



**D-SPIN**

**D-SPIN Report R3.2/4.1  
Final Report on liaison  
activities with other projects  
and initiatives and Roadmap  
Report**

March 2010

D-SPIN, BMBF-FKZ: 01UG0801

Deliverable R-3.2/4.1: Final report on the liaison activities with other projects  
and initiatives and Roadmap report

© All rights reserved by BBAW on behalf of D-SPIN

Editors: Alexander Geyken, Lothar Lemnitzer, Peter Wittenburg

## Summary

D-SPIN is the German counterpart of the European Research Infrastructure project CLARIN (Common Language Resources and Technology Infrastructure, <http://www.clarin.eu/>). The ultimate objective of CLARIN and D-SPIN is to create a federation of existing digital repositories that include language data, and to provide uniform access to those data, wherever there are., CLARIN and D-SPIN aim also at providing 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.

Within the CLARIN federation, the focus of D-SPIN is on German resources, tools and their integration through web services. In addition, some of the project partners are preparing for their role of acting as service centers on the national and pan-European level. Besides these localization efforts, D-SPIN has a special focus in addressing potential users of the infrastructure with the preparation of training material and teaching activities.

This report summarizes the liaison activities of all D-SPIN partners with other projects and initiatives and draws conclusions for a roadmap of further activities to reach the infrastructural goals of D-SPIN.

The D-SPIN/CLARIN efforts have to be embedded in and co-ordinated with other activities dedicated to the building of a technical infrastructure, the preparation of resources, the integration of web services, and the exploration of user needs.

This report is based on the report “liaison activities” provided earlier by WP 5 of the CLARIN project. Many of the contacts made are also relevant for the D-SPIN project. In addition, we report on contacts, which are relevant only for D-SPIN. As such, this report can be considered to be a localized version of the CLARIN report. The roadmap description is specific to the D-SPIN project.

# 1 Background

Research infrastructures are a kind of eco-system that offer embedded lean services at a specific functional layer. There are infrastructures that are offering common services for all kinds of communities (so-called e-Infrastructures such as computer networks, grid services, etc.), or services for selected communities that share certain functionality and services that are dedicated to a specific community. With respect to the latter two levels mentioned we can assume that there will be several overlapping services at the same functional level.



This figure indicates layers of services with at the bottom those that are generic and at the top those that are community specific. Yet it is not clear where exactly the border line is and it will change over time.

Currently, several initiatives are active in specifying and partially implementing such infrastructures. One of the big challenges is to prevent too much overlap in activities. This can only be solved by continuous interactions between and in joint activities of all relevant initiatives. D-SPIN/CLARIN clearly developed this principle to a broad roadmap of activities to play a proper role in building an eco-system of infrastructures in which overlap can be minimized over time and in which cross-fertilization can be optimized.

D-SPIN/CLARIN is guided by the fact that joint workshops and conferences are important, but even more important are project-oriented meetings in small groups and even more productive are joint projects. We have been very active in all these dimensions and describe them in the following chapters.

The relevant projects and initiatives can be classified in the following ways with regard to the D-SPIN/CLARIN perspective:

- a) projects providing services which are interesting or necessary for us to draw on. From the perspective of these projects, our projects can be viewed as a user of these services. A good example of this type is the *Trans-European Research and Education Networking Association* (TERENA; see below and the article “Terena-CLARIN Collaboration” by Peter Wittenburg, Daan Broeder and Dieter van Uytvanck et al in the CLARIN Newsletter #2);
- b) projects and initiatives which also aim at building research infrastructures, but with another subject and / or target group; a good example for this type the *Council of European Social Science Data Archives* (CESSDA);

- c) projects which address the same subject and / or target group but do not necessarily have the aim of establishing a research infrastructure; an example of this type is the eAQUA project;
- d) initiatives which provide frameworks like standards; a good example of this type is the ISO TC 37/SC4 committee.

With projects of type (a) it is useful, from our perspective, to come to stable agreement about the usage of services. CLARIN and D-SPIN profit from the services offered while we hope that these projects and initiatives view our projects as an interesting use case.

With projects and initiatives of type (b) we should discuss common solutions to common problems, to avoid the doubling of efforts and the implementation of solutions which are not compatible with one another.

Projects and initiatives of type (c) should be seen as allies in addressing target groups for the use of the CLARIN infrastructure.

For initiatives of type (d), CLARIN will on the one hand adapt the existing standards and emerging frameworks and on the other hand participate and express requirements which arise in the process of building the infrastructure.

For each liaison activity, the following information will be given:

- details about the **arrangements** with the contact person(s);
- details about the **project or initiative** (Name(s); URL of webpage; coordinator and other key partners relevant for our project; scientific, research and development goals of the project);
- **events** of this project / initiative in which CLARIN participated or CLARIN event in which a representative of this project participated;
- **joint activities** which are an outcome of the liaison;
- short-, middle- or long-term **goals of a co-operation** with the project / initiative;
- **relevance** of this person or project / initiative for D-SPIN

Therefore, this report will be a valuable source of reference and an instrument for monitoring the further development of these contacts.

## **2 Liaison with projects and initiatives**

### **2.1 Projects initiatives which provide services**

#### **2.1.1 e-irg**

##### **Arrangements**

Result of this participation was a joint activity which is documented below. It was first at the Prague meeting (24-25 March, 2009) that D-SPIN could give a clearly motivated statement on the fact that e-Infrastructure initiatives alone (such as Grids) will not be sufficient, since they do not know the community specifics and traditions and therefore cannot provide adequate end-user services.

##### **Details about the project / initiative**

e-IRG (e-Infrastructure Reflection Group, <http://www.e-irg.eu/>) is a political umbrella for groups aiming at coordinating infrastructure activities at European level and discussing aspects of an eco-system of infrastructures.

##### **Events**

D-SPIN actively participated in the Prague conference (see above) to present the CLARIN principles and approaches and it will participate in future meetings.

##### **Joint Activities**

e-IRG and ESFRI established a Data Management Task Force. D-SPIN members edited the metadata section and mainly contributed to the parts about quality assessments and interoperability. The report has been accepted by the e-IRG and ESFRI boards and can operate as a major guideline for all infrastructure initiatives. Further versions of the report have been scheduled.

[http://www.e-irg.eu/index.php?option=com\\_content&task=view&id=241&Itemid=22&show=1](http://www.e-irg.eu/index.php?option=com_content&task=view&id=241&Itemid=22&show=1)

##### **Relevance of the liaison for the D-SPIN project**

Relevant for AP2 (technical aspects)

#### **2.1.2 APA**

##### **Details about the project / initiative**

The APA (<http://www.alliancepermanentaccess.eu/>) is an initiative which wants to put the need for long-term preservation of research data onto the agenda. The Alliance aims to develop a shared vision and framework for a sustainable organizational infrastructure for permanent access to scientific information. The Alliance for Permanent Access (APA) has been established to support the creation of a European Digital Information Infrastructure or in US terms “a cyber infrastructure”. Basically this consists of a series of repositories or archives where the digital record of science (both documents and data) is stored, curated and kept accessible. For universities, research organizations, operational agencies, funding agencies and society at large this is rapidly becoming an issue of crucial strategic importance.

##### **Events**

D-SPIN participated at two annual APA conferences. At the Budapest conference (2008) we were asked to present cost models. At the Den Haag meeting (2009) we made plans for joint activities

which will be mentioned below. At both meetings we could make clear that it is indeed a task of each research infrastructure to take measures that “its” data needs to be curated and preserved.

### **Joint Activities**

Two major activities have been agreed upon: (1) A joint workshop about metadata is currently being planned. This workshop should also cover the aspects of contextual metadata associated with resources, services and controlled vocabularies that can be used across disciplines. This workshop is meant to be co-organized by a number of initiatives and will address all interested infrastructure initiatives. (2) Another joint workshop will address the issue of repository systems, their software setup, their APIs, their costs, etc. It will be co-organized by several other initiatives.

### **Relevance of the liaison for the D-SPIN project**

Relevant for AP2 (technical aspects)

#### **2.1.3 centerNet**

##### **Details about the project / initiative**

This worldwide initiative (cf. <http://digitalhumanities.org/centernet/>) wants to bring centers in the humanities disciplines together to discuss worldwide issues of standardization and harmonization and also to raise the awareness about the needs of curation, preservation and accessibility. centerNet is organized in regions and one D-SPIN member is leading the steering board for continental Europe.

##### **Events**

Yet one cannot speak about concrete activities, but a first summit is planned for 2010.

### **Relevance of the liaison for the D-SPIN project**

Relevant for AP2 (technical aspects)

#### **2.1.4 GEANT Network / TERENA**

##### **Arrangements**

The state of degree of harmonization between the different schemes at European level is so that it can not be directly used by CLARIN, but a joint roadmap has been worked out, so that the currently chosen intermediate solution can be replaced at a certain moment.

##### **Details about the project / initiative**

The purpose of eduGAIN is to provide the means for achieving interoperation between different Authentication and Authorization Infrastructures (AAI). For further details cf. [www.edugain.org](http://www.edugain.org).

Contact with Jaap Kuipers (Surfnet), who is interested in possible co-operations and pilots in the LRT domain. We are in contact with Karel Vietsch and Licia Florio.

##### **Events**

Several meetings were organized between GEANT/eduGain/TERENA and D-SPIN members to discuss solutions for distributed authentication, e.g. 12th TF-EMC2 Meeting, 3-4 December 2008, Utrecht, the Netherlands (cf. <http://www.clarin.eu/events/edugain-meeting>) where MPI Nijmegen presented the requirements of CLARIN with regard to eduGAIN.

## **Goals of the co-operation**

Cross-federation access to service providers, see also the article in CLARIN newsletter #2, page1.

## **Relevance of the liaison for the D-SPIN project**

Relevant for AP2 (technical aspects) and AP7 (legal aspects).

### **2.1.5 DEISA**

#### **Arrangements**

With DEISA a couple of interactions have taken place, and have resulted in concrete projects and plans.

#### **Details about the project / initiative**

DEISA, the Distributed European Infrastructure for Supercomputing Applications, is a consortium of leading national Supercomputing centres that aims at fostering the pan-European world-leading computational science research.

DEISA deploys and operates a persistent, production quality, distributed supercomputing environment with continental scope. It aims at delivering a turnkey operational solution for a future European HPC ecosystem. And by extending the European collaborative environment in the area of supercomputing, DEISA is paving the way towards the deployment and operation of a persistent cooperative European HPC ecosystem, as suggested by ESFRI.

DEISA2, funded by the European Commission in FP7, continues to develop and support the pan-European distributed high performance computing infrastructure established since 2002 within the predecessor project DEISA1 that was funded in the sixth Framework Programme. The DEISA infrastructure is based on a tight coupling of eleven national supercomputing centres from seven European countries, using dedicated network interconnections of GÉANT2 and the projects.

The DEISA consortium is currently consolidating the existing HPC infrastructure and services. Activities and services relevant for Applications Enabling, Operations, and Technologies are continued and enhanced, as these are indispensable for the effective support of computational sciences in the area of supercomputing. In addition, DEISA is extending the service provisioning model towards the inclusion of non-localized Virtual Science Communities. Accordingly, collaborative activities are expedited with further European and international initiatives.

We are via MPI Nijmegen in contact with Stefan Heinzl and Kimmo Koski.

#### **Joint activities**

The REPLIX project is targeting at a solution for safe data replication and synchronization in distributed repository scenarios which take care of data authenticity, maintenance of rights etc. Currently the work is making use of the iRODS software to check whether it is suitable to establish a lean, but robust data replication layer on top of existing repository systems.

<http://www.mpi.nl/research/research-projects/language-archiving-technology/replix>

## **Goals of the co-operation**

Offering distributed computing facilities for humanities research.



## **Relevance of the liaison for the project**

AP2 (technical infrastructure)

### **2.1.6 DELOS**

#### **Arrangements**

Co-operation with DELOS, an association of Digital Libraries

#### **Details about the project / initiative**

DELOS was a Network of Excellence on Digital Libraries partially funded by the European Commission in the frame of the Information Society Technologies Programme (IST), and which is now continuing activities in the form of a scientific association. The main objectives of DELOS are research, whose results are in the public domain, and technology transfer, through cooperation agreements with interested parties.

DELOS is currently working on the development of a Digital Library Reference Model that is designed to meet the needs of the next-generation systems, and on a globally integrated prototype implementation of a Digital Library Management System, called *Delos DLMS*, which will serve as a concrete partial implementation of the reference model and will encompass many software components developed by DELOS partners. They are two major steps in the direction of the DELOS vision.

We are, via the MPI Nijmegen, in contact with Donatella Castelli.

#### **Goals of the co-operation**

Coordinated work on digital library issues, e.g. repositories and metadata.

## **Relevance of the liaison for the project**

AP2 (technical infrastructure)

### **2.1.7 DRIVER**

#### **Arrangements**

Cooperation with the EU DRIVER initiative is planned. CLARIN centres and aggregated services will form nodes in the robust network of DRIVER content providers. It will offer its metadata descriptions so that DRIVER can harvest them and offer them in its services (see also CLARIN Newsletter #2. p. 11).

#### **Details about the project / initiative**

DRIVER is a multi-phase effort whose vision and primary objective is to create a cohesive, robust and flexible, pan-European infrastructure for digital repositories, offering sophisticated services and functionalities for researchers, administrators and the general public.

The congregation of such diverse content and services on the common DRIVER platform is of immense value to both specialized research communities and the general public.

We are, via MPI Nijmegen, in contact with Norbert Lossau and Donatella Castelli.

#### **Goals of the co-operation**

We want to connect humanities repositories to the DRIVER infrastructure.

## **Relevance of the liaison for the project**

AP 2 (technical)

## 2.1.8 nestor (German competence network for digital preservation)

### **Arrangements**

The Institut für Deutsche Sprache (IDS) is partner in nestor, as well as in D-SPIN, and liaises between the initiatives. Christina Bankhardt (IDS) is a member of the nestor group on legal aspects of long term preservation and also works in D-SPIN AP7 on legal issues.

### **Details about the project / initiative**

The German competence network for digital preservation nestor brings together experts from Libraries, archives, museums, and researchers from different disciplines to ensure the long-term preservation and accessibility of digital resources.

### **Goals of the co-operation**

Coordinated development of strategies for long term preservation of language resources.

### **Relevance of the liaison for the project**

AP 2 (technical) AP 7 (legal: preservation of resources affected by intellectual property rights or personal rights)

## 2.1.9 DFN (Deutsches Forschungsnetzwerk)

### **Arrangements**

In Germany various interactions with the DFN eV were carried out which led finally to signing a joint agreement between the CLARIN service provider federation (SPF) and the German national identity federation, i.e. one signature was sufficient to establish a trust domain in which already now many German researchers will be accepted by the service providers with their home identity. Currently 3 German centers are participating in the initial SPF. After tests have been carried out successfully other D-SPIN centers will be integrated.

### **Details about the project / initiative**

DFN offers its users a multitude of customized applications for communication. It is constantly being enhanced by new and innovative applications which are developed and tested in projects and piloting ventures.

DFN provides further training to its users conducting conferences and workshops. Moreover, there are contacts in various competence centres available to provide advice and support with important questions concerning the use of the services.

The association to promote a German education and research network, DFN-Verein, organizes DFN and guarantees its further development and use. DFN-Verein is an acknowledged non-profit association. (cf. <http://www.dfn.de/index.php?id=74989&L=2>).

### **Relevance of the liaison for the project**

AP 2 (technical)

## 2.1.10 Grid Initiatives

### **Arrangements**

D-SPIN had three meetings with typical grid initiatives (cf. <http://web.eu-egi.eu/>, <http://www.unicore.eu/>) to discuss the usefulness of the existing middleware to realize D-SPIN interests. Two meetings were carried out with the developers of UNICORE (Forschungs-Zentrum Jülich) which resulted in deep understanding on how UNICORE could be used for CPU demanding activities and in a roadmap for a joint project. A discussion with EGI at CERN was devoted to also find out what role gLite and various other components could have in the D-SPIN developments. Yet there are no concrete plans for ongoing activities, but intentions for further interactions have been sketched.

### **Joint activities**

Currently a concrete project plan is being worked out together with the UNICORE group to test out the suitability of the software for CLARIN purposes. We expect that there will also be a concrete collaboration with EGI of some sort.

### **Relevance of the liaison for the project**

AP 2 (technical)

## 2.1.11 PARADE/EUDAT

### **Arrangements**

D-SPIN should be a demo case for this project, in particular the web services infrastructure

### **Details about the project / initiative**

Initiative of some communities (among them CLARIN) and computing centers for establishing a pan-European data services landscape. Initiative will participate with a concrete proposal in an EC call.

### **Joint Activities**

Currently this group is working towards a proposal to the EC for a common data service infrastructure. Aside from D-SPIN two strong computer centers from Germany that are also involved in DEISA and PRACE are in leading positions (RZ Jülich, RZ Garching).

### **Relevance of the liaison for the project**

AP 2 (technical)

## 2.1.12 OpenAir

### **Arrangements**

First discussions with the OpenAir initiative have taken place to talk about a “data continuum” that will established an integrated linked domain ranging from primary research data to publications.

### **Details about the project / initiative**

Initiative of major European libraries towards establishing a repository for publications. There is no website yet

**Relevance of the liaison for the project**

AP 2 (technical), AP 5 (curation of metadata for language resources)

## **2.3 Projects / initiatives which establish research infrastructure**

### **2.3.1 Textgrid / Wissgrid**

#### **Arrangements**

Close strategic cooperation in the planning of a German “Agenda 2020” in the field of eHumanities.

#### **Details about the project / initiative**

The joint project **TextGrid** aims to support access to and exchange of data in the arts and humanities by means of modern information technology (the grid). In 2006 development began on a web-based platform, one which will provide services and tools for researchers for analysis of text data in various digital archives - independently of data format, location and software. TextGrid serves as a virtual research environment for philologists, linguists, musicologists and art historians. TextGridRep is a long-term archive for research data in the humanities embedded in a grid infrastructure, which will ensure availability and access to its research data as well as interoperability. The joint research project TextGrid is part of the D-Grid initiative

As a D-Grid project, WissGrid’s objective is to establish long-term organizational and technical D-Grid structures for the academic world. **WissGrid** combines the heterogeneous needs from a variety of scientific disciplines and develops concepts for the long-term sustainable use of the organizational and technical grid infrastructure. In this context, the project aims to strengthen the organizational cooperation of scientists in the grid and to lower the entry barriers for new community grids (cf. <http://www.d-grid-ggmbh.de/index.php?id=102&L=1>).

The Institut für Deutsche Sprache is partner in both the D-SPIN and the TextGrid/WissGrid project.

#### **Goals of the co-operation**

Ensure effective liaison and cooperation and to avoid overlaps of activities between TextGrid and D-SPIN projects; reaching the Humanities community at large.

#### **Relevance of the liaison for the project**

Strategic.

### **2.3.2 eAQUA**

#### **Arrangements**

Close strategic cooperation in the planning of a German “Agenda 2020” in the field of eHumanities.

#### **Details about the project / initiative**

The eAqua-project aims at generating specific knowledge from ancient texts and will provide this knowledge via an open web-portal to the scientific community for future empirical studies. For this purpose researches from the fields of Computer Science and Ancient Science will cooperate to adapt the available text mining technologies to the needs and requirements of the Ancient Studies (cf. [www.eaqua.net](http://www.eaqua.net)).

The Institut für Informatik, department automatic language processing (ASV) is partner in both the D-SPIN and the eAQUA project.

### **Goals of the co-operation**

Ensure effective liaison and cooperation and to avoid overlaps of activities between eAQUA and D-SPIN projects; reaching a small but interesting sub-community.

### **Relevance of the liaison for the project**

Strategic. AP 3 (requirement analysis)

## **2.3.3 DARIAH**

### **Arrangements**

Close co-operation on the European level. Discussing interoperability of the technical architectures of the two projects.

### **Details about the project / initiative**

<http://www.dariah.eu/>; Peter Doorn, DANS

### **Goals of the co-operation**

Ensure effective liaison and cooperation and to avoid overlaps of activities between DARIAH and CLARIN/D-SPIN projects.

### **Joint activities**

In collaboration with all other four infrastructure initiatives from the social sciences and humanities, CLARIN is now preparing a proposal to the EC where the collaboration on cross-disciplinary data services are being targeted. This proposal will intensify the collaboration along concrete projects and in doing so identify common requirements and solutions.

### **Relevance of the liaison for the project**

Strategic.

## **2.3.4. FLARENET/META-NET**

### **Arrangements**

At the recent Barcelona meeting (see below) clear agreements were made about a close synchronization between CLARIN and the META-NET project to take profit from existing components.

### **Details about the project / initiative**

FlareNet will be a European forum to facilitate interaction among LR stakeholders and its deliverables will take the form of policy recommendations to the EC in the area of language resources. In the area of language technology traditionally activities are funded by DG INFSO to overcome current barriers, thus to build cutting edge technology and new resources to train and test this technology. Of course there is much overlap with this community requiring a proper share of the labor and responsibilities. These were addressed at the two FLARENET conferences (2009/2010).

META.NET is a Network of Excellence dedicated to building the technological foundations of a multilingual European information society. The project started in February 2010. This NoE is co-

funded by the 7th Framework Programme of the European Commission through the contract [T4ME](#) grant agreement no.: 249119 (<http://www.meta-net.eu/>)

### **Events**

Common meeting in Berlin, July 2009, with a focus on standards for language resources. Joint CLARIN/FLARENET workshop on standards with D-SPIN participation at the NEERI in Helsinki, September 2009 (see below)

### **Goals of the co-operation**

Avoid overlapping of work. Propagation of our results to a wider, more technical oriented community.

### **Relevance of the liaison for the project**

AP 5 (language resources).

## **2.3.5 CESSDA**

### **Arrangements**

Cooperation with CESSDA (Council of European Social Science Data Archives).

### **Details about the project / initiative**

CESSDA is an umbrella organization for social science data archives across Europe. Since the 1970s the members have worked together to improve access to data for researchers and students. CESSDA research and development projects and Expert Seminars enhance exchange of data and technologies among data organizations, for further details, cf. <http://www.cessda.org>.

### **Events**

Participated and presentation of project in CESSDA WP4 workshop on terminology and nomenclature (Lothar Lemnitzer).

### **Joint activities**

In collaboration with all other four infrastructure initiatives from the social sciences and humanities, CLARIN is now preparing a proposal to the EC where the collaboration on cross-disciplinary data services are being targeted. This proposal will intensify the collaboration along concrete projects and in doing so identify common requirements and solutions.

### **Goals of the co-operation**

Search for common solutions to shared technical challenges.

### **Relevance of the liaison for the project**

AP 5 (language resources).

## **2.4 Projects / Initiatives in the field of language resources and tools**

### **2.4.1. ESF Eurobabel**

#### **Arrangements**

Requirement analysis wrt to language resources with the subprojects of EUROBABEL.

#### **Details about the project / initiative**

ESF Eurobabel projects, cf. <http://www.esf.org/activities/eurocores/running-programmes/eurobabel.html>: “The main purpose of the EUROCORES programme EuroBABEL is to promote empirical research on underdescribed endangered languages, both spoken and signed, that aims at changing and refining our ideas about linguistic structure in general and about language in relation to cognition, social and cultural organization and related issues in a trans-/ multi-disciplinary perspective.”

#### **Events**

Peter Wittenburg and Tamás Váradi attended a workshop on 25th November 2007 in Paris organised by the ESF to prepare the EUROBABEL project (Better Analysis Based on Endangered Languages), where CLARIN was widely promoted as the most suitable infrastructure to take care of the data that arose out of the project.

Ralf Gehrke (Univ. Frankfurt partner) attended the kick-off meeting of the project and promoted D-SPIN, in particular as a repository solution for acquired data. The project will help us with understanding the requirements of the community.

#### **Relevance of the liaison for the project**

AP 3 (requirement analysis)

### **2.4.2. LOEWE Cluster “Digital Humanities”**

#### **Arrangements**

Requirement analysis wrt to language resources with members of the cluster.

#### **Details about the project / initiative**

The LOEWE cluster “Digital Humanities” is planned to be an inter-disciplinary cluster of projects in the field of text and picture corpora and archives with an emphasis on multi-modal corpus management. The participants from the institutions involved (Goethe University Frankfurt, Technical University Darmstadt, Freies Deutsches Hochstift/Goethe Museum, Staedel Museum) cover History of Art, History, Literature, and areas within Linguistics and Information Technology. The cluster proposal is currently under review. The start is scheduled for January 2011.

#### **Events**

Contact has been established with partners from the History of Art and from History who declared their readiness for interviews.

#### **Relevance of the liaison for the project**

AP 3 (requirement analysis)



### 2.4.3. AMICUS

The project AMICUS, *Automated motif discovery in cultural heritage and scientific communication texts*, funded by [NWO](http://ilk.uvt.nl/amicus/) in the Netherlands (<http://ilk.uvt.nl/amicus/>). Cooperation has been established in the topic of standardized linguistic annotation and annotation of topics typical for certain genres, using for this as a base the WebLicht web services infrastructure developed in D-SPIN. Two papers (commonly produced by D-SPIN and AMICUS) describing this work have been accepted for LREC 2010 (<http://www.lrec-conf.org/lrec2010/>) and DH2010. (<http://www.cch.kcl.ac.uk/dh2010/>)

**Events** AMICUS plans a joint workshop with D-SPIN members at a CLARIN Event in October 2010.

#### **Relevance of the liaison for the project**

Through this liaison we ensure a strong link between the linguistic resource community and the eHumanities in the field of literature.

## **2.5 Projects / Initiatives which build and provide standards**

### **2.5.1 ISO**

#### **Arrangements**

Participation in ISO TC37 SC 4, language resources management and in the DIN mirror group. Some D-SPIN consortium members are active participants in the discussion via DIN (Saarbrücken, Stuttgart, IDS Mannheim).

#### **Details about the project / initiative**

The objective of ISO/TC 37/SC 4 is to prepare various standards by specifying principles and methods for creating, coding, processing and managing language resources, such as written corpora, lexical corpora, speech corpora, dictionary compiling and classification schemes. These standards will also cover the information produced by natural language processing components in these various domains. Standards produced by ISO/TC 37/SC 4 should particularly address the needs of industry and international trade as well as the global economy regarding multi-lingual information retrieval, cross-cultural technical communication and information management. The goal of ISO/TC 37/SC 4 is also to ensure that new developments in language engineering, knowledge management and information engineering satisfy the norms of international standardization for:

- developing standards and related documents to maximize the applicability of language resources,
- relating the language resources of different kinds to their applications, and
- enhancing the application of recognized methods and tools in language resources.

For further details, cf. <http://www.tc37sc4.org/>.

Several D-SPIN members are deeply involved in standardization activities in the realm of ISO TC37/SC4; to indicate a few main dimensions in which D-SPIN members are taking leading roles: (1) annotation formats; (2) lexicon format (LMF); (3) model for a “concept” registry; (4) specification for tag sets for various linguistic levels including metadata and (5) model for persistent identifier services. This resulted in concrete activities and implementations already which will be indicated below. Obviously the implicit worldwide discussion process gives D-SPIN members a prominent role and visibility.

#### **Events**

Participation of CLARIN consortium partners at TC37 and SC meetings and ISO meetings, e.g. September 2008. Active involvement of experts from this field in D-SPIN workshops (September 2009 in Berlin, January 2010 in Stuttgart)

#### **Joint Activities**

A number of concrete collaboration projects have been realized between D-SPIN and the ISO TC37/SC4 community to drive ahead standardization and to test out standard proposals:

- the lexicon LMF model has been implemented in the LEXUS tool  
<http://www.lexicalmarkupframework.org/>  
<http://www.lat-mpi.eu/tools/>
- the data category (concept) registry is being implemented in ISOcat;

[http://www.tc37sc4.org/new\\_doc/ISO\\_TC\\_37-4\\_N175\\_SEW-A\\_Global\\_Data\\_Category\\_Registry.pdf](http://www.tc37sc4.org/new_doc/ISO_TC_37-4_N175_SEW-A_Global_Data_Category_Registry.pdf)

<http://www.isocat.org/>

- the persistent identifier model led to the foundation of the EPIC consortium;  
<http://www.pidconsortium.eu/>
- converters for the various annotation standards are being studied and worked out by the WebLicht group;  
<http://weblicht.sfs.uni-tuebingen.de/Weblicht.pdf>
- the metadata categories in ISOcat are defined by the CMDI work in D-SPIN.  
<http://www.clarin.eu/files/wg2-4-metadata-doc-v5.pdf>  
<http://www.isocat.org/>

Thus D-SPIN is leading some implementation and testing work which is so important.

### **Relevance of the liaison for the project**

AP2 and AP5 (standards for data and metadata).

### **2.5.2 TEI**

#### **Arrangements**

Institut für Deutsche Sprache is partner in the TEI, as well as in D-SPIN, and is able to liaise between the initiatives.

#### **Details about the project / initiative**

The Text Encoding Initiative (TEI) is a consortium which collectively develops and maintains a standard for the representation of texts in digital form. Its chief deliverable is a set of Guidelines which specify encoding methods for machine-readable texts, chiefly in the humanities, social sciences and linguistics.

The TEI Consortium is a non-profit membership organization composed of academic institutions, research projects, and individual scholars from around the world, cf. <http://www.tei-c.org/>.

#### **Events**

The annual members meetings in London (November 2008) and Ann Arbor (November 2009) were attended by Andreas Witt (IDS).

#### **Goals of the co-operation**

Adaptation of emerging standards for the D-SPIN language resource (e.g. metadata header).

### **Relevance of the liaison for the project**

AP 5 (language resources and standards)

### **2.5.3 RELISH**

#### **Arrangements**

Participation of the relevant D-SPIN partners (i.e. those who are providing lexical resources) in the ongoing discussion.

### **Details about the project / initiative**

Relish is a transatlantic initiative with the aim of harmonizing the currently used standards for the encoding of machine-readable dictionaries and other lexical resources (i.e. TEI chap. 12, LMF, LIFT).

Two institutions (i.e. MPI Nijmegen and Univ. Frankfurt) are partners in both the D-SPIN and the RELISH project.

### **Events**

#### **Goals of the co-operation**

Adaptation of dictionary encoding to existing/emerging standards in both Europe and the US to enhance interoperability

### **Relevance of the liaison for the project**

AP 5 (language resources and standards)

## **2.6 Other Initiatives and events**

### **2.6.1 CLARA**

#### **Details about the project / initiative**

CLARA (Common Language Resources and their Applications, <http://clara.uib.no>) is a Marie-Curie program funded by the EC which will be used to train young academics in CLARIN/D-SPIN technologies. Several summer schools are being planned where D-SPIN members will play an active role and where D-SPIN technology will be demonstrated. Two D-SPIN partners – Tübingen University and the MPI Nijmegen – are members of CLARA's consortium of 9 academic partner institutions. Positions for early stage researchers (ESR) and experienced researchers (ER) will be available at CLARA's partner institutions to integrate PhD students in concrete research work.

#### **Events**

Summer schools (to come)

#### **Relevance of the liaison for the project**

AP 6 (education)

### **2.6.2 NEERI Conference**

Together with DG Research, DG INFSO, DARIAH, FLARENET, DEISA and EGEE the NEERI conference (2009, cf. <http://www.csc.fi/english/pages/neeri09>) was organized in Helsinki with the goal to invite all ESFRI infrastructure and e-Infrastructure initiatives to interact about certain selected topics of general interest to look for cross-infrastructure synergies. It was one of the most challenging events during the last years and led to a number of concrete activities which will be mentioned below. It also led to a much better and deeper understanding of commonalities and differences between the different communities.

#### **Relevance for the project**

AP 2 (technical), 5 (language resources) and 7 (legal aspects)

### **2.6.3 EGEE User Forum**

The first EGEE user forum in Barcelona (2009, cf. <http://egee-uf4.eu-egee.org/index.php?id=498>) gave the possibility to a number of infrastructure initiatives to present themselves and to get a first understanding of the goals. Despite a highly interesting and partly controversial discussion it was certainly too early to come to joint activities.

#### **Relevance for the project**

AP 4 (Liaison activities)

#### 2.6.4 PARADE meetings

A number of meetings within the PARADE initiative (cf. <http://www.csc.fi/english/pages/parade>) have taken place which was devoted to design the essential pillars of a common data services e-Infrastructure in Europe. PARADE brings together strong data and compute centers and 12 research communities. In these activities D-SPIN was one of the most active groups since it obviously belongs to those communities that have the clearest perspectives about the needs of a common data services infrastructure, thus functionalities which D-SPIN centers can or should not provide. These meetings resulted in a White Paper that was sent to the EC and a D-SPIN member was one of the editors of this White Paper.

#### **Relevance for the project**

AP 2 (technical)



## 2 Conclusions and Roadmap

D-SPIN has done a lot in broad networking and more important not just heading for joint meetings, but always with the intention in mind to come to joint activities that really foster mutual understanding by joint commitments. The disadvantage of this approach is that it costs an enormous effort on all sides and that it takes time. One of the big advantages of such an approach is that young people are trained in these collaborative efforts and adopt these interdisciplinary cultures.

D-SPIN has thus established links with regard to various relevant aspects and wants to continue this policy:

- a firm collaboration with standard organizations also based on concrete projects
- a project oriented collaboration with all e-Infrastructure initiatives such as GEANT, GRID, DEISA, PARADE/EUDAT and e-IRG although this needs to be extended in the grid domain
- a collaboration with the other SSH initiatives although also here concrete collaboration projects need to be defined
- a collaboration with the libraries but also here concrete collaboration projects need to come
- participation in a variety of education and training programs to educate young students in particular

D-SPIN will further look for collaborations, but these only make sense when the domain traditions and solutions will be respected to a certain extent. Grid solutions that may work in the high energy physics may be inadequate for the humanities for example. Only via joint projects of limited size this can be tested and only via such an approach smart APIs can be worked out.



## ***Literature***

CLARIN Newsletters issues 1-4, edited by Dan Cristea and Marko Tadić (available at: <http://www.clarin.eu/newsletter>)