal issues, infrastructures and language processing tools), some of them topically focused and addressing smaller audiences, while others were relatively big events, internally or externally organised by CLARIN consortia. Some highlights:

- Courses or series of seminars; university courses, single or repeated seminars; for example:
  - Corpus Linguistics and Statistical Methods, an online bilingual (Finnish and English) course reported by FIN-CLARIN, which was attended by 88 participants. It was addressed to students in the language sciences and other fields that use corpora, and introduced CLARIN tools.
  - Digital Humanities Seminar at the University of Gothenburg, reported by Swe-CLARIN, had a focus on LT and text mining-related themes and was attended by approx. 100 participants.
- Summer schools and training events; for example:
  - Digital Tools For Humanists, focused on tools and applications dealing with texts

- Venice Summer School on Digital and Public Humanities, focused on historical monuments, artifacts and texts in the interdisciplinary field of digital and public humanities.

  Both events were supported by CLARIN-IT and attended by 35 and 25 participants, respectively.
- Participation of members of CLARIN consortiums in academic conferences, such as
    - LREC 2022 (44 presentations by CLARIN members)
    - The 6th Digital Humanities in the Nordic and Baltic Countries Conference (DHNB 2022); 8 presentations by CLARIN members.
    - The 16th conference for the history of Swedish, attracted 100 participants, 2 CLARIN presentations
    - International Conference on Historical Cryptology, HistoCrypt 2022, more than 50 participants, 3 CLARIN presentations
    - Sprogteknologisk konference 2021, 16/11-2021, 110 participants, CLARIN poster presentation.

Finally, with the appointment of three new ambassadors covering the domains of corpus linguistics and computer science, speech and language pathology, language acquisition and psycholinguistics, in September 2021, the CLARIN Ambassadors programme has gained new momentum (a fourth ambassador was appointed in September 2022, and her activities will be covered in the next report). The new ambassadors - Paul Rayson (Professor in Computer Science at Lancaster University, UK), Satu Saalasti (University Lecturer of Speech and Language Pathology at the University of Helsinki, Finland), and Eva Soroli (Associate Professor of Psycholinguistics at the University of Lille, France) - organised and attended 9 virtual and 6 face-to-face events (including a CLARIN Café, a slot at the LREC conference, and the francophone DH conference in Canada). They have also been very active in disseminating CLARIN in projects and other initiatives, as well as via their social media accounts.

Committees Supporting the Implementation of the Knowledge Infrastructure

Despite the impossibility of organising face-to-face meetings, the activities of the various committees have continued at a steady pace.

The CLARIN Legal and Ethical Issues Committee (CLIC), besides its role of advisor on all issues related to IPR, privacy and data protection and ethical matters, has been working specifically on issues concerning language resources and research under the General Data Protection Regulation. The revision of the 2018 white paper on this topic is ongoing and a follow-up on the 2021 CLARIN Café on the topic of text and data mining exceptions in the new Directive on Copyright in the Digital Single Market is planned for November 2022 (link).

The User Involvement Committee held a virtual meeting in April 2022. UI activities were presented, discussed and disseminated within the various national consortia. A draft report on the national activities is available. It will be finalised in November 2022.

Tour de CLARIN & Impact Stories

The Tour de CLARIN remains a prominent platform for the promotion of the national consortia, their resources, tools, events, and researchers. It has now been extended to also include the work of K-centres. Volume IV of Tour de CLARIN was published in print in 2021. In 2022, new K-centres and national consortia have been invited to be featured and visits are on-going

The introduction of the abovementioned series of impact stories (launched September 2021) also aims to demonstrate and inspire uptake of tools and resources, providing an important platform to showcase projects and outcomes that might otherwise remain unseen by the larger CLARIN community (see section 2.1). For more details, see the Tour de CLARIN and Impact Stories Progress Report 2021-2022 (CE-2022-2110).

Communication and Social Media

The main channels for dissemination are the monthly Newsflash, and the CLARIN accounts for Twitter, LinkedIn, and YouTube. The Newsflash currently has 1992 subscribers. The Twitter channel is used for sharing news, announcing events, and putting particular topics of interest into focus in an easily digestible format. It also supports the dissemination tasks in the EU-funded projects in which CLARIN participates. The CLARIN Twitter channel has now reached an audience of almost 4300 followers (an increase of 8% compared to 2021). The LinkedIn platform is mainly used to share advertisements for jobs at CLARIN headquarters. The CLARIN YouTube channel proved to be particularly successful during the pandemic, as it offered the possibility of re-watching recordings of virtual events and interviews. Among the most popular videos was the showcase on the integration of CLARIN services in EOSC (518 views) and a recording of the talk by Tomaz Erjavec during a CLARIN Café on how to extract useful parliamentary knowledge (Concordancing the Parliament, 239 views).

KPI – Number of views of tutorials and videos

September 2018: 1470 YouTube views over the previous twelve months
September 2019: 3002 YouTube views over the previous twelve months
September 2020: 3920 YouTube views over the previous twelve months
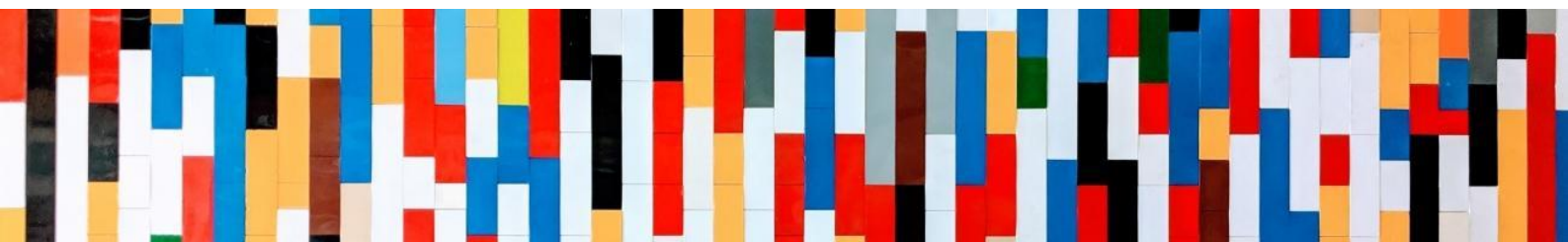September 2021: 6930 YouTube views over the previous twelve months
September 2022: 5255 YouTube views over the previous twelve months
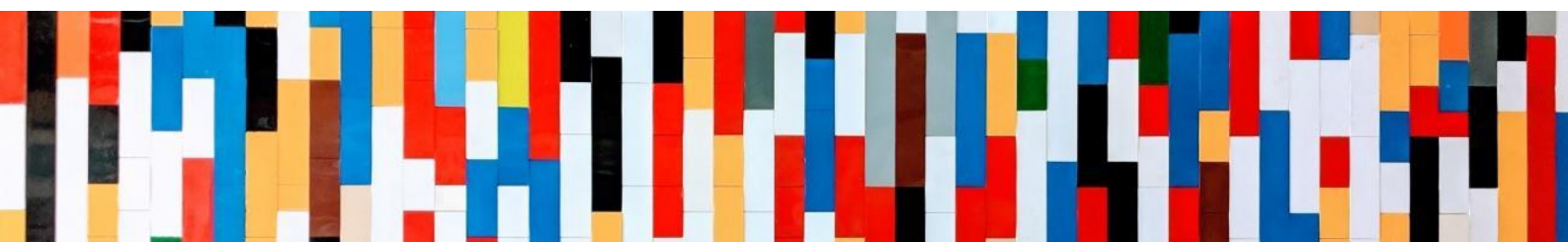
2.2   Technical Infrastructure

Where We Stand

1. Adherence to the FAIR Data Principles is key in the design of all data services
2. Guidance and consultancy is provided on the use of persistent identifiers. This includes topics such as when to use Handles and Digital Object Identifiers and the connection to machine-actionability
3. Procedures for quality assurance and curation, both for tools and (meta)data, are in place. This includes automated checks where possible (e.g. to ensure technical interoperability) and manual intervention where needed (e.g. to assess and improve semantic interoperability)
4. Readiness for integration in the European Open Science Cloud (EOSC): CLARIN has demonstrated its technical maturity by being one of the first RIs to offer integrated services through the EOSC Portal, underlining the high level of interoperability that has been achieved.
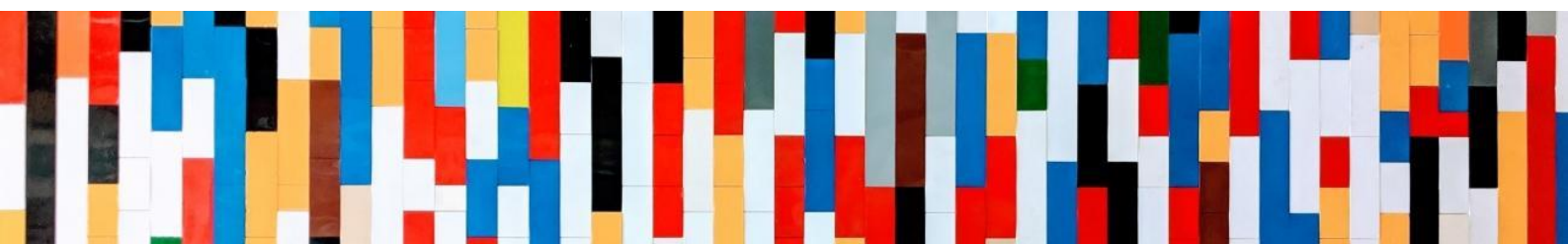
Where We Want to Go

1. Improved discoverability of the tools provided via CLARIN centres and thematically related platforms, such as the European Language Grid and the SSH Open Marketplace. To this end, the synchronisation between relevant sources of information will be enhanced, including the website, the Language Resource Switchboard, the VLO, and tool-tailored search functionality
2. Enhancement of CLARIN Resource Families by integrating multiple resources and tools into an innovative, single virtual environment for data processing, analysis and comparative research
3. Innovative data architecture models with a particular focus on advanced connections between corpus search engines and long-term archiving
4. 'CLARIN for Programmers'
   a. Natural Language Processing services will be advertised more prominently to programming scientists, for instance with well-documented application programming interfaces and example snippets in popular development environments
   b. The results and innovations will be disseminated via web-based tool registries and tutorials for programmers
5. FAIRness and especially interoperability are furthered at several levels:
   a. Among CLARIN centres (e.g. through core metadata recommendations and a general CLARIN gateway service for FAIR digital objects)
   b. With GLAM partners and other data providers and users that can benefit from CLARIN's processing services.

Actions Planned in Work Plan 2022

The following specific activities were planned for the strategic priority area Technology Infrastructure (see CE-2021-1909, section 3.2):

- Furthering FAIRness and especially interoperability at several levels: between CLARIN centres but also with GLAM partners (e.g. Europeana) and other data providers and potential users. Enhanced support will be provided for repository-specific identifiers (e.g. Invenio, PubMed) in the Language Resource Switchboard by relying on the Digital Object Gateway (DOG) functionality
- Enhancing API functionality by:
  - Providing well-documented APIs for the VLO, VCR and the DOG
  - Documenting how these APIs are used within the infrastructure
  - Establishing probes that continuously check the validity of the API output
- Connecting the CLARIN technical ecosystem to relevant 'outsider' systems, such as DataCite for DOIs
- Conceiving a PID gateway service architecture as an important step towards integration of FAIR digital objects
- Contextualising overviews of data, tools, publications, and training material by linking them to discovery platforms such as the SSH Open Marketplace and Zotero
- Informing centres about evolving technical innovations for infrastructures, for example for persistent identifiers, standards, federated content search, and metadata
- Offering more efficient workflows for the creation and curation of metadata for language resources and services
- Setting up a 'CLARIN for Programmers' webpage and experimental online 'open hours' will target newcomers and new CLARIN centres that need technical advice
- Specific plan for funded activities by experts from the network:
  - Re-engineering and migration to the CLARIN ERIC hosting infrastructure of the curation module and link checker
  - Improving the management, browsing, and editing of tool and service metadata
- Technical centres: providing improved documentation on optimal use of persistent identifiers
- Development work on the Virtual Language Observatory, Federated Content Search, Virtual Collection Registry, and Language Resource Switchboard.

Achievements in 2022

Improving the interaction with elements/initiatives/platforms outside the CLARIN Technical Infrastructure formed was a special focus in the Work Plan for 2022. The key elements to achieve this improved interaction are: APIs, high-quality documentation, and pluggable gateway modules.

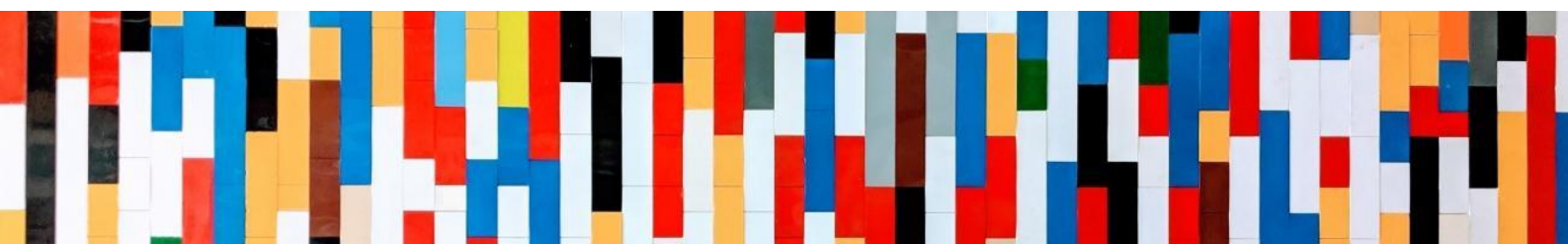## Centres, Data and Services

Regular Centre Assessments

As planned, two assessment rounds were organised. Out of 70 registered CLARIN centres, 24 are currently certified as B-centre. In 2022, there was one new certification and there were three re-certifications. Two more certifications are pending, waiting for the CoreTrustSeal procedure to be concluded.

> **KPI – Number of certified B-centres**
>
> September 2018: 21 certified CLARIN B-centres
> September 2019: 22 certified CLARIN B-centres
> September 2020: 24 certified CLARIN B-centres
> September 2021: 24 certified CLARIN B-centres
> September 2022: **24 certified CLARIN B-centres**

Development of Infrastructural Services

- Digital Object Gateway (DOG): This abstraction layer to access digital objects in a unified way was released as Python library (DOGlib) and as a REST API (DOGapp). It is still being further developed, but can already serve as part of demonstrations, particularly in combination with other APIs
- Virtual Language Observatory: A first development prototype was created for the VLO 5.0, which will be based on a backend API that will be made publicly accessible, which thus allows alternative front-ends to be created
- Virtual Collection Registry: The Virtual Collection Registry (VCR) was connected to several external repository systems in the context of SSHOC, as described in the D3.8 Implementation report and available SSHOC Switchboard and VCR services
- Language Resource Switchboard: Versions 2.4.0 and 2.4.1 of the Switchboard bring improvements to the user interface as well as (experimental) support for multiple input files.
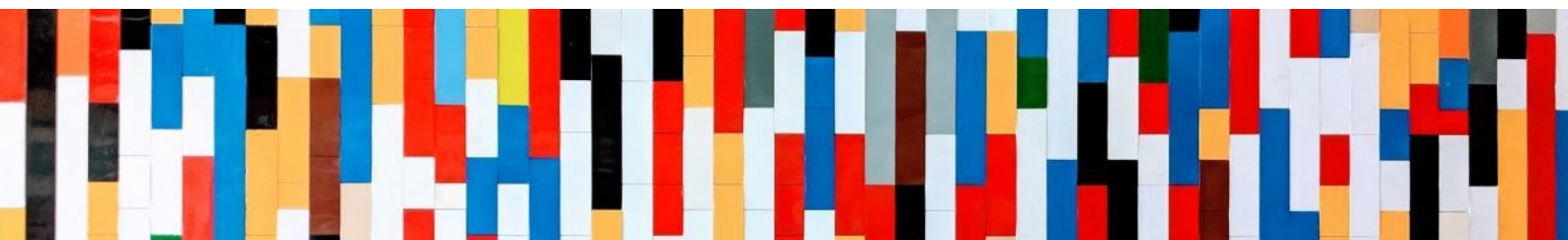
CLARIN for Programmers
Based on the ever-increasing amount of available technical resources for programmers to interact with CLARIN data and services, a dedicated web page was launched. It bundles information about APIs (e.g. for the VCR, the Switchboard and the DOG) and example use cases, often in the form of Python notebooks. One example is the tutorial on the use of notebooks and virtual collections to process newspapers provided through Europeana with CLARIN-PL services.

KPI – Percentage of format recommendations published by centres offering depositing services

September 2018: 8 out of 24 centres = 33%
September 2019: 8 out of 24 centres = 33%
September 2020: 11 out of 24 centres = 46%
September 2021: 18 out of 26 centres = 69%
September 2022: **19 out of 27 centres = 70.4%**

KPI – Percentage of member and observer countries connected to the Service Provider Federation

September 2018: 21 out of 22 countries = 95.5%
September 2019: 23 out of 24 countries = 95.8%
September 2020: 23 out of 24 countries = 95.8%
September 2021: 24 out of 25 countries = 96%
September 2022: 23 out of 24 countries = 95.8%
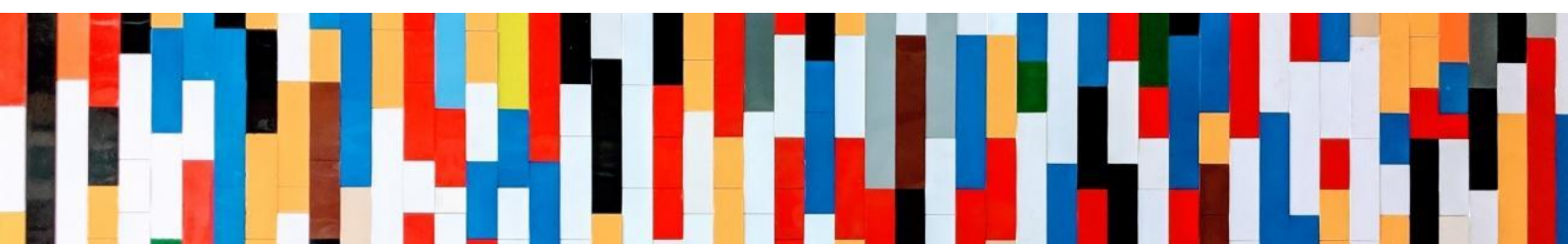
CLARIN Resource Families

In 2022, the CLARIN Resource Families initiative (CRF) has been expanded with overviews of 4 new resource and tool families – corpora of legal texts (32 corpora), language models (86 corpora), sign language resources (35 corpora and 13 lexica) and corpus querying tools (41 desktop tools and 45 online tools). The latter two families have been prepared in the context of the CLARIN Resource Families Project Funding by Henk van der Heuvel (NL) and Martin Wynne (UK). These CRF-funded projects have identified valuable sign language resources and corpus query tools not yet part of the CLARIN infrastructure.

Substantial progress has been achieved in the ongoing curation of metadata issues in existing resource and tool families. The metadata issues, which are collected on GitHub, generally concern missing licence information, annotation formats, or size. As of September 2022, there are 207 open issues remaining while 177 issues have been closed in comparison to the last report in September 2021, when there were 266 open issues remaining with 118 issues closed. In addition, the presentations of some of the existing older families are being overhauled with a greater inclusion of qualitative metadata. New recommendations for describing the metadata of CLARIN tools and resources from the qualitative perspective have been collected. They are based on a survey of the existing guidelines offered by CLARIN B-centres for depositing resources and tools in their repositories. The survey, which was conducted with additional input by the User Involvement Committee, shows that the existing guidelines generally focus on structural aspects of the deposit but neglect qualitative characteristics important for the (re)use of the tools and resources in SSH research. An online training session to present the guidelines to repository administrators is being planned for Q4-2022 .
The CRF initiative is the focus of a book chapter in *CLARIN. The infrastructure for language resources*, which will be published by De Gruyter..

Workshops to Enhance the Technical Infrastructure

- The CLARIN Centre meeting 2022 (hybrid testcase event)
- The CMDI Session number 2, on the topic of vocabularies (virtual event).

2.3 Sustainability

Where We Stand

1. Landscape:
   a. CLARIN is continuously monitoring the need for the evolution and adjusted articulation of the CLARIN Value Proposition
   b. The CLARIN Value Proposition is well aligned with the strategy of the key stakeholders
2. Members and partners:
   a. Emerging national consortia are offered workshops on how to set up a CLARIN infrastructure at the national level and a series of best practice papers has been prepared to help leverage the visibility of the results from national investments
   b. GLAM sector: Structural collaboration with the GLAM sector has been established at the European level
   c. Connection to industry: In many CLARIN member countries, one or more models for collaboration with industry and public-sector parties is either in place or planned.

Where We Want to Go

1. Financial and organisational sustainability:
   a. Consolidation and extension of the current membership base (critical for both organisational and financial sustainability)
   b. Diversification of the financial portfolio in ways compatible with the RI role as a distributor of resources created with public funds
   c. Development of a decommissioning scenario for decentral nodes that lose their local funding, including objective criteria for assessing what constitutes crucial infrastructural resources for which fall-back service options in the wider CLARIN network could be offered
2. Integrating CLARIN within scientific communities:
   a. Extension of the number of disciplinary communities in which CLARIN services are adopted
   b. Pro-active stimulation of compatibility and incorporation of disciplinary results (e.g. from ERC grants and mission-oriented projects) in the CLARIN infrastructure
   c. Alignment of the services offered by CLARIN with disciplinary research agendas, in particular those of the social sciences and humanities
3. Strengthening international collaborations:
   a. The collaboration with non-academic parties and the potential for sustainable relationships is better aligned with successful models at the national level and emerging local initiatives are better supported by CLARIN ERIC
   b. Collaboration with actors from the GLAM sector will be consolidated and new alliances will be sought aiming at opening up cultural heritage sources with digital methods
   c. Viable models for more structural information exchange and collaboration with industrial parties will be articulated and promoted among CLARIN nodes, partly based on the recommendations to be expected from the H2020 project ENRIITC
   d. Joint development of formal models for collaboration among RIs beyond Europe.

Actions Planned in Work Plan 2022

The following specific activities were planned for the strategic priority area Sustainability (see CE-2021-1909 section 3.3):
- Monitor national strategies towards balancing infrastructural investment at national and European level
- Monitor dynamics in the relevant financial frameworks in order to be able to assess the preparedness of countries to continue to invest in RIs and under which conditions
- Develop a deeper insight into the outreach instruments that could lead to wider uptake
- Further explore the strategic positioning of CLARIN as an innovation hub for the data science sector. The potential for collaboration with industry will be explored, including collecting a more comprehensive overview of existing links with industrial parties in the national consortia and the needs for exchange and support that may help building up sustainable relations, as well as an assessment of which industrial niches could be served with access to statistical language models and/or co-development for this topic
- Develop a financial model that underlines the various cost categories, e.g. resources for developing and maintaining technical components versus resources to pay for those who will 'operate' them
- Develop a stronger role for qualitative performance measures and narratives as part of the CLARIN-internal monitoring framework
- Review of the CLARIN Statutes to identify the need for adjustments that could help facilitate the establishment of formal agreements with parties outside of Europe.
- Strengthen outreach and visibility beyond Europe, particularly in Africa and Australia, through existing allies and the CLARIN Ambassador Programme.

Achievements in 2022

The sustainability agenda is a multi-faceted programme, with technical, organisational, political and financial dimensions. These will all be addressed in this section of the progress report, but obviously the key precondition for CLARIN's sustainability is the wider visibility and uptake of CLARIN services in relevant communities of use, both in academia and beyond.

Widening Uptake and Impact

Careful alignment of CLARIN's service offer to the research agendas of the various communities in which there is potential for uptake is ongoing, and the potential for impact along the dimensions of societal relevance has been strengthened through several forms of cooperation across different initiatives, countries, and continents. The following specific strategic steps were taken in 2022 to increase uptake:

- The Ambassadors programme, which proved instrumental for outreach to new communities pre-COVID-19, was relaunched in 2022 with several new ambassadors (see section 2.1 for more details).
- Links with existing research agendas and programmes have been strengthened in order to reinforce the relationship with specific communities of use that could benefit from CLARIN's infrastructural service offer. Through the exchange with experts, opportunities for outreach in new communities have been generated, e.g. through teaching at COST training events, planning thematic CLARIN Cafés, and monitoring the models for resource construction and the preparation of joint project proposals in these communities:
  - European Network for Combining Language Learning with Crowdsourcing Techniques (https://enetcollect.eurac.edu)
  - Distant Reading for European Literary History (https://www.distant-reading.net)
  - NexusLinguarum: Promotes synergies across Europe between linguists, computer scientists, terminologists, and other stakeholders in industry and society, to investigate and extend the area of linguistic data science (https://nexuslinguarum.eu)
  - Multi3Generation (www.multi3generation.eu): The COST action that promotes interdisciplinary research in Natural Language Generation
- A subproject of EOSC Future was initiated in order to develop and apply COVID-19-related metadata labels to records from the life sciences/biomedical domain, the recently launched ParlaMint dataset, as well as a collection of social media data. In addition, ideas for project proposals are being discussed that are suitable for submission in response to Horizon Europe Pillar 2 calls under Cluster 2.
- An update was realised for the dedicated web page specifying how CLARIN contributes to the 2030 Agenda for Sustainable Development adopted by the

General Assembly of the United Nations (UN) in 2015. The page underlines the general potential for direct or indirect impact from access to historical and contemporary language data and analysis tools for research seeking to address societal challenges. These resources can be a vital source of data fuelling the relevant scientific agendas. In addition, they may also support the implementation of policy, a process that is dependent on a good understanding of social behaviour, for which language data can be instrumental.

- To enhance the visibility of CLARIN's impact and thereby inspire audiences to use the service offer, a series of impact stories was initiated in 2021. Published on a [dedicated section of the website](#), the narratives include both longer and shorter stories illustrating CLARIN's link with industry, education and excellence in research. So far, 9 impact stories have been published.

- There are quite some dynamics in the strategies for infrastructural investments, both at the national and the European level. Much attention is being paid to the possible role of Digital Competence Centres and the added value of the European Open Science Cloud (EOSC). As a result of its involvement in a series of EU-funded, EOSC-related projects, such as EOSC Focus (2022-2025), CLARIN is well positioned to timely observe changes that may affect the funding for the national nodes and the ERIC. In the coming two years, more clarity is to be expected on how the changes in the ecosystem may affect countries' preparedness to continue their investments in RIs and under which conditions, and how this may affect CLARIN ERIC's financial sustainability.

- CLARIN's involvement in the H2020 project ENRIITC has worked as a catalyst for the generation of a deeper insight into what could be the pillars of a more mature strategy towards engagement with non-academic partners. With the support of a recently appointed senior advisor, a plan has been developed to enhance the interaction with the public sector and industry with the following foci:
    - Strengthening awareness and buy-in in key stakeholder groups such as regional and local public authorities and policy makers, industry and industry associations/interest groups, GLAM, digital cultural heritage, publishing, educational sector players, and citizens
    - Setting up a task force to prepare a model for the exchange of information with CLARIN nodes regarding interaction with the public sector and industry
    - Raising awareness of CLARIN as a network of expertise and service provision for cutting edge AI/NLP technological research, with the central organisation as a potential broker and as such a possible partner of choice for industry working on NLP applications in large innovation projects funded by the EU and the member states' recovery plans focusing on AI and the digital transformation

- The CLARIN Statutes have been analysed for their suitability as a formal vehicle for the establishment of formal agreements with parties outside of Europe. The most

critical issue seems to be the need for flexibility when it comes to determining what would be a suitable fee level. It has been concluded that the Statutes currently provide sufficient flexibility with regard to this issue.

Extension of ERIC Membership

The United Kingdom has indicated to be prepared to join as a full member. A formal application will only be feasible once the UK has reached agreement with the European Commission on the models for collaboration in the realm of research. Meanwhile, it has been agreed to prolong the special model for observership (which started in 2021) until the end of 2023. The exchange of information with several countries about a model for collaboration (membership, observership, other) has intensified and there are good expectations that this will lead to a formalised model of collaboration in a not-too-distant future. This holds in particular for Spain and Switzerland.

In addition, some regions and institutions have expressed their intention to join the CLARIN community through a special agreement according to Article 18 of the Statutes. CLARIN-CH, the national CLARIN network in Switzerland, has been set up and a consortium of seven Swiss institutes is currently preparing an application to become an observer in CLARIN ERIC. In Spain, the CLARIAH-model has been adopted and an INTELE workshop with representatives of the relevant Spanish ministerial departments, CLARIN ERIC and DARIAH-EU was held on 13 to 14 September 2022 to prepare the last mile for a membership application.

The observership of France was not prolonged, but the collaboration with the French nodes that were active in CLARIN in the past years has remained intense.

KPI – Number of members

In September 2018: CLARIN had 20 members and 2 observers
In September 2019: CLARIN has 20 members and 4 observers
In September 2020: CLARIN has 21 members and 3 observers
In September 2021: CLARIN has 22 members and 3 observers
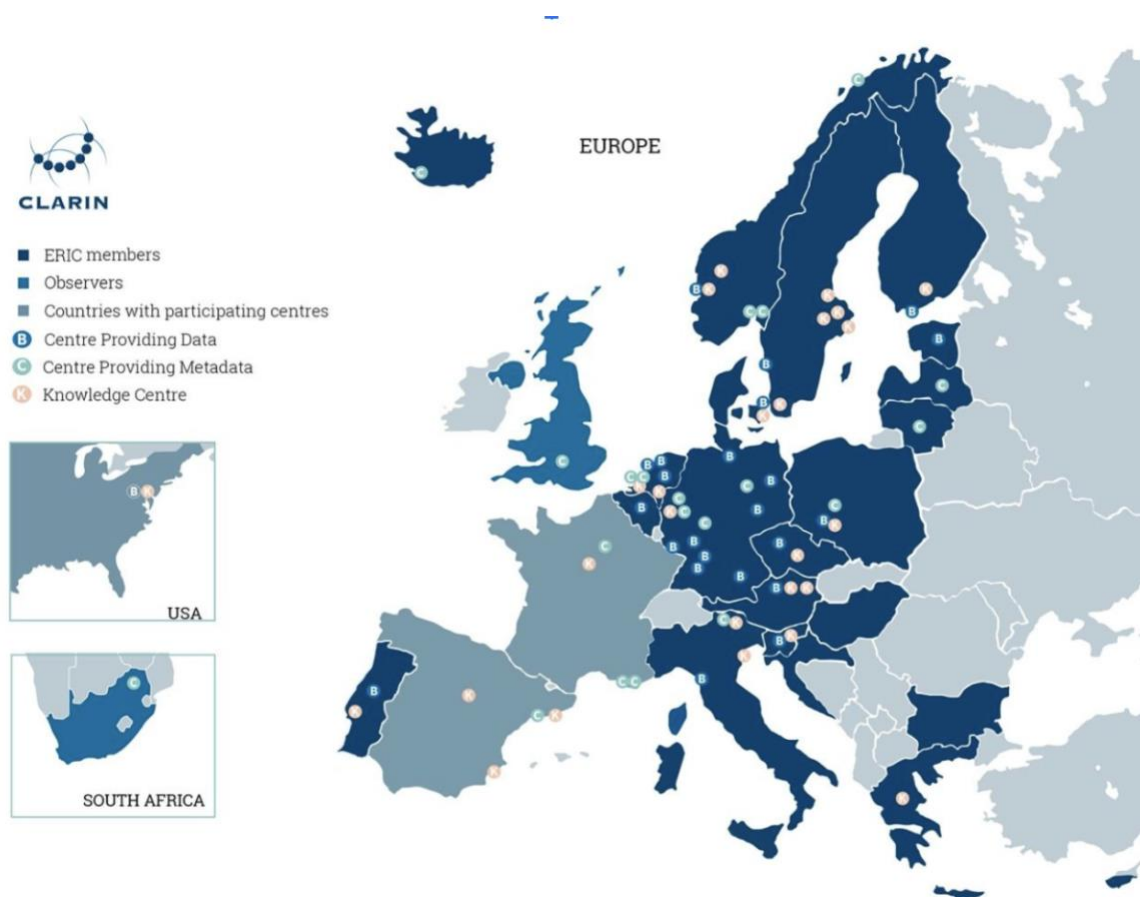In September 2022: **CLARIN has 22 members and 2 observers**

Fig. 5: Map of CLARIN member and observer countries and CLARIN centres (September 2022).

Crossing Borders

CLARIN is increasingly involved in collaborations with other infrastructural initiatives through different models. EU-funded projects play an important role in facilitating the various forms of collaboration. An overview of the various forms of interaction and potential for synergy is presented below.

EU Projects

In the context of EU-funded projects, CLARIN is collaborating with a broad variety of parties across many communities and disciplines. For the multidisciplinary collaboration that is often a crucial pillar of the work in externally funded projects, it is important to be able to consult domain experts, both in the design and planning stages, as well as during implementation and testing. In order to ensure that a pool of candidates that is familiar with the objectives of CLARIN can be consulted, an SSH Expert Panel was established in 2019.

A full overview of CLARIN's participation in EU-funded projects over the past years is provided in CE-2015-0636 (update 2022). The following were active in 2022:

1. Europeana DSI: Series of shorter projects aimed at the development of Europeana as a Digital Service Infrastructure; CLARIN's contribution focuses on facilitating the reuse of digital cultural heritage resources for research purposes; in the context of DSI-4, the last project in the project series (Sept 2015 – Aug 2022), the successful online tutorial on how to use Jupyter notebooks was prepared
2. ERIC Forum: Collaboration with the other ERICs; January 2019 - December 2022
3. SHHOC: Towards a Social Sciences and Humanities Open Cloud; January 2019 - April 2022; CLARIN's contribution to SSHOC is realised by the central organisation and six CLARIN nodes
4. TRIPLE: Towards a discovery platform dedicated to SSH data, researchers and projects at European scale; October 2019 - April 2023
5. ENRIITC: Towards a pan-European network of Industrial Liaison and Contact Officers in order to stimulate the partnerships between research infrastructures and industry; January 2020 - December 2022
6. UPSKILLS: UPgrading the SKIlls of Linguistics and Language Students; September 2020 - August 2023
7. European Language Equality: Roadmap towards digital language equality in Europe in 2030; January 2021 — September 2022
8. FAIRCORE4EOSC; June 2022 - May 2023
9. EOSC Focus; June 2022 - May 2023.

Other SSH ERICs and Platforms

CLARIN ERIC sees a critical need for taking actions to reduce the duplication, overlap and competition between SSH domain players: the ERICs(-to-be) CESSDA, CLARIN, DARIAH, ESS, and SHARE, emerging platforms such as the European Language Grid (ELG) and GoTriple, and initiatives such as ELRC.

Collaboration with DARIAH-EU has been taking place through regular meetings at the level of the two BoDs, as well as in the context of the H2020 cluster project SSHOC (2019-2022) and the follow-up collaboration under the umbrella of SSH Open Cluster, which has been initiated by the ERICs in the SSH cluster and which has been formalised through a Memorandum of Understanding. In the context of the SSH Open Cluster, the launch of the SSH Open Marketplace is a major step. This discovery portal, prepared in the context of the SSHOC project, will be jointly maintained by CESSDA, CLARIN and DARIAH. This collaboration is underpinned by a formal agreement between the three ERICs. This collaboration is also crucial for promoting the specific disciplinary needs of RIs based on distributed data centres in the newly emerging European ecosystem for infrastructures.

In 2022, several cooperative actions took place between CLARIN and DARIAH, such as the alignment of position statements issued, the prolonged maintenance and development of the DH Course Registry (see section 2.1), joint dissemination activities and workshops. In a number of countries (e.g., AT, BE, BG, CZ, DE, FI, GR, NL) CLARIN and DARIAH consortia are or will be jointly funded (sometimes under the name of CLARIAH), which will continue to offer interesting possibilities for bottom-up collaboration initiatives that can be extended to the European level. In particular the collaboration in the education and training action lines has much potential for impact. In 2022, the BoD started the process of collecting the practical and technical challenges stemming from the need to integrate practices of two communities in a local context. This may lead to recommendations for countries where the merging of CLARIN and DARIAH consortia is considered.

The exchange of insights and information among the wider group of RIs that are fuelled by and targeting the SSH domain has proven to be instrumental in ensuring that the SSH perspective will be taken into account in any further steps towards the establishment of EOSC. It has been successfully communicated that out of 1.7 million researchers in Europe, around 500.000 are working in SSH. However, although this figure is acknowledged by the various stakeholders, getting attention for the sometimes deviant perspective of the ERICs in the SSH cluster requires continuous effort.

An interesting case is the Netherlands, where the national CLARIAH consortium applied for funding together with the RIs in the social sciences and the emerging RIs in the humanities domains under the name of SSHOC-NL. This is calling for even more attention for the suitable and sustainable models of collaboration at the level of the SSH cluster.

European Open Science Cloud (EOSC)

In 2020, CLARIN ERIC joined the EOSC Association as a member. Already in 2018, CLARIN demonstrated 'EOSC-readiness' in terms of its level of interoperability and technical maturity. This was illustrated by the fact that CLARIN was one of the first research infrastructures to integrate thematic services in the first release of the EOSC Portal. The rules of participation that will be adopted are being carefully monitored and jointly discussed with the other RI parties, both in the SSH domain and beyond, partly in the context of the cross-cluster collaboration that started in 2019, and as of 2022 through CLARIN's participation in HEU project EOSC Focus (2022-2025), as part of which crucial steps towards the governance and visibility of joint services in the EOSC landscape are being prepared.

e-Infrastructures

Collaboration with global data infrastructures in RDA (Research Data Alliance) has been continued and CLARIN ERIC is participating in the Collaborative Data Infrastructure (CDI) initiated by EUDAT, for which the agreement was signed by the end of 2017. The existing

contacts with OpenAIRE, EGI and GEANT are carefully maintained, as these parties play a crucial complementary role in the emerging Europe's Open Science Policies and, more specifically, the European Open Science Cloud.

GLAM Sector

CLARIN aims to be recognised as a potential partner for initiatives in the GLAM sector (Galleries, Libraries, Archives, and Museums) and is actively looking for projects and platforms to gain more visibility, both to learn more about the specific features of the conditions for research based on digital cultural heritage, and to offer expertise on the various dimensions of service interoperability in the context of a distributed data architecture. In particular, it has become apparent that CLARIN can play a brokering role for identifying possible routes towards solutions for accessibility issues in Europe's digital cultural heritage, which are inherent to the multi-linguality of the collection descriptions.

- The existing collaboration with Europeana and LIBER have proven instrumental in reaching a more mature level of understanding the agendas in this sector and the options for improved alignment and synergy. Collaboration on metadata harmonisation and harvesting have led to visibility of Europeana in EOSC Portal. With LIBER, several joint workshops have been organised in the context of the SSHOC project
- CLARIN was invited to participate in a consultation session organised by the unit in DG CONNECT that is responsible for the EC support for initiatives in the sector that is known as 'Digital Cultural Heritage'
- A workshop CLARIN and Libraries on Interoperability of Text Platforms for Digital Libraries took place in May 2022. It formed the basis for one of the panel sessions at the CLARIN Annual Conference in Prague.

Beyond Europe

As part of the third-party agreement (since 2017), Carnegie Mellon University, USA, is running a certified CLARIN B-centre and K-centre. Additionally, several CLARIN centres are collaborating with Vassar College, Brandeis University, Carnegie Mellon University, and the Linguistic Data Consortium in the framework of the LAPPS Grid project.

South Africa has adopted a national research infrastructure roadmap with funding for SADiLaR, a digital language resources facility. In 2018, South Africa joined CLARIN ERIC as an observer. In 2019, SADiLaR and CLARIN began to explore actions that can lead to further collaboration between CLARIN and other institutes and research groups on the African continent. In addition, SADiLAR established a C-centre with the prospect of becoming a B-centre. There is structural collaboration with the Network for Digital Humanities in Africa and CLARIN's involvement is planned in the network activities aimed at the exchange of information among linguists working on African languages.

Finally, invitations have been received from infrastructural networks and nodes from Australia, Canada, Israel, Mexico, Singapore, and the United States of America to explore the potential for collaboration.

Financial Perspectives

The financial sustainability topic has been addressed as follows:

- The 'decommissioning' scenario for the CLARIN infrastructure and access to the resources in the associated centres which was formulated a few years ago still seems valid. It has been reconfirmed that the reserve budget that has been secured over the past years will allow for a scenario in which there would be a reasonable amount of time (around two years) to ensure a careful transfer of the existing service offer to one or more service providers that have the interest and capacity to adopt the core of the CLARIN service portfolio.

- The in-kind contributions by the national consortia forms the basis for a major part of the CLARIN action lines. The financial volume of these activities is approximately 10 times as high as the funding available for the central work plan. We have started to incorporate this ration explicitly in all relevant reports requested by stakeholders. The explicit reporting may contribute to a wider understanding of the complex dependencies in the model for the financing of multi-country collaboration in the domain of research infrastructures and the corresponding responsibilities.

Implementation of the KPI framework: Next Steps

A roadmap for the implementation of the ESFRI Key Performance Indicators framework has recently been made available. To be prepared for a more structural form of monitoring, experience with measuring certain performance aspects has been built up over the past years. To stimulate measurability of publication output, the CLARIN Zotero database has been updated. The need for incentives that may help to increase the volume of citations and citable papers has led to the decision to open the call for selected papers for the CLARIN post-conference volume to a wider group. As in the longer run the ESFRI monitoring framework will most likely put more weight on the collaboration with non-academic organisations, an exploration was conducted in 2021 of the currently existing collaboration with industry and public-sector parties within the CLARIN countries. A first analysis of the data collected shows that there is much more interaction with non-academic parties than was previously reported. The findings have been fed into a process towards the identification of measurable progress for objectives that may become more prominent in the near future.

2.4   Organisational Development

The central support organisation is one of the pillars for the realisation of CLARIN's mission and the implementation of the strategic agenda. This section describes the capacity and instruments of CLARIN's central organisation and the ambitions for strengthening it.

For the realisation of CLARIN's mission and the implementation of the strategic agenda, a governance model and central support organisation are in place. Starting with a relatively small project organisation for the coordination of activities in around 10 countries, along with a technical development team, it has evolved into a professional organisation with responsibility for a range of central tasks, plus support, for the coordination of activities in around 25 countries. The employment of personnel is arranged through Utrecht University, based on an agreement for housing and secondment. This model comes with good conditions for salaries, pensions, etc., which has enabled CLARIN to recruit staff against attractive conditions and, where appropriate, permanent contracts.

Apart from responsibility for the support of cross-country collaboration and the organisation of events, communication and financial management, the CLARIN Office is also responsible for the administrative and policy support of the General Assembly, the National Coordinators' Forum, the Strategy and Management Board, as well as other committees. There is also in-house capacity for the production of print material, the support for online meetings, and the editing of video recordings, including for online events. As of 2020, the accounting agency BACOM is supporting CLARIN ERIC in all matters regarding financial records and reporting. In order to enhance the available capacity and skill levels for all Office staff members, additional training and development tracks have been undertaken or planned. These tracks can be partly arranged through the offer for employees of Utrecht University.

Where We Stand

1. Current capacity:
   a. Coordination support: A good model has been implemented for collaboration and sharing of responsibilities among the Office team members who work from a service-oriented mindset that contributes to overall trust building.
   b. Technical development: A well-organised team of technical developers is in place, adequately balanced in terms of individual skill profiles and capable of maintaining the high standard of the technical services.
   c. Financial organisational support: Expertise on rules and procedures for ERICs and European projects is well covered by the financial team.
2. Instruments:
   a. Communication: A range of communication channels for outreach and dissemination is in place (Newsflash, website, social media, video channels, printed materials). Dedicated mailing lists for specific communities, such as the network of people with an interest in parliamentary data and the teachers network.
   b. Creation and editing of audiovisual dissemination materials: There is adequate capacity for the basic generation and processing of audiovisual materials and design tasks.

Where We Want to Go

1. Strengthening the existing coordination framework:
   a. Reinforced models for coordination and collaboration between the various bodies and task forces active in CLARIN
   b. Diversification of communication with stakeholders and communities outside of the national consortia
   c. In alignment with sister infrastructures: Ensuring that instruments are in place for capacity development for the building, maintenance and management of the RI nodes, including a common reward system for professionals with a career in the RI landscape
   d. Increased awareness and understanding of the central and decentral roles in ensuring responsibility for resource quality
2. Enriching the Office capacity:
   a. Training & education coordination
   b. Technology watch
   c. A trusted and sustainable model for collaboration with professional designers for front-end website design, development, and updates.

Actions Planned in Work Plan 2022

The following specific activities were planned for the strategic priority area
Organisational Development (see CE-2021-1909, section 3.4):
- Exploration of the option to diversify the communication channels to enhance
  the level of exchange with parties outside the national consortia (e.g.
  independent experts, public sector, policy makers, industry)
- Production of one or more promotional videos, also as a stepping stone for the
  further alignment of and support for communication approaches by the national
  consortia
- Monitoring of the effectiveness of the newly designed website, launched in 2021
- Development of more advanced methods and metrics for:
  - Spotting interest in the CLARIN service offer among new communities that
    have a strong interest in artificial intelligence, data science, and machine
    learning.
  - Measuring impact through publications
- Generating and publishing a Gender Equality Plan (GEP) for CLARIN in
  accordance with what has become a requirement for obtaining funding in
  Horizon Europe
- Adjustment of the division of labour within the BoD and the CLARIN Office to the
  needs of the evolving CLARIN landscape and the profile of the new Executive
  Director. Reinforcement of the overall capacity of the CLARIN Office team and
  the room to liaise with parties outside the national consortia (e.g. independent
  experts, public sector, policy makers, industry).

Achievements in 2022

Most topics in the list above have been addressed within the limitations imposed by the
COVID-19 pandemic. The past period was not optimal for organising and attending
networking events. Nevertheless, new communities have been reached with formats that
have proven useful for attracting specific subcommunities and non-academic
communities. Especially the CLARIN Café format turned out to be a very effective
instrument for online outreach and communication.

The work that is carried out within the CLARIN network as in-kind contribution to the RI is
not always recognised as valuable academic output, which can have negative impact not
only on position of individual contributors, but also in the attention of relevant
stakeholders. CLARIN has therefore started to support the community with pointers to

emerging shifts in evaluation paradigms adopted across Europe. A CLARIN BoD member has become active in the EOSC Task Force for Research Careers, Recognition, and Credit.

At the annual conference a promotional video clip will be launched with the aim of introducing CLARIN's value proposition to new audiences. In the near future other clips zooming in on specific aspects of CLARIN or on the activities in the national consortia will be produced, using the same design style.

The publication of a Gender Equality Plan (GEP) for CLARIN has been prepared. Currently a survey is running among the board members and the staff employed at the central level. The results are meant as evidence underpinning the measures to be described.
The envisaged plan aims to add specificity to CLARIN's unique working environment beyond what is already in place at UU. The focus of this Equality Plan will be on CLARIN's central organisation, though it will also include suggestions for advancing equality within CLARIN more broadly. It will be published on a dedicated webpage.

The newly introduced position of a Chief Operational Officer (COO) as of Q4-2022 will bring some changes to the overall management structure of the central organisation as well as to the Office-internal workflows. These changes are supposed to contribute to a more transparent division of labour within the Office team and clearer communication between the BoD and CLARIN Office. The new management structure is also supposed to address some of the challenges stemming from the growth of the central organisation over the past years, such as the workload for the Office team, and the need to spread the responsibilities of the executive director over more than one person. The COO will closely work together with the new executive director who has been appointed as of 1 January 2023. The new management structure was prepared in the first half of 2022. The implementation, which started around the summer, will be closely monitored in the coming months and year.

3   Other Action Lines

3.1   Governance, Coordination, and Organisation

Governance

The current version of the CLARIN ERIC Statutes was adopted by the General Assembly in 2017 and subsequently submitted to the European Commission for approval, which was announced in April 2018. The current governance structure is summarised in this governance organigram. The implementation is based on a specific model of collaboration between the Board of Directors (BoD), the National Coordinators' Forum (NCF), the Strategy and Management Board (SAMBA, a subcommittee of the NCF), the Standing Committee for CLARIN Technical Centres (SCCTC), and a policy for appointing members of the BoD.

In accordance with the agreed procedure for selecting and appointing new members, Dr. Francesca Frontini was invited to join the BoD for a term of two years, starting 1 January 2021. It has recently been agreed that her position will be prolonged for another term of two years. As of 1 January 2021, the current terms for the NCF chair and vice-chair as well as for the members of SAMBA started. Their terms will end by the end of 2022 and in October or November their successors will be elected.

For the CLARIN Agreements (the documents that specify the contributions that each national consortium intends to make to CLARIN), a simplified template was adopted as of 2019. Up until now, 13 Agreements were formalised based on this new model. CLARIN Agreements are signed for several years, closely following the funding cycles of the respective member consortia, and the procedure can be tailored to the situation in each member and observer country.

Coordination

- The division of responsibilities and load sharing between CLARIN ERIC and the national consortia for the maintenance of the central services and the ambition to stay at the forefront of technical innovation has led to a plan for defining smaller projects (up to 6 PMs) that allow experts from the CLARIN nodes to participate in the development agenda. This contributes to the overall exchange of knowledge and capacity building.
- CLARIN ERIC participates in the acquisition and implementation of funded collaborative projects, with a particular focus on funding opportunities to strengthen the links with other domain-specific RIs in the SSH domain. Such projects often bring room for participation by national nodes as linked third parties. In such cases, CLARIN ERIC is responsible for the coordination and cohesion of the contributions from the national nodes. So far, this model has been adopted in the projects EOSC-hub, EOSC Future, SSHOC and UPSKILLS (see also Section 2.4).
- CLARIN ERIC started a number of initiatives for the exploration of setting up links with parties and communities outside of Europe. As indicated in section 2.3, this internationalisation agenda and the contacts maintained all over the globe have led to several requests to explore suitable modes of collaboration. With the recently improved travel conditions it is to be expected that the exploration of structural international collaboration can be taken up again.

- CLARIN ERIC has been continuously looking for opportunities for CLARIN nodes to contribute to strategic discussions and the visibility of CLARIN in the ESFRI landscape. Important for the visibility of CLARIN also is the participation of BoD members as keynote speakers, teachers and panellists at events such as Meta-Forum, LREC 2022, ESFRI and EOSC meetings. Members of the CLARIN ERIC BoD are often invited to participate in panels and working groups in which the future of the

European landscape for research and RIs is discussed and/or shaped. Franciska de Jong is member of ERIC Forum's first Executive Board, and she represents the SSH cluster in the coordination meetings of the five so-called science clusters (also involving representatives of ESCAPE, EOSC-Life, ENVRI FAIR and PaNOSC). Both Franciska de Jong and Francesca Frontini participate in EOSC Task Forces (the TF for Research Engagement, and the TF for Research Careers, Recognition, and Credit, respectively).

- In order to celebrate CLARIN ERIC's 10-year anniversary in 2022, a book publication was prepared, based on an open call for chapter proposals, edited by Andreas Witt and Darja Fišer. The book will be published in open access by De Gruyter in October 2022.

Funding of Cross-National Activities

CLARIN ERIC's role in stimulating the collaboration across countries is partly taking shape through the support and funding of grass-root initiatives with participants from multiple nodes.

The ParlaMint project, which was launched in 2020 in response to the COVID-19 pandemic, is an important example. It underlines CLARIN ERIC's role in stimulating the collaboration across national teams and has huge potential for impact along multiple dimensions. It will continue to be funded as the CLARIN flagship project until mid-2023. The most recent data release, ParlaMint 2.1, is a multilingual set of around 20 comparable corpora containing parliamentary debates mostly starting in 2015 and extending to mid-2020, with each corpus being about 20 million words in size. (See the ParlaMint project page for more details.) The use of the dataset is stimulated in several ways. In several papers that were submitted for the ParlaCLARIN III workshop held at LREC2022, the use of the ParlaMint data release was reported, and in theHorizon2020 project EOSC Future, the ParlaMint data release is one of the data pillars for a study on how the metadata for life science data on COVID-19 and parliamentary data can be aligned through harmonised ontologies/classifications.

With the travel constraints due to the pandemic, mobility grants and support for face-to-face events was again limited. Support for CLARIN activities therefore concentrated two aspects: support for virtual events organised by the nodes in the network, and support for the further development of resources and collaborations within the infrastructure.
Two main instruments were utilised for this:

- The Support for Virtual Events call: Applicants can ask for advice and hands-on support and everything in between. In the reporting period in total 7 requests have been received that all have led to virtual events run via CLARIN's Zoom account. The support chain is run by the technical team and the communication team.

- The call for [CLARIN Resource Families projects](#) (CRF): In the reporting period, in total 7 proposals for CRF projects were submitted and granted, which can be taken as a strong sign of the vitality of our network.

In the course of 2022, face-to-face events became possible again and event funding was granted to the following workshops (see also Section 2.1):
- CLARIN and Libraries, 9-10 May 2022 ([link](#))
- Teaching with CLARIN@Teaching and Language Corpora, 13 July 2022 ([TaLC](#)).

An overview of all funding and support options can be found on the [funding webpage](#).

3.2   Implementation of the Strategy for 2021-2023

The CLARIN ERIC strategy for 2021-2023 was approved by the General Assembly in November 2020 ([CE-2020-1709](#)). It is the basis for the [Vision and Strategy](#) published on the website.

In the reporting period, the implementation of the Strategy was again somewhat impacted by the COVID-19 dynamics, but in most priority areas the overall progress in the past 12 months has been significant and an inspiring starting point for the work plan of 2023. Priority topics which will require continued attention in the remaining strategy period are:
1. The collection of feedback in order to further enhance the structure and content of the website
2. The further alignment of the CLARIN service offer with that of other platforms in the ecosystem and with the demands emerging from developments in AI
3. The monitoring of evolving stakeholder requirements, including those pertaining to the articulation of an innovation agenda
4. Exploration of the options for stimulating the explicit acknowledgement of the CLARIN infrastructure by users in their publications, as well as more standardised citation of the resources
5. Exploration of suitable frameworks for assessing pathways to socio-economic impact of the CLARIN service offer
6. The reinforcement of the Central Office organisation, for instance through the appointment of a Chief Operations Officer (as of Q4-2022).

3.3   Publication Output and Web Statistics

This subsection concludes with some information on the recent publication output and website visits statistics.

Publications about or References to CLARIN

In order to support the outreach, and to make the publications records easily available for a general audience, a [CLARIN Zotero database](#) was set up in 2016, which was recently restructured.

There have been issues with several methods of gathering bibliographic data on CLARIN-related publications automatically and semi-automatically. In the past year, two promising methods for collecting publications that were applied in 2020 were evaluated:

- First, a broad search on Google Scholar was executed (using the keyword 'CLARIN', minus a set of search terms closely connected to occurrences of irrelevant 'other' CLARIN-named entities) and the results were stored in Zotero. For 2020, the number of results (1440) was again higher than the artificial result limit of 1000 entries. Moreover, many of the hits included also uncited entries from CLARIN repositories that Google Scholar picked up, leading to a biased overview. Therefore, it was decided not to proceed with this method.
- Second, the good recall from Google Scholar was combined with the precision delivered by searching for the common parts of persistent identifiers as issued by CLARIN centres. While this leads to fewer results (672 for 2020, measured in July 2021), cursory inspection of the results indicates that this approach is more accurate in terms of precision. Based on these findings, CLARIN will further fine-tune this reporting method, together with the centres and the consortia, in order to reach a stable and reliable performance indicator for publications that cite CLARIN language resources with a persistent identifier.
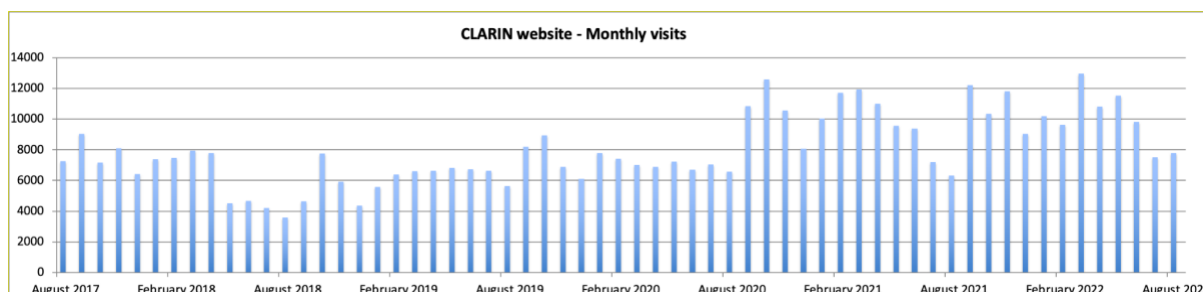
Some highlighted publications for 2021-2022

- Andreas Witt & Darja Fiser (eds.) CLARIN. The infrastructure for language resources, De Gruyter (2022, to appear)
- Jennifer K. Mendoza & Caitlin M. Fausey, Everyday music in infancy. In: *Developmental Science* (2021). DOI: https://doi.org/10.1111/desc.13122
- Laurent Kevers, Stella Retali Medori, A. Ghjacumina Tognotti. A Survey of Language Technologies Resources and Tools for Corsican. Research Report, UMR 6240 CNRS LISA - Université de Corse (2021). DOI: hal-03228733
- Noga Zaslavsky, Mora Maldonado, Jennifer Culbertson. Let's talk (efficiently) about us: Person systems achieve near-optimal compression. In: *Proceedings of the Meeting of the Cognitive Science Society* (2021). DOI: https://escholarship.org/uc/item/2sj4t8m3
- Carrion-Castillo, A., Estruch, S.B., Maassen, B. *et al.* Whole-genome sequencing identifies functional noncoding variation in *SEMA3C* that cosegregates with dyslexia in a multigenerational family. In: *Hum Genet* 140, 1183–1200 (2021). DOI: https://doi.org/10.1007/s00439-021-02289-w
- Michael Adjeisah, Guohua Liu, Douglas Omwenga Nyabuga,Richard Nuetey Nortey,and Jinling Song. Pseudotext Injection and Advance Filtering of Low-Resource Corpus for Neural Machine Translation. In: *Computational Intelligence and Neuroscience* (2021). DOI: https://doi.org/10.1155/2021/6682385
- Bharathi Raja Chakravarthi, Mihaela Gaman, Radu Ionescu *et al.* Findings of the VarDial Evaluation Campaign 2021. In: *Proceedings of the 8th VarDial Workshop on NLP for Similar Languages, Varieties and Dialects* (2021). DOI: http://hdl.handle.net/10138/329644
- Fitsilis, F., & Mikros, G. Development and Validation of a Corpus of Written Parliamentary Questions in the Hellenic Parliament. In: *Journal of Open Humanities Data*, *7*, 18, (2021). DOI: http://doi.org/10.5334/johd.45
- T. Erjavec, M. Ogrodniczuk, P. Osenova *et al*. *The ParlaMint corpora of parliamentary proc*eedings. In: *Language Resources and Evaluation* (2022). DOI: https://link.springer.com/article/10.1007/s10579-021-09574-0
- 17 peer-reviewed publications referring to CLARIN resources in the proceedings of LREC 2022, Marseille. See the CLARIN Zotero page: https://www.zotero.org/groups/562080/clarin/collections/M7Y9TFPE/items/VJ22WJJS/collection
- 20 peer-reviewed peer-reviewed publications in the proceedings of LREC 2022 workshop ParlaCLARIN III. See the ParlaCLARIN Zotero page: https://www.zotero.org/groups/562080/clarin/collections/F4BYDMKB/items/9XGJ74KS/collection
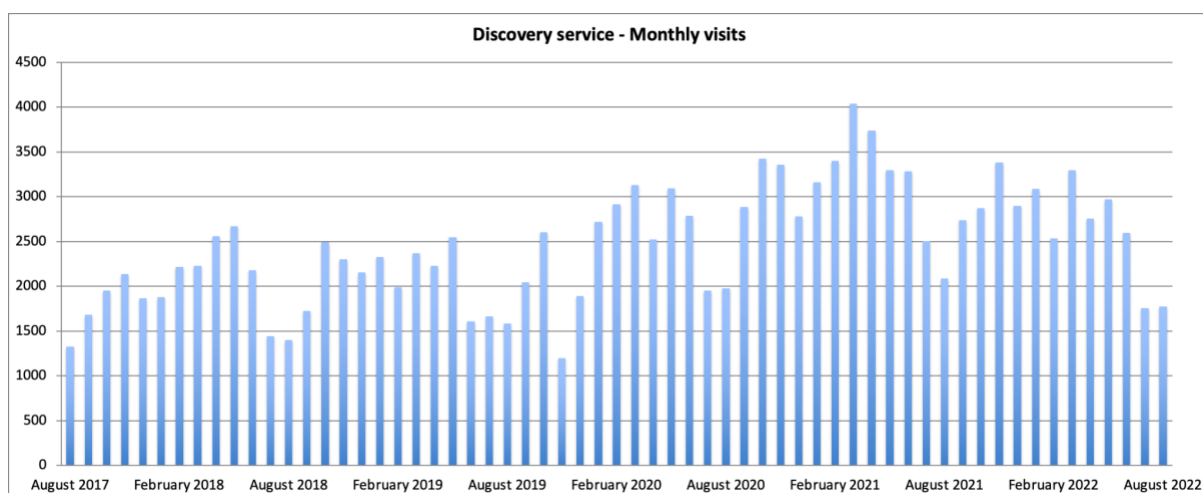
Visits to the Central CLARIN Website

The three charts below display the monthly number of visits to the CLARIN ERIC website and the two most popular central services: the CLARIN Discovery Service and the Virtual Language Observatory. The charts show the following:
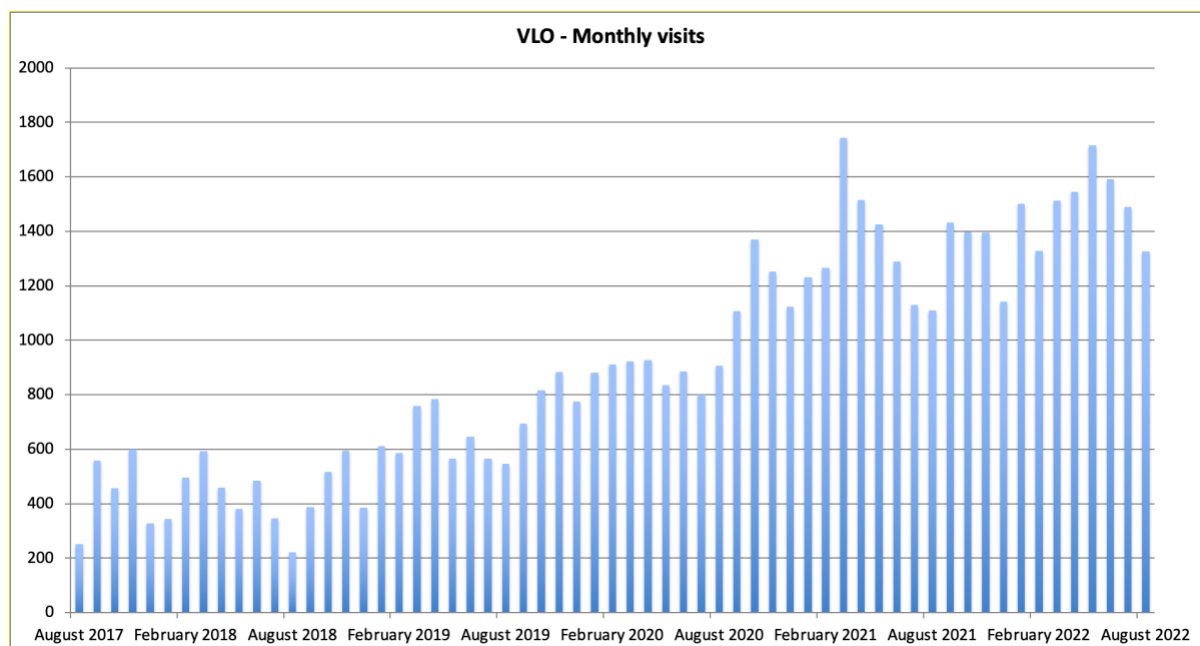

Since August 2021, the number of visits to the CLARIN website has grown by 3.7%

There is a decrease of 14% in the number of visits to the Discovery Service (used during single sign-on to CLARIN service providers). This might be a normalisation after the substantial growth of visits (31.7%) in the previous reporting period during the peaks of the COVID-19 pandemic in 2020/2021.

Since August 2021, the number of Virtual Language Observatory visits has increased by 11.7%. The peaks in May and June 2022 might be related to the tutorial for how to apply Jupyter Notebooks for processing textual resources from Europeana, which was making use of the VLO.

KPI – Average number of website visits per month

September 2018: 78339 visits to the CLARIN website over the previous twelve months. Average per month: 6528

September 2019: 73742 visits to the CLARIN website over the previous twelve months. Average per month: 6145

September 2020: 86745 visits to the CLARIN website over the previous twelve months. Average per month: 7229

September 2021: 119200 visits to the CLARIN website over the previous twelve months. Average per month: 9933

September 2022: **123604 visits to the CLARIN website over the previous twelve months. Average per month: 10300**

4 Concluding Remarks

CLARIN ERIC's results and achievements in the past year demonstrate that technically and organisationally the research infrastructure is in good shape. The local and central CLARIN services that are up and running have grown in number and the scope of the uptake activities and the diversity of audiences has increased significantly. The outreach activities bring vital incentives for use-inspired improvements of the resources and tools. The inevitable switch to virtual meeting and presentation formats has yielded an unanticipated level of understanding of how online interaction can be put to use for the communication with representatives of the CLARIN community more widely and for the outreach to new communities of use.

The emergence of digital technologies and the datafication of society that are transforming all societal and economic sectors bring renewed focus on the nature and role of language in all its aspects and variants, as well as the methodological frameworks that are fuelled by data that is linguistic in nature. This not only enriches the potential scenarios in which CLARIN can be put to use, but also shines a spotlight on the need to support multidisciplinary research agendas, as well as the increased demand for education and training aimed at the development of skills to handle digital data. In several countries, considerable budgets are being made available to facilitate the incorporation of language data as part of AI-based innovation agendas. CLARIN is excellently positioned to play a role in these developments. The incorporation of language models in the CLARIN Resource Familiesand the stimulation of NLP-driven elements in curricula offered to language students are two examples of how CLARIN is contributing to both the technical infrastructure and the knowledge exchange that is needed for the anticipated 'digital transformation'.

The EU projects supporting the further development of CLARIN in various dimensions offer a crucial reinforcement of the consortium's capacity to reach the goals set for the current strategy period, as well as for the longer term. In particular, funding for the exploration and implementation of the various options for collaboration between research infrastructures, both for the SSH cluster as well as RIs from other domains, has helped to select a route for bringing the interoperability paradigms adopted by CLARIN to the next level. Particularly noteworthy is the role of the SSHOC project (2019-2022) in setting an example for how a thematic cluster of RIs can offer discovery tools that fulfil the requirements of service interoperability and contextualisation. The launch of the SSH Open Marketplace can be considered a milestone in the shaping of the RI ecosystem, and the collaboration under the header of SSH Open Cluster is crucial for the next stages of the further integration of the domain-specific federated service offer by RIs and the collaboration at the level of RI clusters into the emerging European Open Science Cloud.