Report AP 9: Dissemination

April 2012
1. Introduction .......................................................... 4
2. Institutions and People ................................................. 4
3. Strategic Decisions ..................................................... 4
4. Activities and Materials Produced ...................................... 5
   4.1 Newsletter .......................................................... 5
   4.2 Flyer ................................................................ 5
   4.3 Events .................................................................. 6
   4.4 CLARIN-D web site .................................................... 6
   4.5 Other web sites ........................................................ 9
5. Strategic Planning .......................................................... 9
1. Introduction

Within the German CLARIN-D initiative, BAS is, in conjunction with IMS of the University of Stuttgart, responsible for the work package AP 9 “Dissemination”. The goal of this work package is to make CLARIN-D visible to the scientific communities, to the CLARIN initiative proper and to the general public.

In the grant contract, measures to achieve this visibility were defined. These measures include the publication of a newsletter, the definition of templates for poster and computer presentations, and setting up a web site for CLARIN-D.

2. Institutions and People

The Bavarian Archive for Speech Signals (BAS) is hosted by the Institute of Phonetics and Speech Communication at Ludwig Maximilian University Munich. Within BAS, the project leader PD Dr. Christoph Draxler is also responsible for BAS contributions to AP 9. Fabian Bross is responsible for the design and editing of the newsletter, which is scheduled to appear every 6 months.

The Institut für Maschinelle Sprachverarbeitung (IMS) is part of Faculty 5 (“Computer Science, Electrical Engineering and Information Technology”) at the University of Stuttgart. The head of the CLARIN-D centre at IMS is Prof. Dr. Jonas Kuhn, the person responsible for the construction, support and maintenance of the CLARIN-D web site is Jens Stegmann.

3. Strategic Decisions

At a preparatory meeting it was decided that CLARIN-D should inform about its aims, contents, status and progress via several communication channels:

• a web site on CLARIN-D, with public and restricted access pages, was to be set up by IMS Stuttgart,

• a newsletter should be published every 6 months by BAS, and

• a paper flyer should be designed and made available by BAS.

These information media should be used by all CLARIN-D partners; furthermore, all CLARIN-D partners need to contribute information about their activities for the sake of the dissemination of information about CLARIN-D.
4. Activities and Materials Produced

4.1 Newsletter

It was decided that the general concept of the newsletter should be to

- present background information on CLARIN-D and its centres,
- report on important CLARIN-D events such as workshops, tutorials and press conferences,
- provide a forum for guest authors from related projects, and
- present use cases from scientific applications with relevance for CLARIN-D.

The BAS published the first newsletter in November 2011. It is written in German, it is 11 pages long, and it is published electronically.

It was one of the first CLARIN-D publications to receive a persistent identifier (PID) – this shows that we are serious about making CLARIN-D a permanent infrastructure!

The newsletter can be found at

http://clarin-d.net/index.php/de/aktuelles/newsletter

4.2 Flyer

After a series of discussions during the CLARIN-D meetings, the contents of the CLARIN-D flyer were fixed. The long discussions were due to a perceived deadlock situation: the flyer should describe the merits of the CLARIN-D infrastructure, but this infrastructure was not in place yet.

To overcome this deadlock, the flyer now focuses on the potential use of CLARIN-D for scientists and researchers in the humanities and the social sciences. Three use cases are presented in a question and answer dialog:

Q: How can I share my data and results with colleagues?

A: Secure authentication in CLARIN-D and standardised data exchange formats allow your colleagues to access your data and the use of web services.

Q: How do I get the output of my tagger into my preferred parser?

A: CLARIN-D offers chaining linguistic tools. This way, users can select from a list of tools the ones that they need and chain them in a useful way – without having to worry about the technical details.

Q: How do I find data that is relevant to my work?
A: In CLARIN-D, data and services are described via metadata. These metadata can be searched, and the search results can be compiled to virtual workplaces for further processing.

The flyer follows the standard format (A4 landscape paper folded in three columns). It was presented at the CLARIN-D project meeting in Saarbrücken in March 2012, and has been distributed to all CLARIN-D centres for further dissemination.

A second flyer, listing the tools and services that are offered by the CLARIN-D centres, is currently being prepared by MPI and Tübingen.

4.3 Presentation materials
BAS created a set of templates for poster and computer presentations in LaTeX-format. These templates are available under

http://www.clarin-d.de/mwiki/index.php/Material

(Note: login required)

4.4 Events
CLARIN-D was present at the annual meeting of the Institut für Deutsche Sprache (IDS) in Mannheim. Every year, the IDS organizes a technology fair, where developers and researchers can demonstrate speech and language processing tools, present new speech databases and corpora, and show how technology can make work easier.

This year, CLARIN-D participated with the following contributions:

<table>
<thead>
<tr>
<th>Centre</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>U Tübingen</td>
<td>WebLicht</td>
</tr>
<tr>
<td>MPI</td>
<td>ELAN</td>
</tr>
<tr>
<td>U Hamburg</td>
<td>Exmaralda</td>
</tr>
<tr>
<td>BBAW and MPI</td>
<td>TLA</td>
</tr>
<tr>
<td>IDS</td>
<td>Folk &amp; Folker</td>
</tr>
<tr>
<td>U Hamburg and IDS</td>
<td>LiS</td>
</tr>
</tbody>
</table>

4.5 CLARIN-D web site

Planning
Assistance was available in the form of the CLARIN-D developers group, which devised a task force that included colleagues from Tübingen, Nijmegen, Hamburg and Stuttgart. They provided support regarding the initial decisions that had to be made, i.e., which content management system to use and which domains to register for the site. Regarding the former, we decided to use the Joomla CMS due
to its open nature, the level of community-driven support and the immense number of extensions available, among others. With respect to the latter, after some discussion the following domains were chosen for the project: clarin-d.de, clarin-d.org and clarin-d.net. Furthermore, the colleagues in Nijmegen offered a subdomain from the web site of the European branch of the CLARIN project: de.clarin.eu. All four should be used for the CLARIN-D web site. Generally, it was decided that the external part of the site would be focussing on issues in public relations, while the protected parts would be used for means of project-internal coordination. With those choices in place, legal peculiarities specific to the hosting of a web site for a federated project like CLARIN-D at a University in the German federal state of Baden-Württemberg had to be sorted out.

In parallel, we began the planning of the layout, navigation and the content of the web site. In accordance with the guidelines by the BMBF, issues in usability and accessibility had to be taken into account, especially regarding the template/stylesheets to be used. Beez 2.0 by Angie Radtke, one of the standard Joomla templates, was identified as a natural candidate due to its adaptability and the fact that it meets the requirements of major accessibility standards like WCAG 2.0. Main themes in this respect are a clear separation of content and layout, logical structure, the provision of anchor links, navigability via keyboard and sufficient color contrasts. Finally, it became clear that we would want to make use of an associated project wiki in addition to the internal part of the web site provided by the Joomla CMS. Other wishes for advanced functionalities brought forth by the colleagues in Tübingen and Nijmegen included a calendar for CLARIN-D related events as well as a forum component for internal discussions. Suitable solutions in the form of Joomla extensions were identified.

**Construction and Testing**

We registered the domains with an external domain hoster, delegated the subdomain in cooperation with Nijmegen, and set up a DNS / network configuration in order to make use of all of them on the same web server. The latter was realized as a virtual machine with more than adequate storage and working memory concerning our purposes (100GB and 4GB, respectively) and could be hosted at the IMS itself as soon as the legal situation was clear. We have been using an Ubuntu 10.04 LTS (Long-Term Support) Server Edition operating system, and the Joomla CMS was to be configured on that machine as a LAMP stack (Linux, Apache, MySQL, PHP) using up-to-date software packages. Concerning the Joomla CMS itself, we chose to make use of the then-current STS (Short-Term Support) version 1.7.x as opposed to the then-still-current LTS version 1.5.x due to plans to migrate quickly to the next LTS version that would be available in early 2012.

The foundational menu structures for the web site and a flanking category system for articles were implemented via standard Joomla mechanisms. The menu structure involved the following main points: a starting page, a detailed project description, information regarding the partners involved (CLARIN-D Centers, Data and Computing Centers, Humanities and Social Sciences), information regarding the activities brought forth by the project (Work Packages, Discipline-specific Working Groups, Infrastructure Working Groups), CLARIN-D in the press, CLARIN-D related events, a link list focussing on other related projects, legal information and also contact information.

The internal section of the site was brought about, too. Joomla extensions like Kunena (forum component) and Jevents (calendar component) were installed and configured for project-related
usage in that part only. In addition to that restricted area, an installation of MediaWiki (the then-current v1.17.x) was set up in order to be used as the CLARIN-D project wiki. Search functionality and login functionality including captchas and account management for the internal part of the web site were included, too.

The colleagues in Tübingen provided adaptations of the original Beez 2.0 template as well as images/logos and also helped with provision of content. Texts for the different sections were adapted from existing sources or had to be written from scratch. Regarding the wiki, some basic content structure to start out with was created. Joomla/Wiki accounts for colleagues, who had signalled an interest in participating in the testing phase, were set up on the onset of the latter on September 9th, 2011. During the testing phase, we received valuable feedback from our colleagues in Tübingen, München, Leipzig, and Nijmegen among others. The full launch of the web site was on September 16th, 2011. This also marked the start of full user support regarding account credentials, questions etc.

**Maintenance**

This will include issues in the extension of the coverage, content and functionalities that followed as well as the administration and support that has to be provided, too.

As suggested by the colleagues in Tübingen, our first priority was to extend the coverage of the site with respect to an English language version and also concerning the content offered. We translated all of the existing textual material into English. Tübingen provided support in improving our initial English draft. Additional menu structures had to be built and to be associated with their German counterparts.

Regarding the content, more detailed descriptions of the work packages had to be provided by the responsible CLARIN-D centers. Furthermore, the leaders of the discipline-specific working groups were asked to contribute information on their groups. This endeavor was coordinated with the colleagues in Leipzig. On a related note, we asked all centers for texts regarding the language resources and services that would be offered in CLARIN-D. All texts were provided in a German and an English version.

We edited and normalized the texts, included them on the web site, and added relevant menu structures, e.g. for „Language Resources and Services“ (Archives, Corpora, Lexical Resources, Tools and Services, WebLicht).

The existing structures and contents had to be reorganized a bit, e.g. a new menu point „News“ (Newsletter, Events, Press Reviews) was introduced. Additional functionalities were included in the site when requested, e.g., extensions for mailing/newsletters (including a respective subscription module available on the site and a mail configuration for clarin-d.de), form creation and inclusion, a document repository, and the embedding of pdf documents for the CLARIN-D newsletter.

Authorization via Shibboleth-based Single-Sign-On was prepared on the web server. However, we refrained from realizing this option for the Joomla CMS and the wiki due to the current situation with the attribute release policy of the majority of the Shibboleth Identity Providers in Germany.
(Our accounts need to be specific to certain individuals, we cannot simply grant access to all academic users from a certain institution.) With respect to changes and maintenance of the MediaWiki installation, we extended functionalities by providing semantic-web based technologies by means of an installation and foundational configuration of Semantic MediaWiki and related extensions, as requested by the colleagues in Saarbrücken.

On general terms, we are providing full administration including updates/upgrades and backups etc. for the underlying operating system, Joomla CMS and the MediaWiki. The Operation System is Ubuntu 10.04.4 LTS at the time of writing with all available updates included. Concerning the CMS, the migration to the new Joomla LTS version 2.5.x was brought about soon after its release in early February 2012. We are running the current version 2.5.4 at the moment. All the extensions are up-to-date. The MediaWiki has also been migrated and is up-to-date with version 1.18.2, too.

A major point that deserves to be mentioned is user support. We created and maintain the accounts of all the researchers associated with CLARIN-D at the centers, the discipline-specific working groups and also the people responsible at the associated high-performance computing centres that have asked for an account. Automatic registration is not an option, because we have to make sure that the people who apply are really associated with CLARIN-D in some meaningful way. Our user support also includes answering the users’ questions and reacting to problems perceived and suggestions brought forward.

Finally, there is a specific CMS delegate at each CLARIN-D centre, i.e., a person who is responsible to bring about changes to the content of the web site via the backend of the Joomla CMS. We provided a tutorial for the CMS delegates at the developers’ meeting in Berlin in mid-January 2012, which introduced them to the Joomla CMS and especially to the backend. This will have to be followed-up upon in the future.

There are plans for a relaunch of the web site, which will take up the suggestions brought forward by the user communities and the CLARIN-D centres. Some general lines will be sketched below in the section on “Strategic Planning”. The new version of the site is scheduled to be launched as soon as possible, but not later than the end of May 2012. Until its launch, only rather conservative changes will be made to the existing site that will pertain to the news section (“newsflash”), language resources and services offered, as well as training and educational material.

4.6 Other web sites

CLARIN-D is actively publishing information about its tools and services to the Virtual Language Observatory.

http://www.clarin.eu/vlw/observatory.php

5. Strategic Planning

The work package AP 9 requires input from all centers with regard to CLARIN-D information dissemination.
In line with a strategy brought forth by Tübingen, we are planning to modify the CLARIN-D web page. One point is that important CLARIN-D news shall be visible in a newsflash immediately when opening the page. Our pertinent services and offers need to be featured prominently on the starting page, too. We also need to include more dynamic content and pictures in general – and in particular on the starting page.

The majority of the texts needs to be reworked in order to be less technical and more appealing to the user communities. We also need more examples. Further features that will be included or featured more prominently are: a digital humanities calendar, helpdesk/support and training materials, a complete list of resources and services, acknowledgement information and a specific area for the user communities / discipline-specific working groups. Furthermore, we plan to collect information from all centres about international and national conferences and workshops at which CLARIN-D members presented their work. This information will also be presented on the web page including a project bibliography.

Last, but not least, we will have to rework the organisational workflow and the responsibilities for the different sections of the site. As we are making use of a modern CMS that features accounts and version control among others, it is possible to work on the different parts of the site in a collaborative way. The CMS delegates will have to play a key role in this in the near future.

Finally, we will ask young researchers who are already using CLARIN-D services, data and tools to present their work at important CLARIN-D meetings, e.g., the M12 milestone in Leipzig in June 2012.
6. Appendices

A1: CLARIN-D web site screenshot
A2: CLARIN-D Flyer
The flyer is printed on heavy semi-matte paper and available in two versions: a PDF version for download, and a cmyk-version for printing.