Common Language Resources and Technology Infrastructure

CLARIN NL

Interim Evaluation
Recommendation Response
17-09-2013

WWW.CLARIN.NL

CLARIN-NL Interim Evaluation April 2009- April 2013
Response to the IAP recommendations

By the CLARIN-NL Executive Board

Contents

Introduction

This document describes the actions taken or planned by the CLARIN-NL Executive Board (EB) in response to the recommendations of the International Advisory Panel (IAP) made in the Interim Evaluation Report.

Response to the recommendations by the IAP

The IAP formulated a number of specific recommendations for the CLARIN-NL project. They are listed here globally categorized by topic, followed by a response of the CLARIN-NL EB.

Dissemination & Outreach

• The successful experience of CLARIN-NL should be shared so that other new efforts all over the world can benefit from it. Further and more extensive outreach should be cultivated. The primary focus for this outreach will be on Europe, but it will also expand to the rest of the world. In countries like China and India, which are investing in NLP, a less complex presentation will suffice.

Response: We will continue to share our experiences with others internationally, as we did before, and will intensify these activities also beyond Europe in the coming years.

Centre assessment and access to resources

• CLARIN-NL should work out a time table for CLARIN Centre assessment including DSA assessment with the relevant parties.

Response: We fully agree with this recommendation. Currently 9 German centres have been certified. On August 30, 2013 a meeting has been held for the Dutch centres. They will sign in for the assessment this fall. It is planned to have most centres certified by the end of 2013. The 5 core centres cover the Dutch language data types pretty well. More university libraries and archives could join as D-centres, but this will have to happen within CLARIAH, since CLARIN-NL has no means left for this purpose.

• CLARIN-NL should provide access to the CLAVAS and TTNWW tools and describe their functionality via the web or other means of dissemination.

Response: The report mentions the URL where TTNWW can be reached on page 30 (http://yago.meertens.knaw.nl/apache/TTNWW/). The CLAVAS URL was not known yet at the time of writing. It is http://openskos.meertens.knaw.nl but will be fully functional only as of August 31, 2013. We have in the meantime also made available links to all kinds of applications, web services, demonstrators and data on the CLARIN-NL website http://www.clarin.nl/node/404 , as a first small step towards the portal. We target to make metadata for the software available by the end of 2013, enabling views on the tools like this one that are generated dynamically.

• CLARIN-NL should provide details about the current state of language resources available to SSH users beyond the level of metadata and should take appropriate actions to promote the use of these resources.

Response: User accessibility is available both via the Virtual Language Observatory (VLO) and via the Metadata Search (currently containing only a sample of all metadata). Access to data and tools must always be possible, in the CLARIN approach, via explicit metadata. A crucial and (almost[1]) obligatory piece of information in the metadata is a PID or URL to the actual resource, or at least –in the case of restrictions –to a website where a license can be concluded to obtain such a resource. So access to the real data is normally provided via the metadata. An action (started earlier this year) is ongoing to check whether for all data metadata are available and visible to users and whether the actual data are indeed accessible via these metadata.

User Involvement
• CLARIN-NL should consider broadening the scope of activities with potential user communities.

Response: In the first phase of the CLARIN-NL project we have concentrated on creating the various ingredients of the Dutch part of the CLARIN infrastructure. In the remaining phase exploitation of the infrastructure becomes more dominant so that actual usage of the infrastructure is becoming an important topic. The portal that is being developed will increase the visibility of the functionality of the CLARIN infrastructure and will be used to stimulate the actual usage. Several user communities for specific functionality have already been set up, and CLARIN-NL will stimulate the setting up of others and facilitate their activities in various ways (financially, logistically). A lot of the dissemination takes place via the CLARIN partners, but this is hard to monitor. All projects should provide use cases which can be shared via de CLARIN portal.

• It is so far not clear who the users are, what they use, and for which purposes. It is also not clear to what extent their needs are being met. CLARIN-NL should develop plans for systematic user evaluations or reports on concrete user experiences.

Response: We fully agree with the recommendation that we should use the results of the CLARIN-NL project that now have become available "as the basis for targeted dissemination activities", thus increasing the actual usage of the infrastructure. Concomitantly we will develop plans for systematic user evaluations in the fall of 2013.

• CLARIN-NL should implement ways of tracking usage (e.g. by weblogs, number of licensees of particular resources, email exchanges helping users of resources, etc.).

Response: We will assess methods for tracking usage and work out a plan for this by Dec 1, 2013. These methods must be such that the measurements are reliable, easy to perform and yield relevant results without hindering the users too much.

Federated Content Search
• CLARIN-NL should actively participate in joint European efforts for tackling the problems of federated content search, since these problems are not specific to CLARIN centres in the Netherlands.

---

[1] In some cases there are only metadata: corresponding data do not exist, or are not accessible, or exist only as analog data.

Response: We fully agree and already do this in the wider CLARIN context and also in projects such as EUDAT. We will monitor existing projects in this domain closely and participate in new international projects that might arise, perhaps in the context of the CLARIN ERIC. Federated content search is not a requirement in CLARIN-NL for centre certification, but it has our full attention.

Sustainability

• More attention should be given to the formulation of a viable business model for the enhancement of the database cultivation and the maintenance of leading edge language technology.

Response: We fully agree, and this topic has our full attention. The issue involves not only the need for money to keep the data and tools available in a sustainable way , but also requires a proper organisation of the user and developer communities involved, an explicit   life cycle model for data and tools, and its application to specific resources. We also closely cooperate with the NL eScience centre on the sustainability of software, which is more problematic than data.

IPR

• Though problems of IPR have been mostly avoided so far in the project, IPR issues should get more attention in order to ensure maximal availability of resources to the research community.

Response: We agree, and will work on this in close cooperation with the CLARIN ERIC IPR committee (where we have one representative). The recent statements by [Neelie Kroes](#)[2] are encouraging in this respect, though of course it still remains to be seen what will come of it. There are various concrete issues, e.g. some IPR limitations on National Library data, the data used in the Namescape project, proper arrangement of the IPR of some text corpora such as the Twente News Corpus, and others. We will investigate to what extent these and similar data can be made available to the CLARIN community. This subject goes beyond the scope of CLARIN-NL and should be dealt with internationally.

---

[2] http://www.youtube.com/watch?v=XH-iRBdv3ho&feature=youtu.be&a

Appendix I: Abbreviations and Acronyms

Acronym	Expansion
CLARIN	Common Language Resources and Technologies Infrastructure
CLARIN ERIC	ERIC legal entity set up for the CLARIN infrastructure
CLARIN-NL	The Netherlands national project that aims to design and construct the Netherlands part of the European CLARIN infrastructure
EB	CLARIN-NL Executive Board
ERIC	European Research Infrastructure Consortium
ES	Evaluation Secretary
IAP	CLARIN-NL International Advisory Panel
IPR	Intellectual Property Rights