Best practice in research data publication

Andreas Hübner
Deutsches GeoForschungsZentrum GFZ

FIDGEO

Fachinformationsdienst
Geowissenschaften der festen Erde

Take home messages

- **Use a domain data repository**
  (don't publish data as supplemental material to a journal article on the publishers website)
- **Whenever possible, use an open licence**
- **Describe your data properly**
- **Provide links to DOIs of related research products**
  (journal articles, data publications, software, samples, …)
- **Cite data publications in your journal articles**
  (use DOI, place citation in the reference list)
- **Publish software along with the source code**

Topics addressed

1. Politics, funders, publisher views
2. FAIR data principles
3. Concerns ….and considerations
4. Repositories
5. Data description and connection to journal article
6. Licenses
7. Open data metrics

1. Politics, funders, publisher views

Politics and funders ask for data sharing

G7 Science and Technology Ministers' Meeting

…promoting increasing access to […] scientific data and publications,…

"…open access is the default setting for research data generated in Horizon 2020."

DFG

…research data should be made available as soon as possible.

G7 Science and Technology Ministers 2016: Tsukuba Communiqué
EU 2016: Guidelines on FAIR Data Management in Horizon 2020
DFG 2015: Leitlinien zum Umgang mit Forschungsdaten

Politics and funders ask for data sharing

Update 2019: Code of Conduct "Guidelines for Safeguarding Good Research Practice"

- describe results clearly and in full
- making the research data, materials and information on which the results are based, as well as the methods and software used, available and fully explaining the work processes.
- Software is made publicly available along with the source code
- Make available in recognised archives and repositories in accordance with the FAIR principles

DFG 2019: **Guidelines for Safeguarding Good Research Practice**
https://www.dfg.de/en/research_funding/principles_dfg_funding/good_scientific_practice/index.html

Publishers ask for data availability

Supporting **data must be made available** to editors and peer reviewers at the time of submission for the purposes of evaluating the manuscript. All manuscripts reporting original research published in Nature journals must include a data availability statement …

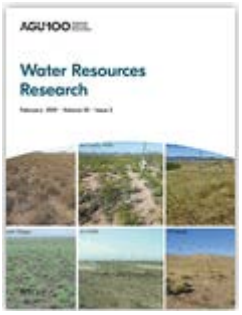
nature
International journal of science

After publication, **all data and materials** necessary to understand, assess, and extend the conclusions of the manuscript **must be available** to any reader of a *Science* Journal.

Science

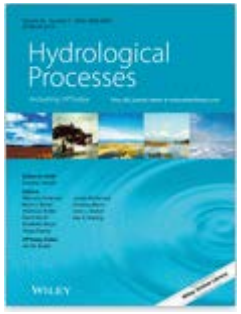
https://www.nature.com/authors/policies/availability.html
http://www.sciencemag.org/authors/science-journals-editorial-policies

Publishers ask for data availability



AGU Data Policy

AGU **encourages authors to identify and archive their data** in approved data centers.



Author Guidelines

If data, scripts, or other artefacts used to generate the analyses presented in the paper **are available** via a publicly available data repository, authors should **include a reference** to the location of the material within their paper.

https://publications.agu.org/author-resource-center/publication-policies/data-policy/
https://onlinelibrary.wiley.com/page/journal/10991085/homepage/forauthors.html#preparingsubmissions

2. FAIR data principles

FAIR data Guiding Principles

Data should be Findable

- globally unique and eternally persistent identifier

Data should be Accessible

- standardized communications protocol

Data should be Interoperable

- machine readable

Data should be Re-usable

- rich and accurate metadata, licence

FORCE11
The Future of Research Communications and e-Scholarship

https://www.force11.org/fairprinciples

https://www.nature.com/articles/sdata201618.pdf

Enabling Findable, Accessible, Interoperable and Reusable Data

in the earth, space, and environmental science

AGU American Geophysical Union · ESIP · RDA Research Data Alliance

Researchers understand and follow expectations related to data management and metadata of the publication

Publishers set standards and follow best practice related to datasets, metadata, accepted repositories and data citation

Scientific repositories are valued for stewardship, data access, improving peer review and digital product quality

Enabling Findable, Accessible, Interoperable and Reusable Data

in the earth, space, and environmental science

AGU American Geophysical Union · ESIP · RDA Research Data Alliance

More details on the project...

Researchers understand and follow expectations related to data management and metadata of the publication

Read what researchers are expected to do...

Publishers set standards and follow best practice related to datasets, metadata, accepted repositories and data citation

Scientific repositories are valued for stewardship, data access, improving peer review and digital product quality

3. Concerns ….and considerations

Concerns...

Lack of time to curate data
Data are only useful if they are understandable, adequate documentation is time-consuming.

Personal investment
Data collection takes time and trouble, other researcher are conceived as freeloaders.

Scooping
Fear of competition and resulting reduced publication opportunities.

Errors
Fear that errors being found in the data.

...and considerations

There are huge benefits to data-sharing, which outweigh the time costs. Shared data can be used for studies that go beyond the original work, with particular benefits when there is pooling of datasets.

And another point, often overlooked, is that uncurated data often become unusable by the original researcher, let alone other scientists, if it is not documented properly and stored on a safe digital site.

Data sharing is also beneficial for reproducibility: researchers will check data more carefully when it is to be shared.

From: Mark J. Costello, Motivating Online Publication of Data https://doi.org/10.1525/bio.2009.59.5.9

Concerns...

Lack of time to curate data
Data are only useful if they are understandable, adequate documentation is time-consuming.

Personal investment
Data collection takes time and trouble, other researcher are conceived as freeloaders.

Scooping
Fear of competition and resulting reduced publication opportunities.

Errors
Fear that errors being found in the data.

...and considerations

Why should others benefit when you have done all the hard work? In fact, when we consider that scientists are usually receiving public money to make scientific discoveries, this line of argument does not appear morally defensible.

But in any case, it is not true that the scientists who do the sharing have no benefits. For a start, they will see an increase in citations, as others use their data.

From: Mark J. Costello, Motivating Online Publication of Data https://doi.org/10.1525/bio.2009.59.5.9

Concerns…

Lack of time to curate data
Data are only useful if they are understandable, adequate documentation is time-consuming.

Personal investment
Data collection takes time and trouble, other researcher are conceived as freeloaders.

Scooping
Fear of competition and resulting reduced publication opportunities.

Errors
Fear that errors being found in the data.

…and considerations

Concerns about being scooped are frequently cited, but are seldom justified.

Indeed, if we move to a situation where a dataset is a publication with its own identifier, then the original researcher will get credit every time someone else uses the dataset.

Data can be embargoed for a certain time, or only part of the data is published: at least the part that was used to support the conclusion that were made in a journal article.

From: Mark J. Costello, Motivating Online Publication of Data https://doi.org/10.1525/bio.2009.59.5.9

Concerns...

Lack of time to curate data
Data are only useful if they are understandable, adequate documentation is time-consuming.

Personal investment
Data collection takes time and trouble, other researcher are conceived as freeloaders.

Scooping
Fear of competition and resulting reduced publication opportunities.

Errors
Fear that errors being found in the data.

...and considerations

Of course there will be errors – there always are. We have to change our culture so that we do not regard it as a source of shame to publish data in which there are errors, but rather as an inevitability that is best dealt with by making the data public so the errors can be tracked down.

From: Mark J. Costello, Motivating Online Publication of Data https://doi.org/10.1525/bio.2009.59.5.9

4. Repositories

Repositories

Repository = (online accessible) database for the recording and publication of research data, texts and other digital objects[1]

re3data.org
REGISTRY OF RESEARCH DATA REPOSITORIES

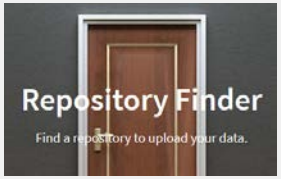
A **demonstration of searching** for research data repositories using the Re3data directory

FOSTER

https://www.fosteropenscience.eu/content/re3data-demo

Find data repositories in the **earth and space sciences domain** that

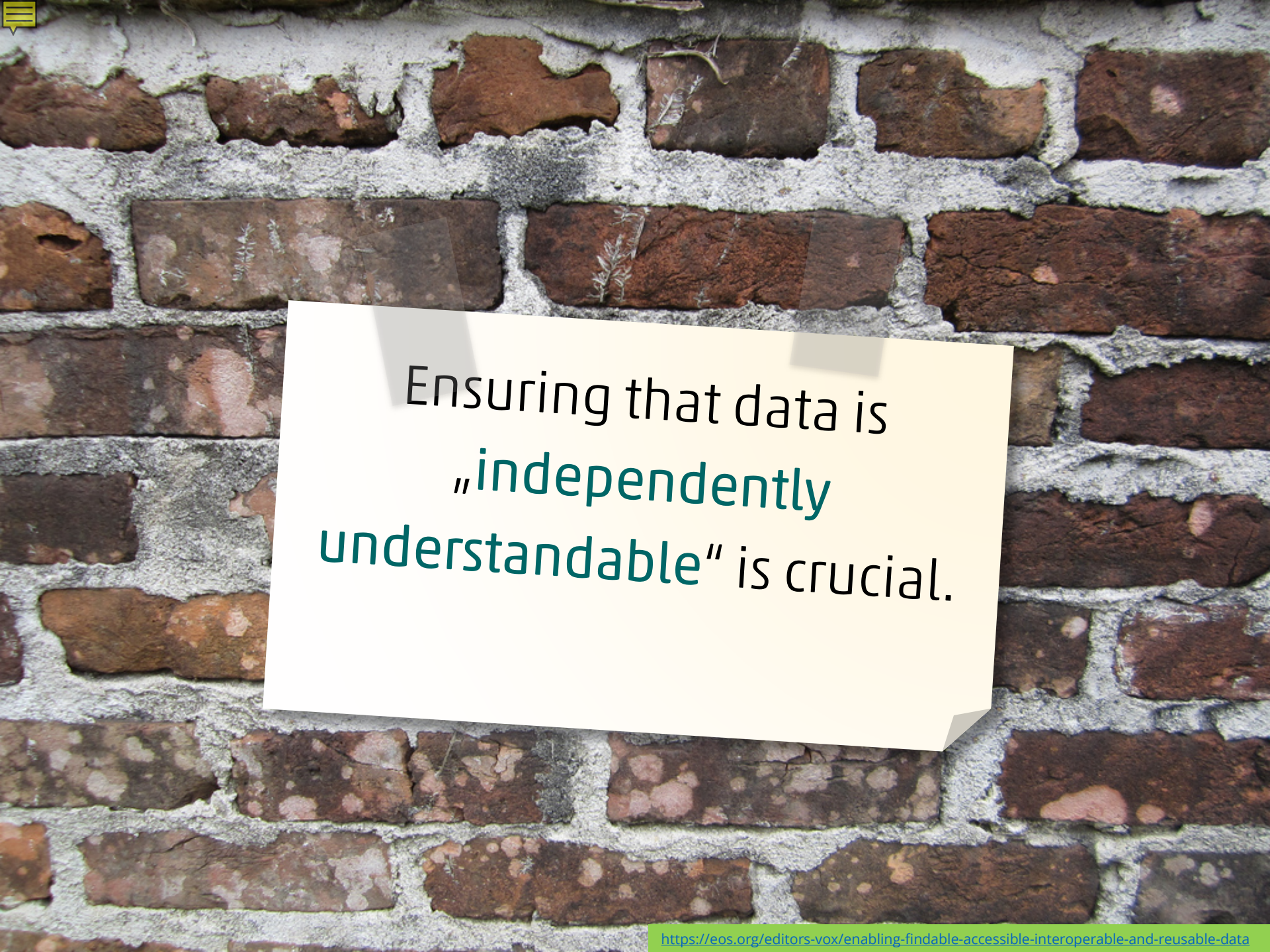
- support open access
- provide persistent identifiers
- accept data for deposit



https://repositoryfinder.datacite.org/

[1] Einstieg ins Forschungsdatenmanagement in den Geowissenschaften, https://doi.org/10.2312/lis.14.01

5. Data description and connection to journal article



https://eos.org/editors-vox/enabling-findable-accessible-interoperable-and-reusable-data

Description of data

Peer-reviewed articles with the **description** of datasets, data collections, data infrastructures, etc.

No data interpretation!

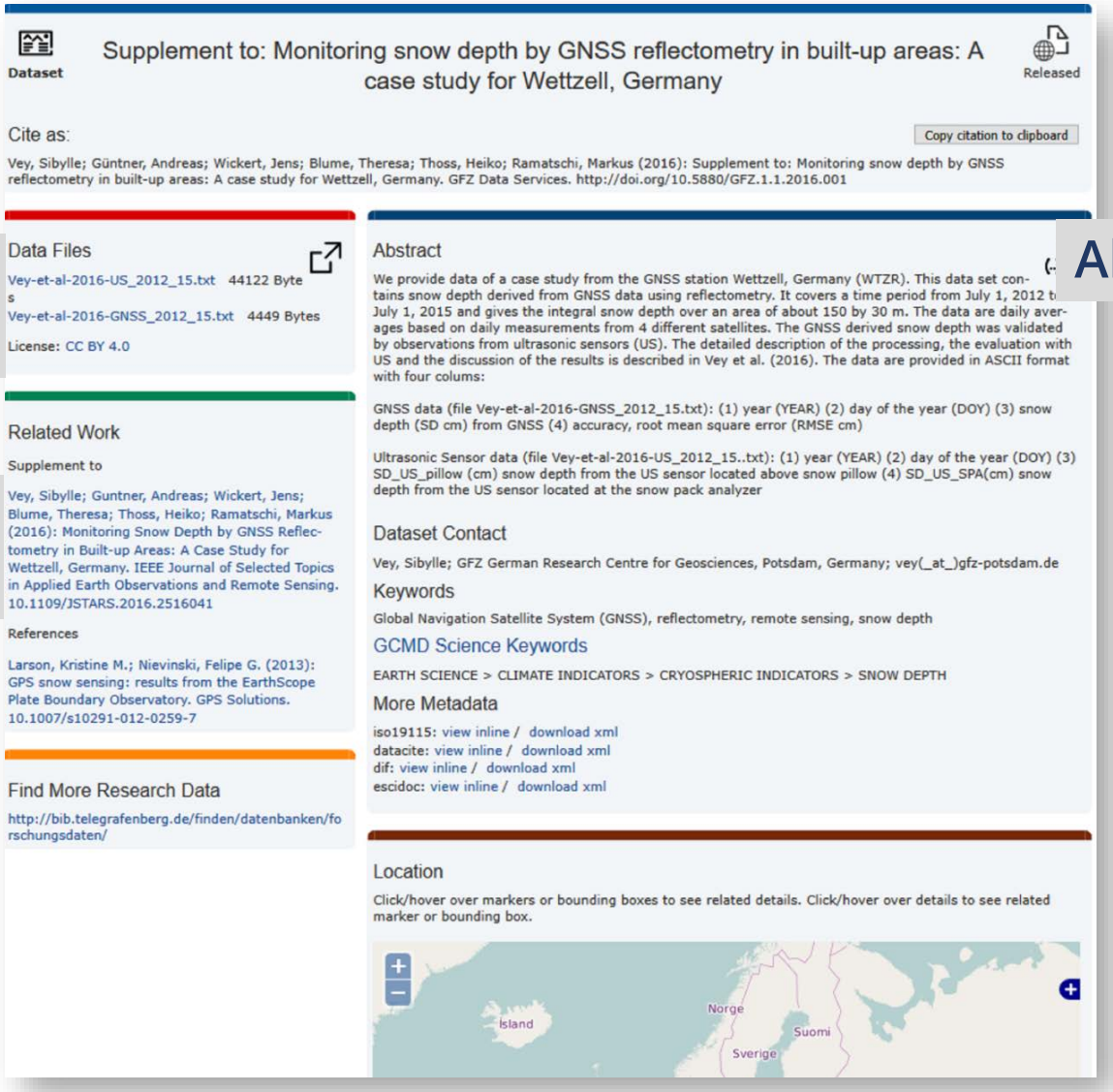
Folie von Kirsten Elger

Data publication example

Links to data files + Licence

Link to journal article

Abstract



Dataset

Supplement to: Monitoring snow depth by GNSS reflectometry in built-up areas: A case study for Wettzell, Germany

Released

Cite as:

Copy citation to clipboard

Vey, Sibylle; Güntner, Andreas; Wickert, Jens; Blume, Theresa; Thoss, Heiko; Ramatschi, Markus (2016): Supplement to: Monitoring snow depth by GNSS reflectometry in built-up areas: A case study for Wettzell, Germany. GFZ Data Services. http://doi.org/10.5880/GFZ.1.1.2016.001

Data Files

Vey-et-al-2016-US_2012_15.txt 44122 Bytes

Vey-et-al-2016-GNSS_2012_15.txt 4449 Bytes

License: CC BY 4.0

Related Work

Supplement to

Vey, Sibylle; Guntner, Andreas; Wickert, Jens; Blume, Theresa; Thoss, Heiko; Ramatschi, Markus (2016): Monitoring Snow Depth by GNSS Reflectometry in Built-up Areas: A Case Study for Wettzell, Germany. IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing. 10.1109/JSTARS.2016.2516041

References

Larson, Kristine M.; Nievinski, Felipe G. (2013): GPS snow sensing: results from the EarthScope Plate Boundary Observatory. GPS Solutions. 10.1007/s10291-012-0259-7

Find More Research Data

http://bib.telegrafenberg.de/finden/datenbanken/forschungsdaten/

Abstract

We provide data of a case study from the GNSS station Wettzell, Germany (WTZR). This data set contains snow depth derived from GNSS data using reflectometry. It covers a time period from July 1, 2012 t July 1, 2015 and gives the integral snow depth over an area of about 150 by 30 m. The data are daily averages based on daily measurements from 4 different satellites. The GNSS derived snow depth was validated by observations from ultrasonic sensors (US). The detailed description of the processing, the evaluation with US and the discussion of the results is described in Vey et al. (2016). The data are provided in ASCII format with four colums:

GNSS data (file Vey-et-al-2016-GNSS_2012_15.txt): (1) year (YEAR) (2) day of the year (DOY) (3) snow depth (SD cm) from GNSS (4) accuracy, root mean square error (RMSE cm)

Ultrasonic Sensor data (file Vey-et-al-2016-US_2012_15..txt): (1) year (YEAR) (2) day of the year (DOY) (3) SD_US_pillow (cm) snow depth from the US sensor located above snow pillow (4) SD_US_SPA(cm) snow depth from the US sensor located at the snow pack analyzer

Dataset Contact

Vey, Sibylle; GFZ German Research Centre for Geosciences, Potsdam, Germany; vey(_at_)gfz-potsdam.de

Keywords

Global Navigation Satellite System (GNSS), reflectometry, remote sensing, snow depth

GCMD Science Keywords

EARTH SCIENCE > CLIMATE INDICATORS > CRYOSPHERIC INDICATORS > SNOW DEPTH

More Metadata

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

Location

Click/hover over markers or bounding boxes to see related details. Click/hover over details to see related marker or bounding box.

Island

Norge

Suomi

Sverige

Contents lists available at ScienceDirect

Tectonophysics

journal homepage: www.elsevier.com/locate/tecto

Properties of granular analogue model materials: A community wide survey

M. Klinkmüller [a], G. Schreurs [a,1], M. Rosenau [b], H. Kemnitz [b]

[a] Institute of Geological Sciences, University of Bern, Baltzerstrasse 1 +3, CH-3012 Bern, Switzerland
[b] Helmholtz-Zentrum Potsdam, GFZ Deutsches GeoForschungsZentrum, Telegrafenberg, D-14473 Potsdam, Germany

sented as grain size distribution curves, in which particle grain size is plotted against cumulative weight percentage (Fig. 2).

The original sieve data have been published open access and are available in Klinkmüller et al. (2016b).

1. Citation in the text

Contents lists available at ScienceDirect

Tectonophysics

journal homepage: www.elsevier.com/locate/tecto

Properties of granular analogue model materials: A community wide survey

M. Klinkmüller [a], G. Schreurs [a,1], M. Rosenau [b], H. Kemnitz [b]

[a] Institute of Geological Sciences, University of Bern, Baltzerstrasse 1 +3, CH-3012 Bern, Switzerland
[b] Helmholtz-Zentrum Potsdam, GFZ Deutsches GeoForschungsZentrum, Telegrafenberg, D-14473 Potsdam, Germany

sented as grain size distribution curves, in which particle grain size is plotted against cumulative weight percentage (Fig. 2).

The original sieve data have been published open access and are available in Klinkmüller et al. (2016b).

1. Citation in the text

References

Heilbronner, R., Keulen, N., 2006. Grain size and grain shape analysis of fault rocks. Tectonophysics 427, 199–216.

Hubbert, M.K., 1951. Mechanical basis for certain familiar geologic structures. Geol. Soc. Am. Bull. 62, 1259–1273.

Klinkmüller, M., Schreurs, G., Rosenau, M., 2016a. GeoMod2008 materials benchmark: The ring shear test data set. GFZ Data Services. http://dx.doi.org/10.5880/GFZ.4.1.2016.002.

Klinkmüller, M., Schreurs, G., Rosenau, M., 2016b. GeoMod2008 materials benchmark: The sieve data set. GFZ Data Services. http://dx.doi.org/10.5880/GFZ.4.1.2016.003.

Klinkmüller, M., Kemnitz, H., Schreurs, G., Rosenau, M., 2016c. GeoMod2008 materials benchmark: The SEM image data set. GFZ Data Services. http://dx.doi.org/10.5880/GFZ.4.1.2016.004.

2. Dataset-DOI in the References

Folie von Kirsten Elger

Contents lists available at ScienceDirect

Tectonophysics

journal homepage: www.elsevier.com/locate/tecto

Properties of granular analogue model materials: A community wide survey

M. Klinkmüller [a], G. Schreurs [a,1], M. Rosenau [b], H. Kemnitz [b]

[a] Institute of Geological Sciences, University of Bern, Baltzerstrasse 1 +3, CH-3012 Bern, Switzerland
[b] Helmholtz-Zentrum Potsdam, GFZ Deutsches GeoForschungsZentrum, Telegrafenberg, D-14473 Potsdam, German

sented as grain size distribution curves, in which particle grai
plotted against cumulative weight percentage (Fig. 2).

The original sieve data have been published open access
available in Klinkmüller et al. (2016b).

References

Heilbronner, R., Keulen, N., 2006. Grain size and grain shape an
Tectonophysics 427, 199–216.
Hubbert, M.K., 1951. Mechanical basis for certain familiar geologic
Am. Bull. 62, 1259–1273.
Klinkmüller, M., Schreurs, G., Rosenau, M., 2016a. GeoMod2008 m
The ring shear test data set. GFZ Data Services. http://dx.doi.org/10.5880/GFZ.4.1.
2016.002.
Klinkmüller, M., Schreurs, G., Rosenau, M., 2016b. GeoMod2008 materials benchmark:
The sieve data set. GFZ Data Services. http://dx.doi.org/10.5880/GFZ.4.1.2016.003.
Klinkmüller, M., Kemnitz, H., Schreurs, G., Rosenau, M., 2016c. GeoMod2008 materials
benchmark: The SEM image data set. GFZ Data Services. http://dx.doi.org/10.5880/
GFZ.4.1.2016.004.

3. Data access via DOI

GFZ Helmholtz Centre Potsdam — GFZ German Research Centre for Geosciences — Impressum

Dataset — Released

GeoMod2008 materials benchmark: The sieve dataset

Cite as: Copy citation to clipboard

Klinkmüller, Matthias; Schreurs, Guido; Rosenau, Matthias (2016): GeoMod2008 materials benchmark: The sieve dataset. GFZ Data Services. http://doi.org/10.5880/GFZ.4.1.2016.003

Data Files

SieveDataOverview.pdf 218020 Bytes
Sieve-data.zip 735235 Bytes
Explanations for the Sieve dataset.pdf 536692 Bytes

License: CC BY 4.0

Related Work

Supplement to

Klinkmüller, M., Schreurs, G., Rosenau, M., & Kemnitz, H. (2016). Properties of granular analogue model materials: A community wide survey. Tectonophysics. doi:10.1016/j.tecto.2016.01.017

Find More Research Data

http://bib.telegrafenberg.de/finden/datenbanken/forschungsdaten/

Abstract

This dataset provides sieve data (grain size distributions) on natural and artificial granular materials used for experimental simulation by the analogue geodynamic modelling community (21 sands and glass beads). The material samples have been collected community-wide and analysed at GFZ Potsdam in the framework of the GeoMod2008 conference benchmark initiative. The context of data collection, details of the material samples and measuring techniques as well as interpretation and discussion of results can be found in Klinkmüller et al. (2016) to which this dataset is supplement material.

An overview of all files of the data set is given in the table SieveDataOverview.

Methods

The data presented here are derived by sieving using a RETSCH Vibratory Sieve Shaker AS 200 basic at GFZ Potsdam's analogue laboratory for tectonic modelling. Mesh sizes used were 630, 400, 355, 224, 125, and 63 micrometer. 1 kg of each sample material has been sieved for 4 hours at maximum Amplitude (3 mm). Laboratory conditions were air conditioned during all the measurements (Temperature: 23°C, Humidity: 45%).

The resulting sieve analysis data are presented as fractions of 1 kg.

Dataset Contact

Rosenau, Matthias; GFZ German Research Centre for Geosciences, Potsdam, Germany; rosen(_at_)gfz-potsdam.de; http://www.gfz-potsdam.de/en/section/lithosphere-dynamics/infrastructure/geodynamics/tectonic-modeling-lab/

Keywords

analogue materials, granular materials, bulk solids, analog models, sandbox, benchmark, Geomod, EPOS, experiment, properties of materials, geological process, materials science

GCMD Science Keywords

EARTH SCIENCE SERVICES > MODELS > PHYSICAL/LABORATORY MODELS
EARTH SCIENCE > SOLID EARTH > TECTONICS

2. Dataset DOI in the References

Folie von Kirsten Elger

Contents lists available at ScienceDirect

Tectonophysics

journal homepage: www.elsevier.com/locate/tecto

Properties of granular analogue model materials: A community wide survey

M. Klinkmüller [a], G. Schreurs [a,1], M. Rosenau [b], H. Kemnitz [b]

[a] Institute of Geological Sciences, University of Bern, Baltzerstrasse 1 +3, CH-3012 Bern, Switzerland
[b] Helmholtz-Zentrum Potsdam, GFZ Deutsches GeoForschungsZentrum, Telegrafenberg, D-14473 Potsdam, German

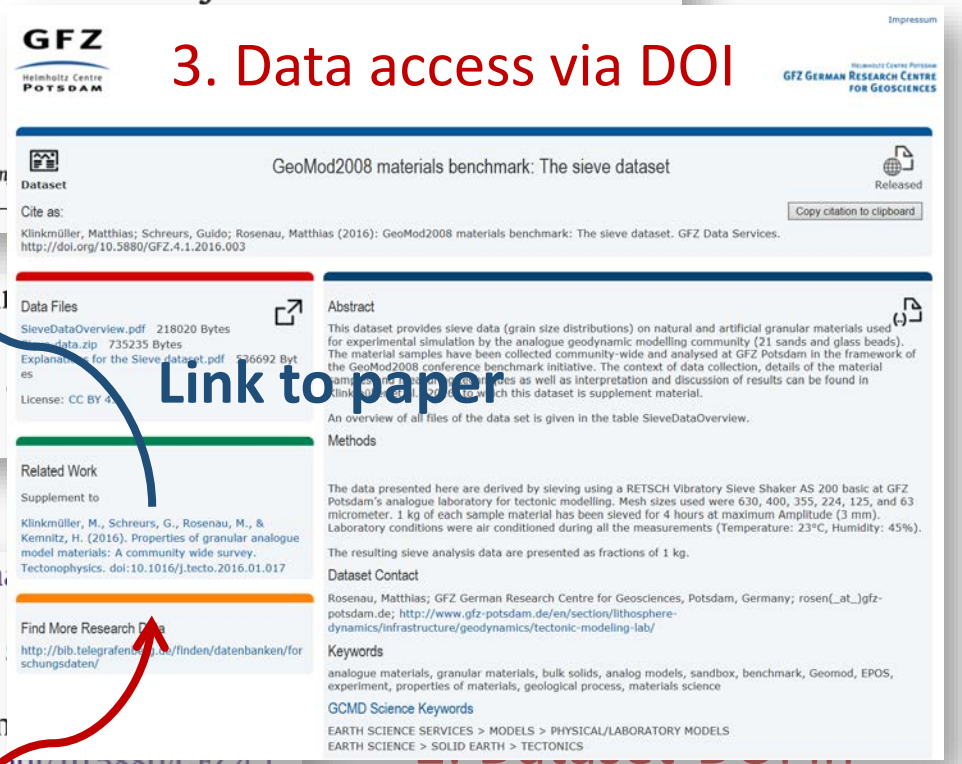
sented as grain size distribution curves, in which particle grai
plotted against cumulative weight percentage (Fig. 2).

The original sieve data have been published open access
available in Klinkmüller et al. (2016b).

References

Heilbronner, R., Keulen, N., 2006. Grain size and grain shape an
Tectonophysics 427, 199–216.
Hubbert, M.K., 1951. Mechanical basis for certain familiar geologic
Am. Bull. 62, 1259–1273.
Klinkmüller, M., Schreurs, G., Rosenau, M., 2016a. GeoMod2008 m
The ring shear test data set. GFZ Data Services. http://dx.doi.org/10.5880/GFZ.4.1.2016.002.
Klinkmüller, M., Schreurs, G., Rosenau, M., 2016b. GeoMod2008 materials benchmark: The sieve data set. GFZ Data Services. http://dx.doi.org/10.5880/GFZ.4.1.2016.003.
Klinkmüller, M., Kemnitz, H., Schreurs, G., Rosenau, M., 2016c. GeoMod2008 materials benchmark: The SEM image data set. GFZ Data Services. http://dx.doi.org/10.5880/GFZ.4.1.2016.004.

3. Data access via DOI



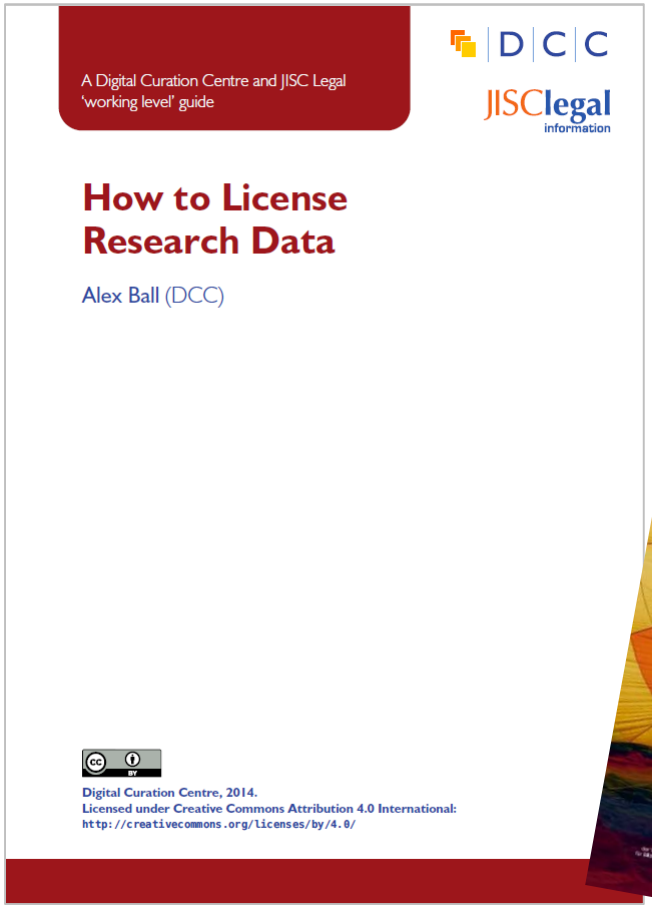
Link to paper

the References

Folie von Kirsten Elger

6. Licenses

Licence your data



https://irights.info/wp-content/uploads/2014/11/Open_Content_A_Practical_Guide_to_Using_Open_Content_Licences_web.pdf

http://www.dcc.ac.uk/sites/default/files/documents/publications/reports/guides/How_To_License_Research_Data.pdf

Licences for software

The use of licenses that are as open and standardized as possible (see e.g. listing of various licenses of the Free Software Foundation or recommendations of the Software Sustainability Institute) can enable both the use of the software and the subsequent use or further development of the program code.

For example :

GNU General Public License

MIT License

Von Victor Siame <vcopovi@wanadoo.fr> - http://www.gnu.org/graphics/gnusvgart.html (http://www.gnu.org/graphics/official%20gnu.svg), FAL, https://commons.wikimedia.org/w/index.php?curid=3653319

7. Open data metrics

The future: open data metrics

Similar to text publications in journals and books, it is useful to know how often published data sets are cited and used.

The initiative Make Data Count is aimed at building social & technical infrastructure for open data metrics. Check out their ebook.