Panel Session on Open and Persistent Access, SDH 2010, Vienna, October 20, 2010.
Convenor: *Steven Krauwer*
Contribution by *Jan Odijk*

Panelists *introduce themselves*
where do they come from and what is their role in the Open and Sustainable Access Universe.

My name is Jan Odijk from the University of Utrecht and I am the project director of CLARIN-NL, the Dutch national project related to the European CLARIN initiative. CLARIN and CLARIN-NL aim to create a research infrastructure for humanities researchers in which data and tools can cooperate seamlessly (are interoperable). From the CLARIN perspective, my interests in open access naturally focus on open access to DATA and TOOLS, and less to publications. Many humanities researchers still keep their data and tools to themselves, and therefore we have recently launched a call in which we urge all humanities researchers to make their data and tools maximally available for other researchers, and we offer them various ways of support for that. See http://www.clarin.nl/node/141 . In the context of the META-NET project, that I am also involved in, similar activities are ongoing, in particular in the creation of an open resource sharing facility called META-SHARE, this time aimed at researchers in and developers of multilingual language and speech technologies

Panelists each describe *what they mean by open and sustainable access*

Most data and tools in research activities are created with public funding and should therefore be made available to the wider research community. The CLARIN infrastructure offers a way for doing this. It makes it possible for researchers to store their data in the infrastructure; it gives guidelines for how to make metadata for the data and tools which can be used by browsing, searching and querying tools and interfaces so that they are accessible and visible to other researchers.

"Open Access" in this context means that researchers must make sure that they have the rights to make data and tools available to other researchers via the CLARIN infrastructure. Concretely, this means that they must make their data and tools available via a server of a (candidate) CLARIN-centre. The ownership of the data, especially if they are third party data (e.g. text corpora) or involve third parties (e.g. recordings of people) can be elsewhere and Open Access does not necessarily imply "open Source", though the latter is desirable whenever possible. I explicitly state that the data and tools must be made available to other **researchers**, which is slightly moiré restricted than to the general public. People are researchers if they are employed at or a student of an organization that has access to the CLARIN infrastructure (including universities, research institutes, libraries, archives, etc.). Possibly special provisions will be made for researchers not related to any such institution (individual researchers, retired researchers, etc.).

"Persistence" is guaranteed by the CLARIN-centres, which must comply with the Data Seal of Approval (http://www.datasealofapproval.org/). The latter guarantees appropriate facilities and procedures for long term preservation (guidelines 6 and 7). In addition, in the context of CLARIN they are required to use Permanent Identifiers to refer to the data and tools.

Panelists each describe what they think is the *most promising way* to arrive at open and sustainable access

It is not sufficient to just request or even require researchers to make their data and tools maximally accessible. They must also be supported in such activities. CLARIN-NL provides such support in many ways:

- It provides guidelines for standards for data both for data and tool formats and their semantics, so that the ideal of making tools and data from various independent sources interoperable can be realized.
- It provides assistance with applying CLARIN selected candidates for standards and best practices.
- It insists that funders require a plan for maximal open access of data and their sustainability as part of a research proposal and that they support such activities (logistically, financially, and with expertise) as part of a research project. We recently discussed this matter with NWO Humanities, who responded positively and have already implemented this requirement fors o-called Investment Projects.
- CLARIN-NL offers support to researchers to arrange IPR matters and ethical issues for data and tools and aims to make available systematic technical facilities for this very likely in collaboration with other infrastructures, which have to deal with similar problems).
- CLARIN-NL offers support itself (financially, logistically, and with expertise) to the curation of research data and tools.
    - We just closed the second call for such curation projects
    - We are working on a plan to offer a service to researchers to curate their data for them (and of course with them)

Panelists each describe what they think is the *biggest threat* for open and sustainable access

Threats:
Each threat also created opportunities, which I will mention as well.
- Researchers will only really comply if they see benefits for themselves. So we must make it possible to show them the benefits and to make them experience the benefits for their own research as soon as possible.
    - We work on that by first making all kinds of demonstrators, but a working, if only partially working CLARIN infrastructure must be set up as soon as possible (and we are working on that too)
    - A second way of making it beneficial to researchers is to make sure that publication of research data and tools gives credit to a researcher, much like making a publication of an article does that now today.
- The tools must be user-friendly, and it should be possible to use them without technical background. It will be quite a challenge to achieve this. Even if the tools are user-friendly, some training and education will be required (after all, even optimally working with a word processor or a spreadsheet requires a training of a few days). Furthermore,

each tool will have limitations, and it is important that the user is aware of these so that he/she uses the tool appropriately.

- o We provide various opportunities for this and attempt to make working with such tools in the CLARIN infrastructure an integral part of the normal humanities curriculum, both at the bachelor and masters level. We are working on a variety of initiatives for that in the Netherlands, and Utrecht in particular.
- Multiple initiatives for open access to data are currently running. We must make sure that we speak with one voice as a community to achieve the goal of maximal open access.
    - o Fortunately, we have close relationships with and partially participate in initiatives such as DARIAH, FLaReNet, META-SHARE, SPARC, etc.