

## **NARCIS: linking CRISs and OARs in the Netherlands. A matter of standards and identifiers**

**Elly Dijk**

Department of Research Information  
Royal Netherlands Academy of Arts and Sciences (KNAW)  
PO Box 95110  
1090 HC Amsterdam  
The Netherlands  
[elly.dijk@bureau.knaw.nl](mailto:elly.dijk@bureau.knaw.nl)

### **Introduction**

In the Netherlands the metadata of the national Current Research Information System (CRIS), the Dutch Research Database<sup>1</sup>, and the metadata of the repositories are brought together in the scientific portal NARCIS<sup>2</sup>. To gather the information from the different systems the OAI-PMH protocol is used.

The main goal of NARCIS is to be the central place for searching all these data (one-stop-shopping), and to create the national showcase of researchers in the Netherlands.

### **The Dutch research information landscape**

#### ***Current Research Information System - CRIS***

In the Netherlands all 14 universities and the Academy have their own CRIS, the so-called Metis-system. The metadata format for Metis is CERIF based<sup>3</sup>. These Metis-systems are not related to each other.

The national CRIS, the Dutch Research Database (NOD; also CERIF based) is partly filled with information from these Metis systems. In the Dutch Research Database the information from the 14 Metis systems comes together. In this database the names of persons and institutions are standardized and connected. Bibliographical information from the Metis systems on research results is not submitted yet.

#### ***Open Access Repositories***

Since 2004 in the Netherlands all universities, the Academy and the National Research Council (NWO) have an Open Access Repository. The National Library (KB) takes care of the long term preservation of the digital publications. Since 2007 also other research institutions have been set up repositories with open access publications.

The institute DANS (Data Archiving and Networked Services) archives research data in the humanities and social sciences. DANS has also set up a repository for these datasets, called EASY.

#### ***Harvesting***

In NARCIS both the Dutch Research Database and the repositories are being harvested, so here one can find information about researchers (expertise), research institutes and research activities as well as (full text) publications and datasets.

#### ***Digital Author Identifier-DAI***

To connect the information in NARCIS coming from all these different databases and repositories the first step has been to assign a Digital Author Identifier (DAI) to researchers in the Netherlands. This identifier acts like a digital glue in realizing the integration of information regardless of how the researcher's name, initials or surname are spelled in the different systems. Within the OCLC library system a thesaurus of author names with corresponding DAIs was created. All the involved institutions have matched the DAIs and author names in their own CRIS (Metis). This work was finished at the end of 2008. The next step is the implementation of the DAI in the repositories of the participants, in the NOD, and in EASY. We expect this to be ready at the end of 2010.

---

<sup>1</sup> See: <http://www.researchinformation.nl>

<sup>2</sup> See: <http://www.narcis.nl>

<sup>3</sup> See: <http://www.eurocris.org/cerif/introduction/>

**Interoperability: OAI-PMH protocol**

To gather the information from the different systems the OAI-PMH protocol is used.

Since the repository community had already implemented this protocol, it was a small step to implement the protocol for all systems included in NARCIS.

🕒 **Repositories and MODS**

Because DC is very confined, the NARCIS community has chosen the Metadata Object Description Schema (MODS)<sup>4</sup> for describing bibliographic metadata elements. There is a strong need for an extensible schema, because of the introduction of the DAI. This Dutch element is no part of standard metadata formats or schemas; MODS offers the possibility to simply add this element.

🕒 **Metadata format for CRIS**

For the harvesting of the CRIS records, a CERIF based format is used. Only the fields that have been placed in the index of the Dutch Research Database are harvested, thus only these fields are searchable within NARCIS. If someone wants to see the complete record, the information is retrieved directly from this CRIS. So the CERIF based XML format is used only for those fields to be stored in the NARCIS index.

**The ideal versus the existing situation: well connected components versus harvesting**

Within ‘Knowledge Exchange’<sup>5</sup> a concept has been developed for connecting information from Current Research Information Systems and the Open Access Repositories, the so-called ‘Academic Information Domain’. There are two kinds of Open Access repositories: repositories with (full-text) publications, and with e-Research, datasets and the software to use it. In this concept the different components (CRISs and OARs) are well connected.

To create a Dutch Academic Information Domain the ideal situation would be: every university has an Academic Information Domain, with well connected components (OARs and CRISs), and all these domains are interconnected. The EC-recommended standard CERIF (Common European Research Information Format) would be ‘the uniting element’ that holds the information systems together, allowing for interoperability.

Neither in the Netherlands nor in a lot of other countries this ideal situation exists. The CRISs and OARs of the institutions are not (yet) related to each other. The reason for this is the fact that in most cases the CRISs are maintained by the departments responsible for research administration, while the OARs are maintained by libraries, or, in the case of e-research, by special archiving institutes.

In this situation harvesting, like we do in the Netherlands, is the best option. And the DAI gives the opportunity to connect partly the information.

**Conclusion**

To interlink the information in the CRISs with (full-text) publications and the e-data of the Dutch repositories we developed the portal NARCIS. In NARCIS we collect information by harvesting the research information, full-text publications and datasets.

The connection is done by the introduction of the Digital Author Identifier. Because of the DAI personal pages can be made with expertise, working address(es), publication list and an overview of (current) research projects.

---

<sup>4</sup> See: <http://www.loc.gov/standards/mods/>

<sup>5</sup> See: <http://www.knowledge-exchange.info/>