Open Polar

A mid-way report toward a unique service

Tamer Abu-Alam

Abstract

Data from the Polar Regions are of critical importance to modern research and decision makers. Regardless of their disciplinary and institutional affiliations, researchers rely heavily on the comparison of existing data with new data sets to assess changes that are taking effect. However, in a recent survey of 113 major polar data providers, we found that an estimated 60% of the existing polar research data is unfindable through common search engines and can only be accessed through institutional webpages. This raises an awareness sign of the need of the scientific community to harvest different metadata related to the Polar Regions and collect it in a homogenous, seamless database and making this database available to researchers, students and publics through one search platform.

This contribution describes the progress in an ongoing project, Open Polar, started in 2019 at UiT The Arctic University of Norway. The project aims to collect metadata about all the open-access research data, articles and other scholarly documents related to the Polar Regions in a homogenous and seamless database. During the first six months of the project, the beta version of the user-interface was established, with a search by map and an advanced search function. An extensive geo-database that includes thousands of polar locations and their geographic information was collected from different sources. The geo-database together with a list of keywords (i.e. on sources, indigenous peoples, languages and other polar-related keywords) will be used in the filtration process.

A Reference Board was formed, and the first board meeting took place in April 2020. The geographic definition of "Polar Regions" was defined in order to include most of the current geographic definitions of "Arctic". The project is still facing some challenges that include for example integration with non-standard data sources who do not use Dublin Core Metadata schema, or are not harvestable through the Open Access Initiative's standard protocol for harvesting (OAI-PMH).

Introduction

The Polar Regions (i.e. the Arctic and Antarctic) are places where global processes interact across the Earth and therefore these regions are considered as barometers to measure the health of the Earth, and as key areas to study the global changes in the Earth's ecosystem (e.g. IPCC, 2007, 2013; Beck et al., 2014). Several recent climate impact assessments on the ecosystem of the Polar Regions (e.g. AMAP, 2011, Forbes, 2011, Turner et al., 2009, 2013, Arctic Council, 2013, Krupnik et al., 2011) draw a consistent pattern of climate-driven environmental, societal, and economic changes in recent decades. These assessments rely heavily on the comparison of existing data with new data sets to assess changes that are taking effect.

Despite its importance, there are massive access barriers to the research data. Tenopir et al. (2011) concluded that 871 (i.e. 67%) out of 1300 researchers pointed to a lack of access to data generated by other researches or institutes in Europe. A similar conclusion was drawn on the polar-related research data, where a findability gap of about 60% was found (Abu-Alam, 2019). Moreover, a study by Johnson et al. (2019) on the Arctic region showed that social sciences and indigenous studies have a findability gap around 84%. Thus, this findability gap is not connected only to the Polar Regions, but similar gaps can be observed on different research disciplines covering different geographic locations. However, the present contribution focuses on the findability gap of polar-related research data.

This contribution describes an ongoing project, Open Polar (https://site.uit.no/open-polar/) started in 2019 at UiT The Arctic University of Norway. The project aims to collect metadata about all the open-access research data and documents related to the Polar Regions in a homogenous and seamless database and making this global output of open-access material available through an inter-active user interface. The project will help to make research data and documents considered to be of relevance to the polar regions more visible and searchable to the end-users and as a result reducing the findability gap.

Open Access research data – a historical view

"Open Access" is a term that has come to mean a set of policies about, as well as methods for publishing, archiving, and disseminating scholarly works. The history of Open Access can be traced back to the end of the 20th century, when university librarians around the world faced a problem known as the "serials crisis". The serials crisis occurred when the subscription costs for publications rose much faster than inflation. In order to balance the limited funds available to the libraries with these subscription high costs, the libraries were forced to make difficult choices between different journals. The serials crisis was the engine behind a larger movement to make scientific research more easily available, and at a lower price.

Initially, the term was related only to free access to peer-reviewed literature (e.g. Budapest Open Access Initiative, 2002). In 2003, through the "Berlin Declaration on Open Access to Knowledge in the Sciences and Humanities", the definition was considered to have a wider scope that includes raw research data, metadata, source materials and scholarly multimedia material.

In line with the goals expressed in the Berlin Declaration, UiT The Arctic University of Norway took the initiative to make open access research documents about the Arctic more visible and searchable through a search service known as "High North Research Documents", launched in 2012 (Longva and Høydalsvik, 2012; Longva, 2011). Some other services dealing with polar research data were established around this time (e.g. Polar Data Catalogue: https://www.polardata.ca/), however the High North Research Documents stands up comparing to other services in terms of the contents, the number of records and its easy to use user-interface.

In 2016 and with the advanced technology in digitalization and cloud storage, the scientific society moved forward in their revolution supporting Open Access. Wilkinson et al. 2016 introduced the "FAIR Data Principles" as a set of guiding principles to make data findable, accessible, interoperable and reusable. Although many research institutes and researchers start to move toward this new digital era by uploading their research data to different institutes' repositories, most of these research data are not findable through standard search engines (e.g. Google Scholar). Consequently, a considerable amount of research data uploaded to institutes' repositories is consistent with the second principle of FAIR (i.e. accessible) but not with the first principle (findable) nor with the third or the fourth principles (interoperable, reusable). Therefore, new services that aim at making research data and documents more visible and searchable are needed.

The need for a new service in the polar sciences

Although the High North Research Documents is considered as an advanced step toward making Arctic-related records more searchable and findable through an easy to use user-interface that allows searching an extensive database that contains more than 1,200,000 open access records, a new service is needed in order to widen the scope to cover all the Polar Regions and to cover both research data and research articles.

A pilot study (Open Arctic Research Index) at the UiT conducted a survey in order to measure the needs of the scientific community for a new service that will provide the community with access to the records published on the Polar Regions through one search platform. The main outcomes (Abu-Alam 2019), as well as the steps that are currently being made to remedy the situation, are described below.

60% findability gap of the polar records

In the survey, a number (i.e. 113) of major and trusted institutes / organizations / research units (hereafter described as "metadata providers" or simply "providers") were mapped. The survey focused only on providers who are dealing exclusively or primarily with polar sciences. The different metadata providers were classified based on their findability through the Bielefeld Academic Search Engine (BASE) and High North Research Documents into three categories:

> 1- Metadata providers not included in BASE and High North Research Documents (58 providers)
> 2- Metadata providers included in BASE and High North Research Documents (21 providers)
> 3- Metadata providers included (partially, as publisher) in BASE and High North Research Documents. These partially included providers do not give full access to all their catalogs (34 providers)

The High North Research Documents and the database of BASE were used here as a basis of our analysis and as examples of common search engines, however, major search engines (e.g. Google) were used to validate the results. The reasons for choosing the High North Research Documents and BASE as examples of the common search engines are that these search engines contain more than 1,200,000 records from the Polar Regions and therefore, they represent highly trusted infrastructures.
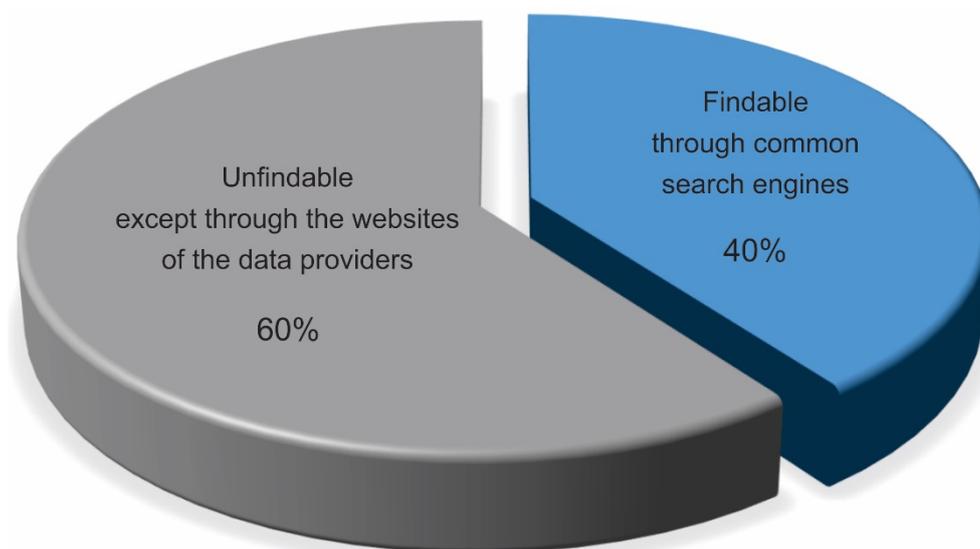

Figure 1. There is a 60% findability gap of the polar-related records based on the records numbers of 113 metadata providers.

The number of records from metadata providers that are not included and partly included in BASE and High North Research Documents are used to indicate the findability gap of the polar records. Interestingly, the numbers of records that are unfindable through common search engines represent around 60% of the total numbers

of the polar records (Fig. 1). This high percentage raises a series of questions. One of these questions is "why is such a high percentage unfindable although the survey targeted major and trusted data providers?" However, discussing such questions is beyond the current contribution. Instead, we will focus here on how to develop a service with the purpose of rendering polar research more visible and retrievable to the scientific community as well as to the interested public, teachers, students and public services.

Arctic versus Polar: broadening the scope

It is worth mentioning here that the sample providers used in the survey are not focused only on research activities in the Arctic area, but the sample covered providers active in both Polar Regions (i.e. Arctic and Antarctic). Therefore, the survey raised a question, namely: should the new service focus only on Arctic Open Access records? Or, should the new service cover all the Polar Regions?

The Arctic and the Antarctic regions are analogue to each other. Many researchers who are working in the Antarctic region are active in the Arctic region as well. Different research funding agents (including European/Scandinavian states funding agents) maintain and operate a large number of observatories and research stations both in the Arctic and Antarctic and as a result the research activities in both regions are linked and relevant (e.g. Ricci & Egerton, 2010; Norsk Polarinstitutt, strategi 2019–2024). In order to provide the scientific communities with a service that covers their needs, the service has to cover all the Polar Regions.

Open Polar: a unique service

The project aims to maximize the number of findable records, therefore metadata from different providers need to be added. The team of Open Polar targets metadata from different national and international metadata providers. By adding the records of these metadata providers, the numbers of the records in the Open Polar's database will be extended to the maximum (as possible), covering a wide range of subjects.

Most of the metadata providers host records from the Polar Regions but also from non-Polar Regions. As a result, a filtration algorithm is needed in order to keep only the relevant records. The filtration algorithm will be conducted by searching the harvested records for one or more keywords. A list of Polar Regions-related keywords will be prepared for this purpose. The relevant records will be saved in a database (hereafter: "Open Polar's database").

Open Polar's database will be structured according to the Dublin Core Metadata Schema. The Dublin Core Schema is a small and simple set of vocabulary terms that can describe any record (i.e. digital or physical record). The schema includes 15 metadata terms, known as the Dublin Core Metadata Element Set. Although the Dublin Core

Metadata is a simple schema, combination of different metadata vocabularies can be used to describe highly complicated records.

The records saved in Open Polar's database will be presented to the end-users using an interactive open-source user-interface. The Open Polar project team decided to use VuFind as a basis of the user-interface. The VuFind is an open-source library resource portal developed by Villanova University's Falvey Memorial Library and maintained by a large community that keeps the source code up to date. The VuFind characterizes by several options that will produce an attractive user-interface. These options include, for example:

- a map search; an option that allows the users to search Open Polar's database using the geographic location of the records;
- search with Faceted Results where the search system allows the user to search from a basic search box and then to be able to narrow down the results by clicking on the various facets of the results;
- advanced search where the users can focus their search to a set of search words in specified metadata fields and/or exclude some records by excluding un-wanted search words;
- save search and organized lists where the users have the ability to save their research results to their own customizable lists. The lists can be retrieved at any time. This helps to eliminate the need for desktop-based citation management software that tends to be too difficult for basic users;
- export the citation of records; the VuFind is able to export the citation of the searched records into several citation format including the APA, Chicago Style and/or the MLA. These exported citations can be saved in a proper format that suite different desktop-based citation management software (e.g. Endnote, RefWorks);
- author biographies where the users can learn more about the authors with contextual information and see all of the records that they have produced;
- URLs to the source: the VuFind is able to present one or more URL to the source of the records which allows the users to retrieve and access the original articles or data (this is an important function since the Open Polar aims to harvest only metadata and this function will allow the users to access the data itself);
- access original data: although we target to add an URL to each record (see the previous point), the VuFind has many APIs to interact with the search, the data and many other features. Users can syndicate the record data with other search services via an OAI server. Users can search using VuFind's algorithms via OpenSearch. And if users want complete access to the indexed data, they can interact with Solr, VuFind's backend search and index engine.

- Internationalization, in the sense that the interface of the VuFind has translations available in several languages including Norwegian, Sámi, Brazilian, Portuguese, Chinese, Dutch, English, French, German, Japanese, Spanish, Arabic and more.
- autocomplete/autosuggestion function is an algorithm that allows the end-users to perform search even if they misspelt the search word.

Location name	Control_name1	Latitude	Longitude
Amundsen Basin	antar*	-60	-115
Attu Station	Alaska	52.84642	173.1861
Attu Mountain	Alaska	52.89139	172.9322
Attu Island	Alaska	52.9025	172.9094
Akimiski island		53.0084	-81.2747
Atlin Lake	British Columbia	59.53194	-133.722
Atlin Lake	Yukon	60.01333	-133.831
Attu	Finland	60.18333	22.33333
Alaska range		62.58333	-153.083
Alpha ridge		63.52611	-151.017
Cape Dorset		64.17851	-76.482
Attu	Greenland	67.93552	-53.5791
Cambridge Bay		69.03665	-105.13
Andfjorden		69.16667	16.33333
Brooks Range	Alaska	69.20222	-143.802
Balsfjord		69.24024	19.22653
Båtsfjord		70.63506	29.72149
Babbage Bay		71.56798	-93.5837
Baffin Island		72.23384	-88.8245
Bylot Island		73.25996	-78.8049

Table 1. An example of the geo-database showing the location names, control_name1 and the latitudes and the longitudes.

In order to allow the end-users to use map search, most of the records in Open Polar's database should have geographic information (i.e. latitude and longitude). However, most of the records are harvested without having their geographic information. As a result, there is a need to create a new algorithm that allows an automatic recognition of the geographic information. In order to do that a database (hereafter: "geo-database"; Table 1) will be established, containing all the existing polar locations with their geographic information. The

algorithm works by searching each record for a polar location (e.g. name of a city, a mountain or a geographic feature). Once the polar location is found, the algorithm will set the pre-defined geographic information of this location to the record's metadata.

A unique service: Reference Board

To ensure that the content of the service is up-to-date and meets the requirements of the scientific community, a reference board consisting of a selected group of researchers and scientists from UiT The Arctic University of Norway and the Norwegian Polar Institute (NPI) has been appointed. The reference board gives advice on strategic matters as well as content selection and functionality aspects of the search engine as such. Since the proposed service is multi-disciplinary, the reference board covers social sciences, humanities, physical and life sciences. The input from the reference board has already proven its worth and helps make Open Polar a unique service in comparison with similar common search engines that do not rely on expert researchers.

A unique service: National-International Board

A board representing different national and international partners and different polar nations is suggested to replace the reference board by the end of the project period and before launching the service. The board can highlight metadata providers from different countries that are not covered by the service and help pave the way for the inclusion of precious records that are currently unfindable. Moreover, the members of the national-international board will help with the dissemination of the service in their respective networks.

Additional subservices

The existing service at the UiT (i.e. High North Research Documents) is used by several universities for educational purposes. However, the way of presenting the records in this service is streamlined mainly for research and not for educational purposes. To establish an attractive and comprehensive service, an educational platform is planned to be added to the Open Polar as a subservice. The needs of this educational platform have recently emerged worldwide due to the pandemic of the Coronavirus (i.e. teaching from home). In this context, we suggest that this subservice can be performed in two steps:

> 1) collect a series of online lectures and talks on the polar sciences and make it available and searchable to students. The Arctic Massive Open Online Course (Arctic MOOC), which is currently under construction at UiT, is a candidate to be included in this step;
> 2) simplify some selected research data to be used for educational purposes.

Other subservices are suggested to be included in the Open Polar. These subservices include (Fig. 2): 1) archiving original data from the Polar Regions; and 2) creating a research platform. Although these subservices, including the education platform, will enrich the final service of the Open Polar, the project team is focusing on the main part of the service (i.e. harvesting Polar Regions-related records and making them more visible and searchable). Therefore, discussing these subservices is beyond the current contribution, but more information about these subservices are found in Abu-Alam (2019).
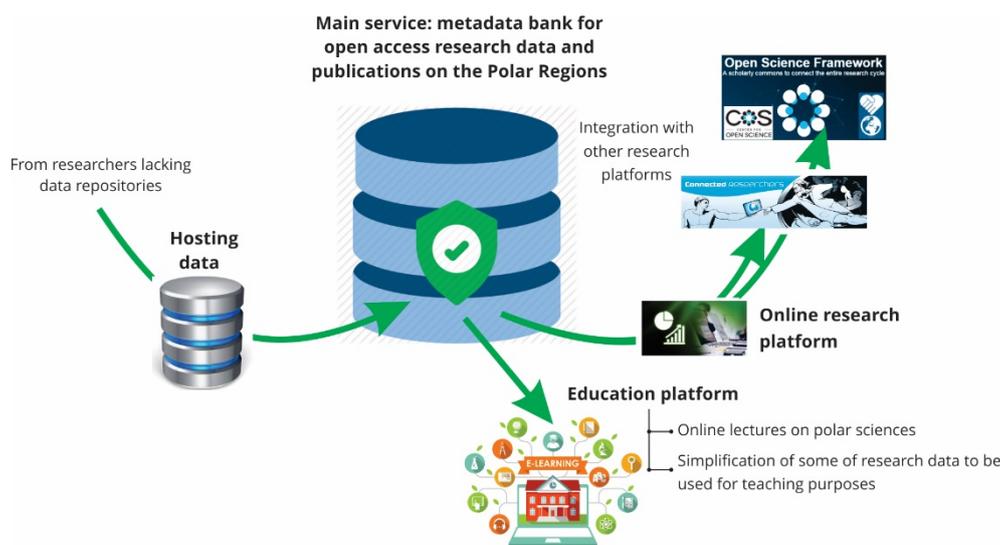
Figure 2. The integration between the main service (i.e. make open-access polar records more visible and findable) and the three additional subservices (modified after Abu-Alam, 2019).

Progress, challenges and planned steps: a conclusion

During the first six months of the project progress has been achieved. The beta version of the user-interface was established and put online (https://james.ub.uit.no/vufind/; Fig 3). In this beta version, the search by map and the advanced search tools are included. An extensive geo-database that includes thousands of polar locations and their geographic information was created (e.g. Table 1). Although the geo-database needs to be cleaned up (e.g. remove unrelated locations and duplications and add some locations with different languages), a simple version of the geo-database was tested successfully. The geo-location database will be used also in the filtration process.

In March 2020, the Reference Board was formed, and the first board meetings took place in April 2020. One of the challenges that face the Open Polar is the geographic definition of the word Polar. The Reference Board decided that the geographic definition of the "Polar Regions" should include most of the current geographic definitions of "Arctic". In addition, the board emphasized that the Open Polar project must consider different national definitions of the "Polar

Regions", including the Norwegian term "nordområdene" (high north), which includes subarctic as well as arctic areas.

Although the Dublin Core Metadata is a standard schema used in the communication among different metadata providers, many providers do not use this schema. Even some of the providers who are using the schema, they do not use it in a proper way (i.e. some information is entered in wrong fields). It is a challenge to use metadata not organized in a proper way. Therefore, a combination of MySQL and Solr technology together with GO and php programming codes will be used in order to read, organize and re-sort the metadata fields.
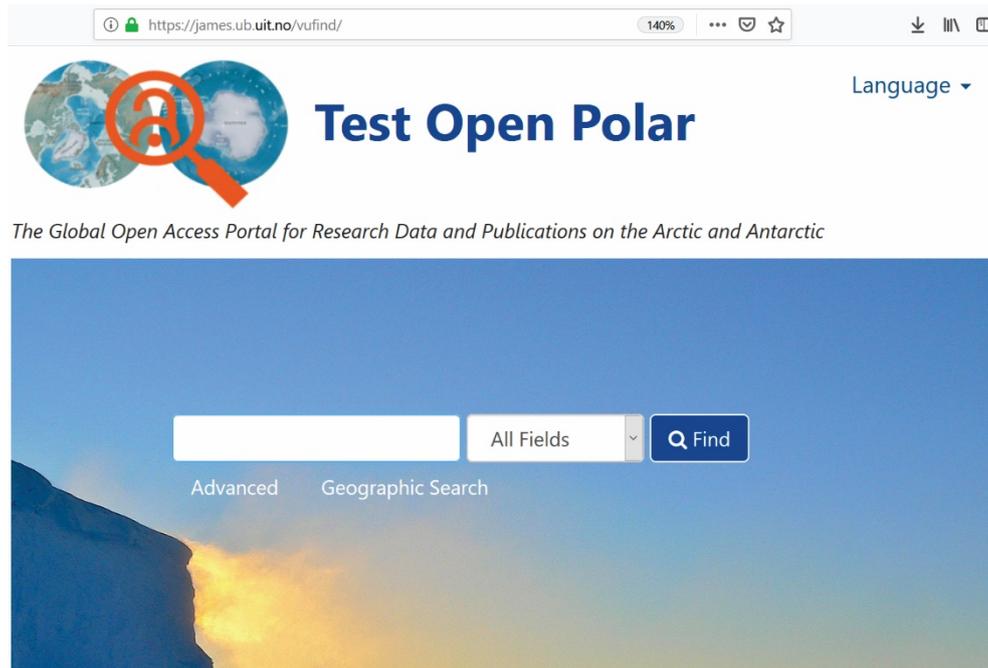
Figure 3. Screenshot shows the beta version of the Open Polar search tool.

It is our plan that by September 2020, all the databases that will be used in the filtration process and identification of geographic information of each record will be completed. In addition, most of the errors and defects in the beta version will be fixed. Before the end of 2020, the filtration algorithm will be tested, and the results will be uploaded to Open Polar's database after adding the missing geographic information. Integration with non-standard metadata sources will be our target for November and December 2020.

Although the Coronavirus pandemic affects our outreach plan for 2020, an extensive promotion plan is planned for the first half of 2021. The promotion plan will be associated with attracting new metadata providers to share their metadata for inclusion in Open Polar. The preliminary plan is to launch the search engine by the second half of 2021.

Open Polar Team

On a daily basis, the Open Polar project depends on a working team. In addition, the project gets occasional advice from the reference

board, whereas major strategic decisions are made by a steering group. As coordinator of the Open Polar project, I would like to thank my colleagues at the Open Polar.

Working team:

- Stein Høydalsvik – the founding father and architect behind both the pilot project OpenARI and the ongoing Open Polar project, UiT university library.
- Leif Longva – central to the High North Research Documents and an important counsellor to Open Polar, UiT university library.
- Karl Magnus Nilsen – taking care of several aspects of the technical parts of the service including filtration algorithm and setting up the user-interface, UiT university library.
- Obiajulu Odu – deals with the technical parts of the service including harvesting metadata using the application programming interface (API), UiT university library.

Steering group:

- Johanne Raade (library director, UiT)
- Per Pippin Aspaas (head of research and publishing support, UiT university library)
- Stein Tronstad (Section leader, NPI)

Reference board:

- Jørgen Berge (UiT)
- Else Grete Broderstad (UiT)
- Alfred Hanssen (UiT)
- Hans-Kristian Hernes (UiT)
- Henry Patton (UiT)
- Lena Seuthe (UiT)
- Øyvind Lundesgaard (NPI)
- Ylva Ericson (NPI)

Bibliography

Abu-Alam T.S. (2019). Open Arctic Research Index: Final report and recommendations. *Septentrio Reports (No. 3).* https://doi.org/10.7557/7.4682

AMAP (2011). AMAP Assessment 2011: Mercury in the Arctic. *Arctic Monitoring and Assessment Programme (AMAP), Oslo, Norway. xiv + 193 pp.* ISBN 978-82-7971-068-4

Arctic Council (2013). Summary for policy-makers. Arctic Resilience Interim Report. *Stockholm Environment Institute and the Stockholm Resilience Centre to the Arctic Council's 8th Ministerial Meeting in Kiruna, Sweden.* http://hdl.handle.net/11374/1629

Beck, I., Huffman, L. T., Xavier, J. C. C. & Walton, D. W. H. (2014). Education and Polar Research: Bringing Polar Science into the Classroom. *Journal of Geological Resource and Engineering 4, 217-221.* https://doi:10.17265/2328-2193/2014.04.004

Forbes, D.L. (2011). State of the Arctic Coast 2010 – Scientific Review and Outlook. *International Arctic Science Committee, Land-Ocean Interactions in the Coastal Zone, Arctic Monitoring and Assessment Programme, International Permafrost Association. Helmholtz-Zentrum, Geesthacht, Germany, 178 p.* http://arcticcoasts.org, ISBN 978-3-9813637-2-2

IPCC (2007). Climate Change 2007: Synthesis report, edited by Pachauri, R. K. & Reisinger, A. *IPCC, Geneva, Switzerland, 104 pp.* https://www.ipcc.ch/report/ar4/syr/

IPCC (2013). "Summary for Policymakers". In Climate Change 2013: The Physical Science Basis, edited by Stocker, T.F., & Qin. D. H. *Cambridge, United Kingdom and New York, NY, USA: Cambridge University Press, 1535 pp.* https://www.ipcc.ch/report/ar5/wg1/

Johnson, N., Druckenmiller, M. & Pulsifer, P. (2019) The Exchange for Local Observations and Knowledge of the Arctic (ELOKA): Working towards interoperability of community-based data platforms. *Polar Data Forum*

Krupnik, I., Allison, I., Bell, R., Cutler, P., Hik, D., López-Martínez, J., Rachold, V., Sarukhanian, E. & Summerhayes, C. (2011). Understanding Earth's Polar Challenges: International Polar Year 2007–2008. *University of the Arctic, Rovaniemi, Finland / CCI Press, Edmonton, Alberta, Canada and ICSU/WMO Joint Committee for International Polar Year 2007–2008, 724 pp.* ISBN 978-1-896445-55-7

Longva, L. & Høydalsvik, S. (2012). High North Research Documents: your source for research documents on the North. *Polar libraries bulletin 68, 7-9.* http://hdl.handle.net/10037/4733

Longva, L. (2011). NAROS: Northern Areas Open Scholarly Documents. http://DOI:10.7557/13.1312

Norsk Polarinstitutt, strategi 2019 – 2024. *Retrieved from* http://www.npolar.no/npcms/export/sites/np/files/vedlegg/strategi.pdf

Ricci, C.-A., Egerton, P. (2010). European Research in the Polar Regions: Relevance, strategic context and setting future directions in the European Research Area. *Edited by the ESF European Polar Board for European Science Foundation, 24 pp.* ISBN: 978-2-918428-30-5

Tenopir, C., Allard, S., Douglass, K., Aydinoglu, A. U., Wu, L., Read, E., Manoff, M. & Frame, M. (2011) Data Sharing by Scientists: Practices and Perceptions. *Practices and Perceptions. PLOS ONE 6(6): e21101.* https://doi.org/10.1371/journal.pone.0021101

Turner, J., Bindschadler, R.A., Convey, P., di Prisco, G., Fahrbach, E., Gutt, J., Hodgson, D.A., Mayewski, P.A. & Summerhayes, C.P. (2009). Antarctic climate change and the environment. *Cambridge: Scientific Committee on Antarctic Research, 526 pp.* ISBN 978-0-948277-22-1

Turner, J., Bracegirdle, T.J., Phillips, T., Marshall G.J. & Hosking, J.S. (2013). An initial assessment of Antarctic sea ice extent in the CMIP5 models. *Journal of Climate 26, 1473–1484.* https://doi: 10.1175/JCLI-D-12-00068.1

Wilkinson, M., Dumontier, M., Aalbersberg, I. & et al. (2016). The FAIR Guiding Principles for scientific data management and stewardship. *Sci Data 3, 160018.* https://doi.org/10.1038/sdata.2016.18