Common Lab Research Infrastructure for the Arts and Humanities

CLARIAH

Jan Odijk

EuroRisNet+ Workshop, Lisbon, 2013-07-05
Overview

• Background
• Goals
• Relation with companies
• Challenges
Background

- **CLARIN** and **DARIAH** on ESFRI Roadmap (2006)
- EU-funded preparatory projects (2008-2011)
- Resulted (inter alia) in **CLARIN ERIC** (since 2012), DARIAH ERIC coming
- 2011 National Roadmap: CLARIN-NL and DARIAH-NL join forces resulting in **CLARIAH**
- partially successful: put on roadmap and obtained ‘seed money’
- 2013 National Roadmap: revised proposal to be submitted
Goals

• A technical research infrastructure in which a **humanities** researcher
  – Can find all data relevant for the research
  – Can find all tools relevant for the research
  – Can apply the tools to the data without any technical background or ad-hoc adaptations
    • Inter alia tools for search in and through the data
  – Can store data resulting from the research
  – Can store tools resulting from the research

• A **distributed** infrastructure based on one or more **centres** per country.
Goals

• Focus in CLARIN on research using **language resources**
• Focus in DARIAH broader
• Focus in CLARIAH on
  – Linguistics and language resources (esp. text)
  – Social History (esp. structured databases)
  – Media Studies (esp. audio-visual data)
Relation with Companies

- Primary user group is academic researchers
- Good support by companies
  - For CLARIN-NL
  - For CLARIAH
- Interest because of technology to extract information from a variety of structured and unstructured (text, AV-data) data (e.g. sentiment/opinion mining, IBM’s Watson)
- Fits in nicely with NL’s Top Sector Policy
- Role of Companies
  - User / reflection group, as potential (secondary) users
  - Implementation of certain aspects as subcontractor
  - Data Providers
Challenges (1)

• CLARIN-NL partially organized as a project, partially as a programme.
• This offered much opportunity to
  – flexibly react to new developments
  – bring in more partners
  – react to ideas and proposals coming from our prospective users.
  – Set up (unanticipated) international cooperation (e.g. with Flanders)
Challenges (1)

• Most other national CLARIN projects are tightly packed projects without this flexibility
  – Given the distributed character of the infrastructure, this makes it difficult to do things together

• Recommendation:
  – National projects should leave enough room for unanticipated international cooperation and other developments
  – their funding agencies should take this into account in their evaluation
Challenges (2): Sustainability

• Work out a business model, supported by the funding agencies, to guarantee the sustainability of services provided by the Centres in the infrastructure after the project has finished
• May require some policy changes with the funders (e.g. allow and even require funding of data management)
Challenges (3): IPR

- Use of copy-righted data for academic research should be allowed without requiring explicit permission
Challenges (4)

- National Roadmaps are not synchronized among countries
  - Big Differences in start-up and running time for each country (e.g. NL 2009, DE 2011, others not yet)
- Can be handled in a distributed infrastructure, but it is not ideal
- A limited form of synchronisation is desirable
Thank you for your attention

For more information

www.clariah.nl
www.clarin.nl
CLARIAH: Industrial Interest