



CLARIN Summary

The CLARIN project is a large-scale pan-European collaborative effort to create, coordinate and make language resources and technology available and readily useable. CLARIN offers scholars the tools to allow computer-aided language processing, addressing one or more of the multiple roles language plays (i.e. carrier of cultural content and knowledge, instrument of communication, component of identity and object of study) in the Humanities and Social Sciences.

CLARIN Key Points

The key points of the CLARIN initiative are:

- Comprehensive service to the humanities disciplines with respect to language resources and technology.
- Technology overcoming the many boundaries currently fragmenting the resources and tools landscape as it is given by institutional, structural and semantic interoperability problems.
- Tools and resources will be interoperable across languages and domains, thus contributing towards addressing the issue of preserving and supporting the multilingual and multicultural European heritage.
- Comprehensive training and education programs that will include university education in the different member states.
- Improve and extensivize web-based collaborations, i.e. creating virtual working groups breaking the discipline boundaries.
- Fostering the development or improvement of standards for language resource maintenance.
- A persistent and stable infrastructure that researchers can rely on for the next decades.

CLARIN Key Technologies

To achieve these challenging goals CLARIN will be built on and contribute to a number of key technologies coming from the major initiatives advancing the eScience paradigm:

- It includes Data Grid technology to connect the repositories and web services the various centers provide;
- It will build on ideas launched by the Digital Library community to create Live Archives, and will further such initiatives;
- It incorporates, and will contribute to, Semantic Web technology to overcome the structural and semantic encoding problems;
- It will incorporate advanced multi-lingual language processing technology that will support cultural and linguistic integration.

Overall Organizational Framework

We propose to create a European Resources Infrastructure that would be based on an open European Federation of strong service centers and repositories that would jointly provide the whole European Humanities (and Social Sciences) community with

- (i) knowledge about the existence of language resources,
- (ii) coordinated creation and archiving of and access to such resources,
- (iii) access to services and tools that would allow scholars to operate on such resources ,
- (iv) bundling of and access to expertise related to specific language processing problems

CLARIN, therefore, will be built on the existing national infrastructures and all the knowledge gathered from the European funded projects in our domain. It is left to the national formation process to specify how the services and repository landscape will be organized in the different European member states. At the European level an efficient umbrella organization has to be setup that will be responsible for the unification and organizations at the European level. CLARIN will also link up closely with appropriate infrastructures in other humanities disciplines. Yet it is not obvious whether there are already appropriate organizations that could take over the integration task at European level and between the humanities disciplines or whether they have to be setup. CLARIN sees itself as a research infrastructure that will offer specialized resources, tools/services and knowledge and would easily join with other complementary initiatives in the humanities area.