

**Leif Longva**

## **NAROS: Northern Areas Open Scholarly Documents**

A living discourse needs to be communicated and disseminated. The Internet is a very powerful tool in this respect. Internet has been around for a while now, but how to utilize the Internet as a communication and dissemination tool, is still evolving (Tananbaum, 2007).

NAROS is a planned service. The intention of NAROS is to utilize the possibilities of the Internet to improve the awareness and the accessibility of scholarly works on topics related to northern areas, thus hopefully paving the way for expanding the arctic discourses. NAROS will collect information on applicable documents through a standard way of automatically harvesting metadata, and utilize the fast growing trend of making scholarly works available through open archives and open access journals. Through the search service of NAROS, researchers, students, and others will have easy access to scholarly documents within the thematic scope of the northern areas.

### **Open dissemination of scholarly works**

The Internet has made it possible to dramatically improve the dissemination of scholarly works. Open access journals give any interested person access to the content. These journals may operate with the same quality control measures as traditional subscription based (or toll access) journals. They differ only in the way they generate revenues, which may be through sponsors or through payment from the authors (or their funders), rather than payments from libraries and readers. The fact that articles in open access journals are openly available to anyone, means that they are more widely disseminated than the toll access journals.

In addition to the growing numbers of open access journals, there is a strong trend in the scholarly community to establish

open archives for dissemination of scholarly works<sup>1</sup>. A growing number of universities, colleges and research institutes have established institutional repositories (IRs), where the research results of the institution are made open and freely available. This is in part motivated by the interest of the institutions themselves, as well as the interest of the individual scholars, to display the production of their research. In part it is motivated by a moral obligation to make publicly available the results of publicly funded research.

Obviously, this is in line with the interests of the scholars as well as the public, to spread the results of publicly funded research to as many readers as possible, as quickly and free of obstacles as possible. Institutions and funders of research has in growing numbers mandated that the results their researchers are producing should be made publicly available in open repositories, or through open access journals<sup>2</sup>.

### **Search tools**

Disseminating scholarly documents through open access journals or open repositories enables anyone to access them. However, searching through general tools like Google, these scholarly documents are competing with the vast and fast growing information load available on the internet. In order to support search and retrieval explicitly of scholarly documents, the open archive community has developed a standard. The standard, called OAI-PMH (Open Archives Initiative Protocol for Metadata Harvesting), specifies how the metadata of an open archive may be harvested<sup>3</sup>. A harvester will automatically collect information on the content of the open archives selected to be harvested. Thus, it is possible to put up a search service, specifically searching the harvested metadata of selected open archives. And the archives may be selected by criteria creating a thematically search tool for scholarly documents.

---

<sup>1</sup> See <http://www.oaister.org/stats.html>

<sup>2</sup> See <http://www.sherpa.ac.uk/juliet/>

<sup>3</sup> See <http://www.openarchives.org/OAI/openarchivesprotocol.html>

The harvesters collect the metadata, i.e. the information on the documents residing in the open archives. The scholarly documents themselves are not harvested. A search result will point to the (persistent) url where the document may be retrieved.

There are several examples of harvesters and search tools established through some selection criteria. To mention two, NORA<sup>1</sup> is a search tool based on harvesting all Norwegian open scholarly repositories, while Avano<sup>2</sup> is a search tool based on harvesting repositories containing scholarly documents within marine and aquatic sciences.

### **NAROS**

NAROS – Northern Areas Open Scholarly Documents – is a pilot project, investigating the viability of the idea of establishing an OAI-PMH harvester where the scholarly documents to be harvested are thematically related to the northern areas. The idea is to include documents in any language, and within any subject, so long as they are scholarly (including journal articles, theses and dissertations, conference papers, books and more, recent material as well as old) and related to the northern areas. The documents may be harvested from open archives, or open access journals, wherever these may be found.

The pilot project will seek to identify the potentials with respect to sources to harvest. Which institutions, research projects, open access journals and other sources contain documents applicable for NAROS? And what is applicable for NAROS? What is “northern areas-related” – what are the northern areas? These questions need to be discussed, in order to work out which sources and which documents to harvest.

### **Sources**

In the pilot project we work along the model of dividing the sources to harvest in two main categories. These are a) sources where all content is applicable for NAROS, and b) sources where

---

<sup>1</sup> <http://www.ub.uio.no/nora/>

<sup>2</sup> <http://www.ifremer.fr/avano/>

some content is applicable. The first category may consist of some research institutes, research projects or open access journals solely focused on the northern areas. Identifying these should be a feasible task. If they do have OAI-PMH-compatible archives, it will be quite effortless to include these in NAROS. If they do not have OAI-PMH-compatible archives, we may hope to stimulate them to establish one, in line with the global trend of open archives.

The second category of potential sources will be the largest category, and will also be the labour-intensive one to deal with. For one, we need to select sources for this category. Any scholarly archive, anywhere in the world may contain a document applicable for NAROS. However, we cannot start by harvesting all sources. We need to start by identifying the most interesting ones. Having selected a set of sources, the content applicable for NAROS needs to be found through a searching algorithm. Identifying good algorithms to use will require manual labour. For the best results, the algorithms need to be adapted to the individual source, since the search options, including the metadata structure, vary from source to source. And of course, the search algorithms must be adapted to the language used in the individual source.

With the use of trial, error and refining the search algorithms, we may work our way to the best algorithms to use for our purpose, individually for each source. In the pilot project we will also investigate how much labour we should devote to a final proofreading of the search results, in order to weed out records matching the search criteria, but still not matching the intentions of NAROS. Likewise, there may always be applicable records in the selected sources that we do not find through our selected search algorithms. These algorithms will therefore need to be tested anew, almost on a continuous basis.

The end report of the pilot project will discuss and describe the amount of labour needed for these manual tasks of the NAROS service.

**The user interface**

The harvested metadata needs to be organised in a database for user friendly browse, search and retrieval purposes. This is not a trivial question. Easy to use web interface is crucial for the success of any web service. NAROS need to define categories according to meaningful criteria for organizing the content, enabling users to browse the content by these categories. Criteria to use may be subject coverage, geographical coverage or document type, to mention just a few. How this best may be done, will need to be examined closer. The options of how to do this will depend on the metadata added to the documents in the archives where they reside.

**A useful service?**

Will NAROS be a useful service for students and researchers, for businesses and public administration, and for the public in general? We believe it will, and we believe that the right timing is now for establishing NAROS. The trend of establishing open access to scholarly documents is strong, and the supply of content for NAROS should therefore be good, and getting better. NAROS may even stimulate more institutions and other sources to establish OAI-PMH-compatible archives.

Openly accessible documents may be found through any web search tool. We believe however that a service that limits the retrieved documents according to the criteria applied by NAROS, will be very valuable for anyone looking for scholarly documents within this scope.

**Reference**

Tananbaum, Greg (2007) "I Hear the Train A Comin' – ALCTS: Part 2", *Against the grain*, Vol 19, No 2, pp 90-93

**Further reading**

Jones, Catherine (2007) *Institutional Repositories: Content and Culture in an Open Access Environment*, Chandos Publishing

Willinsky, John (2006) *The Access Principle : The case for open access to research and scholarship*, Massachusetts Institute of Technology