

**Open data in a scientific context:
From research data management plans
to research data repositories**

Background & Motivation



1998: Research data should be archived for at least **10 years** in the own institute or a supraregional information infrastructure.



2004: The enormous cost of generating scientific research data, archiving and access to (publicly funded) research data was an important aspect in the final document of the "OECD Committee for Scientific and Technological Policy at Ministerial Level".

Since the beginning of first ideas about a research project: It is important **to take the long-term preservation of research data into account** by the scientists as well as the public research promoters which is the key to a successful research data management.

Background & Motivation



Group of Eight 2013: One of the most important international commitments to open science has been signed by the science ministers of the G8 countries in 2013: “...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.”



Horizon 2020: If your Horizon 2020 project is part of the pilot, and your data meets certain conditions, **you must deposit your data in a research data repository** where they will be findable and accessible for others.

National Background

Recommendation of the DFG (2008):

- A **specific organizational concept** has to be defined, which **regulates the sustainable storage of the data**.
- The storage of the research data takes place **within the framework of defined standards**.
- The data are **personally marked** and stored **under the name of the researcher**.
- Each scientist makes his primary research data **available on a supraregional basis as far as possible**.
- The data is **described by metadata**. The **data are quality tested**.

Why to publish research data?

Opportunities for researchers

- Your research will become **more visible**. Publications whose related data are openly available **are significantly more cited**. PIWOWAR AND VISIONS (2013), HENNEKEN AND ACCOMAZZI (2011)
- The publication of research data is increasingly **recognized as a scientific achievement**.
- They **increase the quality and trustworthiness** of research by giving others the **opportunity to verify the results**.
- It is possible to securing own research investments through the establishment of a **blocking period**.

Opportunities for Science

- Data publication opens up **new research potentials** for re-analyzes with **new research questions** or methods by combining data from different sources.
- Data publication **reduces redundant data production in science**, saving time and money.

What are research data?

Research data are the **central object of the scientific knowledge process.**

BÜTTNER ET AL. (2011)

Primary research data are **data obtained during the course of source research, experiments, measurements or surveys.**
They are the **basis for scientific publications.**

DFG (2009)

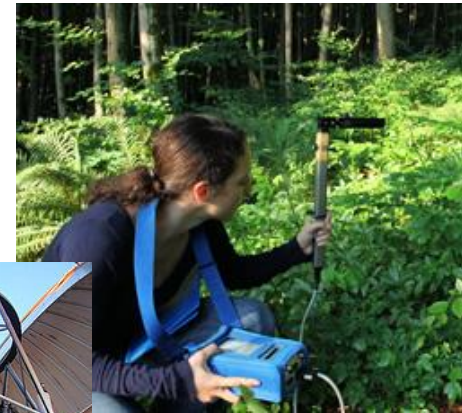
The synonymes:

- **Primary (research) data = Forschungsprimärdaten, Primärdaten**
- **Raw data = Roh- oder Urdaten**

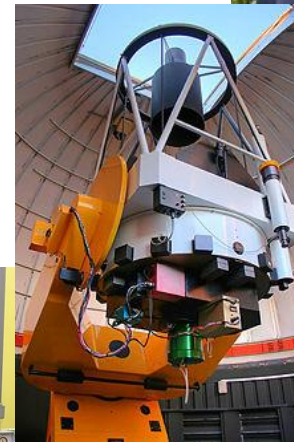
What are research data?

Examples [DREYER, 2012]:

- Measuring data from instrumentes like a telescope
- Data from mass spectrometry (IRMS, SPIN-MIMS usw.)
- Spectograms (AAS, AES, ICP-AES/OES, NMR, NIR, UV/VIS, XRS usw.)
- Full text (creation of critical editions)
- Digital photographs
- databases
- 3-D reconstructions
- Movies, audio files
- vector drawings
- survey data
- Digital maps
- Aerial photos
- Satellite data and images



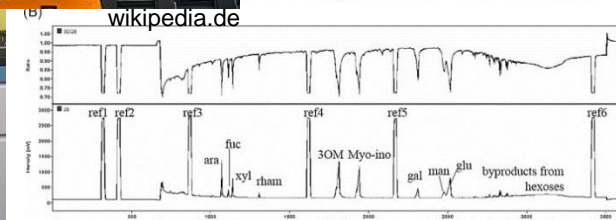
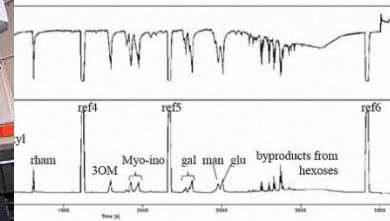
wsf.ch



wikipedia.de



thuenen.de



ir-spektroskopie.de

The Data Management Plan – where to start?

DFG Guidelines for Applications - Project proposals:

"If research data or information is obtained systematically from project funds which are suitable for the subsequent use by other scientists, **please indicate whether and in what way they are made available to others.**

Please take into account the **existing standards and data repositories or archives for your discipline.**"

The Data Management Plan – where to start?

Research issue and first draft of a research grant application: the data management plan.



Planning the method: what are the output data formats?

First draft, ready to publish: what is about legal rights?

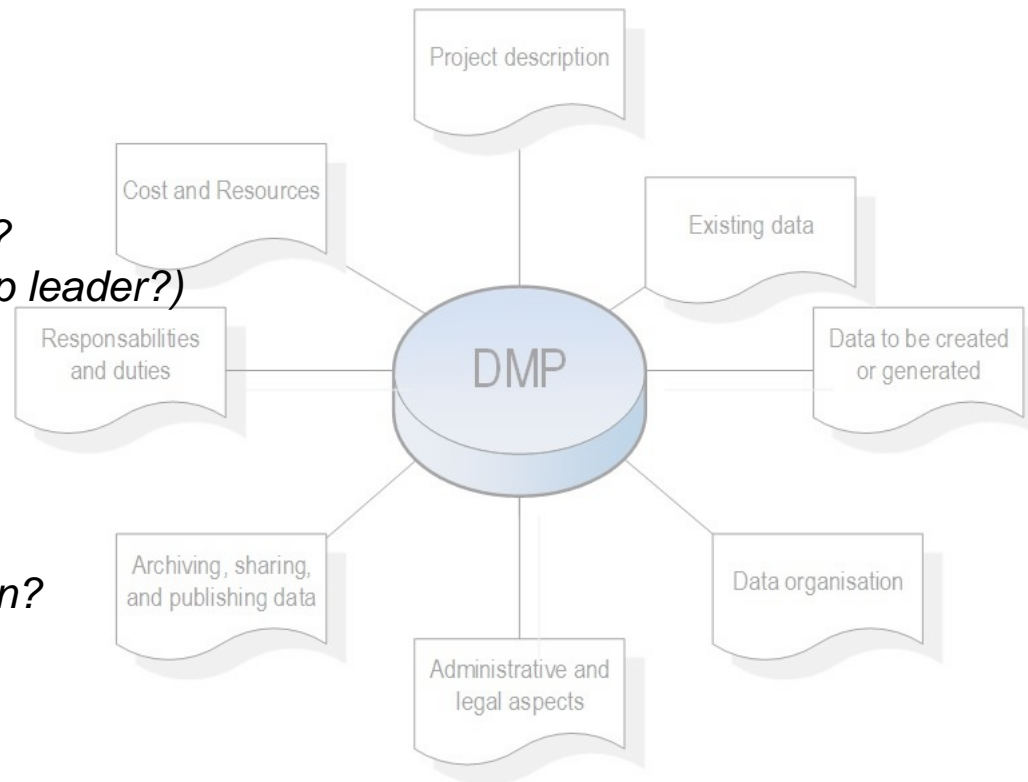


What is about long-term preservation of my data? Where to publish the research data? Access?

The Data Management Plan – where to start?

Data management plan (example)

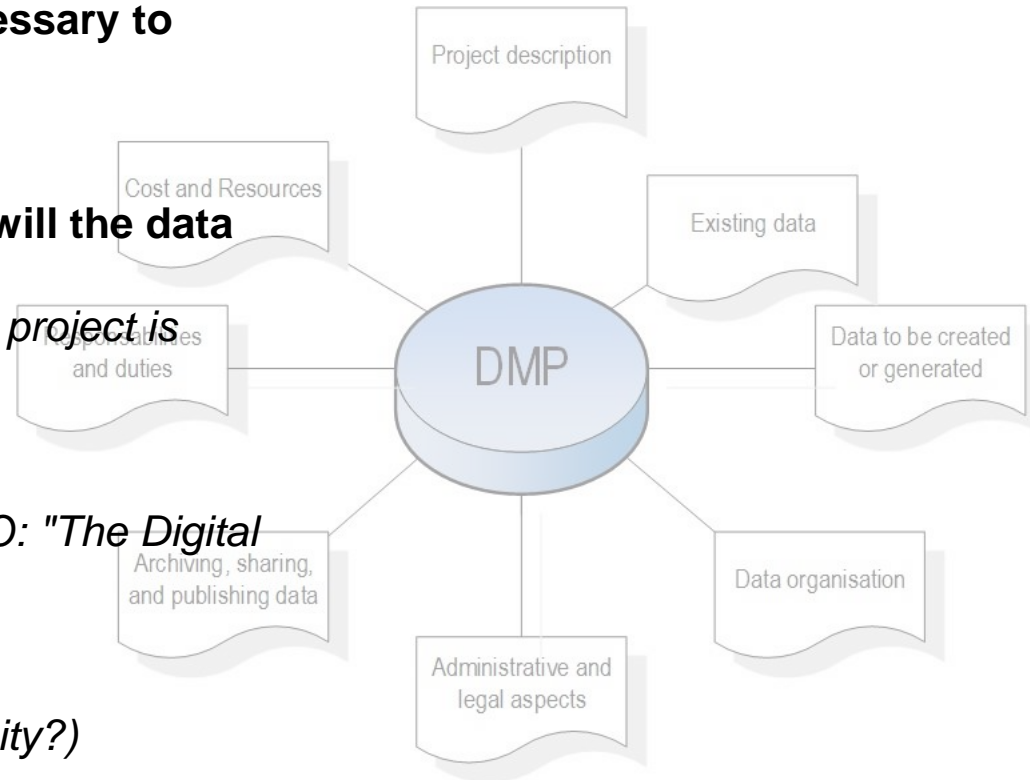
- **What data is generated and used, and how much is the data volume?**
- **Who is responsible for the data?**
(PhD student? Correspondence Author? Supervising Prof., Post-doc, work group leader?)
- **What kind of data?**
(Measurement data? Images?)
- **What data should be archived?**
(Only the data relevant to the publication? Additional data from pre-experiments? Calibration Data?)



The Data Management Plan – where to start?

Data management plan (example)

- **What additional information is necessary to understand the data?**
(For example, in the metadata file)
- **When is the data selection & When will the data be transferred?**
(After publishing in a journal? After the project is completed?)
- **How long should the data be kept?**
(DFG, for example, 10 years; UNESCO: "The Digital Heritage of Humanity")
- **Who is allowed to use the data?**
(Open Access/Data? Specific community?)



Data/file formats

Unlocked - Uncompressed – Not company related or patented

File Format	Recommendation	Avoid
Text	TXT, HTML, RTF, <u>PDF/A</u>	Word, PowerPoint
Tables	CSV, TSV	Excel
Multimedia	Container: MP4, Ogg Codec: Theora, Dirac, FLAC	QuickTime H264
Image	TIFF, JPEG2000, PNG	GIF, JPG

Metadata (mostly as XML-file)

With the help of metadata and a research documentation...

- it is possible to store research data for 10, 20, 50 years in an intelligible form,
- to verify research results and to make research data available to other researchers
- to make research data available for other research questions.

Long-term preservation: where to upload my data?



Registration platform of information.

provides an overview to a scientist where research data are stored or can be stored.

Partners:

- GeoForschungsZentrums (GFZ)
- Institut for library and informationscience (IBI, HU-Berlin)
- Library of Karlsruher Institute of Technology (KIT)

articles

1000 Genomes
Thousand Genomes

Subject(s) Basic Biological and Medical Research Human Genetics Pharmacology
Biological and Biomimetic Chemistry Biology Life Sciences Medicine
Medicine Biological Chemistry and Food Chemistry Chemistry
Natural Sciences

Content type(s) Standard office documents Images Structured graphics
Scientific and statistical data formats Raw data Plain text

Country United Kingdom United States

The 1000 Genomes Project is an international collaboration to produce an extensive public catalog of human genetic variation, including SNPs and structural variants, and their haplotype contexts. This resource will support genome-wide association studies and other medical research studies. The genomes of about 2500 unidentified people from about 25 populations around the world will be sequenced using next-generation sequencing technologies. The results of the study will be freely and publicly accessible to researchers worldwide.

NCBI
National Center for Biotechnology Information

Subject(s) Bioinformatics and Theoretical Biology Biology General Genetics
Human Genetics Medicine Biological and Biomimetic Chemistry
Basic Biological and Medical Research Life Sciences Medicine
Biological Chemistry and Food Chemistry Chemistry Natural Sciences

Content type(s) Scientific and statistical data formats Images Structured graphics
Structured text Plain text Archived data

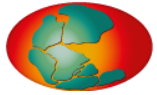
Country United States

The National Center for Biotechnology Information advances science and health by providing access to biomedical and genomic information

Canadas National Aquatic Biological Specimen Bank and Database
NABSB

Subject(s) Zoology Animal Ecology, Biodiversity and Ecosystem Research
Biochemistry and Animal Physiology Biological and Biomimetic Chemistry
Water Research Biology Life Sciences
Biological Chemistry and Food Chemistry Chemistry Natural Sciences
Geosciences (including Geography)

Long-term preservation: where to upload my data?



PANGAEA.

Data Publisher for Earth & Environmental Science

Library system for data from earth system research and environmental sciences.

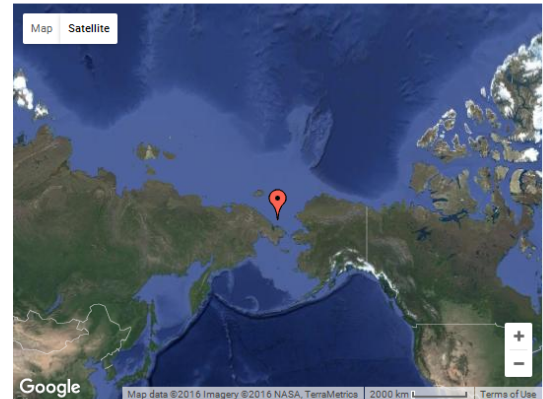
Data can be stored georeferenced in time and spatial localization.

PANGAEA is supported by the Alfred Wegener Institute for Polar and Marine Research (AWI)

Citation: **Tsoy, Ira; Obrezkova, MS; Aksentov, KI; Kolesnik, AN; Panov, VS (2016):** Late Holocene diatom abundances in sediment core HC-11. Dataset #867181 (*DOI registration in progress*), *Supplement to:* Tsoy, I et al.: Late Holocene environmental changes inferred from diatoms in the southwestern part of Chukchi Sea. *Russian Journal of Marine Biology*, **in review**

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

[RIS Citation](#)
[BIBTeX Citation](#)
[Text Citation](#)
[Facebook](#)
[Twitter](#)
[Google+](#)
[Show Map](#)
[Google Earth](#)



Abstract: The study of diatoms in core HC11 collected from the southwestern part of Chukchi Sea, allowed to distinguish 3 diatoms ecological zones, reflecting paleoenvironmental changes during the last 2300 years. The sediment age was based on the sedimentation rates, determined by 210Pb and radiocarbon dating of mollusk shells. The environmental changes of Chukchi Sea revealed by examination of diatoms correlates with global climate changes - the warming of the early and middle Subatlantic and cooling of the late Subatlantic (Little Ice Age). Warming early and middle Subatlantic in the Chukchi Sea was probably stronger than the warming of the late 20th century and was not accompanied by significant changes in sea level.

Coverage: *Latitude:* 67.866670 * *Longitude:* -172.581670
Minimum DEPTH, sediment/rock: 0.005 m * *Maximum DEPTH, sediment/rock:* 1.105 m

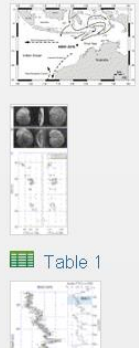
Event(s): **HC-11** [Q](#) * *Latitude:* 67.866670 * *Longitude:* -172.581670 * *Elevation:* -49.0 m * *Recovery:* 1.11 m * *Location:* Chukchi Sea [Q](#) * *Device:* Core (CORE) [Q](#)

Comment: <1 = Additional diatoms recorded after the main count.

Purchase Export Search ScienceDirect Advanced search

- Article outline
Abstract
Keywords
1. Introduction
2. Material and method
3. Results
4. Discussion
5. Conclusion
Acknowledgments
Appendix A. Taxonomic list
Appendix B. Supplementary data
References

Figures and tables



Marine Micropaleontology
Volume 66, Issues 3-4, 20 February 2008, Pages 208-221

Centennial-scale climate variability in the Timor Sea during Marine Isotope Stage 3

Anke Dürkop, Ann Holbourn, Wolfgang Kuhnt, Rina Zuraida, Nils Andersen, Pieter M. Grootes

Choose an option to locate/access this article:
Check if you have access through your login credentials or your institution
Purchase
Get Full Text Elsewhere

http://dx.doi.org/10.1016/j.marmicro.2007.10.002 Get rights and content

Abstract

We present a high-resolution (~ 60-110 yr) multi-proxy record spanning Marine Isotope

- Recommended articles
Tropical warming in the Timor Sea led deglacial...
Climate variability and land-ocean interactions ...
Changes in the thermocline structure of the Ind...
View more articles >>
Citing articles (11)
Related book content

PANGAEA - Related Data
Paleoclimate investigations on sediment core MD01-2378
Map Satellite

Thanks for your attention



Sabine Boccalini, Dipl.-Biol.
University Library
Tel.: 969-4568 (or 2718)
sabine.boccalini@uni-osnabrueck.de

Dr. Marco Gronwald
University Library
Tel.: 969-4322 (or 2085)
marco.gronwald@uni-osnabrueck.de

References

Büttner, S.; Hobohm, H-C.; Müller, L. (2011): Handbuch Forschungsdatenmanagement. Bad Honnef.

DFG (2014): Leitfäden für die Antragstellung – Projektanträge. URL: http://www.dfg.de/formulare/54_01/54_01_de.pdf.

Dreyer, B. (2012): Forschungsdaten und die TIB – Was sind Forschungsdaten? URL: <https://blogs.tib.eu/wp/tib/2012/10/16/forschungsdaten-und-die-tib-was-sind-forschungsdaten>

Henneken, E. A. and A. Accomazzi (2011): Linking to Data – Effect on Citation Rates in Astronomy. arXiv: 1111.3618.

Neuroth, H.; Strathmann, S.; Oßwald, A.; Scheffel, R.; Klump, J.; Ludwig, J. (2012): Langzeitarchivierung von Forschungsdaten – Eine Bestandsaufnahme. Göttingen.
URL: https://nestor.sub.uni-goettingen.de/bestandsaufnahme/nestor_iza_forschungsdaten_bestandsaufnahme.pdf

Paul, E. A. (2007): Soil Microbiology, Ecology, and Biochemistry. Academic Press.

Piwowar, H. A. and Todd J. Visions (2013): Data reuse and the open data citation advantage. DOI: 10.7717/peerj.175.