



CLARIN ERIC status, goals and plans

Steven Krauwer

CLARIN ERIC / Utrecht University

Overview



- The vision (what heaven looks like)
- The mission (how we think that we can get there)
- Guiding principles (our bible)
- Action lines and milestones (what we plan to do)
- Status and structure of CLARIN ERIC
- Plan for 2012
- What about Norway?
- Concluding remarks

Vision – 8 pillars



- Coverage: access to all public language data
- Legal: for research no restrictions other than ethical
- Integration of data: metadata and content search over the whole data space
- Integration of services: combinable advanced tools
- Preservation: sustainable access to what is there
- Ease of access: no technical obstacles for HSS scholars
- Crossing borders: disciplines, countries, continents
- Sustainability: financial, technical, organisational

Vision – a view from heaven



- *give me digital copies of all contemporary documents in European archives that discuss the Great Plague of England (1348-1350)*
- *give me all negative articles about Islam or about soccer in the Slovenski Narod daily newspaper (1868-1943)*
- *find Norwegian TV news interviews that involve speakers with a German accent*
- *summarize all articles in European newspapers of April 2012 about machine translation – in Nynorsk*
- *Show me the pronoun systems of the languages of Alaska*

Mission for the 8 pillars



- Coverage: all EU+A states, all collections, access across borders
- Legal: light licensing, authorization and authentication, IPR
- Integration of data: shared standards for metadata and data
- Integration of services: standards for interoperability
- Preservation: depositing facilities, persistent identifiers
- Ease of access: interfaces, support, training
- Crossing borders: bridges to other communities
- Sustainability: secure future after initial 5 years, based on what happens inside and outside CLARIN

Principles



- Separation of governance and operations
- Light coordination
- Aim at free access
- Countries free in what they do, but not how
- CLARIN ERIC will not own servers, data or tools but facilitate access
- Participation in CLARIN construction activities is open to others than CLARIN ERIC members
- CLARIN will not duplicate work done by others but team up with those doing similar things

Action lines and some milestones (1)



- Coverage: consolidate 9 members, reach out to others, 15 members in 3 years, 20 in 5 years
- Legal: license templates promoted for new and legacy data, talk to legislators about IPR, establish AA in old and new countries
- Integration of data: standards action plan, tools for mapping, tools for curation; identify priority areas for cross border research
- Integration of services: interoperability, identify chainable services, work on showcases
- Preservation: identify at least 1 centre per country, work on change of culture, follow broader data initiatives; in 3 years all data and tools from funded projects deposited

Action lines and some milestones (2)



- *Ease of access*: KSI to support ease of access, awareness, training & support, curricula, centres of expertise; Portal targeting different audiences; Interfaces and visualization
- *Crossing borders*: develop long term vision, but start from existing collaborations; use language as vehicle to collaborate with other disciplines; inter RI, international collaboration; explore industrial collaboration models
- *Sustainability*: demonstrate societal impact; review sustainability models; after 3 years vision and strategy



CLARIN ERIC status and structure

Division of roles



- CLARIN ERIC governs and coordinates the infrastructure across member countries
- National consortia responsible for
 - The operation of centres (one or more per country, fully funded by own government), coordinated by CLARIN ERIC
 - Creation of data, tools and services compliant with CLARIN, nationally funded, selection based on national priorities

Status



- CLARIN ERIC officially created Feb 29 2012
- First official meeting on Apr 18 in Den Haag
- Launching members: AT, BG, CZ, DE, DK, EE, NL, PL and DLU
- More in the pipeline: FI, FR, LV, [HR], [NO]
- Failed to get funding till now: ES, SE
- Pending: PT, GR, LT
- Stalled: SI, RO, HU
- Interest, no action: UK, IT, SK, LU, [TR], [IS], [CH]
- Not joining: MT
- No contact: CY, IE

Structure (1)



- General Assembly
- Scientific Advisory Board
- Board of Directors
- National Coordinators Forum
- Standing Committee for Technical Centres
- Centre Assessment Committee
- Standards Committee
- Dissemination Committee
- Knowledge Sharing Infrastructure Committee
- Legal Issues Committee

Structure (2)



General Assembly (GA):

- **Task:** Main decision taking body
- **Composition:**
 - One voting delegate per member, plus expert if desired;
 - One delegate per observer.
- **Status:** up and running

Scientific Advisory Board (SAB)

- **Task:** Give advice to GA on strategic issues, evaluation of progress and services.
- **Composition:** 5-10 independent experts.
- **Status:** appointments to be confirmed by GA

Structure (3)



Board of Directors (BoD):

- **Task:** Implement decisions by GA and day to day management
- **Composition:**
 - Executive director (Steven Krauwer)
 - Vice director (Bente Maegaard)
 - Ex officio director elected by nat. coordinators forum (Erhard Hinrichs)
 - Ex officio director elected by standing committee for CLARIN technical centres (to be elected)
- **Status:** up and running

Structure (4)



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.
- **Composition:** For each member Head of the National Consortium; heads from non-member consortia may be invited to meetings as observers
- **Status:** First meeting Apr 19, up and running

Structure (5)



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.
- **Composition:** Heads of CLARIN technical centres
- **Status:** Core committee (representative of main hub per member country) now in place, first meeting being organized

Structure (6)



Centre Assessment Committee (CAC)

- **Task:** Formulate criteria for centres and assess quality of individual centres
- **Composition:**
 - Chair: Peter Wittenburg
 - Members: Proposed by members, selected by BoD
- **Status:** Selection ongoing

Structure (7)



Other committees under construction:

- Standards
- Dissemination
- Knowledge Sharing Infrastructure
- Legal Issues
- Possible other committees or working groups for specific tasks

Invitation to join sent out shortly, also to non-members



CLARIN ERIC WORK PLAN 2012

Organisational issues



- Headquarters
 - Formal issues, physical premises, staff, agreement with UU etc.
- Board of Directors
 - Distribution of tasks, agreements with universities, websites
 - By-laws
 - Setting up working structures and make them work (BoD, NCF, SCCTC, SAB, Centre assessment committee)
 - Setting up other committees

CLARIN Construction



- CLARIN Agreements between National Consortium and CLARIN ERIC, specifying national in kind contribution
 - Important exercise, major task, to be concluded end June
- Setting up main technical centres
- Extending with new members and observers, or making collaboration agreements with centres in countries not willing/ready to join

Legal issues



- Working group on legal issues
- License template revisited

Integration of data and services



- BoD preparing plan for integration: ALL to participate
- Collaboration between NCF and SCCTC
 - Select priority areas for integration, for showcasing – making sure to have showcases for users
- Annual conference, in 2012 internal, common understanding of framework, goals and strategy (Sofia, Oct 26-28 2012)
- Standards: working group, workshop, action plan

Preservation & Access



- Each member to appoint at least one centre where users can deposit data and tools for long term preservation and access.
- PID service
- CLARIN Portal
- Educational material
- CLARIN Knowledge centres: criteria, candidate centres identified
- Helpdesk services

Crossing borders & Sustainability



- Collaboration with other research infrastructures (e.g. other SSH RI, e-Infrastructures) and other infrastructure projects (e.g. META-Share)
- Collaboration with international partners
- Establishing success criteria for CLARIN

And what about Norway (1)?



Things that are not possible for NO:

- NO cannot be part of GA or NCF unless they join CLARIN ERIC as Members or Observers
- Without special agreement SSH scholars in NO cannot have access to restricted parts of CLARIN (but no empirical consequences at this stage)

But apart from that we will just pretend NO is part of CLARIN ERIC to the largest possible extent because NO is in the top league

Main communication channel will be the national coordinator (as for consortia in member countries)

And what about Norway (2)?



Things that are possible:

- NO can participate in (and even chair) all other activities and committees (on a self paying basis, just like other national consortia)
- CLARINO Coordinator will normally receive copies of all announcements sent to national coordinators
- Some tasks may be contracted out, members have priority, but NO is not excluded
- If space permits some (1-2) CLARINO participants may be admitted to the CLARIN conference in Sofia (Oct 26/28), on a self paying basis

My main concerns



- Technical: not really
- Many countries still missing, lack of synchronization
- Composition of CLARIN community not balanced (historically grown):
 - Mainly NLP providers
 - Few speech and other modalities
 - Very few libraries
 - Very very few ‘real’ HSS users
- Corresponds what we make to what the users need? (the gearbox syndrome)

Concluding remarks



- The construction of CLARIN has just started, but please note that both the specs and the actual construction are a joint responsibility
- Integration of what we have should be one of the top priorities, but ...
- ... reaching out to the users is at least as important, because if we fail this is where it is most likely to happen!
- I hope and expect CLARINO to be one of our strong players in the construction of CLARIN and I am looking forward to our continued collaboration!