



How we shaped CLARIN ERIC

Steven Krauwer
CLARIN ERIC Executive Director
Utrecht University
s.krauwer@uu.nl



- Common Language Resources and Technology Infrastructure (<http://www.clarin.eu>)
- Basic idea:
 - European federation of digital repositories with language data and tools (text, speech, multimodal, gesture ...)
 - with access to language and speech technology tools through web services to retrieve, manipulate, enhance, explore and exploit data
 - with uniform single sign-on access to archives and tools
 - target audience humanities and social sciences (HSS) scholars (i.e. not industry – at least for now)
 - to cover all EU and associated countries
 - and all languages relevant for target audience

Some typical CLARIN features



- Data and service centres of many different types: university data collections, libraries, archives, ...; mostly public
- Data of different types and modalities (digitized historical documents, books, newspapers, tv news, audio recordings, movies, tweets, sign language recordings, ...)
- Massive fragmentation, both at national and international level; centres mostly unconnected, using own standards
- Users may be interested in form (e.g. linguists) or content (e.g. historians or political scientists)
- Complex IPR conditions: re-purposed data (e.g. digitized novels), ethical considerations (e.g. call centre recordings)
- Language is a sensitive area, very close to cultural identity, as opposed to e.g. astronomical or medical data

Why did we go for an ERIC?



- They are legal entities, not projects, which helps to make them more sustainable, and which allows them to make agreements with other legal entities
- Members are governments, not universities, committing themselves for longer periods of time (min. 5 years)
- Closeness to funding agencies may help to enforce use of standards and sharing of data in projects they fund
- ERICs can participate in EU funding programmes as a single consortium
- VAT was never an issue, but may be important for others

But, to our disappointment:

- Bulk of the funding dependent on funding mechanisms and cycles in participating countries (3-5 years)
- Mostly through competitive Calls for Project Proposals
- Mostly with competition across domains and types of infrastructures
- Lack of synchronicity across countries

CLARIN ERIC



- CLARIN ERIC established by the EC on 29-02-2012, with 9 founding members: AT, BG, CZ, DE, DK, EE, NL, PL, DLU
- More in the pipeline (NO just joined as observer) – but we need all European countries in to make it a real pan-European infrastructure for HSS
- CLARIN ERIC is the governance and coordination body, but will not run or fund operational data services
- Its added value: turning a large collection of unconnected local infrastructure facilities into one single European RI
- CLARIN ERIC member countries pay a modest annual fee to cover the cost of governance and coordination
- The bulk of the national funding goes into operations at the national level
- Important principle: the autonomy of national consortia

The autonomy principle



- Countries will each set up and fund a national CLARIN consortium, that will provide data and linguistic services and create new data and tools according to their own research priorities, as they are responsible for their own languages
- National consortia receive only national funding (i.e. no central funding by CLARIN ERIC)
- No cross-border funding other than the annual fee
- It is up to the countries to decide how to shape and fund their CLARIN consortia and how to relate them to other activities at the national level (e.g. research programmes, digitisation programmes, etc)
- CLARIN ERIC does not tell countries **how much to spend** or **what to do**, but **how to do it**, e.g. standards to be used, quality requirements for centres, open access

The lightness principle



- In all EU countries HSS is a low-budget domain, so that the threshold to join should be kept low and manageable (fee based on % of EU GDP in 2010, with fixed maximum)
- No surprises: fees fixed for 5 years
- CLARIN ERIC is a light ERIC:
 - The cost of central governance and coordination should be kept low: a small and decentralized head office
 - Technical operations should be placed where the competence is and where the facilities are: in the member countries, and not at the central CLARIN ERIC level
- As CLARIN ERIC does not fund the countries decision taking should be based on consensus rather than on the power of money. The governance structure reflects this.

General Assembly (GA)



- **Task:** Main decision taking body
- **Composition:**
 - One voting delegate per Member country, plus expert if desired;
 - One delegate per Observer (not voting).
- **Appointed by:** Ministries in Member countries
- **Special roles:** President, Vice President, both elected from and by Member delegates.
 - The President chairs the meetings of the GA (replaced by the Vice President if absent).
- **Meetings:** At least once per year (and maybe more in the start up phase of the ERIC). President has regular consultations with Executive Director.
- **Rationale:** ERIC needs decision taking body where all its members are represented

Board of Directors (BoD)



- **Task:** Implementation of policies and decisions by GA.
- **Reports to:** GA.
- **Composition:**
 - Executive Director, full-time, working at the headquarters, responsible for day to day management. Is responsible for the liaison between the ERIC and the scientific field and proposes strategies for the ERIC according to scientific developments and user needs.
 - 2-4 Directors, part-time (0.2-0.4), working at own institution, out of which two ex officio members (see National Coordinators' Forum, and Standing Committee for CLARIN technical Centres below).
- **Appointed by:** GA.
- **Special roles:** Chair (Executive Director), Vice Chair (appointed by GA).
- **Meetings:** Frequent. Executive Director has regular meetings with President.
- **Rationale:** ERIC needs an executive body to implement GA decisions.

Scientific Advisory Board (SAB)



- **Task:** Give advice to GA on strategic issues, evaluation of progress and services.
- **Reports to:** GA.
- **Composition:** 5-10 independent experts.
- **Appointed by:** GA.
- **Special roles:** Chair, appointed by GA.
- **Meetings:** Annual.
- **Rationale:** GA needs independent advice on scientific matters.

National Coordinators' Forum (NCF)



- **Task:** Act as main liaison between CLARIN ERIC and national CLARIN Consortia, and is responsible for implementation of GA's decisions at the national level, maintaining coherence and consistency across the CLARIN Infrastructure and between Members.
- **Reports to:** BoD.
- **Composition:** For each Member the head of the national consortium.
- **Appointed by:** Member countries.
- **Special roles:** Chair (elected by NCF, and is ex officio member of the BoD), Vice Chair (elected by NCF).
- **Meetings:** At least twice a year, but in practice more often
- **Rationale:** The technical operation of the CLARIN services and the creation of content is carried out and funded by the members through national CLARIN consortia. A body is needed where those responsible for the consortia get together and ensure implementation of the GA's decisions in a coherent and consistent way across members. A thin ERIC does not have the resources to communicate directly with individual members of national consortia, and should not bypass those who are responsible for the execution of national CLARIN projects.

Standing Committee for CLARIN Technical Centres (SCCTC)



- **Task:** Gives advice to CLARIN ERIC and National Coordinators in order to ensure consistency, coherence and stability of infrastructure services.
- **Reports to:** BoD and National Coordinators.
- **Appointed by:** Member countries.
- **Composition:** For each main centre the head of the centre; BoD can invite heads of other centres to join SCCTC.
- **Special roles:** Chair (elected by SCCTC, and is ex officio member of the BoD), Vice Chair (elected by SCCTC).
- **Meetings:** At least twice a year, but in practice more often
- **Rationale:** Given the critical role of the main centres in the backbone of CLARIN a special body is needed where the heads of these centres get together directly and identify problems and solutions related to the operation of the infrastructure across members. As the centres operate under the responsibility of the national consortia the SCCTC gives advice to BoD and NCF.

Concluding remarks



- After 14 months we have no evidence that the distributed thin ERIC structure doesn't work, but what works for 10 countries may not work for 30 (e.g. consensus!)
- As a consequence of the autonomy principle CLARIN is better described as an evolving mosaic than as a jig-saw puzzle based on a predefined picture
- Even if interoperability can be ensured by using common standards, coverage is very uneven and differs from language to language
- In the absence of a large central pot, and given our commitment to not increasing the fees we plan to attract EU funding to arrive at more harmonized coverage and better cross-border integration