HGF Alliance: Remote Sensing and Earth System Dynamics

Data Publication

Kirsten Elger & Roland Bertelmann, GFZ German Research Centre for Geosciences, Potsdam
Why are we speaking about research data?

Research data is fundamental for scientific research.

There are increasing expectations by the scientific community as well as by funding agencies and the public to make publicly funded research results and data free and open accessible without any constraints.

What many researchers fear:
- too much work with no benefit
- data publications were deleted from reference lists by journal editors
- they mis-interpret or mis-use my data
- someone will publish my data before me
We examined the citation history of 85 cancer microarray clinical trial publications with respect to the availability of their data. The 48% of trials with publicly available microarray data received 85% of the aggregate citations. Publicly available data was significantly (p = 0.006) associated with a 69% increase in citations, independently of journal impact factor, date of publication, and author country of origin using linear regression.
Recent international steps towards open access of scientific research data

- Berlin Declaration on Open Access to Knowledge in the Sciences and Humanities (2003): „Open access contributions include original scientific research results, raw data and metadata, source materials, digital representations of pictorial and graphical materials and scholarly multimedia materials.”

- German Alliance of Science Organisations (2010): Principles for the Handling of Research Data: „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”

- G8 Science Ministers Statement (June 2013): „to the largest extend and with the fewest constraints possible, publicly funded scientific reserach data should be open [...] while acknowledging the legitimate concerns of private partners.“

- Followed by the EU Implementation of the Open Data Charter and national initiatives like, e.g., the „Digital Agenda“ of the Federal Government of Germany, ....
Principles for the Handling of Research Data

German Alliance of Science Organisations

Partners

- Alexander von Humboldt Foundation
- Deutsche Forschungsgemeinschaft (DFG, German Research Foundation)
- Fraunhofer-Gesellschaft
- German Academic Exchange Service (DAAD)
- German Academy of Sciences Leopoldina
- German Rectors' Conference (Hochschulrektorenkonferenz - HRK)
- Helmholtz Association
- Leibniz Association
- Max Planck Society
- Wissenschaftsrat (German Council of Science and Humanities)
• **Open science**, the unrestricted access to scientific publications and cultural heritage, is an ongoing and future trend in the scientific landscape worldwide. Research publications and other digital objects such as research data and scientific software will thus be publicly available on the internet.

• The Helmholtz Association was one of the initial signatories of the „Berlin Declaration on Open Access to Knowledge in the Sciences and Humanities“ in 2003. This commitment towards open access was then formally approved by its Assembly of Members (assembly of the directors of the Helmholtz Centres): „Publications from the Helmholtz Association shall in future, without exception, be available free of charge, as far as no conflicting agreement with publishers or others exists.“ (Resolution of the Assembly of Members, 27 September 2004).

• An Open Access Policy of the Helmholtz Association is ensuring that publications originating from funded projects will be made freely available to the public on the internet.
....there is no way out: we will have to share our data but... with acknowledgement
Statement of Commitment, Jan 2015
Coalition on Publishing Data in the Earth and Space Sciences

This statement of commitment signals important progress and a continuing commitment by publishers and data facilities **to enable open data in the Earth and space sciences**.

**Scholarly publication is a key high value entry point in making data available, open, discoverable, and usable.** Most publishers have statements related to the inclusion or release of data as part of publication, recognizing that inclusion of the full data enhances the value and is part of the integrity of the research. **Unfortunately, the vast majority of data submitted along with publications are in formats and forms of storage that makes discovery and reuse difficult or impossible.**

Earth and space science data should, to the greatest extent possible, be stored in appropriate domain repositories that are widely recognized and used by the community, follow leading practices, and can provide additional data services. We will work with researchers, funding agencies, libraries, institutions, and other stakeholders to direct data to appropriate repositories, respecting repository policies.

Key commitment of publishers and unions:

- To promulgate metadata information and domain standards, including in the online directory, to help simplify and standardize deposition and reuse.
- To promote referencing of data sets using the Joint Declaration of Data Citation Principles, in which citations of data sets should be included within reference lists.
- To include in research papers concise statements indicating where data reside and clarifying availability.
- To promote and implement links to data sets in publications and corresponding links to journals in data facilities via persistent identifiers. Data sets should ideally be referenced using registered DOI’s.
Data publications – persistent and citable
Data publication with assigned DOI

- **citable**
  DOI have emerged as the leading system for text and data publication (COPDESS 2015).

- **persistent**
  long-term data access guaranteed (by the publisher) despite servers being changed or switched off or people change affiliations and email addresses.

- **with metadata and data description**
  essential for data re-use and discovery, a comprehensive data description should be made a condition for assigning a DOI to a dataset.
What do I need for a data publication

1. Research data (ideally file-based)
2. Metadata for data discovery (authors, title, abstract, location, etc.)
3. Structural/technical metadata for data re-use (data format, processing steps, used instruments or sensors, etc.)

→ Digital object identifier (DOI)
The Collisional Orogeny in the Scandinavian Caledonides (COSC) scientific drilling project focuses on mountain building processes in a major mid-Paleozoic orogen in western Scandinavia and its comparison with modern analogues. The transport and emplacement of subduction-related highgrade continent-ocean transition (COT) complexes into the Baltoscandian platform and their influence on the underlying orogenetic evolution will be studied in a section provided by two fully cored 2.5 km deep drill holes. This operational report concerns the first drill hole, COSC-1 (ICDP 5004.1-A), drilled from early May to late August 2014.

COSC-1 is located in the vicinity of the abandoned drill site, close to the town of Järv, inland, Sweden and was planned to sample a thick section of the Sverige nappe and to penetrate its basal thrust zone into the underlying lower grade metamorphosed belt. Despite substantial technical problems, the drill hole reached 2702.8 m in only 95 days and 100% core recovery was achieved. Surprising was the homogeneity of the Sverige nappe rocks, the unexpected thickness of its basal thrust zone (> 500 m) and that the drill hole, therefore, did not penetrate the bottom of the thrust zone. However, lower grade metamorphosed rocks were encountered in the lowest part of the drill hole together with tens of meters thick mylonites that are, unexpectedly, rich in large garnets.

The drill core was documented on-site and XRF scanned off-site. During various stages of the drilling, the borehole was documented by comprehensive downhole logging. This operational report provides an overview over the COSC-1 operations from drilling preparations to the sampling party and describes the available datasets and sample material.

Keywords:
SOLID EARTH, ROCKS/MINERALS/CRYSTALS, geoscientific information, caledonides, COSC, deep hydrocarbons, dynamics, earth flow, tectonics, ICDP-2011/00, microbiology, norway, orogen, scandes, scandinavias, seismic, seeder, earth science

GCMD Science Keywords:
EARTH SCIENCE > SOLID EARTH > ROCKS/MINERALS/CRYSTALS > METAMORPHIC ROCKS > METAMORPHIC ROCK FORMATION

More Metadata
iso10417: view online / download xml
datacite: view online / download xml
ddref: view online / download xml
w3id: view online / download xml

Related Work
In the reference:

Supplement to

References

Structural Metadata

Metadata of the Data Tables

<table>
<thead>
<tr>
<th>Sites</th>
<th>column Name</th>
<th>Data Type</th>
<th>Description</th>
<th>Validation Text</th>
</tr>
</thead>
<tbody>
<tr>
<td>EXPEDITION</td>
<td>Numeric</td>
<td>expedition number</td>
<td>integer value</td>
<td></td>
</tr>
<tr>
<td>SITE</td>
<td>Numeric</td>
<td>site number</td>
<td>integer value</td>
<td></td>
</tr>
<tr>
<td>NAME</td>
<td>Text</td>
<td>site name or locality</td>
<td>text string of max. 40 characters</td>
<td></td>
</tr>
<tr>
<td>PLATFORM</td>
<td>Text</td>
<td>platform identifier, C=Chikyu, J=Joides, M=Mission Specific, R=Drill Rig</td>
<td>text string of max. 1 character</td>
<td></td>
</tr>
<tr>
<td>LATITUDE_DEG</td>
<td>Integer</td>
<td>decimal degrees of site latitude (latitude of hole 'A')</td>
<td>integer value between 90 and 90</td>
<td></td>
</tr>
<tr>
<td>LATITUDE_MIN</td>
<td>Double</td>
<td>decimal minutes of site latitude (latitude of hole 'A')</td>
<td>real value</td>
<td></td>
</tr>
<tr>
<td>LATITUDE_DIR</td>
<td>Text</td>
<td>direction latitude</td>
<td>text string of max. 1 character</td>
<td></td>
</tr>
<tr>
<td>LONGITUDE_DEG</td>
<td>Integer</td>
<td>decimal degrees of site longitude (longitude of hole 'A')</td>
<td>integer value between -180 and 180</td>
<td></td>
</tr>
<tr>
<td>LONGITUDE_MIN</td>
<td>Double</td>
<td>decimal minutes of site longitude</td>
<td>real value</td>
<td></td>
</tr>
<tr>
<td>LONGITUDE_DIR</td>
<td>Text</td>
<td>direction site longitude</td>
<td>text string of max. 1 character</td>
<td></td>
</tr>
<tr>
<td>DATE_START</td>
<td>Date</td>
<td>date of site start</td>
<td>date in U</td>
<td></td>
</tr>
<tr>
<td>DATE_END</td>
<td>Date</td>
<td>date of site end</td>
<td>date in U</td>
<td></td>
</tr>
</tbody>
</table>

ICDP – data labels

...essential for data re-use
Formats for data publication

1. Data supplement to scientific articles
   - data description in the scientific article

2. Article in a Data Journal
   - data description is a peer-review article

3. Data publication with adjoint data report
   - data description in a data report
1. Widely known: data supplements to scientific articles

- Links to datasets
- Link to original article with data description
Data Supplements -2

• required by many journals → but only open access if the journal is an open access journal

• problem: data supplements have been lost after a journal changed its publisher

We recommend...

• to publish data supplements in open access data repositories

• synchronous to the publication of the scientific article with cross-references between the article and the dataset
2. New development: Data Journals

Peer-review articles with the description of datasets, data collections, data portals, etc.
Data Journals: Example ESSD

- Copernicus Open Access Journal for the publication of datasets and data collections
- Peer-review of articles and adjoined datasets
- **No scientific interpretation** of the data
- Data storage in appropriate domain repositories (and not in ESSD)
- **Indexed in the Web of Science since March 2015** (as first data journal)
3. Data Reports – GFZ examples

Institutional Report Series have long traditions, as important sources of information. Today: online accessible and citable with DOI...

- Flexible format – „enhanced data description“
- More than a README but less than a data article
- Project-specific design if required
- One data report could serve for several datasets
Example: data publication with data report

MINAS, temporary MINi ArrayS within the frame of IPOC

Scientific Technical Report STR14/03 - Data

DOI 1 for Report

DOI 2 for Data

Data report for enhanced data description
Data publication for EDA
PANGAEA data publication: Stettner et al. (2015) Retreat of top cliff of Kurungnakh Island, Lena Delta, Siberia, Russia, 2010-2014, with links to shapefiles, doi:10.1594/PANGAEA.846164

Citation:
Stettner, Samuel; Bartsch, Annett; Widholm, Barbara; Heim, Birgit; Günther, Frank; Morgenstern, Anne; Roth, Achim; Chetverova, Antonina; Fedorova, Irina (2015): Retreat of top cliff of Kurungnakh Island, Lena Delta, Siberia, Russia, 2010-2014, with links to shapefiles. Alfred Wegener Institute - Research Unit Potsdam, doi:10.1594/PANGAEA.846164

Abstract:
Thawing-induced cliff top retreat in permafrost landscapes is mainly due to thermo-erosion. Ground-ice-rich permafrost landscapes are specifically vulnerable to thermo-erosion and may show high degradation rates.

Within the HGF Alliance Remote Sensing and the FP7 PAGE21 permafrost programs we investigated how SAR and optical remote sensing can contribute to the monitoring of erosion rates of ice-rich cliffs in Arctic Siberia (Lena Delta, Russia).

We produced two different vector products:
1) Intra-annual cliff top retreat based on TerraSAR-X (TSX) satellite data (2012-2014):
High-temporal resolution time series of TSX satellite data allow the inter-annual and intra-annual monitoring of the upper cliff-line retreat also under bad weather conditions and continuous cloud coverage. This published SAR product contains the retreating upper cliff lines of a 1.5 km long part of eroding ice-rich coast of Kurungnakh Island in the central Lena Delta. The upper cliff line was mapped using a thresholding approach for images acquired in the years 2012, 2013 and 2014 for the months June (2013, 2014), July (2013, 2014), August (2012, 2013, 2014) and September (2013, 2014). The cliff top retreat vector product is called ‘upper_cliff_TerraSAR-X’. While the 2014 cliff lines show a clear retreat of 2 to 3 m/month, the cliff top lines for 2012 and 2013 are...

Related to:

Coverage:
Latitude: 72.328150 * Longitude: 126.284335

Download dataset as tab-delimited text (use the following character encoding: UTF-8: Unicode (PANGAEA default))

<table>
<thead>
<tr>
<th>Date/Time</th>
<th>Date/time end</th>
<th>Content</th>
<th>URL link</th>
</tr>
</thead>
<tbody>
<tr>
<td>2010-08-05 2014-08-19</td>
<td>4-year cliff top retreat based on optical satellite data (2010-2014)</td>
<td><img src="URL" alt="Link" /></td>
<td></td>
</tr>
<tr>
<td>2012-06-01 2014-09-30</td>
<td>Intra-annual cliff top retreat based on TerraSAR-X (TSX) satellite data (2012-2014)</td>
<td><img src="URL" alt="Link" /></td>
<td></td>
</tr>
</tbody>
</table>
Product Guide: TerraSAR-X derived Cliff Top Retreat

Introduction and product overview

Data and Methods

Processing of Optical Image Data

Processing of SAR Data

Results and References

Data Products

**Biosphere:**
- Forest
- ...

**Geosphere:**
- Earthquakes
- Volcanoes
- Landslides
- Urban subsidence

**Hydrosphere:**
- Soil moisture
- Ocean currents

**Cryosphere:**
- Land ice
- Sea ice
- Permafrost
- Snow

Proposed cross-reference of data publications on the HGF EDA Website
Cryosphere > Permafrost (upper part)

Permafrost is perennial frozen ground for at least two consecutive years. It is acknowledged by the World Meteorological Organisation (WMO) and the United Nations Framework Convention on Climate Change (UNFCCC) as Essential Climate Variable (ECV) and is being monitored for decades, e.g., within the Global Terrestrial Network for Permafrost (GTN-P)...

Remote sensing can monitor these indicators and identify hotspots of surface change. Consequently, it can advise on extension of in-situ monitoring networks and places in-situ measurements into a wider spatial context and supports modelling of subsurface conditions.

Radar remote sensing operates through cloud cover and during polar night allowing high-temporal monitoring of disturbance, of the frozen or non-frozen state of the ground, of frost-heave and ground subsidence, of waterbodies and lake and river ice, as well as the dynamics of surface soil moisture and vegetation.
Cryosphere > Permafrost (lower part)

Available products

**Disturbance:**
Permafrost landscape instability and geohazards express themselves as landslides and (thermo-) erosion (**disturbance**) that can be monitored by satellite remote sensing.

- **Top-cliff erosion:** change detection of the top-cliff erosion line along coastal cliffs in the Lena River delta based on TerraSAR-X data from 2013-2014 ([Link to PANGAEA Stettner et al. 2015](https://doi.org/10.17898/3HA.A6FFF7))

- **Next product when it is available**

**Subsidence:**
Seasonal ground subsidence occurs every summer due to the melting of ice in the upper ground. The magnitude of subsidence can be used to derive the ice content in the upper ground that seasonally thaws in summer (active layer).

- **Next product when it is available**

More product groups to come: **ground frozen/ non-frozen state, surface moisture, lake dynamics, lake ice (as applicable)**
Cryosphere Work Packages

WP C7: Sea ice classification and thickness retrieval

**Responsible:** Prof. Dr. Lars Kaleschke (Universität Hamburg)

WP C8: Freeze/thaw, soil moisture, and terrain disturbances in permafrost regions

**Responsible:** Dr. Birgit Heim (AWI)

**Data products:** *Disturbance: permafrost cliff erosion doi: 10.1594/PANGAEA.846164*

WP C9: Permafrost thaw-season subsidence

**Responsible:** Dr. Julia Boike, Prof. Dr. Wolfgang Hubberten (AWI)
conclusions

• Free and open access to scientific results and research data is required by the scientific community, funding agencies, and the public

• Data publications can be included in reference lists

• Best and safest way: data publication with DOI in appropriate (theme specific) open access data repositories (e.g. permafrost products in PANGAEA, Geosphere Products at GFZ, sea ice in WDS Climate at DKRZ...)

• Cross-reference at the HGF-EDA website

• Next steps, timeline