

Open Research Data, Data Portals and Data Publication – an Introduction to the Data Curation Landscape

by Kirsten Elger^{1*}, Boris Biskaborn², Heinz Pampel¹ and Hugues Lantuit²

Abstract: During the past decade, the relevance of research data stewardship has been rising significantly and data publication has become more familiar. Preservation of research data for long-term use, including its storage in adequate repositories has been identified as a key issue by the scientific community as well as by research agencies and the public. In practice, however, the current state of data sharing and re-use requires considerable improvement. This paper reviews recent developments in this area, and aims to provide some guidance to the increasing diversity of newly developed digital solutions, such as data journals, online data repositories, and citable digital object identifier (DOI) for datasets. We examine the differences and similarities between different examples of Arctic-related data management, including the newly created database of the Global Terrestrial Network for Permafrost GTN-P, a Canadian example of a (meta)data portal (Polar Data Catalogue), and examples of data repositories (e.g., PANGAEA, Nordicana D) and data journals (e.g., Earth System Science Data). We also describe the newly established Registry of Research Data Depositories (re3data.org) as a convenient resource for individual researchers to get an overview on and identify an appropriate repository for their scientific datasets as well for funding agencies during the evaluation process of the data management plan of research proposals.

Zusammenfassung: In den letzten zehn Jahren ist die Bedeutung des Zugangs und der Nachnutzung von Forschungsdaten gestiegen. Neue Publikationsstrategien für Forschungsdaten stellen sicher, dass wissenschaftliche Daten dauerhaft in geeigneten Daten-Repositorien gespeichert und zugänglich gemacht werden können. Auch wenn die Umsetzung dieser Publikationsstrategien von der wissenschaftlichen Community und von Forschungsförderorganisationen als zentrale Herausforderung für das Wissenschaftssystem benannt wird, ergeben sich in der Praxis noch viele Herausforderungen. Dieser Artikel gibt einen Überblick über aktuelle Entwicklungen im Bereich des Forschungsdatenmanagements. Exemplarisch werden einige Beispiele für den offenen Zugang vorgestellt und Publikationsstrategien für Forschungsdaten beschrieben. U.a. werden Aufgabe und Dienstleistung von Daten-Repositorien, Daten-Journalen sowie Daten-Portalen im Bereich der Arktisforschung erläutert. Exemplarisch werden folgende digitale Forschungsdateninfrastrukturen vorgestellt: Das kanadische (Meta)data-Portal Polar Data Catalogue, die neue Datenbank des Global Terrestrial Network for Permafrost GTN-P (Metadaten und Daten), die Daten-Repositorien PANGAEA und Nordicana D sowie das Daten-Journal Earth System Science Data (ESDD). Darüber hinaus wird der Service des Registry of Research Data Depositories (re3data.org) vorgestellt. Dieses internationale Verzeichnis unterstützt Forschende und Forschungsförderer bei der Identifikation von geeigneten Daten-Repositorien zur Speicherung und Zugänglichmachung ihrer Forschungsdaten.

INTRODUCTION

Research data are fundamental for scientific research. Especially in the Geosciences, where observational data are not reproducible, and many historical data sets are important

doi:10.2312/polfor.2016.009

¹ GFZ German Research Centre for Geosciences, Telegrafenberg, 14473 Potsdam, Germany;

* Corresponding author: <kelger@gfz-potsdam.de>

² Alfred Wegener Institute, Helmholtz Centre for Polar and Marine Research, Telegrafenberg, 14473 Potsdam, Germany.

This paper represents a supplement to the International Conference “Our Climate – Our Future: Regional perspectives on a global challenge”, 6–9 October 2014 in Berlin, Germany.

Manuscript received 3 June 2015; revised version 17 January 2016; accepted 21 March 2016.

benchmarks, e.g., in the context of climate change, data curation and publication meets the needs for scientific reproducibility as well as offering potential for re-use.

For more than a decade, the free and open access to scientific results and research data has become more and more important, both by political expectations and technical possibilities. Especially the advent of the internet and new digital possibilities enabled the collection of high quantities of research data and often triggered a movement from empirical science towards data-driven science. Today, many datasets are available via the internet. Very often, however, these datasets are not used to their full extent because they are not systematically archived or made readily accessible or are not sufficiently described (COPDESS 2015). Another impediment to free and open data exchange is the reluctance of researchers to share their data in the absence of appropriate credit for the large investment of time and intellectual effort that went into collecting, processing, and describing the data, as well as concern that their data may be misused or misinterpreted.

This article reviews recent developments in the data curation landscape, and aims to provide some guidance to the increasing diversity of newly developed digital solutions, such as data journals, online data repositories, and citable digital object identifier (DOI) for datasets followed by presenting some exposed examples. We neither aim to provide a complete overview on the constantly increasing number of data repositories and portals, nor will we recommend or favour any of them. After a brief overview on the latest political developments, we introduce the key components for open research data (e.g., metadata, digital object identifier, different possibilities for data publication) and further examine the differences and similarities between them by introducing different examples of Arctic-related data management, including the newly created database of the Global Terrestrial Network for Permafrost GTN-P, the Polar Data Catalogue, a Canadian example of a (meta)data portal, and examples of data repositories (e.g., PANGAEA, Nordicana D) and data journals (e.g., Earth System Science Data). We also introduce the newly established Registry of Research Data Depositories (re3data.org) as a convenient resource for individual re-searchers to identify an appropriate repository for their scientific datasets as well for funding agencies during the evaluation process of the data management plan of research proposals. Note that all acronyms are listed (Tab. 2) at the end of this paper.

Politics

Following-up the BERLIN DECLARATION ON OPEN ACCESS TO KNOWLEDGE IN THE SCIENCES AND HUMANITIES (2003)

there are more and more initiatives to foster the free and open access not only to scientific journal articles but also to research data and metadata. Especially during the last about five years, there is an increasing expectation by research funding agencies and the public that close attention should be given to data management, and many research agencies and organisations require that grant proposals explicitly address the plans for metadata and data archiving e.g., the European Commission, the German Research Foundation (DFG), the International Arctic Science Committee IASC, etc.). Internationally, the largest impact and a testimony for open science was the G8 Science Ministers Statement from June 2013. They claim that *“to the greatest extent and with the fewest constraints possible, publicly funded scientific research data should be open [...] whilst acknowledging the legitimate concerns of private partners”* and that *“increasing free access to peer-reviewed, published research results will require sustainable solutions”* (G8 SCIENCE MINISTERS, 2013). This statement was followed by several national and international initiatives like the “EU Implementation of the Open Data Charter” (EUROPEAN COMMISSION, 2013a), which requires, e.g., the use of open formats, semantic interoperability, to ensure data quality and documentation, and a clear definition of intellectual property rights, e.g., by using open licences for scientific data, etc. The “Digital Agenda 2014–2017 of the Federal Government of Germany” supports a *“comprehensive open access strategy designed to enhance incentives and ensure more efficient, ongoing access to publicly funded research publications and data”* (BMW i et al. 2014).

Several other countries adopted open science policies, like, e.g., G8 Open Data Charter – Canada’s Action Plan (GOVERNMENT OF CANADA 2014), the memorandum on “Public Access to the Results of Federally Funded Research” (OFFICE OF SCIENCE AND TECHNOLOGY POLICY 2013) and the “US Open Data Action Plan” (US GOVERNMENT 2014), or the GERMAN ALLIANCE OF SCIENCE ORGANISATIONS (2010) with a clear statement in favour of open research data in the preamble: *“Quality-assured research data are a cornerstone of scientific knowledge and [...] can often serve as the basis for further research. [...] Preserving research data over the long-term and making them available therefore does not only serve the verification of prior results, but also, to a large extent, the obtaining of future ones. It is a strategic task to which science and the humanities, politics as well as other parts of society, must contribute.”*

An important step for open data in Arctic research is the “Statement of Principles and Practices for Arctic Data Management” of the International Arctic Science Committee that was released in April 2013 (IASC 2013). It states that all research projects seeking endorsement by IASC must adhere to the principle of full and open access to data, and must make metadata (basic descriptive information of collected data) available in an internationally recognised standard-format to an appropriate catalogue or registry. The policy further states that IASC should actively encourage adherence to the principles and may withdraw project endorsement if necessary (IASC 2013). The full open access to scientific results and research data, together with a comprehensive data management plan, is also required for every project funded within the new EU framework program Horizon 2020 (EUROPEAN COMMISSION 2013b)

and other funding agencies (e.g., National Science Foundation NSF, German Research Foundation DFG).

COPDESS and the Statement of Commitment

Founded in October 2014, a new initiative joins together major Earth and Space Science publishers with primary Earth and Space Science data repositories and related consortia, the Coalition on Publishing Data in the Earth and Space Sciences (COPDESS 2015, HANSON et al. 2015). To mark the launch of the new initiative they published a “Statement of Commitment” that signals *“an important progress and a continuing commitment by publishers, data facilities and unions to enable open data in the Earth and Space Sciences”* (COPDESS 2015). By the end of 2015 it is already signed by 38 of the leading publishers and data centres in Earth and Space Sciences, including Science, Nature, AGU, EGU, Springer, Copernicus, Elsevier, NSIDC, ICSU and many more. By committing to the proper citation of scientific datasets in scientific articles (see below), this statement is a major step forward for the acknowledgment and recognition of the important scientific work to make research data publicly available together with a comprehensive description of the published datasets and clear indications about where to find them. In addition and often as a consequence of their signature of the “Statement of Commitment”, many institutions and publishers have released policies for full and open access to research data.

The consortium has pointed out that even though it is widely acknowledged that *“scholarly publication is a key high value entry point in making data available, open, discoverable, and usable.”* and *“most publishers have statements related to the inclusion or release of data as part of publication [...] the vast majority of data submitted along with publications are in formats and forms of storage that makes discovery and re-use difficult or impossible.”* They further recommend that research datasets should preferably be stored in *“appropriate domain repositories”* (COPDESS 2015) and are currently developing an online registry for these.

The key message in favour of acknowledging and promoting data publications and important incentive to convince scientists to share and publish their data, is the commitment *“to promote referencing of data sets using the Joint Declaration of Data Citation Principles”* (DATA CITATION SYNTHESIS GROUP 2014), in which *“citations of data sets should be included within reference list”* (COPDESS 2015).

In addition they agreed to make sure to *“include in research papers concise statements indicating where data reside and clarifying availability”* and *“to promote and implement links to data sets in publications and corresponding links to journals in data facilities via persistent identifiers”*, ideally by registered DOI’s (COPDESS 2015).

DATA PUBLICATION WITH DIGITAL OBJECT IDENTIFIER (DOI)

Publishing research datasets with assigned digital object identifier (DOI) has emerged as convenient solution for publishing citable and persistently accessible research data. By rules of

the International DOI Foundation the registration of a DOI requires the submission of at least a minimum set of metadata. In the following we will give a brief introduction to DOI, metadata and different formats for the DOI-referenced data publication.

Digital Object Identifier (DOI)

A DOI is an “online reference (digital), pointing to (identifying) a resource (object). The DOI system links, through a directory, references and web addresses of an object to a “landing” page providing information on access and metadata about that object – at a minimum its creator, title, publisher, year of publication, and DOI. This allows DOIs to provide a stable, persistent, resolvable reference taking users to an object, even if web addresses or other references to the location of an object, or its content, change” (HORTON 2015). Using DOIs brings stability to data referencing in the digital era where scientific sources and references are much more than printed paper (e.g., databases, websites, audio, blogs, video, social media, etc.). A DOI persistently directs to the related text or data set, independent of website changes, servers getting switched off or other changes of sources.

Digital Object Identifiers are well established for scientific articles since the early 2000s. The concept to also cite data sets, archived in data repositories, with a persistent DOI, has been developed in Germany within the STD-DOI project funded by the German Science Foundation between 2004-2007 with the participation of the Technical Information Library Hannover (TIB), the GFZ German Centre for Geosciences, PANGAEA, OKRZ, and the German Space Agency DLR (KLUMP et al. 2006). Following this was the foundation of DataCite in December 2009, which is a world-wide nonprofit organisation for DOI for research data, under the realms of the “International DOI Foundation”. Among other ways to persistently identify references to digital objects, DOIs have emerged as the leading system for text and data publication (COPDESS 2015).

Nevertheless, even if a DOI guarantees the discovery of research data, it may not guarantee the data quality or enable data re-use. Therefore, high-quality research data must be accomplished by metadata for data discovery and re-use. Assuming the scientifically correct and careful acquisition and processing of a dataset, the scientists, the data repository or the publisher should make sure that the dataset is accompanied by an adequate and sufficient description suitable for data discovery and re-use.

Metadata

To meet the requirements for intelligent openness standards that published data sets must be intelligible and usable by others for both, verification of research results and data re-use, data sets must be supported by explanatory metadata (ROYAL SOCIETY LONDON 2012). Metadata, or data about data, is “structured information that describes, explains, locates, or otherwise makes it easier to retrieve, use, or manage an information resource”. (NATIONAL INFORMATION STANDARDS ORGANIZATION 2004).

There is a general difference between structural metadata which is essential for data re-use (e.g., the information of instruments or sensors used to measure the data, applied functions or processing steps, quality control, etc.) and the more descriptive metadata for data discovery (e.g., the information of the existence of a dataset, the brief description of the data and the project, i.e. in form of an abstract, measurement period, contact information of the data originator and distributor, and information about where to obtain the datasets). Metadata for data discovery aims to discover research data via search engines, data portals, etc., whereas structural metadata is essential for data re-use and often directly attached to the datasets (e.g., in form of a README file).

Metadata Standards

Using international metadata standards, that mainly exist for metadata for data discovery allow database interoperability and are a sign of quality. Prominent metadata schemes for geo-referenced datasets and services are, e.g., the ISO19115 Schema accessed on 28 December 2015, the DATACITE (2014) Metadata Schema, or outcomes of the European INSPIRE initiative, the “Infrastructure for Spatial Information in the European Community”. To facilitate data discovery and re-use, it is strongly recommended to not only provide the minimum set of obligatory metadata, that is usually very small in number and designed to fit to every type of data, but to make sure to describe the datasets as good as possible, e.g., by using the recommended fields in the Datacite Metadata Schema 3.1 (DATACITE 2014, see Tab. 1).

In addition, there are more and more initiatives to define standards for structural metadata. These are always discipline-specific and often developed in large collaboration projects, or global networks, e.g., as metadata forms, protocols or data models that are often accomplished by detailed instruction for the data collection in the field. Examples are the GTN-P metadata forms (BURGESS 2000), the ADAPT standard protocols of the “Canadian permafrost research programme Arctic Development and Adaptation to Permafrost in Transition” or the recommendations for seismic metadata of the INTERNATIONAL FEDERATION OF DIGITAL SEISMOGRAPH NETWORKS (FDSN, 2014) that are a combination of structural metadata and metadata for data discovery and are now recommended for all seismic networks worldwide.

Data Publication formats

Research data may be published supplementary to journal articles, with a descriptive article in one of the new Data Journals, or as independent entities (KATZ & STRASSER 2015) with accompanying structural metadata, e.g., in form of a readme or a data report. For curation purposes and data re-use it is always recommended to store the data in open access data repositories and not to submit them to the journals as supplementary material. This is also a recent recommendation of publishers within the COPDESS consortium. An open access repository enables data re-use even if the corresponding articles was not published in an Open Access Journal.

Category	Sub-category
Resource information	DOI, publisher, publication year , licences
People involved	Authors: name , affiliation, role (e.g.: data manager, distributor, editor, hosting institution, project leader, project member, researcher, research groups, etc.), authors-id type and number (ORCID-ID), Contact person: name, email, affiliation, Contributors: name, affiliation, email, authors-id and number (authors and contributors may be people or institutions),
Description	Abstract , methods, table of contents, series numbers, others
Keywords	Free keywords and controlled vocabulary via thesauri to fulfil metadata standards
Spatial & temp. coverage	Visual control via a mapping tool and the possibility to enter bounding boxes and or point values, each with different temporal domains and descriptions,
Dates	e.g. created, embargo until, valid from ... to
Related reference	e.g. for documentation, supplements, references, new versions, etc.

Tab. 1: Metadata for data discovery used in the Data Repository of the GFZ German Research Centre for Geosciences. Bold fields are obligatory, the others “recommended for data discovery” in the DATA CITE (2014) Metadata Scheme 3.1. While the metadatabase is designed to produce DataCite, ISO19115, and other standards, the user interface in which the scientists enter their metadata uses a language that is understandable for scientists (e.g. creators are named authors, geoLocationPoints or geoLocationBoxes are inserted via the mapping tool or direct entry of coordinates, etc.). Every field that may be automatically generated (e.g., the URL of the metadataset or rights, dates of submission or acceptance) and is not visible for the scientist.

Tab. 1: Übersicht über die Metadaten für Data Discovery, die im Repository des Deutschen GeoForschungsZentrums GFZ erhoben werden. Es handelt sich sowohl um die Pflichtfelder (fett) als auch die für die Data Discovery empfohlenen zusätzlichen Metadatenfelder, wie sie im DATA CITE (2014) Metadaten Schema 3.1 definiert werden. Das GFZ Datenrepository unterstützt darüber hinaus auch ISO19115 und andere Metadatenstandards. Bei der Gestaltung des Metadatenerhebungsformular wurde besonderer Wert auf die Nutzerfreundlichkeit gelegt. Das Formular ist in einer für Wissenschaftler verständlichen Sprache angelegt, was sich z.B. in der Bezeichnung der beteiligten Wissenschaftler als Autoren (und nicht als “creators”) widerspiegelt. Auch die Eingabe der geografischen Koordinaten der Datensätze erfolgt mit Hilfe einer interaktiven Karte und es wurde Wert darauf gelegt, dass automatisch generierbare Informationen, wie z.B. die URL einer gewählten Lizenz oder auch das Datum der Einreichung eines Datensatzes, automatisch bezogen wird und nicht durch die Wissenschaftler eingetragen werden müssen.

METADATA PORTALS, REPOSITORIES, DATA JOURNALS, DATA REPORTS – A DATA PUBLICATION TOOLBOX

Metadata Portals

Metadata portals are the most important source to get information about projects, data, activities, and people involved (by metadata for data discovery). They usually have search engines and provide standardised metadata that may be exchanged with other portals via standard application programming interfaces (API), like, e.g., the Open Archives Initiative Protocol for Metadata Harvesting (OAI-PMH). Metadata portals provide information about where to find the data and aim to give as much information about data and project that a potential user is able to decide whether the described datasets are fitting his needs or not before accessing or downloading the data. Metadata portals are very often the entry point to data repositories (e.g., PANGAEA, NSIDC Data Search) and exist

for the whole range from collections of information stored in Excel sheets and published via a list on a website to large inter-institutional object-oriented databases, like, e.g., the Canadian Polar Data Catalogue with various search options and partly access to the datasets (see below). New developments are overarching national and international metadata portals that are retrieving their metadata not from individual scientists, but by harvesting it from other metadata portals and data repositories. Examples for these metadata catalogues of research data collections are, e.g., B2FIND, which was developed within the EU FP7 project EUDAT or the new “Arctic Data Explorer of the National Snow and Ice Data Center NSIDC”. The US initiative “Earth Cube” aims to develop a common cyberinfrastructure for the purpose of collecting, accessing, analysing, sharing, and visualizing all forms of data and related resources for solid Earth, hydrosphere, atmosphere, and space environment.

Data Repositories

Data repositories are digital research infrastructures for archiving and distributing research data sets. They “ensure a maximum of accessibility, stability and reliability to facilitate working with and sharing of research data.” (PAMPEL et al. 2013) and often have a data discovery metadata portal as entry point. The landscape of data repositories is very heterogeneous. Datasets are accompanied by descriptive metadata and often have links to related scientific articles. Data repositories may or may not provide DOIs to their datasets and the completeness of metadata is quite variable. There are institutional, disciplinary, multidisciplinary, and project specific data repositories (PAMPEL et al. 2013), however, it is recommended to archive datasets preferably in open access, theme-specific or institutional data repositories that provide DOIs to the archived datasets, hence making them citable (COPDESS 2015, EUROPEAN COMMISSION 2013b).

Data Journals

Data journals publish articles about original research datasets, collections or data portals. The articles mainly describe datasets, data collections or databases without giving an interpretation of the data itself. Most data journals, e.g., Earth System Science Data (ESSD) or Nature’s Scientific Data follow a distributed way for the actual data; i.e., the data are not accessible attached to the journal article but stored in a data repository and will be reviewed together with the scientific article. Data journals are rapidly evolving in all scientific disciplines and some have already been indexed in the Web of Science (e.g., ESSD).

Data Reports

For many datasets, the standardised metadata fields for data discovery are not suitable alone for an appropriate data description and many datasets are accompanied by readme files for the technical description of the datasets. These readme files, however, are usually only accessible after downloading and unzipping a dataset. Discovering in the readme file that the dataset does not fit the needs of the researcher,

e.g., because the temporal coverage is not fitting, requires the deletion of the dataset and a second or third trial and often leads to the decision of not using the data set at all.

An additional option for the description of datasets is to publish the dataset description in form of a data report. They have a flexible format, are fully readable before downloading the datasets (in contrast to many readme files) and especially attractive for datasets that are not large enough to be described in a data article.

For many years, German Research Centre for Geosciences (GFZ) publishes “Scientific Technical Reports (STR)” as a report series which is electronically persistent available and citable with assigned DOIs. Typical contents for STRs were PhD theses or project reports. In 2011, this series was opened for the description of datasets as Scientific Technical Report Data. These data reports offer a full and consistent overview and description to all relevant parameters of linked published datasets. They have a flexible format that is applicable to descriptions of datasets for all scientific fields and are internally reviewed. Whenever possible, we are developing templates for different types of datasets (e.g., for descriptions of seismological experiments realised with instruments of the Geophysical Instrument Pool Potsdam (GIPP, e.g., ASCH et al. 2014).

Furthermore, it is possible to publish one data report for several published datasets that are all referring to the same data report as description (e.g., different runs of a climate model). With the newly developed project-specific design of data reports and specific content of landing pages, (e.g., ICDP Operational Reports and datasets, LORENZ et al. 2015a, b) they have qualified as helpful tool to fill the gap between basic metadata and restricted readme information on the one hand and preparing extended journal articles on the other hand.

EXAMPLES FOR DATA DISCOVERY AND PUBLICATION

Polar Data Catalogue: Canadian portal for interdisciplinary Arctic Research

The Canadian Polar Data Catalogue (PDC) is an online portal for data discovery (and partly data access) covering a wide range of scientific fields from natural sciences to policy, health, and social sciences (FRIDDELL et al. 2014a, 2014b). It has been developed during the International Polar Year (IPY) 2007/2008 and is co-financed by the ArcticNet at the University Laval and the Canadian Cryospheric Information Network (CCIN) at the University of Waterloo. Each scientist who is funded by ArcticNet or any of its partners or is doing research at any field station supported by ArcticNet partners, is supposed to submit their metadata for data discovery to PDC on an annual base. PDC metadata follows ISO19115 standards (see FRIDDELL et al. 2014a for an overview) and metadata standards of the Federal Geographic Data Committee (FGDC, listed in ELGER et al. 2014).

In June 2015, the portal has more than 2100 metadata entries. In addition, PDC hosts about 250 datasets from the International Polar Year 2007/2008 and almost 28.000 Radarsat 1 satellite images from both Polar Regions, Radarsat-2 satellite

mosaics of Antarctica, and sea-ice charts of the Canadian Ice Service.

In addition to full-text and advanced search options with various parameters, the catalogue has a geospatial web-based mapping system that can be used in combination with various search options (Fig. 1). The geospatial search tool in PDC has been developed with consultation of northern communities who were especially interested to search for projects that have taken place in their regions. This interest also led to the development of a PDC-lite version that is fully operational under low internet bandwidth connections that are abundant in many parts of the Canadian North (see ELGER et al. 2014 for an extended description).

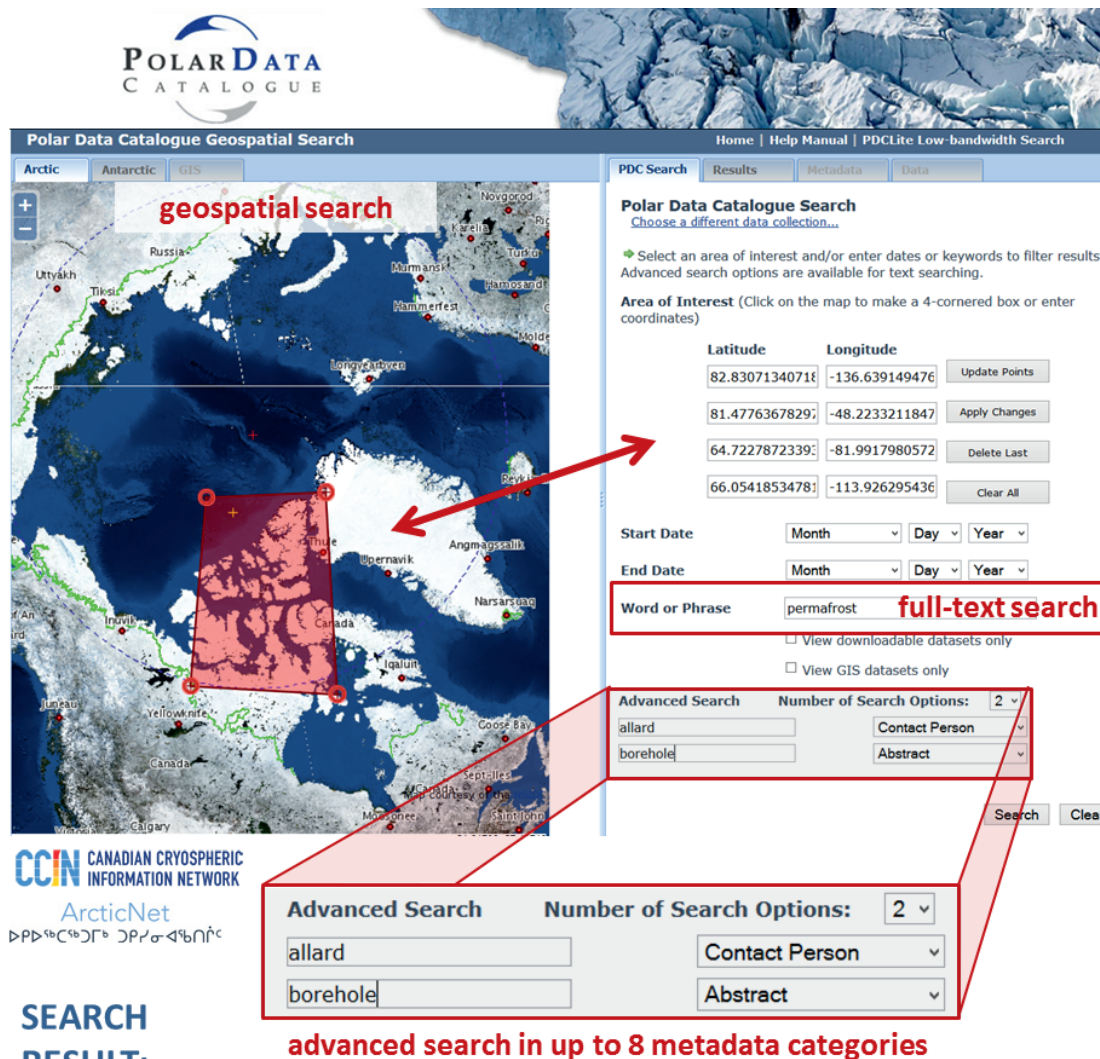
The new Database of the Global Terrestrial Network for Permafrost GTN-P – an example for a thematic database

Permafrost is formally defined as perennial frozen ground that remains at or below 0 °C for at least two consecutive years. Some 25 % of the global land mass is underlain by permafrost, and is increasingly vulnerable to degradation as a result of global climate change. The Global Terrestrial Network for Permafrost (GTN-P) is the prime international program concerned with permafrost monitoring. Initiated in 1999 by the Global Climate Observing System (GCOS) and the Global Terrestrial Observing System (GTOS) of the World Meteorological Organisation (WMO) and managed by the International Permafrost Association (IPA), GTN-P is dedicated to long-term monitoring of permafrost temperature and active-layer thickness, with the goal of obtaining a comprehensive view of the spatial structure, trends, and variability of changes in the active layer and permafrost in both Hemispheres.

GTN-P has two international components with very different data in temporal and spatial resolution:

- (1) The Thermal State of Permafrost (TSP) programme is a circumpolar network of more than 1000 boreholes ranging from 1 m to about 1000 m depths located in the polar regions of both Hemispheres and in mountainous permafrost areas (Fig. 2 for GTN-P sites in the northern Hemisphere). The temperature data have a high temporal resolution (hourly to daily measurements, annual for deep boreholes, BISKABORN et al. 2015a).
- (2) The Annual Thaw Depth (ALT) is mainly observed by the Circumpolar Active Layer Monitoring (CALM) program. This network incorporates 240 sites in Arctic, sub-Arctic, Antarctic, and mountainous regions. About 70 % of the sites are located in Arctic and Subarctic lowlands underlain by continuous permafrost. Discontinuous and mountainous permafrost areas contain 20 % and 11 % of ALT sites respectively. The distribution of ALT sites is not uniform, a circumstance attributable to historical circumstances and logistical constraints. The active layer is measured within grids with side lengths between 10, 100, and 1000 m on an annual basis (BISKABORN et al. 2015b).

Since the beginning, GTN-P adopted an open data policy and developed standardised metadata forms for both programmes (BURGESS et al. 2000). However, more than a decade after the implementation of GTN-P, the data was not used as much as it was intended. Reasons for this were, e.g., the large vari-



advanced search in up to 8 metadata categories



ISO 19115 metadata

link to datasets, e.g. in Nordicana D

citation information, title, description, etc.

Fig. 1: Example from the Canadian Polar Data Catalogue. The map-based geospatial search is linked with the geographic coordinates and allows search for a certain region. Alternatively and additional, users may use the full-text search or the advanced search with specific search in up to eight categories (e.g., contact person, responsible parties, abstract, title, etc.). The results show classical metadata for data discovery in ISO19115 and FGDC standards and show links to the data sets if applicable. An overview on all ISO19115 metadata fields in PDC is given in FRIDDELL et al. (2014a). The figure is a combination of different snapshots from PDC websites accessed in June 2015.

Abb. 1: Darstellung einiger Funktionalitäten des kanadischen Polar Data Catalogue. Die Suche von Metadatenfeldern erfolgt entweder über die Karte als Geodatendienst (links oben) oder durch die Eingabe von Suchparametern. Neben der Volltextsuche gibt es auch erweiterte Sucheinstellungen, die die Suche innerhalb von acht Kategorien erlaubt (Kontaktperson, Autoren, Kurzbeschreibung, Titel, usw.). Die Metadaten folgen Metadatenstandards gemäß ISO19115 und FGDC und beinhalten Verknüpfungen („link“) zu den vorhandenen Datensätzen, sofern diese existieren. FRIDDELL et al. (2014a) geben eine vollständige Übersicht über die im Polar Data Catalogue geführten ISO19115 Metadatenfelder. Die Abbildung ist eine Kombination verschiedener Momentaufnahmen der Websites mit zusätzlichen grafischen Elementen und wurde im Juni 2015 erstellt.

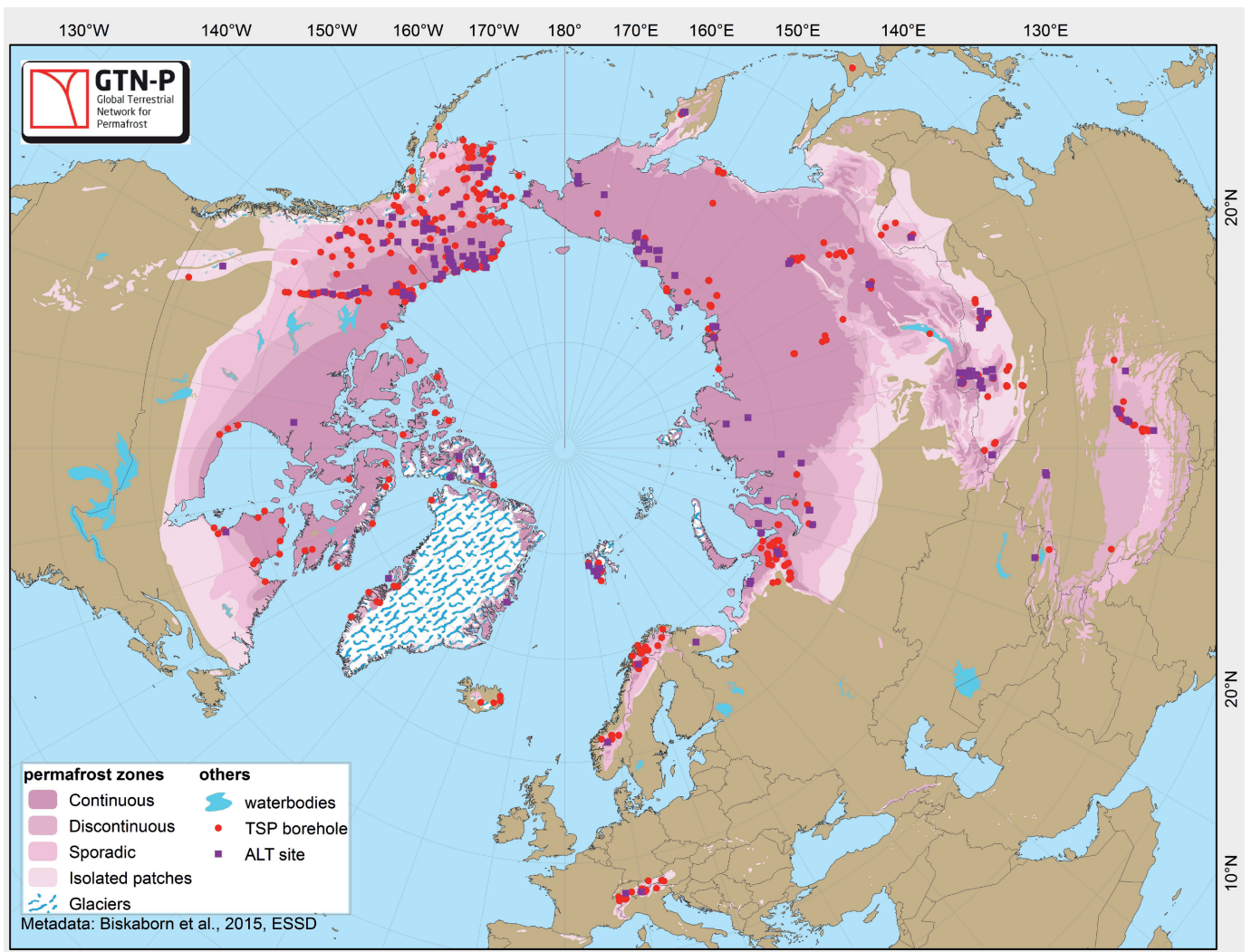


Fig. 2: Monitoring sites of the Global Terrestrial Network for Permafrost (GTN-P) in the Northern Hemisphere and their location within permafrost zones, as defined by the International Permafrost Association IPA. Red dots show permafrost boreholes of the Thermal State of Permafrost (TSP) programme, purple dots active-layer thickness (ALT) monitoring sites of the Circumpolar Active Layer Monitoring (CALM) programme.

Abb. 2: Übersicht über die Messstationen und -flächen des Global Terrestrial Networks for Permafrost (GTN-P) der nördlichen Hemisphäre und deren Lage in den jeweiligen Permafrost Zonen, welche von der International Permafrost Association IPA definiert wurden. Rote Punkte markieren Permafrost-Bohrlöcher, die im Rahmen des Thermal State of Permafrost (TSP) Programms gebohrt wurden. Violette Punkte zeigen die Lage langjähriger Messfelder der Tiefe der sommerlichen Auftauschicht des Permafrostes an, welche im Rahmen des Programms Circumpolar Active Layer Monitoring (CALM) angelegt wurden.

ability of data description in different countries (even despite the availability of standardised metadata sheets) and the lack of a central repository for borehole data (especially for the TSP component). While CALM data was accessible via the central CALM website (as mainly as Excel spreadsheets and txt files), borehole temperatures of the TSP network could only be partially accessed via national data portals, e.g., the Norwegian NORPERM database (for TSP data from Norway and Svalbard) or the National Snow and Ice Data Centre (NSIDC) and the Permafrost Laboratory at the University of Alaska Fairbanks (for many boreholes in Alaska and Russia). In November 2011, the necessity to build a joint database for both components of the Global Terrestrial Network for Permafrost was identified and the EU FP7 program Changing Permafrost in the Arctic and its Global Effects in the 21st Century (PAGE21) provided the frame for developing a GTN-P governance structure and a state-of-the-art data management system. This central GTN-P database was released in April 2014 and is

fully described in BISKABORN et al. (2015b) and at the corresponding website (see Tab. 1). It contains data, metadata, a search engine with specific filters for permafrost researchers and an automated data visualisation including quality control. Metadata information is entered into the Data Management System following the ISO19115 documentation by developing a metadata frame-work suitable for permafrost monitoring sites. Upload of both, data and metadata, is coordinated and quality checked by 36 GTN-P National Correspondents from 25 different countries.

PANGAEA: Data Publisher for Earth and Environmental Science

World Data Systems (WDS) reformed the former World Data Centres (WDC) of the International Council of Science (ICSU), which were created to host and distribute scientific

data sets during the 1957–1958 International Geophysical Year. WDCs originally were mono-disciplinary, e.g., WDC Climate, WDC Mare, etc. but as a whole covered a broad range of disciplines from the natural and social sciences to humanities. As all data held in WDC's were available for the cost of copying and sending the requested information, they were the first "open access" data centres. Today, many former WDC's changed to the multi-disciplinary ICSU World Data System. The World Data System (WDS) is an interdisciplinary body of the ICSU with the vision to promote long-term stewardship of,

and universal and equitable access to, quality-assured scientific data and data services, products, and information across a range of disciplines in the natural and social sciences, and the humanities.

PANGAEA is a certified member of the ICSU-WDS system and one of the first data repositories, assigning DOI to data sets. It is hosted by two German institutes: the Alfred-Wegener Institute Helmholtz Centre for Polar and Marine Research (AWI) and the Centre for Marine Environmental Sciences

PANGAEA*
Data Publisher for Earth & Environmental Science

data description section

Data Description

Citation: Duguay, Claude R; Soliman, Aiman; Hachem, Sonia; Saunders, William (2012): Circumpolar and regional Land Surface Temperature (LST), version 1, with links to geotiff images and NetCDF files (2007-2010). *University of Waterloo, Canada*, doi:10.1594/PANGAEA.775962

Related to: Hachem, Sonia; Allard, Michel; Duguay, Claude R (2009): Using the MODIS land surface temperature product for mapping permafrost: an application to northern Québec and Labrador, Canada. *Permafrost and Periglacial Processes*, 20(4), 407-416, doi:10.1002/ppp.672

Hachem, Sonia; Duguay, Claude R; Allard, Michel (2011): Comparison of MODIS-derived land surface temperatures with near-surface soil and air temperature measurements in continuous permafrost terrain. *The Cryosphere Discussion*, 5, 1583-1625, doi:10.5194/tcd-5-1583-2011

Langer, Moritz; Westermann, Sebastian; Bolke, Julia (2010): Spatial and temporal variations of summer surface temperatures of wet polygonal tundra in Siberia - implications for MODIS LST based permafrost monitoring. *Remote Sensing of Environment*, 114(9), 2059-2069, doi:10.1016/j.rse.2010.04.012

Soliman, Aiman; Duguay, Claude R; Saunders, William; Hachem, Sonia (2012): Pan-Arctic Land Surface Temperature from MODIS and AATSR: Product Development and Intercomparison. *Remote Sensing*, 4(12), 3833-3856, doi:10.3390/rs4123833

Westermann, Sebastian; Langer, Moritz; Bolke, Julia (2012): Systematic bias of average winter-time land surface temperatures inferred from MODIS at a site on Svalbard, Norway. *Remote Sensing of Environment*, 118, 162-167, doi:10.1016/j.rse.2011.10.025

New version: Duguay, Claude R; Soliman, Aiman; Hachem, Sonia; Saunders, William (2014): Circumpolar and regional Land Surface Temperature (version 2) with links to geotiff images (2007-01 to 2013-12). *University of Waterloo, Canada*, doi:10.1594/PANGAEA.836729

Project(s): ESA Data User Element - Permafrost (DUE-Permafrost)

Coverage: Median Latitude: 66.470000 * Median Longitude: 156.640000 * South-bound Latitude: 61.700000 * West-bound Longitude: 78.500000 * North-bound Latitude: 71.250000 * East-bound Longitude: -128.500000

Event(s): DUEPermafrost_Alaska * Latitude: 69.600000 * Longitude: -148.600000 * Device: Satellite remote sensing (SAT) * Comment: position describes the center of area
DUEPermafrost_Central_Yakutia * Latitude: 61.700000 * Longitude: 130.800000 * Device: Satellite remote sensing (SAT) * Comment: position describes the center of area
DUEPermafrost_Laptev_Sea * Latitude: 71.250000 * Longitude: 131.000000 * Device: Satellite remote sensing (SAT) * Comment: position describes the center of area

Comment: The Land Surface Temperature (LST) products and services identified by users for the pan-Arctic (25 km resolution) scales include weekly and monthly averages from 2000 to 2010 from which annual averages can also be calculated. The LST processing integrates the LST level 2 products from MODIS and AATSR distributed by NASA and ESA, respectively. Post-processing functions supply University Waterloo-level-3 weekly and monthly LST products for regional (1 km) and pan-Arctic (25 km) scales. The pan-Arctic product, with a spatial resolution of 25 km, is produced by spatial averaging of 1-km observations.

MOD11_L2 and MYD11_L2 LST (Version 5 from NASA Terra and Aqua satellites) and ATS_NR_2P (from ESA Envisat satellite) products at 1 km resolution are used as input data to generate pan-Arctic and regional products. The original geo-located LST observations are characterized by an irregular distribution based on the satellite orbits. The Northern Hemisphere EASE-Grid Lambert Equal Area Azimuthal projection with a sphere datum (with a radius of 6371.228 km) was selected as the standard projection for the operational pan-Arctic and regional products. Original MODIS and AATSR LST level 2 observations are projected using the EASE-Grid coordinate system and interpolated to a regular EASE-Grid with 1 km spacing using triangulation. The EASE-Grid projection was chosen since this is the system adopted by the GlobSnow project and for most snow and ice products distributed by NSIDC. Local time is calculated using UTC acquisition time and longitude. UTC is extracted from ADS information for AATSR data and from the file name of MODIS level 2 (Terra and Aqua) products, yielding a temporal accuracy of ± 15 minutes, which is found to be sufficient for weekly and monthly products.

data download section

Download dataset as tab-delimited text (use the following character encoding: ISO-8859-1: ISO Western (PANGAEA default))

1 <input type="checkbox"/>	2 <input type="checkbox"/>	3 <input type="checkbox"/>	4 <input type="checkbox"/>	5 <input type="checkbox"/>	6 <input type="checkbox"/>	7 <input type="checkbox"/>
Event	Date/time start	Date/time end	URL data	Comment	Format	Size [kByte]
DUEPermafrost_panarctic <input type="checkbox"/>	2005-02-01	2009-12-31	Link	panarctic AATSR LST, monthly 2005-2009, 25 km grid	geotiff	36178
DUEPermafrost_panarctic <input type="checkbox"/>	2005-02-01	2009-12-31	Link	panarctic AATSR LST, monthly 2005-2009, 25 km grid	netCDF	96495
DUEPermafrost_panarctic <input type="checkbox"/>	2008-01-01	2008-12-31	Link	panarctic AATSR LST, weekly 2008, 25 km grid	geotiff	165303
DUEPermafrost_panarctic <input type="checkbox"/>	2008-01-01	2008-12-31	Link	panarctic AATSR LST, weekly 2008, 25 km grid	netCDF	690165
DUEPermafrost_panarctic <input type="checkbox"/>	2000-03-01	2010-09-30	Link	panarctic MODIS LST, monthly 2000-2010, 25 km grid	geotiff	128077
DUEPermafrost_panarctic <input type="checkbox"/>	2000-03-01	2010-09-30	Link	panarctic MODIS LST, monthly 2000-2010, 25 km grid	netCDF	294732

Fig. 3: Example of a PANGAEA landing page for a data collection of remotely sensed land surface temperature products from the ESA Data User Element DUE Permafrost Project (DUGUAY et al. 2012). The metadata fields in the data description section (upper part) include information about how to cite the datasets (citation), cross-references to related journal articles or other datasets (related to), the project name, spatial coverage, a data/ project description section (comment), the parameter description (here: bands, periods, time series), licence for working with the data, and the file size. The data download section (lower part) has information on the temporal domain and links to data download in different formats (geotiff, NetCDF) and the download link to the Product guide with the full description of the dataset (not shown in the figure).

Abb. 3: Beispiel einer DOI Landing Page des PANGAEA Datenrepositoriums. Bei dem gezeigten Datensatz handelt es sich um eine Kollektion von Fernerkundungsprodukten (Land Surface Temperature), die im Rahmen des ESA Data User Element (DUE) Permafrost Projekt entstanden sind (DUGUAY et al., 2012). Die Metadaten im beschreibenden Teil oben beinhalten neben der Zitationsvorgabe auch Querverweise zu wissenschaftlichen Artikeln oder anderen relevanten Datensätzen, den Projektnamen, die geographische Abdeckung der Datensätze, eine kurze Projektbeschreibung, die Lizenz zur Nachnutzung der Daten und die Dateigröße. Im unteren Teil sind die Verknüpfungen ("link") zu den Datensätzen gegeben, welche in zwei unterschiedlichen Formaten (geotiff und NetCDF) verfügbar sind. Darüber hinaus gibt es hier einen "link" zu der vollständigen Produktbeschreibung (nicht in der Abbildung sichtbar).

(MARUM). The free and open access international data library aims at archiving, publishing and distributing geo-referenced data from Earth system research with guaranteed long-term data availability. The datasets are accomplished by standardised metadata (ISO19115) and can be fully read before downloading the datasets (Fig. 3). The policy of data management and archiving follows the Principles and Responsibilities of ICSU World Data Systems and the OECD “Principles and Guidelines for Access to Research Data from Public Funding”. Authors submitting data to the PANGAEA data repository for archiving agree that all data are provided under a Creative Commons License. Each dataset can be identified, shared, published, and cited by using a DOI. After editorial review, datasets are archived by data curators as supplements to journal articles or independent datasets or data collections. At the end of 2015, there are more than 34.000 dataset DOIs registered by PANGAEA.

Nordicana D

Nordicana D is a DOI-referenced, open access, online data repository that was launched in February 2013 at the Centre d'études Nordiques (Centre for Northern Studies, CEN), Québec, CA. Until December 2015 they have published 23 data collections, mostly with long-term meteorological time-series from climate stations and (shallow) boreholes in the Canadian North, provided by CEN scientists and project partners. Nordicana D is following a versioning strategy, i.e. whenever there is an update to the datasets, e.g., by adding another year to the data or relocating stations, etc., a new version number will be published for the same DOI. As required by the DOI assignment, all version numbers are accessible and the website provides a history of changes for each version (Fig. 4). For each Nordicana D collection standardised metadata for data discovery are archived in the Polar Data Catalogue (see above for details in the respective section).

The description of the datasets allows easy identification of a suitable dataset before downloading the data. Data download is possible for different temporal resolutions, from measured data every hour, to daily, monthly, and annual averages. It has started with mostly micro-meteorological data from local climate stations (air and ground temperatures, wind speed and direction, net radiation, snow height), but also, more recently, other types of data. These are, e.g., permafrost soil geochemical and geophysical data from the ADAPT project, and lemming monitoring data, with additional data entries in progress, including for circumpolar diatoms and microbial DNA.

Earth System Science Data (ESSD) – the first data journal for the Geosciences

Launched in 2008 as one of Copernicus' Open Access Journals, ESSD is an international, interdisciplinary journal for the publication of articles on original research data (sets), furthering the re-use of high-quality data of benefit to Earth system sciences (PFEIFFENBERGER & CARLSON 2011). It is an ideal place to publish scientific articles on research datasets or collections, data portals and repositories as well as new-developed nontrivial statistical and other methods. Any interpretation of the datasets or comparison of the described

methods with others is beyond the focus of the journal. ESSD is open for various formats, ranging from regular articles, to brief communications (e.g., on additions to datasets) and commentaries, to review articles and special issues.

ESSD uses a distributed approach with the datasets being stored in recognised data repositories with an assigned DOI (e.g., PANGAEA) and cross-referenced in the respective ESSD article. The interactive public peer-review with discussion papers is also open for comments from the scientific community and includes the review of the published datasets or portals described in the articles, especially with respect to the data description and usability. On March 17, 2015, ESSD was the first data journal that has been indexed by Thomson-Reuters' Web of Science (ESSD news 2015).

RE3DATA.ORG – A REGISTRY OF RESEARCH DATA REPOSITORIES

With the increasing relevance for appropriate storage of research data, more and more data repositories and archives are evolving across all scientific fields. Funding agencies require the formulation of a data management plan for each research proposal in which they often want to know in which repository the newly measured or produced data will be stored. Since 2012 re3data.org (Registry of Research Data Repositories) offers researchers, funding organisations, libraries and publishers a comprehensive overview of the heterogeneous landscape of research data repositories. Project partners in re3data.org include the GFZ German Research Centre for Geosciences, Humboldt-Universität zu Berlin, Purdue University, and Karlsruhe Institute of Technology (KIT). The work of re3data.org has been funded by the German Research Foundation (DFG) in Germany and the Institute of Museum and Library Services (IMLS) in the United States. The re3data.org registry cooperates with other Open Science initiatives like BioSharing and OpenAIRE.

In January 2016 re3data.org lists more than 1,400 research data repositories from around the world covering all academic disciplines (Fig. 5). The databases, repositories and data infrastructures are described, following the “Metadata Schema for the Description of Research Data Repositories” that has been developed within the project (RÜCKNAGEL et al. 2015). Besides the detailed description, re3data.org summarises the properties of a repository into a user-friendly icon system helping researchers to easily identify an adequate repository for the storage of their datasets (PAMPEL et al. 2013, and see Fig. 6). More than 5,000 unique visitors are using the registry per month. At present, an average of ten repositories is added weekly to the registry.

The registry offers two search possibilities: free text search through a simple search box and filters for more specific searches (Fig. 7). In the list of results each record includes the name of the repository, the subjects covered, a brief description of the content and a set of icons visualizing key properties of the repository (Fig. 6). A comprehensive view of the descriptive record of the repository can be obtained by clicking on the name of the repository in the search results. It is also possible to simply browse through the list of indexed data repositories. An Application Programming Interface (API) enables machine access to the registry.

Nordicana D1:

DOI : 10.5885/44985SL-8F203FD3ACCD4138

Environmental data from Northern Ellesmere Island in Nunavut, Canada



Liens Polar Data Catalogue (Métadonnées) / Polar Data Catalogue links (Metadata)

[CCIN 79 \(Ensemble / Overview /\)](#) -- [CCIN 9810 \(Lac A / Lake A /\)](#) -- [CCIN 11675 \(Lac C1 / Lake C1 /\)](#)

Links to PDC

Statut / Status

En revision / Under review

Historique des versions / Version history

[Version 1.2](#) (2002-2014) - Mise à jour le 1er octobre 2014 / Updated October 1, 2014
[Version 1.1](#) (2002-2013) - Mise à jour le 12 septembre 2013 / Updated September 12, 2013
[Version 1.0](#) (2002-2012) - Mise à jour le 4 février 2013 / Updated February 4, 2013

Version history

Sites de mesure / Measurement sites

Site	Latitude	Longitude	Altitude (m)	Profondeurs des températures de sol (m) Ground temperature depth range (m)	Mise en service Startup date	Fermeture Closing date
Plus d'info More info Ellesmere Parks Canada (ELLEPAR)	83.094	-74.163	20.0	0.20	1er juillet 2002	1er juillet 2010
Plus d'info More info Ellesmere Ward Hunt (ELLWARH)	83.093	-74.130	5.0		2005 August 12, 2005	

Site overview

Data download: hourly, daily, monthly, yearly

Les fichiers téléchargés en format ZIP contiennent un fichier d'instruction et un fichier de données en format texte (ASCII).
 SVP! Citez toujours les données lorsque vous les utilisez.
 Download ZIP file contains a readme file and a data file in text format (ASCII).
 Please! Always quote citation when using data.

Site	Données / Data	Début / Begin	Fin / End	Données enregistrées / Recorded data	Journalière / Daily	Mensuelle / Monthly	Annuelle / Yearly
Ellesmere Parks Canada (ELLEPAR)	Humidité de l'air moyenne Average air humidity	29 juillet 2002	8 juillet 2010	Obtenir/Get	Obtenir/Get	Obtenir/Get	Obtenir/Get
	Température moyenne de l'air Average air temperature	29 juillet 2002	8 juillet 2010	Obtenir/Get	Obtenir/Get	Obtenir/Get	Obtenir/Get
Ellesmere Parks Canada (ELLEPAR)	Température moyenne du sol Average ground temperature	29 juillet 2002	8 juillet 2010	Obtenir/Get	Obtenir/Get	Obtenir/Get	Obtenir/Get
	Hauteur moyenne de neige au sol Average snow height	29 juillet 2002	8 juillet 2010	Obtenir/Get			

Download section

Fig. 4: Commented elements of a DOI landing page for the online data repository Nordicana D (number D1). It contains climate station data from 2002 to 2014 from three sites in Northern Ellesmere Island in Nunavut, Canada (CEN, 2015). In addition to the required metadata for data discovery (map, citation, abstract, key references, contributors) it provides links to additional metadata in the Polar Data Catalogue, the version history, the site overview table with key information on the time series, and the data download section at the bottom. Each data download is a life query into the database with the possibility to choose different temporal resolutions for data download (hourly: measured data or daily, monthly or annual averages). The dataset is updated on an annual basis. The figure is a compilation of different snapshots and explanations from the Nordicana D website, accessed in June 2015.

Abb. 4: Kommentierte Ausschnitte einer DOI Landing Page des kanadischen Datenrepositoriums Nordicana D (Datensatz D1). Der Datensatz enthält mikro-meteorologische Daten (Lufttemperatur, und -feuchtigkeit, oberflächennahe Bodentemperaturen) von drei Messstationen im nördlichen Ellesmere Island in Nunavut, Kanada (CEN, 2015). Die DOI Landing Page enthält die obligatorischen Metadaten für Data Discovery, wie eine Karte, die Zitationsvorgabe, die Kurzbeschreibung („abstract“), Querverweise zu wissenschaftlichen Artikeln mit weiterführenden Informationen über die hier publizierten Datensätze und die beteiligten Wissenschaftler. Darüber hinaus gibt es Querverweise zu den entsprechenden Metadaten im Polar Data Catalogue und eine Übersichtstabelle mit detail-ierten Informationen zu jeder einzelnen Station. Im unteren Bereich befinden sich die direkten Verknüpfungen („link“) zum Herunterladen der Datensätze. Jeder Datensatz wird in unterschiedlicher zeitlicher Auflösung angeboten: stündliche Werte (wie sie vom Sensor aufgezeichnet wurden), sowie tägliche, monatliche und jährliche Mittelwerte. Jedes Jahr kommt eine weitere Jahreszeitreihe hinzu (und die DOI bekommt eine neue Version). Die Abbildung ist eine Kombination verschiedener Momentaufnahmen der Websites mit zusätzlichen grafischen Elementen und wurde im Juni 2015 erstellt.

RDR indexed by re3data

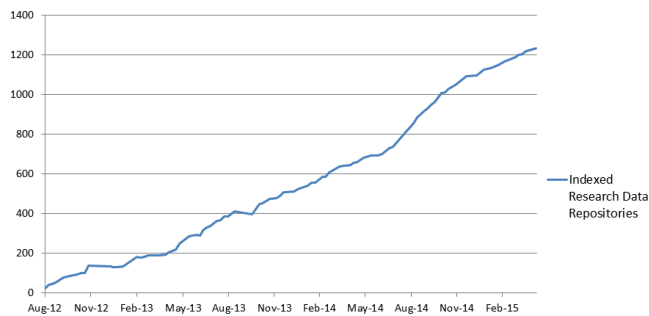


Fig. 5: Number of research data repositories indexed in re3data.org between August 2012 and May 2015.

Abb. 5: Anzahl der in re3data.org gelisteten Forschungsdaten Repositorien zwischen August 2012 und Mai 2015.

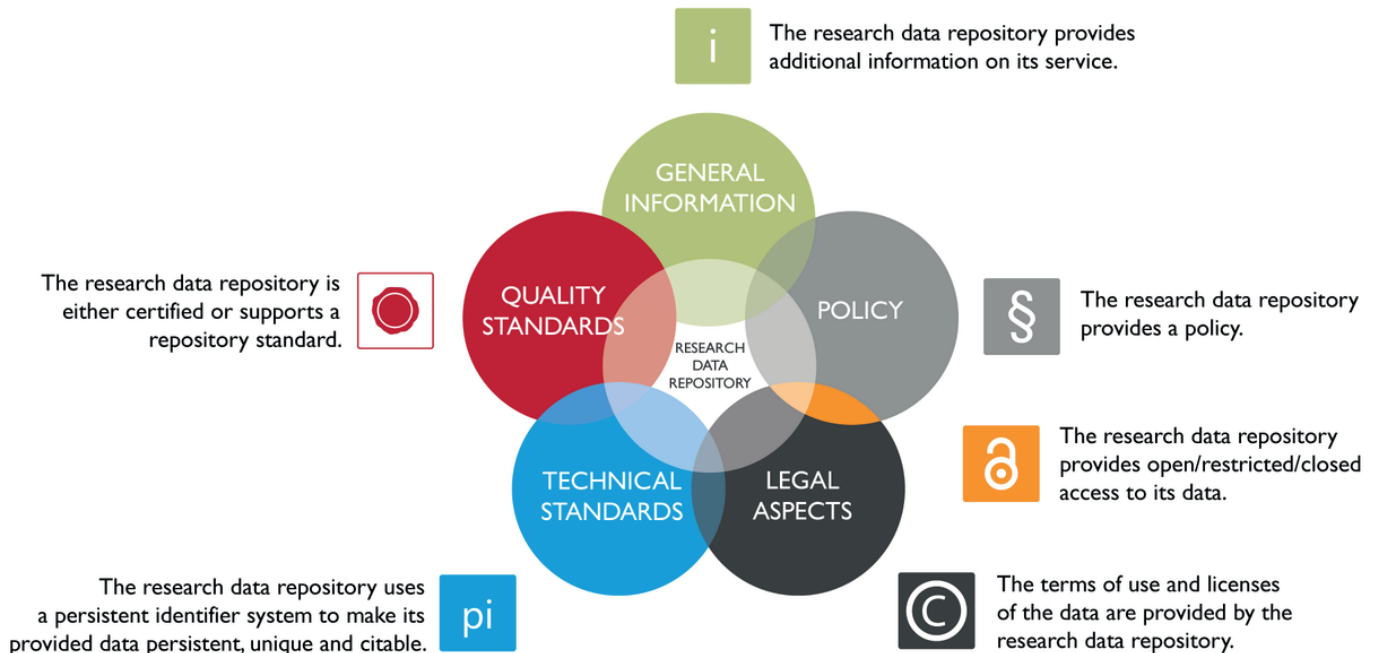


Fig. 6: The icon system of re3data.org provides quick information about the availability of metadata, if the repository provides a data policy, legal aspects for re-use (e.g., licences), the type of persistent identifier provided by the repository, and whether the repository is certified or supports repository standards (see also Fig. 7).

Abb. 6: Die Symbole von re3data.org bieten einen schnellen Überblick über die Daten-Repositorien. Durch die Symbole werden folgende Informationen graphisch angezeigt: Sind die Repositorien beschrieben? Haben sie eine Data Policy auf ihrer Website? Sind die Nachnutzungsregeln geklärt (Lizenzen)? Vergeben sie persistente Identifikatoren für ihre Datensätze und wenn ja welche? Ist ein Repository zertifiziert und wenn ja nach welchem Standard? (vgl. Fig. 7).

The screenshot shows the re3data.org search interface. At the top, there is a navigation bar with links: Home, Search, Browse, Suggest, FAQ, About, Schema, API, Contact, and Imprint. Below this is a search bar with the text 'geosciences' and a 'Search' button. Underneath the search bar are three filter sections: 'Subject' with an 'Add subjects' dropdown, 'Content Type' with an 'Add content types' dropdown, and 'Country (of the responsible institutions)' with an 'Add countries' dropdown and a 'Germany' button. Below these are three checkboxes: 'Certificates' (checked), 'Open Access' (checked), and 'Persistent Identifier' (checked). A 'Reset filter' button is on the right. Below the filters, it says '3 results (filtered) (1 - 3)' and 'Sort by weight'. The first result is 'PANGAEA Publishing Network for Geoscientific and Environmental Data'. It has several icons (i, a, c, doi, r, s) and a list of subjects: Atmospheric Science and Oceanography, Biology, Geochemistry, Mineralogy and Crystallography, Geology and Palaeontology, Geophysics, Geophysics and Geodesy, Geosciences (including Geography), Life Sciences, Natural Sciences, and Oceanography. It also lists content types: Archived data, Audiovisual data, Images, Plain text, and Standard office documents, and the country Germany. A description at the bottom states: 'The information system PANGAEA is operated as an Open Access library aimed at archiving, publishing and distributing georeferenced data from earth system research. The system guarantees long-term availability of its content through a commitment of the operating institutions.'

Fig. 7: Search options and features of re3data.org. In addition to a simple full-text search, several filters may be used. The first overview on the results already provide information on subjects, content types and countries for each repository, as well as the icons visualising key properties for each repository. An extended overview is available after selecting a result.

Abb. 7: Beispiel eines Suchergebnisses auf re3data.org mit Erläuterungen der einzelnen Optionen. Zusätzlich zur Volltextsuche, können verschiedene Filter gesetzt werden (erweiterte Sucheinstellungen). Das Suchergebnis zeigt eine kurze Übersicht über die jeweiligen Repositorien, die bereits die wichtigsten Parameter wie wissenschaftliche Disziplin, Inhaltstypen und das Land, welches das Repositorium betreibt, aber auch die re3data.org Symbole enthält. Bei der Auswahl eines der Suchergebnisse öffnet sich eine Unterseite mit der vollständigen Übersicht über das Repositorium (unten in der Abbildung).

A growing numbers of funders and research institutions recommend the use of re3data.org for the identification of appropriate repositories, e.g., during the composition of a research proposal. For example: Projects, funded under HORIZON 2020 are advised by the European Commission’s “Guidelines on Open Access to Scientific Publications and Research Data in Horizon 2020” (EUROPEAN COMMISSION 2013b) to use re3data.org.

Operators of data repositories can suggest their infrastructures to be listed in re3data.org by filling in an online application form. A repository is indexed by re3data.org when a minimum set of requirements are met. These are, e.g., that the mode of access to the data and repository as well as the terms of use must be clearly explained on the repository websites and the repository must have a focus research data. Before a new record of a research data repository is published in re3data.org, all gathered information is reviewed by a second team member. An editorial board is currently being formed to add a scientific quality check via an online form. The repository has to inform users about the mode of access to the data and state

the terms of use on the repository websites, since these statements are essential pieces of information for researchers either searching for or depositing data.

From 2016 on re3data.org will be included into DataCite. Bringing this service together with DataCite, who mints and manages DOIs for datasets, will yield new opportunities to explore in combining a registry of data repositories with information about persisted datasets to create new value for the research community.

DISCUSSION AND CONCLUSIONS

The increasing relevance and importance of open access to research data for the scientific community, funding agencies and the public together with new technical opportunities in the digital era triggered the development of an increasing number of data repositories, metadata portals and data journals, enabling both, the discovery and persistent and citable

access to research data sets in all scientific fields. Based on examples for Arctic data sets, we have described the functionality and differences of different possibilities to make research data open available via the internet as well as for data journals.

Re3data.org, the Registry of Research Data Repositories offers guidance in the constantly increasing number and very heterogeneous landscape of research data repositories. The structured description of each repository is accomplished by a user-friendly icon system, helping researchers to easily identify an adequate repository for the storage of their datasets. The inclusion of re3data.org into DataCite from January 2016 on emphasizes its relevance and the necessity for this service.

We recommend to always choose a DOI-referenced discipline-specific or institutional data repository to make research datasets publicly available. DOI-referenced datasets are citable, permanent accessible, and accompanied by at least a minimum set of metadata for data discovery, improving the usability of a published data set. Seeing the recent developments towards metadata portals combining metadata of several databases (e.g., NSIDC Data Explorer), we recommend to not only submit the obligatory metadata to the registration agency, data repository or portal, but to make sure that the metadata provides a usable description of a published dataset, e.g., by adding the recommended fields of the DataCite Metadata Scheme (DATA CITE 2014).

For curation purposes and data re-use datasets should preferably be stored in open access data repositories and not be submitted to the journals as supplementary material, as it is recommended by the COPDESS consortium. An open access repository facilitates data re-use even if the corresponding articles were not published in an Open Access Journal.

Moreover, by signing the Statement of Commitment (COP-DESS 2015), many publishers have agreed to list dataset DOIs in reference lists of journal articles. Attaching DOI to research datasets and citing them in reference lists is a major step towards the recognition of the important work of data producers for the scientific community and bring data providers the recognition they deserve (INTERNATIONAL DOI FOUNDATION, 2012).

In addition, because reference lists are always targets for evaluation metrics, it is also possible to assess the number of citations for DOI-referenced datasets, which provides an independent proof for the re-use of a published dataset. Consequently, whenever possible, we recommend following the recommendations of the “Joint Declaration of Data Citation Principles” (DATA CITATION SYNTHESIS GROUP 2014) and always cite DOI-referenced data publications in journal articles and list them in the reference lists, and not, as it was common practice for many years, in the acknowledgments.

ACKNOWLEDGMENTS

We like to thank Birgit Heim from the Alfred Wegener Institute in Potsdam and Warwick Vincent from the University Laval and the Centre d'études nordiques in Québec, Canada for valuable discussions and feedback that greatly improved the manuscript. We also thank two anonymous reviewers for their careful review and constructive comments.

- Asch, G., Tilmann, F., Schurr, B. & Ryberg, T. (2014): MINAS, temporary MINi ArrayS within the frame of IPOC.- Scientific Technical Report - Data 14/03, Potsdam: Deutsches GeoForschungsZentrum GFZ: 1-13, doi: 10.2312/GFZ.b103-14036.
- Berlin Declaration on Open Access to Knowledge in the Sciences and Humanities (2003): Available online: <<http://oa.mpg.de/lang/enuk/berlin-prozess/berliner-erklarung/>> (accessed on 05 July 2015).
- Biskaborn, B.K., Lantuit, H., Dreßler, A., Lanckman, J.P., Jóhannsson, H., Romanovsky, V., Cable, W., Sergeev, D., Vieira, G., Pogliotti, P., Nötzli, J. & Christiansen, H.H. (2015a): Quality assessment of permafrost thermal state and active layer thickness data in GTN-P.- GEOQuébec 2015, Conference Paper, Québec.
- Biskaborn, B.K., Lanckman, J.P., Lantuit, H., Elger, K., Streletskiy, D.A., Cable, W.L. & Romanovsky, V.E. (2015b): The new database of the Global Terrestrial Network for Permafrost (GTN-P).- Earth Syst. Sci. Data 7: 245-259, 2015, doi: 10.5194/essd-7-245-2015.
- BMW et al. (2014): Bundesministerium für Wirtschaft und Energie, Bundesministerium des Innern und Bundesministerium für Verkehr und digitale Infrastruktur (2014): Digitale Agenda 2014-2017.- August 2014. URL: <<http://www.bmwi.de/BMWi/Redaktion/PDF/Publikationen/digitale-agenda-2014-2017.property=pdf,bereich=bmwi2012,sprache=de.rwb=true.pdf>> (accessed on 10 June 2015).
- Burgess, M., Smith, S., Brown, J., Romanovsky, V. & Hinkel, K. (2000): The Global Terrestrial Network for Permafrost (GTNetP): permafrost monitoring contributing to global climate observations.- Current Research 2000 E14, Geol. Survey Canada, Ottawa, Canada, URL: <http://wmsmir.cits.mcan.gc.ca/index.html/pub/geott/ess_pubs/211/211621/cr_2000_e14.pdf> (accessed on 23 December 2015).
- CEN (2014): Environmental data from Northern Ellesmere Island in Nuna-vut, Canada, 1.2 (2002-2014).- Nordicana D1, doi: 10.5885/44985SL-8F203F-D3ACCD4138.
- COPDESS (2015): Coalition on Publishing Data in the Earth and Space Sciences: Statement of Commitment from Earth and Space Science Publishers and Data Facilities.- URL: <<http://www.copdess.org/statement-of-commitment/>> (accessed on 22 December 2015).
- Data Citation Synthesis Group (2014): Joint Declaration of Data Citation Principles.- M. MARTONE (ed) San Diego CA: FORCE11, URL: <<https://www.force11.org/datacitation>> (accessed on 10 June 2015).
- DataCite (2014): DataCite Metadata Schema 3.1 XML Schema.- doi: 10.5438/0011.
- Duguay, C.R., Soliman, A., Hachem, S. & Saunders, W. (2012): Circumpolar and regional Land Surface Temperature (LST), Version 1, with links to geotiff images and NetCDF files (2007-2010).- University Waterloo, Canada, doi: 10.1594/PANGAEA.775962.
- Elger, K., Vincent, W.F. & Barnard, C. (2014): Chapter 11: Knowledge capture and data management (p. 233-254).- In: E. Topp-Jørgensen et al. (eds): INTERACT - Management planning for Arctic and northern alpine research stations – Examples of good practices. DCE – Danish Centre for Environment and Energy, Aarhus University, 1-324, doi: 10.2312/GFZ.LIS.2014.001.
- ESSD News (2015): ESSD indexed by Thomson Reuters as first data journal in Earth sciences.- URL: <http://www.earth-system-science-data.net/about/news_and_press/2015-03-17_essd-indexed-in-the-wos.html> (accessed on 23 June 2015).
- European Commission (2013a): EU Implementation of the G8 Open Data Charter.- URL: <http://ec.europa.eu/information_society/newsroom/cf/dae/document.cfm?doc_id=3489> (accessed on 10 June 2015).
- European Commission (2013b): Guidelines on Open Access to Scientific Publications and Research Data in Horizon 2020. Version 1.0, 11 December 2013.- URL: <http://ec.europa.eu/research/participants/data/ref/h2020/grants_manual/hi/oa_pilot/h2020-hi-oa-pilot-guide_en.pdf> (accessed on 06 May 2015).
- Friddell, J., LeDrew, E. & Vincent, W.F. (2014a): The Polar Data Catalogue: Data Management for Polar and Cryospheric Science.- 70th Eastern Snow Conference, Huntsville, Ontario, Canada, 4-6 June 2013.
- Friddell, J., LeDrew, E. & Vincent, W.F. (2014b): The Polar Data Catalogue: Best practices for sharing and archiving Canada's Polar Data.- Data Sci. J. 13(0): PDA1-PDA7, doi: 10.2481/dsj.IFPDA-01.
- G8 Science Ministers (2012): G8 Science Ministers Statement; June 12, 2013.- URL: <<https://www.gov.uk/government/news/g8-science-ministers-statement>> (accessed on 10 June 2015).
- GERMAN ALLIANCE OF SCIENCE ORGANISATIONS (2010): Schwerpunktinitiative “Digitale Information”: Principles for the Handling of Research Data; June 24, 2010.- URL: <<http://www.allianzinitiative.de/en/core-activities/research-data/principles.html>> (accessed on 10 June 2015).
- Government of Canada (2014): G8 Open Data Charter – Canada's Action Plan.- URL: <<http://open.canada.ca/en/g8-open-data-charter-canadas-action-plan>> (accessed on 10 June 2015).
- Hanson, B., Lehnert, K. & Cutcher-Gershenfeld, J. (2015): Commit-

Acronym	Meaning	Web address
ADAPT	Canadian permafrost research programme Arctic Development and Adaptation to Permafrost in Transition	< http://www.cen.ulaval.ca/adapt/protocols/adapt.php >
AGU	American Geophysical Union	< https://sites.agu.org/ >
Arctic Data Explored	“Arctic Data Explorer” of the National Snow and Ice Data Center NSIDC	< http://nsidc.org/acadis/search/ >
ArcticNet	Network of Centres of Excellence of Canada	< http://www.arcticnet.ulaval.ca/ >
AWI	Alfred Wegener Institute Helmholtz Centre for Polar and Marine Research	< www.awi.de >
BioSharing	A curated, searchable portal of inter-related data standards, databases, and policies in the life, environmental and biomedical sciences	< https://biosharing.org/ >
B2FIND	Discovery service based on metadata steadily harvested from research data collections from EUDAT data centres and other repositories.	< www.eudat.eu/services/b2find >
CALM	Circum Polar Active Layer Monitoring	< http://www.gwu.edu/~calm/ >
CCIN	Canadian Cryospheric Information Network at the University of Waterloo	< https://www.ccin.ca/home/ >
CEN	Centre for Northern Studies, Québec, CA	< http://www.cen.ulaval.ca/en/page.aspx?lien=index >
COPDESS	Coalition on Publishing Data in the Earth and Space Science	< http://www.copdess.org/ >
DataCite	International not-for-profit organization, which aims to improve data citation	< www.datacite.org >
DFG	German Science Foundation / Deutsche Forschungsgemeinschaft	< http://www.dfg.de/en/index.jsp >
DKRZ	Deutsches Klimarechenzentrum	< https://www.dkrz.de/ >
DOI	Digital Object Identifier	< www.doi.org >
EarthCube	Earth Cube	< www.earthcube.org >
EGU	European Geosciences Union	< http://www.egu.eu/ >
ESSD	Earth System Science Data	< http://www.earth-system-science-data.net/ >
EU	European Union	< europa.eu >
EUDAT	Collaborative Pan-European infrastructure providing research data services, training and consultancy	< www.eudat.eu >
FDSN	International Federation of Digital Seismograph Networks	< www.fdsn.org >
FGDC	Federal Geographic Data Committee	< https://www.fgdc.gov/ >
GCOS	Global Climate Observing system	< http://www.wmo.int/pages/prog/gcos/ >
GTOS	Global Terrestrial Observing System	< http://www.fao.org/gtos/ >
GFZ	Helmholtz Centre Potsdam German Research Centre for Geosciences	< http://www.gfz-potsdam.de >
GIPP	Geophysical Instrument Pool Potsdam	< http://www.gfz-potsdam.de/gipp >
GTN-P	Global Terrestrial Network for Permafrost	< http://www.gtnp.org >, < http://gtnpdatabase.org/ >
IASC	International Arctic Science Committee	< http://iasc.info/ >
ICDP	International Scientific Drilling Program	< http://www.icdp-online.org/home/ >
ICSU	International Council for Science	< http://www.icsu-wds.org/ >
IMLS	Institute of Museum and Library Services, US	< https://www.imls.gov/ >
INSPIRE	Infrastructure for Spatial Information in the European Community	< http://inspire.ec.europa.eu >
IPA	International Permafrost Association	< http://ipa.arcticportal.org/ >
ISO19115	ISO Standards Catalog	< http://www.iso.org/iso/home/store/catalogue_ics/catalogue_detail_ics.htm?csnumber=53798 >
KIT	Karlsruhe Institute of Technology	< http://www.kit.edu/ >
MARUM	Centre for Marine Environmental Sciences	< https://www.marum.de/ >
Nature’s Scientific Data	Nature.Com Scientific Data	< http://www.nature.com/sdata/ >
NORDICA D	The Nordica D collection	< http://www.cen.ulaval.ca/nordicanad/en_index.asp >
NORPREM	Norwegian data base for TSP from Norway and Svalbard	< http://www.tspnorway.com/20090301%20NEWS%20NORPERM%20DATABASE%20open.htm >
NSF	National Science Foundation	< http://www.nsf.gov/ >
NSIDC	National Snow and Ice Data Center	< http://nsidc.org/acadis/search/ >
OAL-PMH	Open Archives Initiative Protocol for Metadata Harvesting	
OECD	Organisation for Economic Co-operation and Development	< http://www.oecd.org/ >
OpenAIRE	Network of Open Access repositories, archives and journals that support Open Access policies	< https://www.openaire.eu/ >
PAGE21	Changing Permafrost in the Arctic and its Global Effects in the 21st Cent.	< www.page21.eu >
PANGAEA	Data Publisher for Earth & Environmental Sciences	< www.pangaea.org >
Permafrost Laboratory	Permafrost Laboratory at the University of Alaska Fairbanks	< http://permafrost.gi.alaska.edu/ >
PDC	Polar Data Catalogue: Canadian portal for interdisciplinary Arctic Research	< https://www.polardata.ca/ >
re3data.org	Registry of research data repositories	< www.re3data.org >
TIB	Technical Information Library Hannover	< https://www.tib.eu/en/ >
TSP	Thermal State of Permafrost	< http://ipa.arcticportal.org/activities?catid=0&id=15 >
WDS	ICSU World Data System	< https://www.icsu-wds.org/ >
WMO	World Meteorological Organization	< http://public.wmo.int/en >

Tab. 2: List of all used acronyms and associated websites.

Tab. 2: Liste aller verwendeter Abkürzungen und zugehörigen Webseiten.

- ting to publishing data in the Earth and Space Sciences.- *Eos* 96; doi: 10.1029/2015EO022207.
- Horton, L. (2015): Digital Object Identifiers: Stability for citations and referencing, but not proxies for quality. April 23, 2015, The London School of Economics and Political Science.- URL: <<http://blogs.lse.ac.uk/impactofsocialsciences/2015/04/23/digital-object-identifiers-stability-for-citations>> (accessed on 6 October 2015).
- IASC (2013): Statement of Principles and Practices for Arctic Data Management; April 16, 2013.- URL: <http://www.innovation.ca/sites/default/files/Rome2013/files/IASC%20Statement%20on%20Arctic%20Data%20Management_2013.pdf> (accessed on 6 October 2015).
- International DOI Foundation (2012): DOI handbook, version 5.- URL: <<http://www.doi.org/hb.html>> (accessed on 30 December 2015).
- International Federation of Digital Seismograph Networks FDSN (2014): FDSN recommendations for seismic network DOIs and related FDSN services [WG3 recommendation].- doi: 10.7914/D11596.
- Kratz, J. & Strasser, C. (2014): Data publication consensus and controversies (version 3; referees: 3 approved).- *F1000Research*, 3:94; doi: 10.12688/f1000research.3979.3.
- Klump, J., Bertelmann, R., Brase, J., Diepenbroek, M., Grobe, H., Höck, H. & Wächter, J. (2006): Data publication in the open access initiative.- *Data Sci. Journal* 5: 79-83; doi: 10.2481/dsj.5.79.
- Lorenz, H., Rosberg, J.-E., Juhlin, C., Bjelm, L., Almqvist, B., Berthet, T., Conze, R., Gee, D.G., Klonowska, I., Pascal, C., Pedersen, K., Roberts, N.M.W. & Tsang, C.F. (2015a): COSC-1 operational report - Operational data sets.- GFZ German Research Centre for Geosciences, <<http://dx.doi.org/10.1594/GFZ.SDDB.ICDP.5054.2015>>.
- Lorenz, H., Rosberg, J.-E., Juhlin, C., Bjelm, L., Almqvist, B., Berthet, T., Conze, R., Gee, D.G., Klonowska, I., Pascal, C., Pedersen, K., Roberts, N.M.W. & Tsang, C.F. (2015b): COSC-1 operational report Explanatory remarks on the operational data sets.- Deutsches GeoForschungsZentrum GFZ, doi: 10.2312/ICDP.2015.001.
- National Information Standards Organization (2004): Understanding Metadata.- NISO Press, Bethesda, MD 20814 USA, URL: <<http://www.niso.org/publications/press/UnderstandingMetadata.pdf>>.
- Office of Science and Technology Policy (2013): Increasing Access to the Results of Federally Funded Scientific Research.- URL: <https://www.whitehouse.gov/sites/default/files/microsites/ostp/ostp_public_access_memo_2013.pdf> (accessed on 6 October, 2015).
- Pampel, H. & Dallmeier-Tiessen, S. (2014): Open Research Data: From Vision to Practice. - In: S. BARTLING & S. FRIESIKE (eds): Opening Science. The Evolving Guide on How the Internet is Changing Research.- Collaboration and Scholarly Publishing, Heidelberg, Springer, 213-224, doi: 10.1007/978-3-319-00026-8.
- Pampel, H., Vierkant, P., Scholze, F., Bertelmann, R., Kindling, M., Klump, J., Goebelbecker, H.-J., Gundlach, J., Schirnbacher, P. & Dierolf, U. (2013): Making Research Data Repositories Visible: The re3data.org Registry.- *PLOS ONE* 8, e78080 (2013), doi: 10.1371/journal.pone.0078080.
- Pfeiffenberger, H., & Carlson, D. (2011): *Earth System Science Data* (ESSD) - A Peer Reviewed Journal for Publication of Data.- *D-Lib Magazine* 17(1/2), doi: 10.1045/january2011-pfeiffenberger.
- Rücknagel, J., Vierkant, P., Ulrich, R., Kloska, G., Schnepf, E., Fichtmüller, D., Reuter, E., Semrau, A., Kindling, M., Pampel, H., Witt, M., Fritze, F., van de Sandt, S., Klump, J., Goebelbecker, H.-J., Skarupianski, M., Bertelmann, R., Schirnbacher, P., Scholze, F., Kramer, C., Fuchs, C., Spier, S. & Kirchhoff, A. (2015): Metadata Schema for the Description of Research Data Repositories .Version 3.- doi: 10.2312/re3.008.
- Royal Society London (2012): Science as an open enterprise.- The Royal Society Science Policy Centre report 02/12.- ISBN: 978-0-85403-962-3. URL: <https://royalsociety.org/~media/Royal_Society_Content/policy/projects/sape/2012-06-20-SAOE.pdf> (accessed on 30 December 2015).
- US Government (2014): U.S. Open Data Action Plan; May 9, 2014.- URL: <https://www.whitehouse.gov/sites/default/files/microsites/ostp/us_open_data_action_plan.pdf> (accessed on 6 October 2015).