

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 contraints.

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





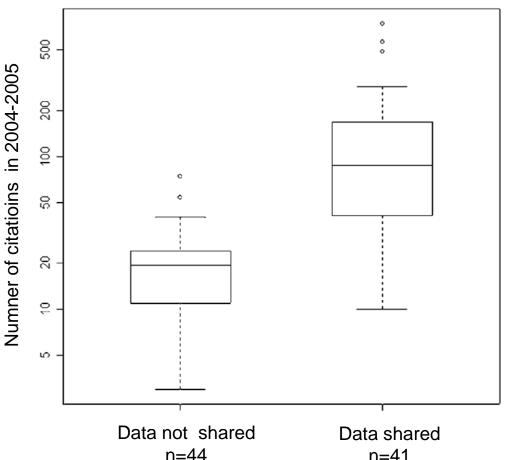


Sharing Detailed Research Data Is Associated with **Increased Citation Rate**

Heather A. Piwowar*, Roger S. Day, Douglas B. Fridsma

Department of Biomedical Informatics, University of Pittsburgh School of Medicine, Pittsburgh, Pennsylvania, United States of America

"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."



doi:10.1371/journal.pone.0000308

n = 41

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









- Alexander von Humboldt Foundation
- Deutsche Forschungsgemeinschaft (DFG, German Research Foundation)
- Fraunhofer-Gesellschaft



- German Academy of Sciences Leopoldina
- German Rectors' Conference (Hochschulrektorenkonferenz HRK)
- Helmholtz Association
- Leibniz Association
- Max Planck Society
- Wissenschaftsrat (German Council of Science and Humanities)









MAX-PLANCK-GESELLSCHAFT









Helmholtz Open Science Policy

- 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.

2013





....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.

Signed by: American Astronomical Society, American Geophysical Union, American Meteorological Society, Biological and Chemical Oceanography Data Management Office, Woods Hole, Oceanographic Institution (BCODMO), Center for Open Science, CLIVAR and Carbon Hydrographic Data Office (CCHDO), Community Inventory of EarthCube Resources for Geosciences Interoperability, (CINERGI), Consortium of Universities for the Advancement of Hydrologic Science, Inc. (CUAHSI), Continental Scientific Drilling Coordination Office (CSDCO), Copernicus Publications, Council of Data Facilities, Elsevier, European Geosciences Union, Geological Data Center of Scripps Institution of Oceanography, Geological Society London, ICSU World Data System, Incorporated Research Institutions for Seismology (IRIS), Integrated Earth Data Applications (IEDA), John Wiley and Sons, Magnetics Information Consortium (MagIC), Mineralogical Society of America, Neotoma Paleoecology Database, National Snow and Ice Data Center, Nature Publishing Group, OpenTopography, Paleontological Society, Proceedings of the National Academy of Sciences, Rolling Deck to Repository (R2R), Science, Springer, UNAVCO

Statement of Commitment, Jan 2015

Coalition on Publishing Data in the Earth and Space Sciences

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)
- 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)





Example for metadata for data discovery



DOI- "LANDING PAGE"

HELMHOLTZ CENTRE POTSOAM
GFZ GERMAN RESEARCH CENTRE
FOR GEOSCIENCES



COSC-1 operational report - Operational data sets



Cite a:

Lorenz, Henning: Rosberg, Jan-Erik; Juhlin, Christopher; Bjelm, Lelf; Almovist, Bjarne; Berthet, Théo; Conze, Ronald; Gee, David G.; Klonowska, Iwona; Pascal Christophe; Pedersen, Karsten; Roberts, Nick; Tsang, Chinfu (2015): COSC-1 operational report - Operational data sets. GFZ German Research Centre for Geosciences. http://dx.doi.org/10.1594/GFZ.SDDB.ICDP.SO54.2015

Data Files

This dataset contains files with restricted (R) a ccess. You may download or apply for access at the following contacts:

- Lorenz, Henning
 Uppsala University, Department of Earth Sciences, Geophysics
 henning.lorenz@geo.uu.se
- COSC Consortium http://cosc.icdp-online.org

Supporting information: Lorenz, H.; Rosberg, J. E.; et al. (2015): COSC-1 operational report Explanator y remarks on the operational data sets. Deutsches GeoForschungsZentrum GFZ doi:10.2312/ICDP.201 5.001

(R) All Data

Sites 2427 Bytes Holes 15133 Bytes Core Runs 85575 Bytes

Core Sections 300426 Bytes Core Boxes 59763 Bytes

Core Overviews 61279327 Bytes (R) Lithological Descriptions

(R) Sample Request (R) Core Samples taken

Mud Samples taken 20781 Bytes (R) Multi Sensor Core Logging

(R) XRF logging Borehole Measurement Campaigns 4966 Bytes Borehole Measurement Runs 12358 Bytes

(R) Borehole Measurement Files
(R) Composite Borehole Log Plots

Drilling Time Breakdown per Day 11110 Bytes
Drilling Time Breakdown of Tasks 102353 Bytes
Drilling Technical Parameter 35538 Bytes
Used Drill Bits 2981 Bytes

License: CC BV 4.0

End of moratorium: /2017-03-01

Abstract

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 onto the Baltoscandian platform and their influence on the underlying allochthons and basement 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 5054-1-A), drilled from early May to late Aurust 2014.

COSC-1 is located in the vicinity of the abandoned Froß mine, close to the town of Åre in Jämtland, Sweden and was planned to sample a thick section of the Seve Nappe and to penetrate its basal thrust zone into the underlying lower grade metamorphosed allochthon. Despite substantial technical problems, the drill hole reached 2502,8 m driller's depth and nearly 100 % core recovery was achieved. Surprising was the homogeneity of the Seve 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 metasedimentary rocks were encountered in the lowermost part of the drill hole together with tens of metres 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, geoscientificInformation, caledonides, COSC, deep hydrosphere, dynamics, europe, heat flow, himalaya, ICDP-2011/03, microbiology, norway, orogen, scandes, scandinavia, selsmic, sweden, earth science

GCMD Science Keywords

EARTH SCIENCE > SOLID EARTH > ROCKS/MINERALS/CRYSTALS > METAMORPHIC ROCKS > METAMORPHIC ROCK FORMATION

More Metadata

iso19115: view inline / download xml datacite: view inline / download xml dif: view inline / download xml escidoc: view inline / download xml

Location

Latitude: 63.4063 Longitude: 13.203057

related work

download

data files

(for data

publidation)

Related Work

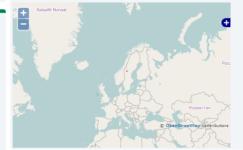
IsReferencedBy

Lorenz, H.; Rosberg, J. E.; et al. (2015): Operational report about phase 1 of the collisional orogeny in the scandinavian caledonides scientific drilling project (COSC-1). Deutsches GeoForschungsZentrum GFZ doi:10.2312/ICDP.2015.002

Supplement to

Lorenz, H.; Rosberg, J.-E.; et al. (2015): COSC-1 – drilling of a subduction-related allochthon in the Palaeozoic Caledonide orogen of Scandinavia. Scientific Drilling doi:10.5194/sd-19-1-2015

References



title citation

description/ abstract

keywords

spatial/ temporal coverage





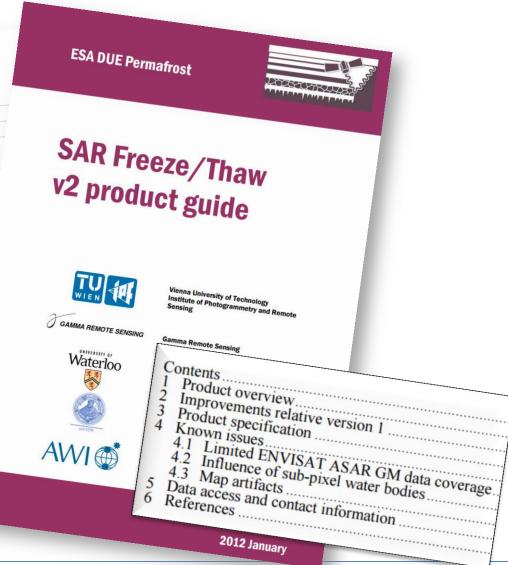
Structural Metadata

Metadata of the Data Tables

Sites

ites				Validation Text
	Column Name	Data Type	Description	integer value
	EXPEDITION	Numeric	expedition number	integer value
-	SITE	Numeric	site number	text string of max. 2
2	SILE	Text	site name or locality	characters
3	NAME	Text	platform identifier, C=Chikyu J=Joides, M=Mission Specific	, lext string
	PLATFORM	Text	R=Drill Rig	character integer value betw
4		Integer	decimal degrees of site latitude (latitude of hole 'A'	
5	LATITUDE_DEG		decimal minutes of site latitude (latitude of hole 'A'	') real value
6	LATITUDE_MIN	Double		text string of macharacter
7	LATITUDE_DIR	Text	direction latitude decimal degrees of site	
,		Integer	longitude (longitude of ho	ole integer value and 180
8	LONGITUDE_DEG		decimal minutes of site	real value
9	LONGITUDE_MI	N Double		text string character
	O LONGITUDE_DI	R Text	direction site longitude	date in U
1	DATE_START	Date	date of site start date of site end	date in l
	DATE_END	Date	date of site enti-	

ICDP - data labels







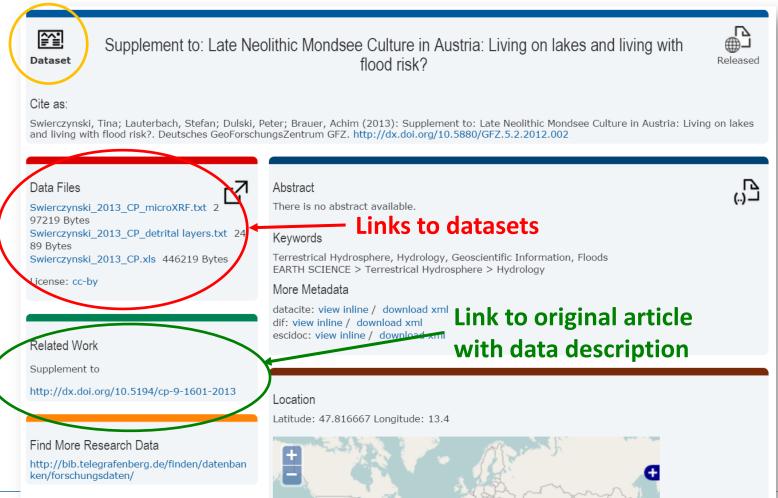
Formats for data publication

- 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







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



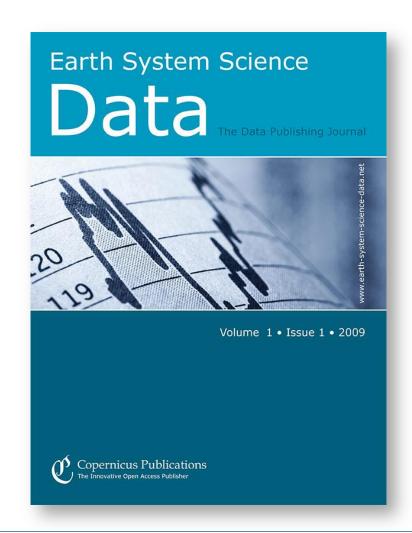


2. New development: Data Journals



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

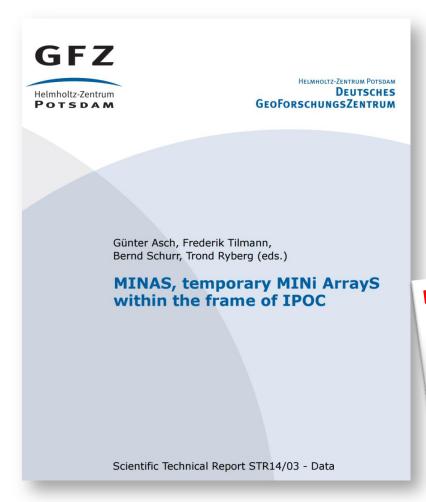
Institutional Report Series have long traditions, as important sources of information. Today: online accessible and citable with DOI...

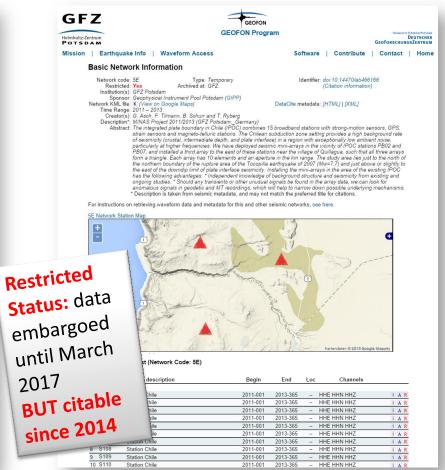
- Flexible format "enhanced data description"
- More than a README but less than a data article
- Project-specific design if required
- One data report could serve for several datasets





Example: data publication with data report





DOI 1 for Report

DOI 2 for Data



→ Data report for enhanced data description

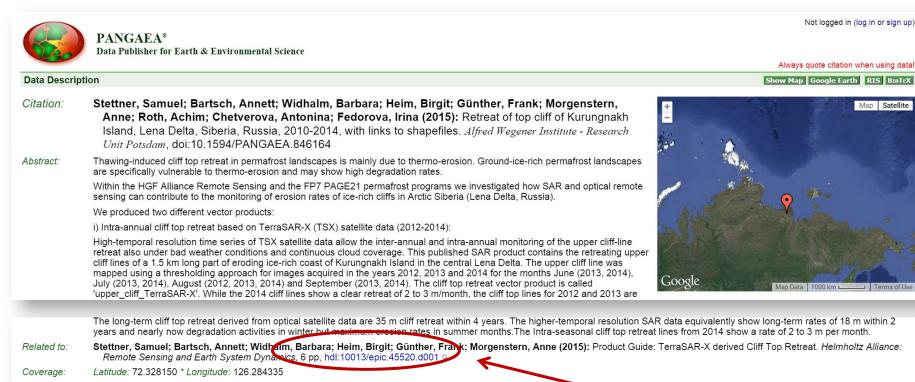


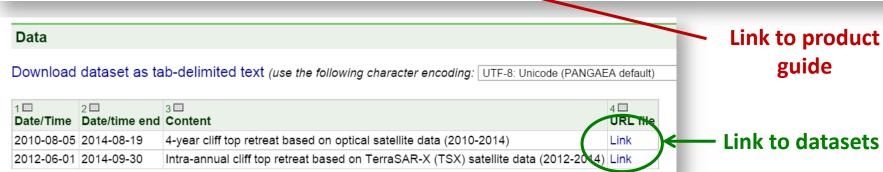
Data publication for EDA





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

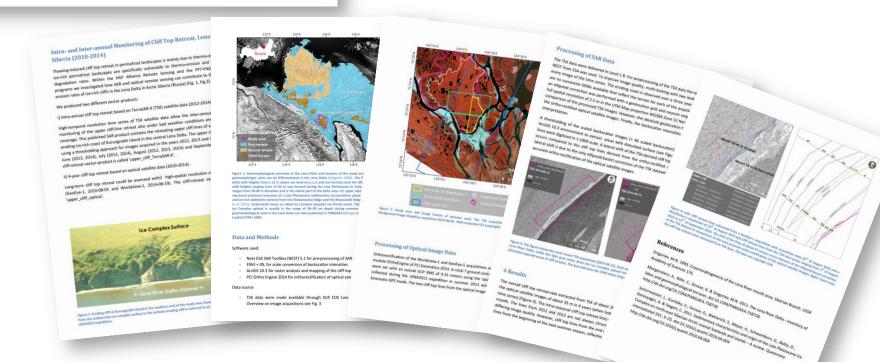


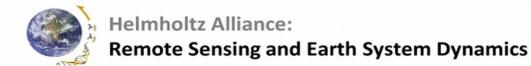


PANGAEA data publication: Product Guide to Stettner et al. (2015) doi:10.1594/PANGAEA.846164

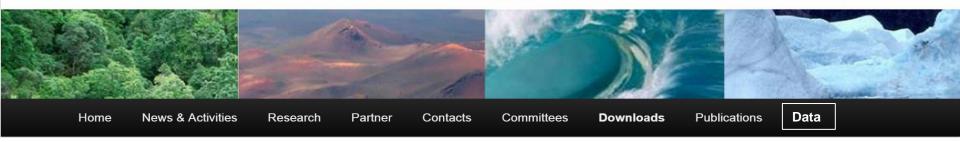


- 6 pages
- Introduction and product overview
- Data and Methods
- Processing of Optical Image Data
- Processing of SAR Data
- Results and References









Data Products



Biosphere:

- Forest
- ...



Hydrosphere:

- Soil moisture
- Ocean currents



Geosphere:

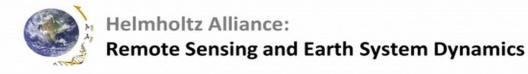
- Earthquakes
- Volcanoes
- Landslides
- Urban subsidence



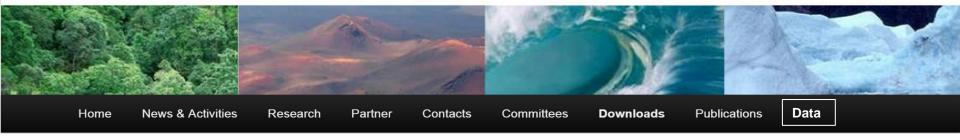
Cryosphere:

- · 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 Meterological 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)...



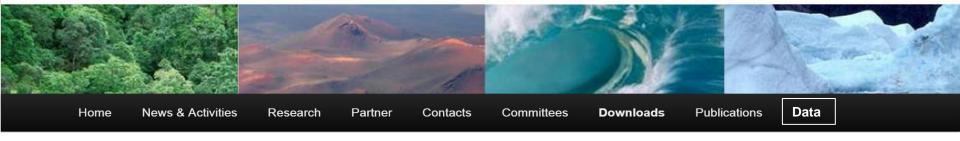
Remote sensing can monitor these indicators and identify hotspots of surface change. Consequently, it can advise on extension of in-situ monitoring networks and places in-situ measurements into a wider spatial context and supports modelling of subsurface conditions.

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 (<u>Link to PANGAEA Stettner et al. 2015</u>)
- Next product when it is available

Subsidence:

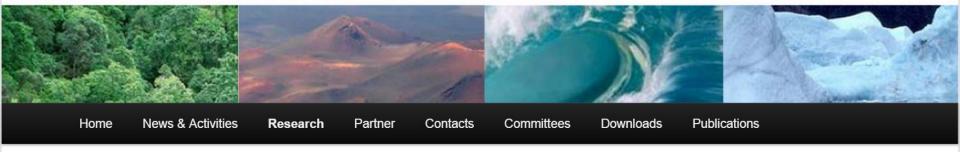
Seasonal ground subsidence occurs every summer due to the melting of ice in the upper ground. The magnitude of subsidence can be used to derive the ice content in the upper ground that seasonally thaws in summer (active layer).

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

Internal EDA Teamsite

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



