

Sharing and Archiving Research Data

Markus Stocker

September 12, 2017

Outline

- Open Data imperative
- How to share data
- Data repositories
- Credit for publishing data

Open Data

- Research data are assets
- Increasingly recognized and valued research output
- Valuable to others, in research as well as industry
- Critical for reproducibility
- Mandates to open access publicly funded data
- Anyone free to use, reuse, and redistribute
- Possibly subject to attribution

Why Sharing Data

- Get credit for your work
- Share it with *your future self*
- Enable reproducibility and repurposing of your data
- Increase visibility of your work
- Access to data (and software) seen positively in peer-review
- Because you are required to do so
- Because “research data is not mine”

<https://www.linkedin.com/pulse/repeat-me-research-data-mine-wainer-lusoli>

How to Share Data

- It is challenging to share data well
- Making data usable by others in other contexts is difficult
- Study and follow existing principles and guidelines
- Share your data using a high-quality repository

FAIR Principles

- Findable
 - ▶ Persistently identified, described, indexed
- Accessible
 - ▶ Retrievable by identifier, open protocol, accessible metadata
- Interoperable
 - ▶ Represented using formal, accessible, shared FAIR vocabulary
- Re-usable
 - ▶ Meet community standards, associate with provenance, usage license

<https://www.force11.org/group/fairgroup/fairprinciples>

Five Star Data

- * Publish your data on the web, open license, any format
- ** Publish your data as structured data
- *** Publish your data in a non-proprietary open format
- **** Use identifiers to refer to your data, e.g. units, parameters
- ***** Link your data to other data, to provide context

<http://5stardata.info/en/>

Data Identification

- Persistent identification of data is important
- Idea is similar to identification of articles
- Provides an unambiguous reference to data
- Used to link or refer to data, e.g. in text
- Data are increasingly identified by DOI
 - ▶ e.g. 10.1594/PANGAEA.787617
- DataCite is the global provider of DOIs for research data

Data Citation

- Always cite data used in research
- Citing data is often very easy
- Similar to citing literature
- Data repositories often support citation export
- Exports in standard formats, supported by tools




PANGAEA.


Data Publisher for Earth & Environmental Science



Citation:

Nürnberg, Dirk; Müller, A; Schneider, Ralph R (2000): Paleo-sea surface temperature calculations in the equatorial east Atlantic. *PANGAEA*,  <https://doi.org/10.1594/PANGAEA.787617>,

Supplement to: Nürnberg, D et al. (2000): Paleo-sea surface temperature calculations in the equatorial east Atlantik from Mg/Ca ratios in planktic foraminifera: A comparison to sea surface temperature estimates from U37K', oxygen isotopes, and foraminiferal transfer function. *Paleoceanography*, **15(1)**, 124-134,  <https://doi.org/10.1029/1999PA000370>

 **Always quote above citation when using data!** You can download the citation in several formats below.

[RIS Citation](#)

[BisTeX Citation](#)

[Text Citation](#)

[Facebook](#)

[Twitter](#)

[Google+](#)

[Show Map](#)

[Google Earth](#)

Data Repositories

- re3data.org knows of 1920 research data repositories
- Covers major subjects and many special domains
- Quality varies widely
- Good quality data repository may meet many principles
- Support for data identification, citation, curation, standardization
- About 150 repositories are DSA or WDS certified



PANGAEA.

Data Publisher for Earth & Environmental Science

Not logged in

SEARCH SUBMIT ABOUT CONTACT

Submit Data



Welcome to PANGAEA® Data Publisher

Our services are generally open for archiving, publishing, and re-usage of data. The World Data Center PANGAEA is member of the ICSU World Data System.

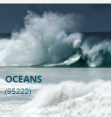
ALL TOPICS

Search for measurement type, author name, project, taxa,...



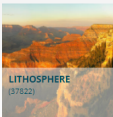
TOPICS

MAP



OCEANS

(95222)



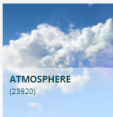
LITHOSPHERE

(37822)



BIOLOGICAL
CLASSIFICATION

(23459)



ATMOSPHERE

(23920)



PALEONTOLOGY

(22461)



ECOLOGY

(13419)



BIOSPHERE

(6284)



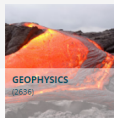
LAND SURFACE

(5420)



CHEMISTRY

(3456)



GEOPHYSICS

(2636)



Latest News

2017-05-24
SAD NEWS



On the 27th of April our colleague Mathias Weinreb unexpectedly deceased. Since several years

Mathias has given IT support in the fields of hydroacoustics and terminologies. We lost a respected co-worker and friend. The PANGAEA Team

2017-05-08
RE:PUBLICA 2017



PANGAEA presents on the re:pública 2017 conference in Berlin to the broad public: How is data gathered, archived for publication, and reused by other

scientists? We are looking forward to meeting you on this conference as part of the submarine track!

► Show all 6 news items...

Featured Data

Kim, J-H; Buscail, R; Fanget, A-S et al. (2017):
(Table 1) SPM samples and sites along the Rhône Rive

PANGAEA

- Open-access data publisher for earth and environmental science
- Archives, publishes, and distributes geo-referenced research data
- Can be used by any researcher to use, archive, publish data
- Published data are freely available
- Data and metadata canonicalized in relational database
- Data published in standard formats, accessed with standard protocols
- Data are identified and citable, using DataCite DOIs
- Data retrieval supported by full-text faceted search
- Cross-linking between data and articles
- Access to data warehouse
- Data undergo editorial process
- Ensures high usability of published data

Other Repositories

- re3data.org knows of 64 repositories in oceanography, e.g.
 - ▶ Data Portal German Marine Research
 - ▶ Marine Geoscience Data System
 - ▶ World Data Service for Oceanography
- Generic data repositories (not just data)
 - ▶ figshare
 - ▶ zenodo
 - ▶ ResearchGate

Credit

- Perhaps primary motivation to publish data
- Credit not as “advanced” as for articles, no d-index
- Article about the data used as workaround
- Demanded co-authorship on articles that use data
- Claim contributions to your ORCID record

Take aways