# High North Research Documents: Your Source for Research Documents on the North

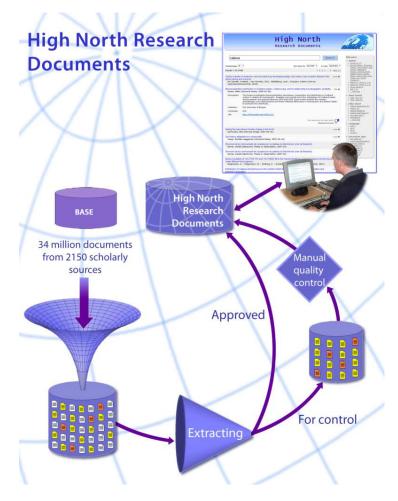
by *Leif Longva* and *Stein Høydalsvik* University Library of Tromsø, Norway

High North Research Documents (High North RD) is a freely available database providing access to research documents of relevance within the thematic scope of the high North. Developed by the University Library of Tromsø and launched in late January 2012, the nearly 150,000 records included as of April 2012 cover documents written in any language on any subject related to the high North. High North RD is an extremely valuable tool for researchers and students, as well as public policymakers, decision makers in the public or private sector, NGOs and interest groups, and anyone else interested in the high North.

## **Background**

Open access to research documents is a global trend. Numerous higher education and research institutions all over the world have established open archives to make research documents freely available. Similarly, the number of open access journals is growing steadily. Open access research books and monographs are also starting to appear.





High North RD takes advantage of the freely available metadata from these sources to extract documents relevant to the high North. We do this with the kind help and cooperation from Bielefeld University Library and its Bielefeld Academic Search Engine (BASE), which the Bielefeld Library developed and has run since 2004. Using the common standard, Open Archives Initiative – Protocol for Metadata Harvesting (OAI-PMH), BASE now harvests almost 2,200 open sources containing research documents from around the world. As of April 2012, BASE has metadata from approximately 35 million records.

#### How?

High North RD is an overlay service on top of what is harvested by BASE. High North RD extracts all relevant documents through a simple but intelligent textmining algorithm. The extraction process is mostly done automatically, but does involve a manual control. The extracting algorithm will be further developed and improved. As the corpus of open access research literature grows and is included in BASE, so will the High North RD keep growing in volume.



## **High North RD** continued from page 7

The method of automatic extraction from heterogeneous metadata is, to our knowledge, new. Our method means that High North RD will cover all subject fields and include documents in any language. However, because English-language documents currently predominate, we need to further develop the extracting algorithm for various other languages.

As an automatic textmining process will never be perfect, it is important to make use of the knowledge our users hold and engage them in improving the quality of High North RD. The service includes a feedback function by which users are encouraged to identify documents not relevant for the high North so they can be removed.



#### A propos de quelques particularités d'un dialecte algonquin "

Explication des indications typographiques. Dans les exemples quand la clarté l'exige les éléments morphologiques sont separés par des traits et nous notons en italique ceux que nous voulons nettre en évidence.

Nous plaçons entre parenthèses les lettres euphoniques et nous

Nous plaçons entre parenthèses les lettres cuphoniques et nous crivons en petites capitales les éléments dont nous ne voulons ou ne pouvons préciser la signification.

Les pronoms personnels et les éléments pronominaux sont notés en caractères gras et suivis d'un indice numérique qui précise la personne, les numéros 1, 2 et 3 correspondant respectivement aux trois personnes du singulier; 4, 5 et 6 aux trois personnes du pluriel, 3', 3', 3', aux nuances que traduisent les pronoms latins hic, siec, lile, et 3', au français on; zero indicant une valeur pronominale indécise qui n'est pas determinée en elle même, ou que nous n'avons

### Postglacial Colonization and Parallel **Evolution of Metal Tolerance in the** Polyploid Cerastium alpinum

Anna-Britt Nyberg Berglund Faculty of Natural Resources and Agricultural Sciences Department of Plant Biology and Forest Genetics Uppsala Department of Natural Sciences Mid Sweden University Sundsvall

Doctoral thesis Swedish University of Agricultural Sciences Uppsala 2005

Image: University of Washington, Digital Collection, American Indians of the Pacific Northwest Collection; Tlingit woman selling baskets and other crafts on sidewalk, Alaska.

Article: Hedatuz, Basque culture and science digital library; Léon, Albert (1925), Revista Internacional de los Estudios Vascos (16). pp. 68-73

Doctoral thesis: Epsilon Open Archive, Swedish University of Agricultural Sciences; Anna-Britt Nyberg Berglund (2005).

#### What?

High North RD is made up of any and all kinds of research-based documents. Examples of some of the important types of documents you will find are doctoral dissertations, peer-reviewed journal articles, books and book chapters, research reports and conference papers, and master theses. You will also find photos, as well as film and sound clips presenting personal interviews or lectures, or documenting life in the North in past or present times.

Our methodology ensures that no matter where in the world a document is produced or published, as long as it is open access and harvested by BASE, and as long as it is relevant to the high North, it should be found in High North RD. Documents on the Antarctic area are included as well, as many of these materials may relate to issues concerning the North. One notable advantage for users who are searching or browsing the database is that the documents have been preselected as relevant to northern regions.

### **Lessons Learned**

As mentioned, the reason BASE is able to harvest from open access sources worldwide is that these sources commonly use the OAI-PMH standard. Unfortunately, some of the sources include records without any fulltext documents, or they include records whose documents have restricted access. If such records are not easily identifiable in the metadata, which they seldom are, it means that non-open access records will be



# High North RD continued from page 8

harvested along with the open access ones. Thus, although we try our best to weed them in both BASE and High North RD, you may sometimes find records without any open access documents.

So far we have received very positive feedback from users, which is encouraging. We are convinced that modern and efficient communication between scholars and the need to disseminate research results to society at large, means that open access to research is the future. Another benefit is the ease with which open access metadata lends itself to creating tailor-made services. At this time, we are therefore not deterred by the fact that High North RD is limited to open access documents.

#### The Road Ahead

Since High North RD covers all areas of the circumpolar North, knowledgeable contributers from other institutions and other geographic areas and other interested persons are invited to join a team of editors. The task for this team will be to further improve the extracting algorithm so that it better identifies relevant documents. We expect that the workload for the editorial team would be very limited. Please contact us if you are interested in joining the High North RD editorial team.

Please also send us feedback, especially about any documents that should be removed. We need your help in making the High North RD an even better service.

—This work is licensed under a Creative Commons Attribution 3.0 Unported License.



