



CLARIN-NL HELPDESK

Paul van Caspel

helpdesk@clarin.nl

Wat is de CLARIN Helpdesk?



- Het is een punt waar CLARIN-*NL* subprojecten vragen kunnen stellen mbt. de door CLARIN-*NL* vereiste en voorgestane technische eisen en aanbevelingen.
 - Metadata
 - AAI
 - PIDs
 - Webservices
 - Resourceformaten
- De CLARIN-*NL* helpdesk beantwoord vragen en verwerkt de antwoorden in de helpdeskinformatiesystemen voor toekomstig gebruik

Opzet CLARIN-NL helpdesk



- CLARIN-NL helpdesk portaal
 - Lijst met informatiebronnen
 - Frequently Asked Questions

- CLARIN-NL helpdesk e-mail: helpdesk@clarin.nl

- Infrastructuurspecialisten

Case:



- Mijn project is zojuist gestart en ik ben verantwoordelijk voor de metadata productie. Waar kan ik algemene informatie vinden over metadata?

✓ <http://trac.clarin.nl/>

Case:



Welcome to the Clarin-NL Wiki

In this section, generic information on Clarin topics can be found. Currently, collections of links to relevant documentation are presented with a short description of its content. This information is presented in chapters that reflect the division in Clarin topics made elsewhere:

- [Metadata](#)
- [ISOcat](#)
- [PIDs](#)
- [Webservices](#)
- [Formats and Standards](#)
- [AAI](#)

The Clarin-NL helpdesk also has a [Frequently Asked Questions](#) section. Any requests for changes or additions can be submitted by e-mailing the Clarin-NL [✉ helpdesk](mailto:helpdesk).

Metadata

⇒ <http://www.clarin.eu/files/wg2-4-metadata-doc-v5.pdf>

Title: Metadata Infrastructure for Language Resources and Technology
Date/Version: 2009-02-04 - Version 5

Content: This document gives an overview about how metadata descriptions are used until now, what the deficits of the current infrastructures are and which lessons we as community learned from about a decade of experience. Based on this the requirements for a new CLARIN approach are being worked out. This document will be discussed in the appropriate working groups and in the Executive Board. It will be subject of regular adaptations dependent on the progress in CLARIN.

⇒ <http://www.clarin.nl/system/files/BestPracticeGuide-V4.docx>

Title: Best Practice Guide for using CLARIN metadata components
Date/Version: 2010

Content: The Dutch CLARIN project "Creating and using CLARIN metadata components" was the first to actually test the use of components and to try to create metadata descriptions for resources available in two Dutch language resource centers: the Institute for Dutch Lexicology (INL) and the Meertens Institute. This "Best Practice Guide" is the result of this project. It will however in the future be extended with new experiences gained by new projects that will make use of the CMDI.

⇒ <http://www.clarin.eu/files/metadata-CLARIN-ShortGuide.pdf>

Title: Component Metadata
Date/Version: 2009-02

Content: In this A4 shortguide, some introductory information is provided on Component Metadata following the generic shortguide layout of "What is it?", "What is it for?", "Who can use it?", "When can it be used?" and "How does it work?".

⇒ <http://www.clarin.nl/system/files/clarin-md-component.pdf>

Title: Metadata and DCR
Date/Version: 2010-03-25

Content: In this presentation given at an ISOcat workshop in Utrecht, shortcomings were discussed in "traditional metadata" and benefits of "component metadata". To guarantee interoperability while using different components, data categories are discussed. Furthermore, a bigger picture is provided through some diagrams and finally, an overview of building and using components in practice is provided.

⇒ http://www.clarin.nl/system/files/LREC2010_Metadataproject_FdV-v1.1.ppt

Title: Creating & Testing CLARIN Metadata Components - A CLARIN-NL project
Date/Version: 2010-05-18

Content: This presentation was given at LREC in Malta and featured the topics "What is CMDI?", "What is the goal of our project?", "How to go from a resource to harvestable metadata?" and "Findings of the project and future challenges". This presentation has an accompanying text document that can be found ⇒ [here](#).

Case:



- Ik heb de documentatie doorgenomen en er is een profiel in de ComponentRegistry dat bijna voldoet aan onze eisen op een aantal belangrijke punten na. Wat is de beste manier om ervoor te zorgen dat het 100% matcht?
 - x <http://trac.clarin.nl/>
 - ✓ <http://trac.clarin.nl/trac/wiki/WikiStart/FAQ/>

Case:



Welcome to the Clarin-NL helpdesk FAQ

In this section, previously asked questions on Clarin topics can be found. This information is presented in chapters that reflect the division in Clarin topics made elsewhere:

- [Metadata](#)
- [ISOcat](#)
- [PIDs](#)
- [Webservices](#)
- [Formats and Standards](#)
- [AAI](#)

If the answer to your question is not here, please send an e-mail to the Clarin-NL [✉ helpdesk](mailto:helpdesk)

Metadata

Question	Answer
What is metadata?	It is data about data: information describing properties of linguistic resources. Think of the size of a corpus, the recording date of a speech file, the purpose for which annotations were created.
What is a metadata scheme?	A fixed set of elements for the description of resources. Think of the traditional filing cards in the library, specifying the writer and title of each book.
What metadata schemes are there for the description of linguistic resources?	Quite a few, examples are: Dublin Core (⇒http://en.wikipedia.org/wiki/Dublin_Core), OLAC which is an enriched version of Dublin Core (⇒http://www.clarin.eu/glossary/term/259), IMDI (⇒http://www.clarin.eu/glossary/term/137), the TEI header (⇒http://www.clarin.eu/glossary/term/136), ...
So what metadata scheme is being used within CLARIN?	Good question. In fact there is not such a thing as a single CLARIN metadata scheme. Practice showed that using a particular scheme for a large community (e.g. the humanities) often results in a mismatch between the chosen elements and the needs of the user.
If there is no single metadata scheme, how should I describe my resources to be compatible with the CLARIN infrastructure?	CLARIN proposes a component-based approach: you can combine several metadata components (sets of metadata elements) into a self-defined scheme that suits your particular needs. Of course you can share your profile with others (in fact we strongly advise that). If sharing the full profile is not an option, you still can use common components, e.g. a component to describe a sound recording. In case that still does not address your needs, it is even possible to create components yourself.
So where do I find more information about creating components, profiles and using profiles created by others?	⇒http://www.clarin.eu/toolkit
Where can I find all details, references and the background for this component-based metadata concept?	Check out the CLARIN specification document: ⇒http://www.clarin.eu/files/wg2-4-metadata-doc-v5.pdf
But if I deliver CLARIN metadata today, can you make my metadata available to the broad public?	Not yet - bringing all metadata descriptions together ("harvesting"), making them searchable ("indexing") and citeable ("creating persistent identifiers") is an important part of the infrastructure that CLARIN is building. As with all infrastructure these things require a solid base to build on. That base is currently being constructed, but this also means that there is currently no simple-to-use method for accessing the new CLARIN metadata.
How can I provide CMDI metadata over OAI-PMH?	See ⇒http://www.clarin.eu/node/3014
So what do I do if I would like to make my metadata available to the public right away?	Until the infrastructure for the component-based metadata is fully in place you can use one of the following metadata schemes: OLAC - see ⇒http://linguistlist.org/olac/olac.html IMDI - see ⇒http://www.mpi.nl/imdi/ Data described in one of both formats will be made available and searchable via the CLARIN catalog and the Virtual Language Observatory. Apart from that we will ensure that these metadata descriptions will be converted to CLARIN component-based metadata.
OK, my IMDI or OLAC metadata (describing linguistic resources) is ready, how to proceed now?	Send a mail to ✉dieter.vanuytvanck@mpi.nl and we will incorporate your metadata as soon as possible
What is this metadata harvesting thing? Are you growing potatoes?	No, we are not. It is a term used for gathering metadata descriptions from several locations and storing it in a central database. You can find the results of such a harvesting process at ⇒http://catalog.clarin.eu/ (click on OLAC data providers)

Case:



- Het antwoord op mijn vraag is niet af te leiden uit beschikbare documentatie of eerder gestelde vragen. Wat nu?
 - x <http://trac.clarin.nl/>
 - x <http://trac.clarin.nl/trac/wiki/WikiStart/FAQ/>
 - ✓ helpdesk@clarin.nl