Title: **Towards a Dutch Academic Information Domain**
Authors: **Arjan Hogenaar and Wilko Steinhoff**
KNAW, Research Information Department

1. Introduction

The development of repositories in The Netherlands has started January 2003 with the kick-off of the Digital Academic REpositories (DARE) programme. DARE has had considerable impact on the Dutch information infrastructure. Before DARE, every university was developing its own system to deposit the outcomes of scientific research. The interrelations between these repositories were a point of minor concern.

No wonder, the first big issue within DARE was the development of a Dutch information infrastructure for repositories. The SURF Foundation was responsible for the co-ordination of the programme, in which all Dutch universities, the Royal Netherlands Academy of Arts and Sciences (KNAW), the National Library (KB) and the Netherlands Organisation for Scientific Research (NWO) participated.

As a result, all the participants managed to implement a repository. These repositories were (and still are) harvestable using the OAI-PMH protocol. Repositories improve visibility, but the really exciting advancement is the aspect of the harvestability of these repositories. All repositories may act as contact-providers (data layer), making the metadata of their output harvestable by service-providers (services layer). DARE has chosen to follow the open, international standards to ensure interoperability.

In order to show the world the possibilities of the concept, SURF and the KNAW have decided to develop a first service on top of the repositories: DAREnet.
The DAREnet website is giving access to the metadata of publications that are available via the repositories of the DARE-partners. So, DAREnet is not harvesting the objects (publications) themselves. The DAREnet model, using Open Source applications and protocols, is now being used all over Europe, thanks to the DRIVER-I project.

2. NARCIS

Open Access to scholarly information is of course essential for the world-wide sharing of and building upon completed research. But access to publications, although very important, is not enough to serve the scientific community.
Nowadays, a researcher wants to have access to all kinds of research related information. In this respect, important information sources are research data (the raw research results) and research information (description of researchers with their expertise, research institutions, research projects and research programmes).

Until recently, these types of research related information were strictly divided. Normally the university libraries were responsible for the acquisition and dissemination of publications, while the research administration departments were involved in the registration of the research information.

Therefore, in co-operation with the academic institutions in The Netherlands, the KNAW has taken the initiative to develop a totally new service, called NARCIS.
NARCIS, National Academic Research and Collaborations Information System, is a Dutch portal which combines research information with information from OAI repositories and descriptions of research data. Besides, NARCIS offers an overview of scientific news items taken from various sources.

3. Academic Information Domain (AID)

The importance of the different types of information relevant to scientist has been recognized by 'Knowledge Exchange (KE)', a collaboration of DFG, JISC, DEFF and SURF. KE has developed a new concept for connecting research information and the contents of Open Access Repositories (OAR): the Academic Information Domain (AID). In this new view, the OAR consist of both research data (primary scholarly material) and publications (secondary scholarly material).

The major difference between NARCIS and this AID is the fact that in the AID concept the several information sources are interrelated. So, to come to a Dutch AID much work will have to be done to introduce and maintain the relationships between research information, OAR and datasets.

4. How to realize the interlinking of the information sources within the AID?

For the realization of the interlinking, an important precondition is the control over the several information systems. The KNAW is responsible for the maintenance of the Dutch Research Database (NOD). The NOD is the national database for research information.
The KNAW is also responsible for the DAREnet and NARCIS services, which include access to the OAR. The institute DANS is responsible for the archiving of datasets in the fields of the humanities and social sciences. DANS is an institute of the KNAW. Therefore, the KNAW has the opportunity to set up such an AID. The challenge is the way of interlinking the information.

Research descriptions, research publications and datasets have one aspect in common: the name of the author (researcher). So, the author name is an interesting candidate for the interlinking. Unfortunately, author names are not always spelled on the same way.

DAI

So, instead of using the name of the author for the interlinking, it is better to use a Digital Author Identifier (DAI). This DAI is an identifier giving access to all kinds of variations in spellings the names of an individual author. The Dutch scientific institutions will have attributed DAI's to the names of their researchers by the end of 2007. In the beginning of 2008, the DAI-author name(s) relationships will be implemented into the three information sources.

Persistent Identifiers

Introducing the DAI is not enough to create a stable, reliable information system. As said before, systems like DAREnet and NARCIS don't harvest the objects (publications) themselves. They harvest the metadata of the objects. The objects remain in the local repositories. This means that a change in the location in the repository or even a transport of the object from one repository to another could cause dead links. Luckily, the National Library is responsible for the long term preservation (in the e-depot) of the objects in the repositories.

Using persistent identifiers, one will be able to identify resources independently of the localisation of these resources. In other words: persistent identifiers (PI's) don't give immediate access to a URL where one can locate an object. Instead, a resolver is needed to provide a requester with the actual location(s) of the wanted object[1].

---

[1] This resolver can also be used to locate a specific object within the e-depot of the National Library, for instance if this object is no longer available within a local repository.

It is foreseen that DANS will be involved in the assigning of the persistent identifiers to information objects (datasets or research publications). In Holland, it is decided to use the URN:NBN:nl system that has been developed by the CDNL and CENL[2]. The persistent identifiers will be attributed to the objects by the individual research institutions. The actual resolver will be built by DANS.

Harvesting metadata records with the DAI and PI
Simple Dublin Core metadata elements have been used for a long time in Dutch repositories. Of course, simple DC cannot be used in the harvesting of the identifiers. To start with, a stripped MODS version will be included in a DIDL document (together with the DC metadata record) in order to harvest the identifiers.

5. Technical elaboration of the Dutch AID

Although the Dutch AID is not fully functioning yet, the main structure has already been developed. In the KNAW, the research information metadata, the research publications metadata, and the metadata of the datasets are being harvested by the Open Source harvester Sahara (see figure 1). Thereafter, the data is stored in a Lucene/SRU enabled Open Source index, called Teddy. NARCIS is being used as a front-end to search the index.
The DAI is needed in order to present search results deriving form the three information sources into one single result page.
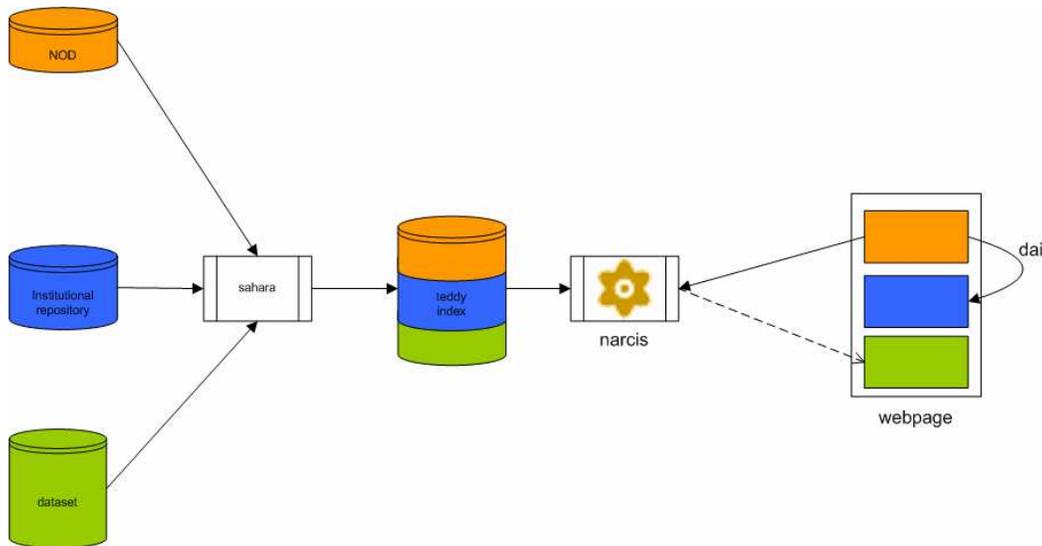

Fig. 1:
The open source Sahara OAI-PMH harvester harvests metadata from the Dutch Research Database (NOD), the institutional repositories (IR) and (in the future) the datasets. The data is stored in a lucene/sru enabled open source index called Teddy. By means of a web-based front end called Narcis, the end user is able to search the index. Narcis will then merge the search results into one result page, using the Digital Author Identifier (DAI) as digital glue.

---

[2] This URN:NBN system has already been implemented by the German National Library.

6. AID in a European perspective

In other European countries there is also a growing awareness of the fact that the relationships between research information, research data, and research results are of vital scientific importance.

After the realization of a simple European information infrastructure in the DRIVER I project, the DRIVER II project (started December 1, 2007) will focus on the development of a modern infrastructure that will help researchers to set up (virtual) communities.

All the original Dutch DARE-partners are participating in this European project. Within the project, they have special interest in the creation of enhanced publications[3], the development of object models for enhanced publications and in the building of a demonstrator by which the added value of these interrelated information sources can be showed.

7. Literature

Waaijers, L. The DARE Chronicle: Open Access to Research Results and Teaching Material in the Netherlands. Ariadne 53, October 2007. http://www.ariadne.ac.uk/issue53/waaijers

Dijk, E; Baars, Chr.; Hogenaar, A.; Meel, M. van. NARCIS: the Gateway to Dutch Scientific Information. Accessed on November 27, 2007, available at: http://elpub.scix.net/data/works/att/233_elpub2006.content.pdf

Razum, M; Simons, E.; Horstmann, W. Exchanging Research Information. Accessed on November 24, 2007, available at: http://www.driver-repository.be/media/docs/KEIRstrandreportExchangingResearchInfoFINALFeb07.pdf

Author Identification Project in The Netherlands. Live Serials UKSG. Accessed on November 27, 2007, available at: http://liveserials.blogspot.com/2007/04/author-identification-project-in.html

EPICUR : Uniform Resource Name (URN)- Strategy of the German National Library. Accessed on November 27, 2007. Available at: http://www.persistent-identifier.de/english/3352-EPICUR_policy.php
Identifier URN:NBN:DE:1111-200606309.

8. Acronyms

KNAW   Royal Netherlands Academy of Arts and Sciences
DARE   Dutch Academic REpositories
NWO   Netherlands Organisation for Scientific Research
KB   National Library (of Holland)
CDNL   Conference of Directors of National Libraries
CENL   Conference of European National Librarians
NARCIS National Academic Research and Collaborations Information System
DRIVER Digital Repository Infrastructure Vision for European Research
DFG   Deutsche ForschungsGemeinschaft
JISC   Joint Information Systems Committee
DEFF   Danmarks Elektroniske Fag- og Forskningsbibliotek
NOD   Dutch Research Database
AID   Academic Information Domain
OAR   Open Access Repositories
URN   Uniform Resource Names
NBN   National Bibliography Numbers
MODS   Metadata Object Description Schema
DIDL   Digital Item Declaration Language

---

[3] Enhanced publications are publications coupled to materials like primary data, simulations or research descriptions. They can be described as compound information objects that are aggregations of distinct information units that, when combined, form a logical whole [DRIVER II definition]