D-SPIN, a CLARIN partner project

Ralf Gehrke
Johann Wolfgang Goethe University, Frankfurt

EuroBABEL Launch Conference
Berlin, September 13, 2009

The starting point

- Much language data in archives, but they are only known to insiders.

- The archives are mostly unconnected.

- Every archive has its own standards for storage and access.

- Mostly only simple file retrieval is offered.

- Available tools are hard to use for non-specialists.

CLARIN

- "Common Language Resources and Technology Infrastructure"

- http://www.clarin.eu

- "CLARIN is committed to establish an **integrated** and **interoperable** research infrastructure of language resources and its technology. It aims at lifting the current fragmentation, offering a **stable**, **persistent**, **accessible** and **extendable** infrastructure and therefore enabling eHumanities."

CLARIN

- **integrated**: the resource and service centres are connected via Grid technology and form a virtually integrated domain
- **interoperable**: the resources and services will be based on Semantic Web technologies to overcome format, structure and terminological differences
- **stable**: the resources and services are offered with a high availability
- **persistent**: the resources and services are planned to be accessible for many years so that researchers can rely on them
- **accessible**: the resources and services are accessible via the web; different access methods and training possibilities are offered tailored to the needs of the communities making use of them
- **extendable**: the infrastructure is open so that new resources and services can be added easily

CLARIN

- The ultimate objective of CLARIN is to create a European federation of digital repositories that include language-based data, to provide uniform access to the data, wherever it is, and to provide existing language and speech technology tools as web services to retrieve, manipulate, enhance, explore and exploit the data. The primary target audience is researchers in the humanities and social sciences and the aim is to cover all languages relevant for the user community.

CLARIN National Groups

- 171 member institutions in 32 countries
- "While all mentioned bodies act at the European level, every national CLARIN group is expected to define an appropriate national coordination structure."

D-Spin

- „Deutsche Sprachressourcen-Infrastruktur" (German Infrastructure of Language Resources)

- http://www.sfs.uni-tuebingen.de/dspin

- Purpose: Implementing the infrastructure on a national level and contributing actively to the development of CLARIN

- Funded by the German BMBF (Bundesministerium für Bildung und Forschung), launched in 2008

D-SPIN project partners (1)

- Max-Planck-Institut für Psycholinguistik Nijmegen (Peter Wittenburg)
- Seminar für Sprachwissenschaft, Universität Tübingen (Prof. Dr. Erhard Hinrichs)
- Institut für Deutsche Sprache Mannheim (Dr. Marc Kupietz)
- Berlin-Brandenburgische Akademie der Wissenschaften (Dr. Alexander Geyken)
- ASV - Institut für Informatik, Universität Leipzig (Prof. Dr. Gerhard Heyer)

D-SPIN project partners (2)

- Institut für Vergleichende Sprachwissenschaft, Universität Frankfurt (Prof. Dr. Jost Gippert)

- Deutsches Forschungszentrum für Künstliche Intelligenz Saarbrücken GmbH (Thierry Declerck)

- Institut für maschinelle Sprachverarbeitung, Universität Stuttgart (Prof. Dr. Hinrich Schütze)

- Angewandte Sprachwissenschaft und Computerlinguistik, Justus-Liebig-Universität Gießen (Prof. Dr. Henning Lobin)

Working Packages (APs)

- AP 1: Management

- AP 2: Technical infrastructure

- AP 3: Planning of resources and applications w.r.t. users from the humanities

- AP 4: Connections with national and European initiatives

Working Packages (APs)

- AP 5: Adaptation and integration of resources and tools
- AP 6: Training and instruction
- AP 7: IPR and ethical issues, license models
- AP 8: Persistence

AP 3: Resources and applications

- Establishing contact to potential user groups and elicitation of user needs and expectations
- Support for selected projects outside of D-SPIN by use the infrastructure being developed by D-SPIN

Thank you!

Ralf Gehrke

gehrke@rz.uni-frankfurt.de