

State of the Infrastructure

Bente Maegaard, Franciska de Jong

Vice Executive Director / Executive Director

9 October 2018



Overview

Strategic planning 2018 - 2020

Achievements in 2018

Performance measurement

Governance



ANNUAL CONFERENCE 2018

Pisa, Italy

Strategic Planning 2018-2020

Four priority areas

- Uptake by researchers
- Technical infrastructure
- Knowledge Sharing Infrastructure
- Sustainability

CLARIN's Strategic Planning webpage:

<https://www.clarin.eu/content/vision-and-strategy>

Uptake by researchers

Where we stand

- The communities supported by CLARIN: researchers, students, educators, developers, policy makers, citizen-scientists.
- Current availability and maturity of the infrastructural support varies greatly across languages and domains in (digital) humanities, social sciences, and beyond.
- For various research scenarios specific workflows are available.

Uptake by researchers

Where we want to go

- User requirements and user experience:
 - Further insight into requirements that would enable researchers to integrate the use of CLARIN resources in the emerging Advertise **success stories** based on use of the infrastructure and methodological frameworks.
 - Create a high-level **task-based** inventory of datasets and tools.
 - Expand the detailed overviews and access pages to the **Families of Resources**.
- Promotion within current and new user communities:
 - **Wider outreach** to user communities in the (digital) humanities, social sciences, and beyond through existing international networks, associations and fora.
 - Foster trans-national collaboration and knowledge sharing among researchers working resources and tools for data types relevant for a **wide range of disciplines**.
 - Announce **improved functionality** realized in response to user feedback
 - Advertise **success stories** based on use of the infrastructure and **support visibility** of researchers that contribute.
- Monitoring and evaluation:
 - Monitor the **user experience** through surveys with potential **new user** communities.
 - Explicit support for **scientific reproducibility** and **results reliability** by ensuring the transparency of tools on offer, in line with the Open Science agenda.

Uptake by researchers

Achievements in 2018 (highlights)

- The **CLARIN Resource Families** have become a salient thread in the outreach to new communities. It has inspired work on requirements and on the visibility of data sets.
- **Tour de CLARIN** has become a flagship activity for promoting the national consortia, their resources, tools, events and researchers within and outside the CLARIN network.
- At least one **UI event** has been organised in each of the CLARIN member countries (except in Croatia which joined in March 2018 only), and even in France (an observer country



Technical infrastructure

Where we stand

- Adherence to the **FAIR principles** (findability, accessibility, interoperability and reusability) is leading in the design of all data services.
- Many tools and workflows have been incorporated into the infrastructure, while there is **potential for further integration**.
- 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.
- On request CLARIN's Legal Issues Committee (**CLIC**) provides guidance/expertise on the matters of Intellectual Property Right issues and licenses, privacy protection, and ethical issues.

Technical infrastructure

Where we want to go

- Working with CLARIN tools:
 - Aim to describe all relevant datasets and tools with high-quality CMDI metadata to optimize their retrieval via the Virtual Language Observatory (**VLO**).
 - All suitable corpus search engines at the CLARIN centres should be connected to the Federated Content Search (**FCS**) service.
 - **Clear recommendations** are available as to the choice of standards for datasets, as well as clear guidelines for developers on how to make tools CLARIN-compatible.
- Workflows and interoperability
 - Develop and support high-quality **matching functionality** that couples specific data to relevant tools
 - Dedicated support for projects that come with requirements related to **multilinguality**.
 - Increased usability of web services for **non-technical researchers**.
- Access regulations and repositories
 - Provide at least one CLARIN-compatible **depositing service** for each country.
 - Promote **depositing** of data resulting from **public funding**.

Technical infrastructure

Achievements in 2018 (highlights)

NB. Detailed presentation on Day 3 by Dieter Van Uytvanck

Centres, data, services

Regular centre assessments

As planned, 2 assessment rounds were organised. Out of 47 registered CLARIN centres, 21 have been certified as B-centre.

Access to persistent identifier services

CLARIN ERIC became a non-allocating member of the European Persistent Identifier Consortium (EPIC).

Closer integration of data and tools

Achieved in part by connecting more tools to the *Language Resource Switchboard* and by performing related checks and experiments.

Knowledge Sharing Infrastructure

Where we stand

- The CLARIN website is *the* channel for the dissemination of all types of information and audiences.
- There is support for training, networking and outreach, for which grants can be made available: CLARIN for Researchers.
 - Various types of **workshops** to build strategic connections between tools and software developers on the one hand and users on the other hand.
 - A **VideoLectures** channel offers tutorial materials for some of the resources and tools in the CLARIN infrastructure
- A network acknowledged **K-centres**
- **Mobility Grants** can be applied for to enhance the integration within CLARIN at the level of human expertise.



Knowledge Sharing Infrastructure

Where we want to go

- Central website:
 - Take into account the diversity in the communities and enlarge the volume of **visual formats**.
 - Support the promotion of advanced CLARIN tools for human language processing across SSH domain and beyond.
- Training and education:
 - Provide **dedicated training material** (user guides, video tutorials, showcases, exercise materials) for the key components of the CLARIN infrastructure.
 - Develop best practices for enhancing **MA and PhD programmes** to teach about how to use the CLARIN tools and resources, and promote integration of CLARIN-related topics into academic curricula.
 - Extend the use of **mobility grants** and funding schemes for **User Involvement events** and **workshops** to reinforce the connection between the creators of CLARIN tools and data and the various communities of use.

Knowledge Sharing Infrastructure

Achievements in 2018 (highlights)

- Annual conference: more participants, more papers, PhD session
- The number of **K-centres** has grown from 9 to 11; 2 more applications under assessment.
- More promotion for **mobility grants**
- Instruments for *training*: VLO screencast, webinars and  
- A workshop is planned by the end of 2018 on the impact of the new General Data Protection Regulation (**GDPR**) on personal data in the field of language resources and tools. Collaboration with the CLIC committee and invited legal experts.
- **Citable papers** on how CLARIN complies with FAIR and FACT: <https://www.clarin.eu/faq/how-can-i-cite-clarin>

Sustainability

Where we stand

- In 2018 the **revised statutes** entered into force.
- At the end of 2017 CLARIN ERIC had 19 **members** (18 countries and 1 international organisation), **plus** 2 observers and 1 third party.
- The CLARIN website contains an **information page** for newcomers and potential new members.

Sustainability

Where we want to go

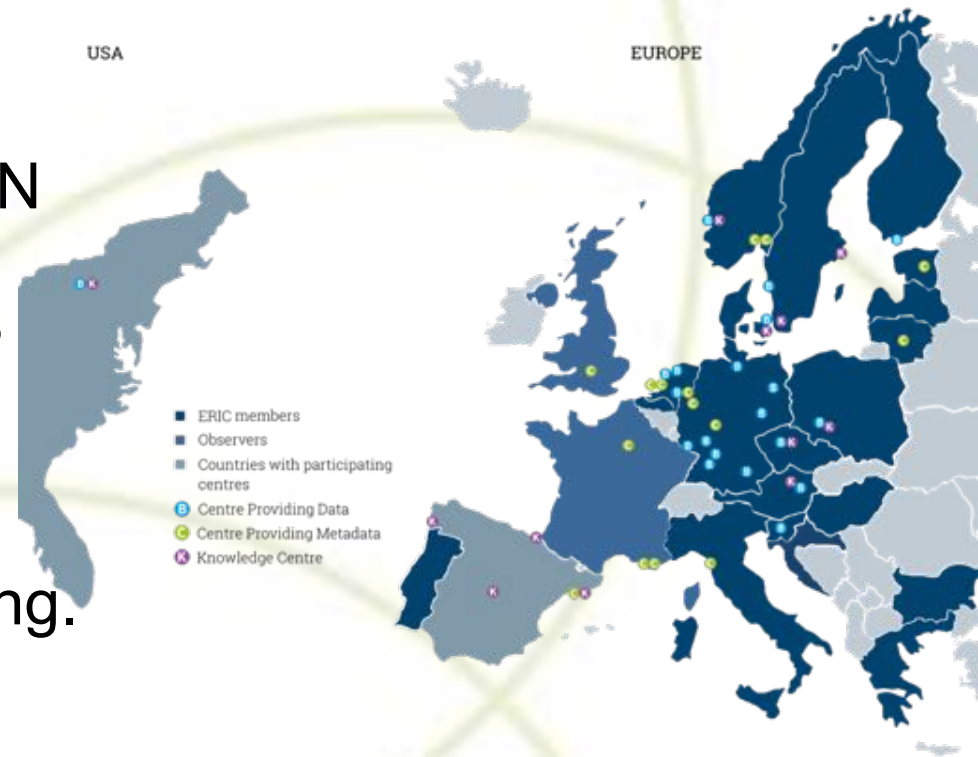
- **Support for national teams** in identifying possibilities for the acquisition of funding for activities at the national level that contribute to the integration of (aspects of) the country's infrastructure into the CLARIN research infrastructure.
- **Alliances** with organisations that may be beneficial for the acquisition of income on top of the membership fees will be further explored and extended, in particular with the aim to reinforce the link with other research infrastructures in SSH and the emerging European Open Science Cloud.
- 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.

Sustainability

Achievements in 2018 (highlights)

Extension of the ERIC membership

Croatia (HR) became the twentieth member of CLARIN ERIC. A workshop is being planned for December 2018 consortia from the countries that recently joined and for representatives of countries that have an interest in joining.



Sustainability

Crossing Borders

- Active contribution to the discussion regarding the EU landscape of RIs and the establishment of European Open Science Cloud (**EOSC**).
- **Current EU projects**
 - PARTHENOS, a thematic cluster around European RIs in the humanities;
 - DSI-4 - Collaboration with Europeana;
 - EOSC-hub that started in January 2018.
- **New EU projects (start in 2019)**
 - ERIC Forum, collaboration with the other ERICs;
 - SSHOC, towards a Social Sciences and Humanities Open Cloud; till April 2022. The CLARIN contribution to SSHOC will be realised by the central organisation and six CLARIN nodes;
 - Room for focus on migration studies, requirements for qualitative data analysis, support for processing interviews and heterogeneous data.



Sustainability

Alliances

- CLARIN maintains strong collaborative ties with sister RIs, in particular with DARIAH-EU
- Agreements aiming at collaboration are in place with:
 - META-NET/Cracker
 - LIBER
 - EUDAT-CDI
 - **new:** ELRA

Sustainability

Performance measurement

- Key Performance Indicators (**KPIs**) – draft proposal (see slides)
- **Visits** to the central CLARIN website (see slides)

Governance

- **New statutes** have been prepared and approved; new model of collaboration between BoD and NCF has been implemented; **SAMBA** established (see slides)

Key Performance Indicators

- Until now CLARIN progress has been measured in relation to the strategic planning and/or workplans.
- But current thinking is that one should rather relate to the objectives of the RI as set out in the statutes.
- In 2017 Steven Krauwer and Bente Maegaard made an overview of all the performance indicators they and anyone else could come up with, 70-80.
- In 2018: careful check if they are candidate KPIs: i.e. checking if they are measurable in a reasonably straightforward way, checking their relationship to the objectives etc, and comparison with approaches adopted by other RIs and related organisations.
- Difference between **Offer** and **Use**. Tried to focus on **Use**. Collaboration is a third aspect.

KPIs

(E indicates that a KPI is mostly applicable at the ERIC level, N at the National level, and E/N at both levels)

- Number of members (E)
- Number of certified B-centres (N)
- Number of publications (written by CLARIN and non-CLARIN people) (N)
- Average number of website visits /month (E)
- Federated login in all countries and for all relevant users (N)
- Collection of standards and mappings (E)
- Number of certified K-centres (N)
- Number of face to face participants in workshops (E/N)
- Number of views of tutorials and videos (E/N)
- Collaboration with RIs (E)
- Global relations, international collaboration (E)
- Good collaboration with libraries, archives, museums, e.g. number of national consortia including at least one member of this type or having collaboration. (N)

KPIs next step

- Has been discussed with the NCF.
- Final version is under development.
- Your NCF representative has a link to the preliminary document if you are interested.
- Fast comments can still be integrated.

NCF Strategy and Management Board (SAMBA)

- Subcommittee of the NCF
- Platform for exchange between the NCF and the BoD
 - For two reasons: size of NCF, NCF chair no longer part of BoD
- Membership, 2 members from each group
 - *Group 1*: Germany, Italy, and The Netherlands. (EH, **JO**)
 - *Group 2 - Western Europe, Scandinavia and Baltic countries*: Denmark, DLU, Estonia, Finland, Latvia, Lithuania, Norway, Sweden. (**KL** (chair), KV)
 - *Group 3 - Central Europe and Southern Europe*: Austria, Bulgaria, Croatia, Czech Republic, Greece, Hungary, Poland, Portugal, and Slovenia. (KS, MP)
- Tasks
 - Prepare discussions and decisions of the NCF in strategic and financial matters
- Meetings: 4 per year

National Coordinators' Forum



NCF Chair **Maciej Piasecki**



NCF Vice-chair **Krister Lindén**

Strategy and Management Board



Erhard Hinrichs



Maciej Piasecki



Krister Lindén



Kiril Simov



Jan Odijk



Kadri Vider

Board of Directors



Franciska de Jong



Dieter Van Uytvanck



Bente Maegaard



Darja Fišer

General Assembly



Ursula Brustmann, *president*

see you @

bmaegaard@hum.ku.dk

f.m.g.dejong@uu.nl

