

Title	CLARIN Strategic Planning 2018-2020
Version	2.0
Author(s)	BoD
Date	29-11-2017
Status	Approved
Distribution	BoD, NCF, GA
ID	CE-2017-1041



Contents

1	Introduction	2
1.1	Vision, mission, landscape.....	2
1.2	Where we stand	3
1.3	Where we need to go.....	4
1.4	Structure of the strategy	4
2	Uptake by Researchers	4
2.1	Enhancing outreach.....	5
2.2	Enhancing the usability of services	6
3	Technical Infrastructure	7
3.1	Integration of data and services, interoperability.....	7
3.2	Advanced tools and workflows	8
3.3	Discovery and deposition of data.....	9
3.4	Legal aspects.....	9
4	Knowledge Sharing Infrastructure	10
4.1	Central website.....	10
4.2	Training, workshops, tutorials.....	10
4.3	Knowledge centres	11
4.4	Mobility grants	12
5	Sustainability.....	12
5.1	Governance	13
5.2	Financial perspectives	13
5.3	Membership	15
5.4	Compliance with the ESFRI framework.....	16
5.5	Risk Assessment and Mitigation, KPIs	17

1 Introduction

The first strategic plan for CLARIN ERIC was made in 2012, and the most recent update of this is [CE-2015-0657](#). It is now time to review and renew the strategy planning.

As agreed in 2016 the strategy will be structured along four dimensions: Uptake by Researchers, Technical Infrastructure, Knowledge Sharing Infrastructure and Sustainability. The overall aim is to consolidate what has been achieved and to further develop CLARIN's visibility so that more researchers become aware of the possibilities, to enhance the services for access to language resources and tools, and through a programme for collaboration, support and training. Finally focus will be put on sustainability, organisationally as well as financially. This document presents a strategic plan for three years, starting in 2018.

1.1 Vision, mission, landscape

The CLARIN ERIC vision, mission and value proposition¹ have served as a source of inspiration and guiding principles for the definition of the goals, activities and tasks to be described. In line with the role of CLARIN as an ESFRI research infrastructure for the SSH² domain, the mission and vision are primarily oriented towards the role of CLARIN for research communities and the sectors they collaborate with.

Vision and mission

Currently the vision and mission is expressed as follows:

Vision: "All digital language resources and tools from all over Europe and beyond are accessible through a single sign-on environment for the support of researchers in the humanities and social sciences". Mission: "Create, and maintain an infrastructure to support the sharing, use and sustainability of language data and tools for research in the humanities and social sciences."

We will keep monitoring CLARIN's role in the overall landscape and the best way of articulating and communicating the vision and mission to the outside world.

Landscape

The landscape in which CLARIN is positioned and on which it is supposed to have impact is a crucial factor, both for the planning and for the measurement of success of the activities undertaken.

Disciplinary orientation

¹ The CLARIN value proposition (CLARIN-PLUS deliverable [D5.4](#), 2016) describes in detail the current offer of CLARIN. A public version of the report has been published in 2017 (<https://www.clarin.eu/value-proposition>).

² Throughout this document SSH will be used as the abbreviation for the research domain of the humanities and the social sciences, in line with the terminology used in the [ESFRI Roadmap 2016](#).

CLARIN stimulates and supports the adaptation, enhancement, reuse and repurposing of available language data, thereby enabling scholars in the (digital) humanities and social sciences and beyond to open new research avenues within and across disciplines that address one or more of the multiple societal roles of language. Language is an important carrier of cultural content and information, both synchronically and diachronically, but it also plays a role as a reflection of scientific and societal knowledge, as an instrument for human communication, as one of the central components of the identity of individuals, groups, cultures or nations, as an instrument for studying human cognition and expression, and as an object of study and preservation. Scholarly communities that have research agendas aiming at a deeper understanding and processing of the multiple layers and modalities in language data can benefit from the longstanding expertise and resources built up in the research domains that have been brought together in CLARIN, including computational linguistics, digital humanities, language and speech technology, cognitive science, artificial intelligence and computer science.

Open Science

CLARIN contributes to realizing the vision that underlies the emerging European policies towards Open Science, interconnecting researchers across national and disciplinary borders by offering seamless access to data and services in line with the FAIR and Responsible Data Science principles. This contribution is coloured by the particular characteristics of CLARIN as (i) a distributed infrastructure rather than as a stand-alone facility and (b) an infrastructure for a domain with a wide diversity in research cultures.

Stakeholders

The construction and operation of an infrastructure involves many different stakeholders, each with their own interests and expectations, ranging from individual researchers, research institutions, data archives, citizen scientists, infrastructure service providers, funding bodies and governments, to sectors that are not primarily academic, such as the data industry, the GLAM sector (galleries, libraries, archives, and museums) – just to mention a few. They are all taken into account in the CLARIN strategy.

1.2 Where we stand

Since the establishment of CLARIN as ERIC in 2012, the participating countries have developed a model for the integration of online services, the interoperability of data and tools, and cooperation across borders. Today CLARIN is a highly distributed infrastructure with over 40 centres in more than 20 countries, with a central service platform that gives access to language resources and analysis tools developed in member countries via single sign-on, and with a growing user base in a variety of scholarly domains and other communities of use. Each CLARIN ERIC member is responsible for its own national part of

the infrastructure and has, to some extent, its own national agenda for what needs to be achieved and how. At the same time, the national nodes share their data, services and expertise with the other members and with the central level. All this has become feasible due to the careful international coordination and a professional central organisation.

1.3 Where we need to go

The strategic goals for the period 2018-2020 reflect what is seen as most urgent for the future stages:

- (i) to stimulate further uptake of CLARIN by developing and implementing optimized models for outreach, to reach a better understanding of what researchers need and to focus on the functionality they need to have at their fingertips and the knowledge they are looking for,
- (ii) to extend the coverage and scope of what is currently operational both technically and in terms of the instruments for knowledge sharing, and to improve the robustness and the match with user requirements of the language technologies offered,
- (iii) to ensure the long-term sustainability of the integrated assets.

Where relevant the strategic plan differentiates between the elements that are to be developed centrally and the ones that are taken up in the national consortia.

1.4 Structure of the strategy

As agreed in 2016 and stated in the introduction, the structure for the strategy 2018-2020 comprises four priority areas, each addressed in a separate section with subsections where relevant:

- Uptake by Researchers
- Technical Infrastructure
- Knowledge Sharing Infrastructure
- Sustainability

Each (sub)section follows this structure:

- Analysis of the current stage
- Objectives that would address the needs and gaps
- Strategy towards the objectives

2 Uptake by Researchers

CLARIN aims to offer a central entry point for researchers interested in language resources and technologies and stimulate CLARIN-wide uptake activities. In addition, the members stimulate uptake in their countries and through their centres.

2.1 Enhancing outreach

Analysis of the current stage

With the local and central CLARIN services up and running, these services are widely used by researchers in linguistics and language technology, and, in most countries, by scholars in various subdisciplines of digital humanities and social sciences. Moreover, the potential for extended outreach is in place. However, recent events organized to collect feedback from researchers have demonstrated that several prospective communities of use require good explanation of the infrastructure capacities and optimization of the visibility of CLARIN resources, in order for the CLARIN potential to fully materialize in their fields. Furthermore, since CLARIN is for the most part an open and distributed infrastructure, CLARIN has limited central knowledge of which users access which items and for what purpose.

A number of disciplines have been identified that they could benefit more from CLARIN's infrastructure than they currently do: Literary Studies, History, Journalism and Media Studies, Ethnography and Anthropology, Migration studies, Political Studies, Sociology and Psychology. CLARIN has different audiences or types of users that need to be approached and catered for differently: researchers, students, educators, developers, authorities. Not surprisingly, CLARIN offers easy access to researchers from the Natural Language Processing (NLP) and corpus linguistics community, but with the emerging new data-intensive research paradigms and with the potential for more impact-driven research in multidisciplinary settings the scholarly needs keep evolving and require careful monitoring. Many research fields could benefit from a good overview of datasets on specific topics that are available in multiple languages. While it is a typical feature of the European Research Area that comparable datasets exist in multiple regions, the current search and analytical functionality could be optimized for the support of comparative research across linguistic, national, cultural and temporal borders.

Objectives

- a. Deeper understanding of the characteristics and goals of researchers who are using the infrastructure.
- b. Wider outreach to user communities in the (digital) humanities and social sciences through existing international networks, associations and fora.
- c. Insight into requirements that would enable researchers to integrate the use of CLARIN resources in the emerging multidisciplinary research agendas and methodological frameworks.
- d. Increased awareness in a wider range of audiences of what CLARIN has on offer.

Strategy

- a. Identify user requirements for accessing the services, tools and resources within the infrastructure and monitor the user experience through surveys and focus groups with potential new user communities.

- b. Promotion of any improved functionality realized in response to user feedback.
- c. Develop better statistics of CLARIN usage by user groups along disciplinary and other dimensions.
- d. Foster trans-national collaboration and knowledge sharing among researchers working on the development and analysis of resources and tools for data types relevant for a wide range of disciplines in (digital) humanities and social sciences.
- e. Set up an advisory structure that can improve the potential for effective outreach and impact in the social sciences.
- f. Arrange presence at the key scientific, educational and promotional events in the relevant disciplines.

2.2 Enhancing the usability of services

Analysis of current stage

Early successes with setting up the infrastructure need to be followed up with steps to make it more user-friendly. The current availability and maturity of the infrastructure varies greatly across languages and domains. Existing metadata for the tools and resources do not always cover all the details needed for specific research tasks.

Objectives

- a. A unified search portal that gives access to all the resources and tools available within the infrastructure.
- b. An adequate communication channel between communities of use and the developers of CLARIN services.
- c. A basic set of tools and resources for all languages in the CLARIN infrastructure.
- d. Explicit support of the scientific reproducibility and results reliability by ensuring the transparency of quality and characteristics of the tools in the infrastructure.
- e. Improved visibility of the researchers that contribute to and/or use the infrastructure, as well of their resources and tools, and the institutions they work from.

Strategy

- a. Evaluate the capability of the infrastructure for particular scenarios of use.
- b. Monitor the model for communicating user requirements and experiences between communities of use and the SCCTC. Create a high-level task-based inventory of datasets and tools that are readily available for online usage.
- c. Identify existing resources and tools suited for integration in the infrastructure in the near future.
- d. Advertise success stories (impact-factor publications, major awards) based on use of the infrastructure.

3 Technical Infrastructure

Over the past years, CLARIN has constructed a sound and stable technical basis to support the sharing of language data and tools across institutional, disciplinary and international borders. Besides enriching and strengthening this infrastructure, the time is now ripe to work towards a next level of deepened interoperability within the CLARIN ecosystem. This requires efforts on the side of the tool and the data providers, which will be closely coordinated with the National Coordinator's Forum and the Standing Committee for CLARIN Technical Centres.

3.1 Integration of data and services, interoperability

Analysis of current stage

A considerable collection of language data and tools exists, spread over hundreds of institutions all over Europe. When these are provided through certified CLARIN centres, the findability, accessibility, interoperability and reusability (FAIR principles³) is guaranteed. Hence CLARIN strives to maintain the current strong network of centres and to set-up new CLARIN centres – each of them providing high-quality data sets and tools, with equally high-quality metadata.

Objectives

- a. All relevant datasets and tools are described with high-quality CMDI metadata and can therefore be found via the Virtual Language Observatory.
- b. Single sign-on for direct access or requests for access to datasets or tools that are password-protected.
- c. Clear recommendations are available as to the choice of standards for datasets.
- d. A single web interface through which all openly available language resources can be queried.
- e. Dedicated support for projects that come with requirements related to multilinguality.

Strategy

- a. Each CLARIN centre provides metadata for its data and tools. CLARIN ERIC closely monitors the quality of the provided metadata and organizes detailed feedback and targeted curation campaigns, in close collaboration with the national consortia. Activities include consultation with technical experts, the SCCTC and representatives of communities of use (see also section 2).

³ See <https://www.force11.org/fairprinciples>

- b. Password-protected data and tools that rely on a local username and password are connected to the CLARIN single-sign on system for user-friendly authentication. The authorization remains in the hands of the hosting centre.
- c. An inventory of standards and formats in use at CLARIN centres will be matched with the recommendations from the standards committee, as to provide a good overview of recommended practices. This overview will be updated over time.
- d. All suitable corpus search engines at the CLARIN centres will be connected to the Federated Content Search by setting up a compliant endpoint.

3.2 Advanced tools and workflows

Analysis of current stage

There is a wide variety available of tools for natural language processing. Many of them currently work in isolation. It is often hard as a researcher to find the right tool to process a certain dataset. For a provider of a tool it is not always clear what is the optimal way to integrate it into CLARIN.

Objectives

- a. High-quality data/tool matching for web applications.
- b. Improved interoperability between web services.
- c. Improved quality of language processing tools in terms of performance and robustness in the perspective of their application in research scenarios.
- d. Increased usability of web services by non-technical researchers.
- e. Clear guidelines for developers on how to make tools CLARIN-compatible.

Strategy

- a. Focus on the low-hanging fruit in terms of tool interoperability by connecting web applications in CLARIN centres to the Language Resource Switchboard. Provide documentation, tutorials and hands-on support to make CLARIN web applications (including full workflow systems) compatible with the Language Resource Switchboard. Investigate the existing workflow applications and select the most mature systems. Stimulate connecting web services to these systems. Require compatibility of the workflow web application with the Language Resource Switchboard.
- b. Promote connecting existing web services to at least one prioritized workflow system.
- c. Give maximal visibility to CLARIN tools that are deeply integrated, via single sign-on, the LR Switchboard and/or the chosen workflow systems.
- d. Set up a coordination structure for the resources and tools based on feedback from communities of use.

3.3 Discovery and deposition of data

Analysis of current stage

Over 20 CLARIN centres are providing depositing services in 13 CLARIN member or observer countries; one third party also provides depositing service. Most of these are fully certified by CLARIN and the Data Seal of Approval. While providing a Data Management Plan (DMP) is becoming more and more a requirement to acquire funding for scientific research, there is still a lot of variation in the nature of such DMPs. Moreover, despite DMPs being in place, many research data sets are still not deposited in trusted repositories with a long-term perspective.

Objectives

CLARIN-compliant repositories for the depositing of all language data emerging from publicly funded projects in member countries that remain accessible for researchers and referable through Persistent Identifiers.

Strategy

- a. Provide at least one CLARIN-compatible depositing service for each member country.
- b. Promote the depositing of research language data at CLARIN repositories.
- c. Benefit from the participation in H2020 cluster project PARTHENOS in order to work towards widely shared standards for DMPs.

3.4 Legal aspects

Analysis of current stage

Currently CLARIN has a Legal Issues Committee (CLIC) that is responsible for the guidance/expertise on the matters of IPR (Intellectual Property Right) issues and licenses, privacy protection, and ethical issues. CLIC is actively involved in activities and discussions related to the EU Copyright reform, supporting Open Access (OA) and Open Science.

Objectives

- a. Compliance of language data and tools that are shared via CLARIN infrastructure with the technical requirements and standards, and clarity about legal status.
- b. User understanding on the legal rules and implications regarding text and data mining across Europe.

Strategy

- a. Encourage data and tools providers to use OA licenses, if applicable.

- b. Providing legal consultation/guidance, and general instructions/best practices to the community members via CLARIN Legal Issues Committee on request.

4 Knowledge Sharing Infrastructure

The Knowledge Sharing Infrastructure is meant to work as 'glue': it consists of a set of measures and facilities aimed at securing the continuous transfer of knowledge between the main players involved in the construction, operation and use of the infrastructure. Below we describe four of the main components of the knowledge sharing infrastructure.

4.1 Central website

Analysis of current stage

The structure and look & feel of the current website is the result of a redesign and implementation process that is still ongoing. There is room for further optimization through the use of visual formats and the continuous monitoring of the content and the user experience.

Objectives

High-quality information (explanation, showcases, demos) is offered about the organization of the infrastructure, available resources and tools, and the activities and events within CLARIN. Improved user experience, increased satisfaction for the search and discovery of information, all with attention for the diversity in the communities of use.

Strategy

- a. Invest in the CLARIN website as a convincing information platform to demonstrate the nature, quality and applicability of all tools available via CLARIN.
- b. For the information provision a diversity of formats will be deployed: ranging from reports, best practice papers, FAQ, and blogpost from relative outsiders, to video recordings and screen casts.
- c. Wherever feasible/relevant information provision will be enhanced with graphics, clips, etc. Experiments with the introduction of navigation tools that use visuals will be set up.

4.2 Training, workshops, tutorials

Analysis of current stage

CLARIN has been organising various types of workshops, both decentrally and centrally, that have delivered strategic connections (with those who are in charge of building and

operating new CLARIN consortia and centres, and with specific communities of use) and insight into the effectiveness of a range of formats. Some events have been filmed and recordings are now available through the newly started Videolectures channel. Tutorial materials exist, but not for all key resources and tools in the CLARIN infrastructure. Surveys with out-of-domain experts have pointed to the need for more guidance for using the tools and resources and best practices for various scenarios of use. The possibility to disseminate information through the educational system has not been thoroughly explored.

Objectives

- a. A series of CLARIN central workshops, well balanced with the events organized by national consortia.
- b. Coverage of all relevant themes by one or more tutorials by 2020 and where feasible with recordings that are available on the website.
- c. Develop training material (user guides, video tutorials, showcases, exercise materials) for the key components of the CLARIN infrastructure.
- d. Increased awareness about the availability of advanced CLARIN tools for human language processing.

Strategy

- a. Build on already existing workshop formats, e.g., for newcomers and the thematic workshops on special data types. Investment in training will be tuned to the collaboration with DARIAH-EU and the relevant activities in PARTHENOS.
- b. Create a webpage that (i) offers a framework for the sharing of training material and (ii) gives access to all educational materials (including recordings of tutorials) with a dedicated navigation structure.
- c. Develop best practices for enhancing MA and PhD programmes to teach about the digital research methods needed to effectively use the CLARIN tools and resources.
- d. Work towards integration of CLARIN-related topics into university curricula.

4.3 Knowledge centres

Knowledge centres can be established for a diverse range of topics, either based on language (“the K-centre for the Danish language”), modality (“the K-centre for speech analysis”), domain (“the K-centre for Dutch historical texts”), technology (“the virtual K-centre for machine learning”, “the K-centre for OCR”), data types (“the K-centre for treebanking”), etc.

Analysis of current stage

There are 8 acknowledged K-centres.

Objective

A K-centre for each of the relevant themes at the end of 2020.

Strategy

Identification of relevant themes and the priorities will be continued in consultation with NCF and other stakeholders. Possible hosts will be approached and pointed to the possibility for funded workshop to prepare the preparatory steps.

4.4 Mobility grants

Mobility grants consist of funding that allows researchers and technical staff working at a CLARIN institute to visit another institute, typically for up to 1 week, to do research, give instruction or be instructed in the use, construction or operation of the CLARIN infrastructure. This instrument enhances the integration within CLARIN at the human resources level, and thus reduces the risk of creating a two-speed CLARIN.

Analysis of current stage

Application procedure for mobility grants are in place, but the funds are not exhausted.

Objectives

Increased awareness, and consequently higher use of mobility grants.

Strategy

More active promotion of the mobility grants.

5 Sustainability

Given the stage of maturity that CLARIN has reached, attention for consolidation and sustainability along various axes has become of vital importance. Of course the reuse of resources available and the uptake of the services offered are vital, and the strategies described above are thus crucial pillars for the sustainability of CLARIN. In addition to issues of governance, membership and finances, the models of coordination and collaboration need to be adequate for securing the infrastructure in place. Special attention is needed for the collaboration with the institutes responsible for the maintenance of data repositories that represent the national heritage, as well as alignment with the dynamics in the ESFRI landscape and the expectations regarding societal impact.

5.1 Governance

Analysis of current stage

A review of the statutes from 2012 has been the basis for a proposal for revision that will be ready for approval by the end of 2017. With the growth of the consortium the need for diversity in the implementation of the relation between the individual consortia and CLARIN ERIC is becoming apparent. In consultation with the NCF a review of the model of the CLARIN agreements has been initiated.

Objective

A more effective governance underpinned by revised statutes and an implementation of the governance that take into account the need for diversity across countries.

Strategy

The analysis of the statutes, comparison with other ERIC statutes, and consultation of relevant bodies within CLARIN will soon be finalized. After discussion with the EC revised statutes will be submitted to the GA and the Commission for approval and publication.

A new model for CLARIN agreements that addresses the diversity in requirements across countries will be proposed, refined and implemented.

5.2 Financial perspectives

National funding

Analysis of current stage

The funding for the national activities varies from country to country. In some of the member countries there is lack of local funding, or the continuity of funding is uncertain. The sustainability implications that may follow from such a situation differ according to the stage of infrastructural development. If no substantial development has been made available at all, the main consequence will be a lack of perspective on the integration of language resources in the CLARIN RI.

If the existing level of funding for a CLARIN country decreases, the results of earlier investments that are critical for the RI would need to be secured. This pertains to both the national infrastructure and the wider CLARIN infrastructure.

Objectives

- a. Support for national teams in identifying possibilities for the acquisition of funding for activities at the national level that contribute to the integration of (aspects of) the country's infrastructure into the CLARIN RI;

- b. In countries that have to face budget cuts: a fall back option for crucial components of the national research infrastructure that can help to sustain the results obtained. This can either be sought within the country or elsewhere in the consortium.

Strategy

- a. CLARIN ERIC will initiate/participate in proposals for funding that give room for contributions that strengthen the national research infrastructures
- b. CLARIN ERIC will develop an 'adoption strategy': a model for shared responsibility for data and services that are considered crucial for the overall CLARIN mission. The model should outline how and when (a selection of) services and data repositories in a specific country could be archived, duplicated or transferred to centres/institutions in other countries with more secure funding, to mitigate the impact of potential budget cuts. This includes the adoption of 'orphaned' resources, tools or services which are no longer (or have never been) properly curated and made accessible.

Central funding

Analysis of current stage

Since the establishment of CLARIN ERIC the income based on fees from participating countries has been growing continuously. Most likely the income will remain stable in the coming years, but some countries have reported that in the longer run changes in the local support for sustainable funding have to be anticipated.

CLARIN ERIC will never run out of funding at once. The members always allocate their contributions for a number of years, typically 3-5 years, sometimes with a gap in between two periods. As a result, a drastic budget reduction would become evident with a 2 to 3 years notice period. This will give a sufficient amount of time to develop, negotiate alternatives for securing the research data and services and/or downscale the central activities.

Objective

- a. Alliances with organizations that may be beneficial for the acquisition of income on top of the membership fees.
- b. Insight in the most critical aspects of a number of scenarios that could occur in order to be able to adjust effectively to a drastic budget reduction.

Strategy

CLARIN ERIC will engage with/initiate consortia that are eligible for applying for funding.

As part of the scenario-based approach CLARIN ERIC will develop:

- a. an overview of governance issues that are to be addressed if the central funding would decrease below a certain minimum;

- b. a minimal technical maintenance program that would make it feasible to sustain the backbone of the existing infrastructure;
- c. an overview of partners (other RIs, ERICs, GLAM parties, etc.) that are willing to participate in a joint exploration of services that could be shared in order to help to sustain aspects of an infrastructure.

5.3 Membership

New members

Analysis of current stage

CLARIN ERIC currently has 19 members (18 countries and 1 international organization), two observers and one third party. The size of the consortium is adequate in terms of coverage of language resources and expertise, and the of volume of the basic funding, but in view of the European added value it is considered desirable that all EU countries are formally linked to CLARIN. The CLARIN website contains an information page for newcomers and potential new members.

Goal

CLARIN aims to extend its membership and would welcome non member countries who want to contribute. CLARIN also wants to attract more third parties.

Strategy

The BoD will continue the communication with candidate countries. This website section with information for interested countries will be further developed. Specific events will be organised for newcomers/potential members. The annual budget caters for at least one workshop and for mobility grants.

Membership continuation

Analysis of current stage

CLARIN ERIC currently has 19 members (18 countries and 1 international organization), two observers and one third party. None of the members have discontinued their membership, while observer status has not been discontinued except by opting for full membership.

Objectives

CLARIN ERIC aims at retaining all its existing members. Less observers should imply that an observer country has opted for membership.

Strategy

CLARIN ERIC will continue to provide and extend its core infrastructure services to members and observers. The CLARIN Value Proposition is meant to draw attention to the benefits of CLARIN membership. The CLARIN agreement template will be revised as to make the expectations regarding the contributions from members clearer and to establish an agreement procedure that better adjusted to the diversity of members' and observers' local policy context.

5.4 Compliance with the ESFRI framework

Analysis of current stage

The position of CLARIN ERIC as Landmark on the ESFRI Roadmap 2016⁴ comes with expectations and criteria for compliance. The ESFRI framework for monitoring the Landmark initiatives is currently under development. CLARIN has been anticipating this monitoring process and is preparing for an assessment that is likely to cover at least the following aspects: collaboration across borders, training and education of those who can use and/or shape RIs, visible societal impact.

Objectives and activities planned

Coordination of the collaboration across borders

- a. Coordination of collaboration across sectors. In the case of CLARIN this implies reinforcement of the potential for collaboration with the GLAM sector (Galleries, Libraries, Archives, Museums) and industry, as well as international collaboration.
- b. Contribution towards the establishment of an academic reward framework that merits contributions to infrastructure and the use of it in cooperation with EU and national research agencies.
- c. Integration and support of Europe's policies towards Open Science.
- d. Cooperation with other RIs, in particular the ones serving SSH communities, such as DARIAH and CESSDA
- e. Cooperation with European and global initiatives for data sharing and virtualization.

Training and education of those who can use and/or shape RIs

In section 4.2 above the objectives and approach towards the training of current and future users of the CLARIN RI have been outlined. An additional objective is to increase the options for the training of analytics skills for SSH students for working with language data and to raise the interest and skill level among developers to engage in the maintenance and future development of infrastructural initiatives for language resources. This will be taken up in close collaboration with other ERICs.

⁴ Cf. <http://www.esfri.eu/roadmap-2016>

Visible societal impact⁵

The societal impact of research based on language data is ubiquitous and goes back to ancient times. Language learning is just one example of an educational application domain with societal impact. Language models for text mining applications, speech recognition, machine translation, spelling support, etc. have become commodities nowadays, and further societal impact can be expected from the collaboration with artificial intelligence, cognitive science, and data science at large. With the contemporary volume of computer-mediated communication, language can now also be claimed to be a data type (big or not so big) that captures cultural and societal phenomena. In the collaboration with various SSH domains, e.g., political science, psychology, memory studies, migration studies, a clear demand is emerging for language data analytics that offer results that can help citizens, policy makers and other professionals to take the step from (big) data to (big) decisions. Analysis results need to be explainable for humans⁶ and the analysis steps cannot be offered as a black box. These observations form the starting point for the CLARIN strategy aimed at increasing the visibility of the potential for societal impact and at developing partnerships with the data science sector. The objectives are:

- a. CLARIN language resources integrated in the innovative frameworks for language education.
- b. Contributions to the development of novel methodological frameworks that can be deployed for the analysis of societal and cultural phenomena in a responsible manner.
- c. Insight in the potential for partnerships with data science companies.

In addition to the already ongoing effort towards increasing the potential for impact, in the period 2018-2020 the following activities are foreseen:

- a. Supporting existing and emerging collaborations aimed at using digital language resources for language learning.
- b. Collaboration with leading representatives of the social sciences with the aim to develop models for how to integrate heterogeneous data types and conceptual frameworks that can help to validate the outcomes of research in the social sciences that is (partly) based on language data.
- c. Widening of the local networks of societal and industrial partners and involving them in the design of student projects.

5.5 Risk Assessment and Mitigation, KPIs

This part of the planning is will be elaborated in 2018. It will be based on:

⁵ For the SSH domains, ESFRI has recently adopted 'Societal and Cultural Innovation' as a guiding concept for impact assessment.

⁶ In the context of machine learning this is referred to as 'explainable AI' (XAI).

- a. An update of the risk section of CE-2015-0657
- b. Input from the CLARIN-PLUS deliverables D4.1 Report on risk management for e-Infrastructures and D6.1, a report on financial and organisational sustainability, including a section on Key Performance Measurement (to be finalized in December 2017).