

The Danish CLARIN project

- building a researcher's toolbox, a digital research infrastructure for the humanities

Hanne Fersøe
University of Copenhagen
Centre for Language Technology
hannef@hum.ku.dk



Background – the EU context

- European Strategy Forum on Research Infrastructures
 - ESFRI Roadmap process start 2001, report 2006, update 2008
 - Recommendation for 6 SSH-projects (Social Sciences and Humanities)
 - CLARIN is one of the recommended SSH projects
- Funding model and legal instruments developed at EU level
 - Funding for preparatory phases for 34 projects from the Roadmap given in the 7th Framework Programme
 - A Council Regulation to provide a legal instrument for the construction and deployment phases for pan-European RIs has been proposed
- Development of national Roadmaps
 - National participation in construction and deployment of pan-European RIs must be funded nationally
 - The 27 EU member countries have agreed to develop national roadmaps, appr. 10 are completed.



Background – The national Danish Context

The Danish Ministry of Science, Technology and Innovation commissioned the Danish Council for Strategic Research to **survey the needs and propose a strategy for future research infrastructures**. The report was published in December of 2005.

The Danish Government has formulated 18 initiatives called the **Globalization Strategy**. The report was published in April 2006. Within the focus area 'More Competition and better Quality in Public Research' one of the initiatives is to **reserve a research grant pool for research infrastructures**.

A **call for RIs**, according to the 2005 report, was published with deadline in September 2007, 200 mio. DKK p/y, 3 years

The Danish Agency for Science, Technology and Innovation is currently working on a **national Roadmap**



The Danish CLARIN project - outline

The Danish CLARIN Consortium has been granted 15 mill. DKK (2 mill. €) for the years 2008-10

The consortium partners are led by the University of Copenhagen, CST

- University of Copenhagen (3 research groups)
- University of Southern Denmark (1 research group)
- University of Aarhus (2 research groups)
- Copenhagen Business School (1 research group)
- Society for Danish Language and Literature
- Danish Language Council
- The Royal Library
- The National Museum

The grant is for **construction** of a national RI for Humanities.

The consortium will collaborate with EU-CLARIN where possible.

The consortium will propose a strategy for the exploitation phase following the construction.

The Danish CLARIN consortium aims to be part of the pan-European CLARIN RI in the construction and exploitation phases.



Danish CLARIN - The Project

Thematically organised work packages

- Basic written language resources, i.e. collection and annotation of written texts:
 - Current and older Danish texts
 - General language and specialised sublanguages (Danish)
 - Literary texts and common knowledge (Danish)
 - Multilingual, aligned texts
- Language resources and tool for spoken language, i.e. collection and annotation of spoken texts:
 - Dialogue, video, read-aloud
 - Mark-up according to different methods and principles
 - Development of accompanying tools for search and retrieval



Danish CLARIN - The Project

- Technological language resources, i.e. a constructed data collection:
 - Modification of existing technological resources and development of new ones
 - Traditional and electronic dictionaries
 - Dictionaries and semantic nets for computers
 - Linking between different dictionaries and between dictionaries and corpora.
- The technical infrastructure of the Danish CLARIN platform, i.e. a digital repository with a web user interface, managing
 - Access rights to users based on user verification mechanisms
 - Access rights to content based on resource profiling
 - Search and retrieval facilities
 - A personal work space
 - Communication facilities



Strengths

- The consortium is very strong and to the point with four universities, The Royal Library, The National Museum, and two cultural institutions -> a good basis for true humanities projects.
- The consortium has an good combination of the necessary skills and experience: humanities, language technology, language resources, computer science.
- The result will not merely be a technical platform – it will be a platform with content.



Future challenges

- Getting more humanities scholars to use it and accept it, thus demonstrating the usefulness of the toolbox to other scholars.
- Getting massive amounts of data added to the platform from a wide range of fields within the humanities and social sciences.
- Funding for the exploitation phase.
- Development of an instrument to encourage inclusion of additional resources.

