Archiving and collecting Arctic datasets: Open Arctic Research Index

Tamer Abu-Alam, Karl Magnus Nilsen, Obiajulu Odu, Stein Høydalsvik

UiT The Arctic University of Norway

karl.magnus.nilsen@uit.no

UiT

THE ARCTIC UNIVERSITY OF NORWAY

Collecting data from major collaborators and metadata providers (e.g. BASE, Polar Data Catalogue, Norwegian Polar Institute, etc.).
This option will be used if the metadata provider allow using OAI-PMH service

Custom-made harvesting of different databases after a permission from the metadata providers.
This option will be used if the metadata provider does not allow using OAI-PMH service

Getting feedback and suggestions/ modifications for new records or existing records from the end-users.
These suggestions will not be added to the database before approval from the admin of Open ARI.

Temporary database

Harvesting all the databases and the existing records on the Arctic and the Antarctic (as many as we can)

Filtering the temporary database to find the relevant records

Five stages of filtration

First stage: Filter out none relevant journals and source institutes

Second stage: Filter out the broken records (e.g. records missing links, text, irrelevant, etc.)

Third stage: Compare the records in the temporary database with those that we have in the final database (from the previous filtration process). Remove the duplications and keep only the most updated version.

Fourth stage: Approve the records from certain sources (e.g. from institutes concerning with polar data or published in journals related to polar issues)

Fifth stage: Check for certain keywords

Semi final database

Normalizing metadata and checking the quality (automatic/manual work to reduce the error)

Final database

/ Abstract

The number of digital repositories containing publications and datasets on the Arctic region are increasing enormously. Users want relevant information according to their query with minimum interval of time. Scholars are compelled to search the individual repositories to get their desired documents.

Open Arctic Research Index (Open ARI), a planned service at UiT - The Arctic University of Norway, aims to collect and index all the openly available Arctic-related publications and datasets in a single open access metadata index. By providing a simple search dialog box to the index, users can search all these repositories and archives in a single operation.

The project investigates how such a service can support researchers in their research by making results from Arctic research more visible and better retrievable based on a standardized, interdisciplinary metadata set. The project started by clarifying the need for a new technical solution to collect all the published material using algorithms that allow the best way of filtering relevant records. We have defined 115 possible national and international collaborators who can feed the Open ARI with content. The team will analyze the success opportunities and the challenges in order of planning a full-scale management model.

Show search results as timeline
Search by multiple keywords

Search by geographic location
Show search results on map

- Use different social media channels
- Increase the audience by distributing announcements in journals, newsletters and websites of different organizations and associations dealing with polar science
- Distribute flyers in well attended conferences and meetings

Outreach activities to broaden our audience and getting feedback

Posting these data to an easy searchable front end website with different interesting options

To take the OpenARI to the next level - which is not only hosting metadata, but to be a platform to define research gaps and create new forefront knowledge, we may prepare a web application where scientists can discuss, collaborate and use different databases.

/ Classification of metadata providers

The Metadata providers can be classified into three categories based on their relation with Bielefeld Academic Search Engine (BASE) which is our main metadata provider.

1. Metadata providers not included in BASE
2. Metadata providers included in BASE
3. Metadata providers included (partially - as publisher) in BASE. These providers do not give access to all their catalogs

Metadata providers not included in BASE

Metadata providers included in BASE as data providers (sources)

Metadata providers included in BASE as publishers and not as data sources.

Only some catalogs included in BASE

All catalogs included in BASE

Open ARI

BASE database

/ Mapping the potential metadata providers

115 metadata providers were mapped according to the previous classification. The following figure shows some of these providers. The statistics show numbers of the polar records in these metadata provides.

Open Data (Canada)
Polar Data Catalogue (PDC)
University of Alberta - datasets
Centre Etude Nordique
Dept. Fisheries and Oceans Canada
ELDAT
SeaDataNet
ECOMET
Arctic Data Centre
Svalbard Integrated Arctic Earth Observing System
Institute of Marine Research
Norwegian Satellite Earth Observation Database for Marine and Polar Research
Joint WMO/IOC Technical Commission for Oceanography and Marine Meteorology
INTAROS
Norwegian Marine Data Centre (NMDC)
Finnish archives
ARCTIC SPACE CENTER
Environment Climate Data Sweden
EnviDat
Group on Earth Observations System of Systems
Global Terrestrial Network for Permafrost (GTN-P)
biodiversity
Netherlands Polar Data Center
Italian Arctic Data Center
Russian Federation Service's Arctic and Antarctic Research Institute
Chinese National Arctic and Antarctic Data Centre
Arctic Research Mapping Application
Global Cryosphere Watch
Ocean Data and Information System
GEOTRACES
Arctic Spatial Data Infrastructure
NSF Arctic Data Center (NSF ADC)
National Snow and Ice Data Center
UCAR NCAR - Earth Observing Laboratory
NASA Earth Observing System
Rolling Deck to Repository (R2R)
Gulf of Alaska Data Portal
NCEI - National Oceanographic Data Center
Environmental Data Initiative
CLIVAR and Carbon Hydrographic Data Office
ICPSR
NCEI National Climatic Data Center
NCEI - National Geophysical Data Center
Scholars Portal Dataverse U. of Toronto
Arctic Institute of North America
Yukon Research Centre
Norwegian Meteorological Institute
The University Centre in Svalbard
The Norwegian Meteorological Institute
DataversNO-arkivet
The Geological Survey of Denmark and Greenland
Swedish National Data Service
UK Polar Data Centre
British Oceanographic Data Centre
Australian Ocean Data Network
Australian Antarctic Data Centre - AADC
Arctic Data archive System
Conservation of Arctic Flora and Fauna
Polar Research Institute of China
Korea Polar Research Institute
OBIS - Ocean Biogeographic Information System
KNB Data Repository
Global Biodiversity Information Facility

Pie charts

Denmark 0.26% Sweden 0.079%
Canada 4.19%
USA 5.06%

Japan 0.299%
Australia 1.21% USA 0.97% Canada 0.58%
EU 3.21% Norway 0.099%

UK 18.15% USA 22.36% Canada 1.14%
Belgium 20.04%
EU 38.02%

International 90.36%

Germany 93.9%

Percentage of records not included in BASE (distributed based on countries or regions)

Percentage of records partially included in BASE as publishers (distributed based on countries or regions)

Percentage of records included in BASE (distributed based on countries or regions)

A histogram shows the numbers of the polar records in the different metadata providers.

Legend:
- Metadata providers not included in BASE
- Metadata providers partially included in inkluderl i BASE
- Metadata providers included in BASE
- BASE

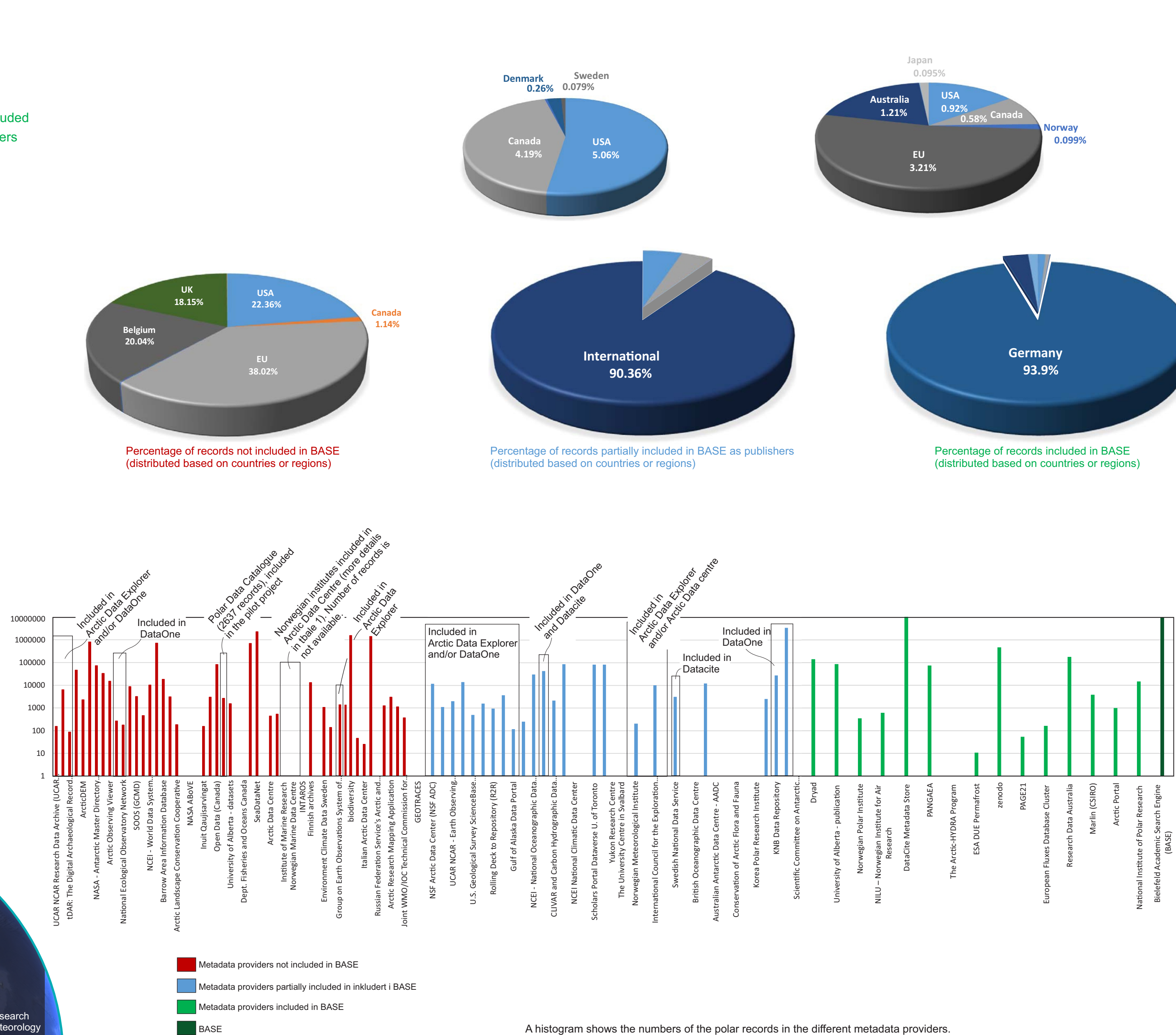## / Mapping the needs and solutions toward a better service

In order to provide the end users with the best service that will help them in their research and study, we have prioritized fifteen requirements that will be included in the OpenARI.

Some of these requirements are:

- Add locations value for each record (e.g. latitudes and longitudes)
- Search using a map
- Show search results on a map
- A proper map projection centered around the North and the South Pole
- Show search results as a timeline
- Use autocomplete search words algorithm.
- Multilingual user interface
- Social media sharing of records
- Increase the number of the records in the database by adding more records from other metadata providers.

/ Progress and time frame

The first stage of the Open ARI project was a six month pilot project, running from Sep. 2018 to Feb. 2019. The main goal of the pilot project was to assess the possibilites to build a metadata index for publications and datasets related to the Arctic, and to estimate the resources needed. The recommendations from the final report of the pilot project suggest that such a service will be of great value, because a large part of the relevant publications and datasets can only be found through their own repositories. The report also strongly recommends to extend the scope of the project to include the Antarctic region in addition to the Arctic. From here on the scope will be the Polar regions. Based on the recommendations from the report it has been decided to go forward with a full scale project. This project will start up some time in the fall of 2019.

UNIVERSITY OF TROMSØ

UiT / THE ARCTIC UNIVERSITY OF NORWAY