Peer reviewed data publication (= providing quality data for future reuse)

Sünje Dallmeier-Tiessen, Hans Pfeiffenberger

Alfred-Wegener-Institute for Polar and Marine Research, Helmholtz Association - Germany

Contact: suenje.dallmeier-tiessen@awi.de

Codata 21 – Scientific Information for Society - From Today to the Future

Agenda

- Providing Data: Open Access to Data?

- Data Publishing
  - Prerequisites and Obstacles

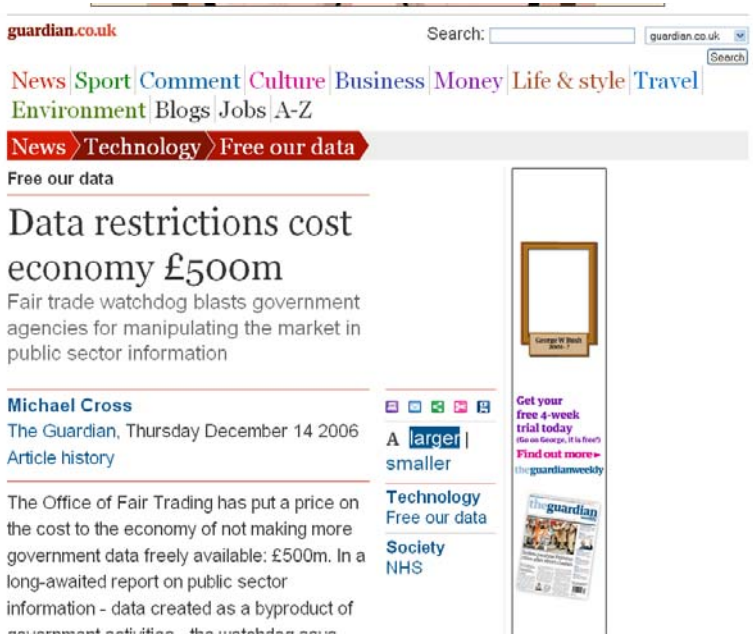
- Earth System Science Data – A Data Publishing Journal
  - Aims
  - Manuscript and Review Guidelines
  - Structure of the Article

- Summary, Outlook

Open Access to Data – Why?

- Costs

- *Reuse and Reconstruction*:
  - ➔ Further Interpretation ➔ further advancement of science
  - ➔ No cost-intensive repetitions of measurements
  - ➔ Independent quality assessment


guardian.co.uk

Search:

News | Sport | Comment | Culture | Business | Money | Life & style | Travel | Environment | Blogs | Jobs | A-Z

News › Technology › Free our data

Free our data

Data restrictions cost economy £500m

Fair trade watchdog blasts government agencies for manipulating the market in public sector information

Michael Cross

The Guardian, Thursday December 14 2006

Article history

The Office of Fair Trading has put a price on the cost to the economy of not making more government data freely available: £500m. In a long-awaited report on public sector information - data created as a byproduct of

A larger | smaller

Technology
Free our data

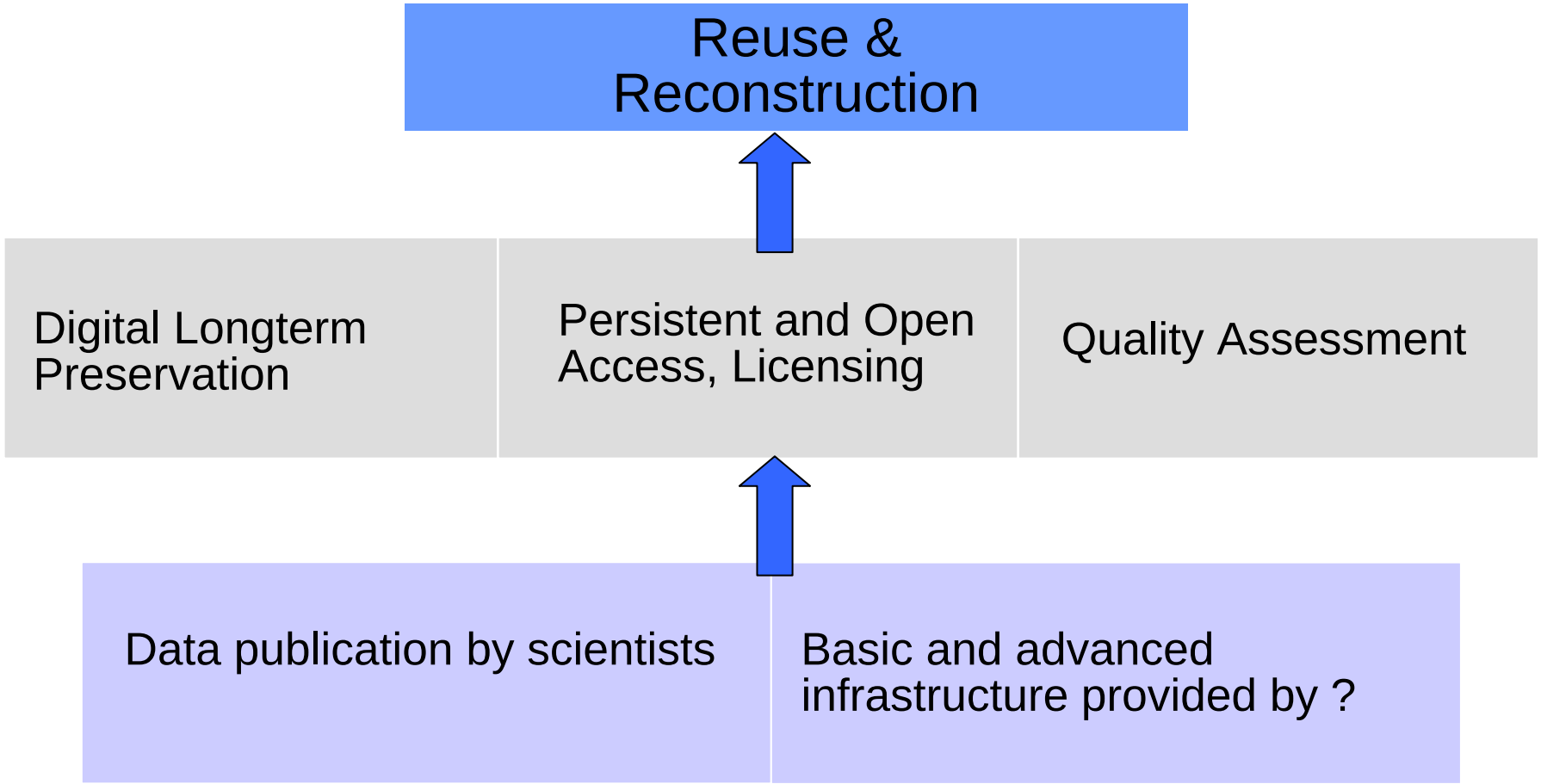
Society
NHS

Get your free 4-week trial today

Find out more

But: dependent on discipline and types of data – some may be restricted due to personal rights etc.

Open Access to Data – Structure



Data „publication“ so far: Infrastructures used

- Own computer, publication on request via Email
- Supplement to „traditional“ publications
  - Sometimes restricted access
  - Not peer-reviewed, often no full documention of instrumention and processing
- Websites: project or institutional websites
  - Website shut down after project ends? URL changes…
- Repository (no certification):
  - No metadata, technical and instrumental background difficult to understand
- Repository (certified):
  - Reuse enabled via documentation and persistent access

➔ but no reward for scientists

Data Publication: Earth System Science Data

Key problem:

missing reward for data publication,

missing quality assurance

Solution:

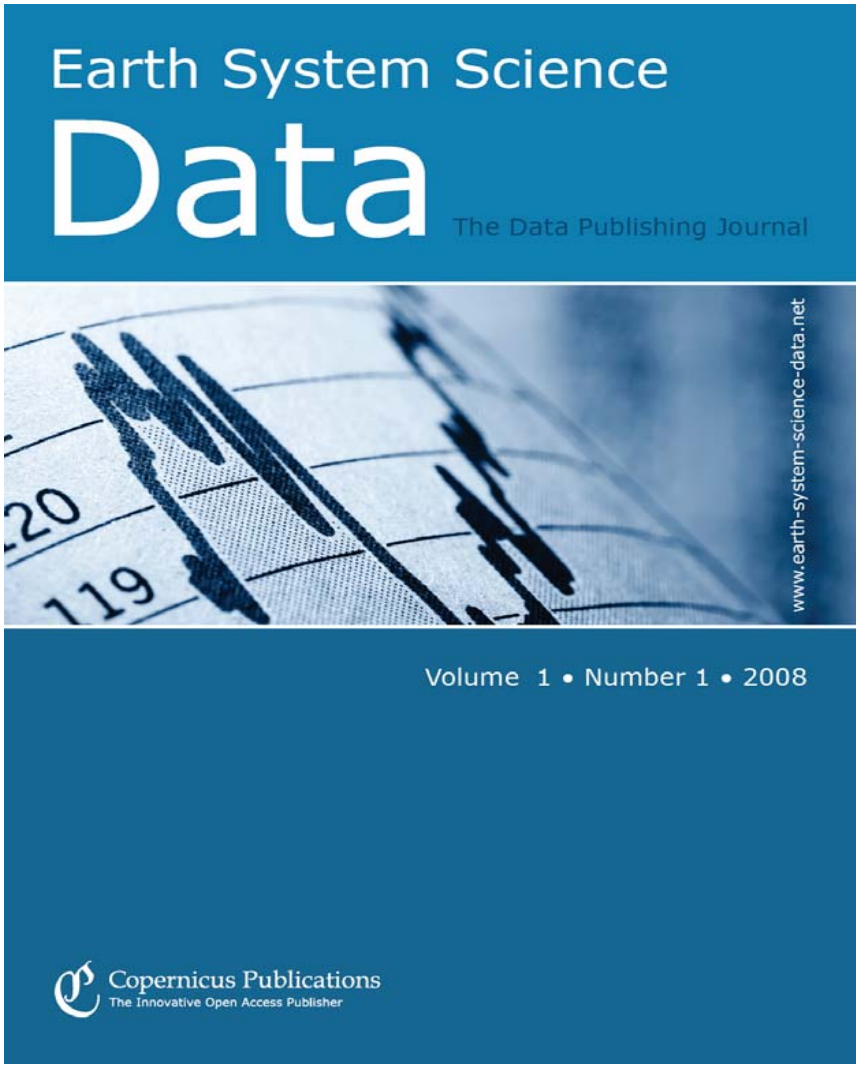
Making use of the established publication process

➔ Citable publication

➔ Peer Review

Earth System Science Data

➔ *To foster reuse of the data*



Who is who…

Advisory Board:

Prof. Dr. Paul J. Crutzen

Prof. Dr. Sydney Levitus

Prof. Dr. Alexander Petrovich Lisitzin

Editors in Chief:

David Carlson

Hans Pfeiffenberger

Publishing House

Copernicus Publishers – OA Publisher, EGU

➔innovative peer-review process

Copernicus Publications
The Innovative Open Access Publisher

Earth Syst. Sci. Data Discuss., 1, 1–13, 2008
www.earth-syst-sci-data-discuss.net/1/1/2008/

Earth System Science Data Discussions

Earth System Science Data Discussions is the access reviewed discussion forum of *Earth System Science Data*

Compilation of ozonesonde profiles from the Antarctic Georg-Forster-Station from 1985 to 1992

G. König-Langlo and H. Gernandt

Alfred Wegener Institute for Polar and Marine Research, Bussestraße 24, 27570 Bremerhaven, Germany

Received: 29 July 2008 – Accepted: 5 September 2008 – Published: 22 September 2008

Correspondence to: G. König-Langlo (gert.koenig-langlo@awi.de)

Published by Copernicus Publications.

Abstract

On 22 May 1985 the first balloon-borne ozonesonde was successfully launched by the staff of Georg-Forster-Station (70°46′ S, 11°41′ E). The following weekly ozone soundings mark the beginning of the continuous investigation of Germany to study the vertical ozone distribution in the southern hemisphere.

In 1985 these ozone soundings have been the only record showing the change of vertical ozone distribution in the southern polar stratosphere in September and October. The regular ozone soundings from 1985 until 1992 are a valuable reference data set since the chemical ozone loss became a significant feature in the southern polar stratosphere.

The balloon-borne soundings were performed at the upper air sounding facility of the neighbouring station Novolazarevskaya, just 2 km apart from Georg-Forster-Station. Till 1992, ozone soundings were taken without interruption. Afterwards, the ozone sounding program was moved to Neumayer-Station (70°39′ S, 8°15′ W) 750 km further west.

Data coverage and parameter measured

Repository-Reference: doi:10.1594/PANGAEA.547983
Available at: http://dx.doi.org/10.1594/PANGAEA.547983
Coverage: East: 11.8300; South: −70.7700
Location Name: Georg-Forster-Station, Antarctica
Date/Time Start: 1985-05-22T05:19:00
Date/Time End: 1992-01-29T01:19:00

ESSDD
1, 1–13, 2008

Antarctic ozonesonde profiles

G. König-Langlo and H. Gernandt

Parameter	Short Name	Unit	Comment
Altitude	Altitude	m	height above mean sea level
Date/Time	Date/Time		universal time code (UTC)
Longitude	Longitude		at launching point
Latitude	Latitude		at launching point
Ozone, partial pressure	O3	mPa	
Pressure, at given altitude	PPPP	hPa	
Temperature, air	TTT	deg C	
Wind direction	dd	deg	
Wind velocity	ff	m/sec	

1 Instrumentation

Since the end of 1974 balloon-borne electrochemical ozone sondes have been regularly launched, i.e. once or three times per week, at the Aerological Observatory Lindenberg (52.22° N, 14.12° E) of the Meteorological Service of GDR. The ozone sensor was developed by Rönnebeck and Sonntag (1976) on the basis of the Brewer-Mast principle (Brewer and Milford, 1960). The sondes of the type OSE-2, OSM-2, OSR (since April 1981), OSE-3 (since September 1986) and OSE-4 (since August 1989) were manufactured in cooperation with the Scientific Instrument Laboratory of the Academy of Sciences in Berlin. Prototypes of the sonde took part in the International Ozone Sonde Intercomparisons at Hohenpeissenberg in 1970 (Attmannspacher and Dütsch, 1970) and in 1978 (Attmannspacher and Dütsch, 1981).

The ozone sensors were adapted to the Russian balloon-borne radiosonde system Meteorit/RKS-5 for data transmission, receiving and processing of ozone partial pressure. The standard data, temperature and humidity, were obtained from the radiosonde RKS-5. Pressure, wind direction and wind speed were derived from radar measurements of the ground station Meteorit tracking the balloon flight.

ESSDD
1, 1–13, 2008

Antarctic ozonesonde profiles

G. König-Langlo and H. Gernandt

The ozone partial pressure measurements were achieved by a small electrically driven gas sampling pump which forces ambient air through a sensing solution of an electrochemical cell (Brewer-Mast principle) which generates an electrical current proportional to the mass flow rate of ozone.

For balloon-borne ozone profile measurements a pump correction has to be applied in order to compensate the decreasing pump efficiency with increasing height and changing air temperature. Both, an inadequate pump correction and an erroneous estimate of residual ozone above the height of balloon burst may contribute to the overall measurement error of the ozone profile. Usually an independent column ozone observation X_D by spectrometer measurement is compared with the integrated column ozone X_S between the ground level and the height of balloon burst plus estimated residual ozone above that level to adjust the recorded profile values. The correction factor is

$$C = X_D / X_S.$$

Systematic differences and random errors of the electrochemical ozone sonde, type OSR, has been estimated by analysing 20 tandem ozone soundings at the Aerological Observatory Lindenberg in 1982 (Feister et al., 1985). Random errors are at their maximum of about 10 to 13% in the troposphere and above 32 km, and reach a minimum of 2 to 5% between 20 and 28 km. The mean random error is 11.5% in the troposphere, 7% in the stratosphere beneath the ozone maximum height (ca. 22 km), and 5.6% above that height.

2 Data Provenance and Structure

The first permanently operated German research base – later named Georg-Forster-Station – was established in 1976 in the Schirmacher Oasis at 70°46′ S, 11°41′ E. Since then the station was permanently used and operated as an annex to the Russian station Novolazarevskaya until 1987, and then as a German Antarctic station named after

ESSDD
1, 1–13, 2008

Antarctic ozonesonde profiles

G. König-Langlo and H. Gernandt

2.1 Subsets of data

The dataset compiles 426 individual subsets of data, one for each sounding. Figure 2 presents an overview of the ozone measurements from the whole dataset.

The data were taken while the balloons were ascending. The soundings were terminated at the burst level of the balloons. The ascent velocity was about 5 m/s. During summertime – when the stratosphere was relatively warm – frequently the balloons reached levels above 30 000 m. During wintertime the low temperatures in the stratosphere (below −80°C) lead to lower burst heights. Regular launches were performed once per week and during spring (September, October) up to three soundings per week. Occasionally strong winds or technical problems (1989) made the measurements impossible.

2.2 Related datasets

Related datasets which were used with the data from Georg-Forster-Station:

Syowa:

Irregular ozone soundings since 1966. These data are used as a very important reference for the pre-ozone hole period in the Dronning Maud Land region (Gernandt et al., 1996).

Halley:

Ozone soundings gained in 1987 were used for dynamical studies of ozone variations (Gernandt, 1995).

ESSDD
1, 1–13, 2008

Amundsen-Scott South Pole Station:

Ozone soundings gained in 1987 were used for dynamical studies of ozone variations (Gernandt, 1995).

Koldewey-Station:

Ozone soundings gained since October 1991 data are used for bipolar comparisons (Gernandt et al., 1998; Rex et al., 2000).

Neumayer-Station:

After January 1992 the ozone soundings at Georg-Forster-Station were continued at Neumayer-Station (70°39′ S, 8°15′ W) 750 km further west (König-Langlo et al., 2006). The results of both stations can be regarded as one time series, see Fig. 3. It is the longest nearly continuously measured time series of balloon-borne ozone observations since the chemical ozone loss became a significant feature in the stratosphere.

The combined time series of both stations show the development of the stratospheric ozone layers since 1985 until now. Height-dependent trend and forcing studies were performed for the combined time series of balloon-borne ozone soundings (Gernandt et al., 1998).

3 Data access

The whole data set is published at http://dx.doi.org/10.1594/PANGAEA.547983. It contains the 426 doi-references of each single sounding. The whole dataset can be downloaded as one zipped file of tab-delimited textfiles. Each textfile includes all necessary metadata and a table containing all parameters of one sounding. No special software is needed to access the data. Special software is offered to convert or visualize the

ESSDD
1, 1–13, 2008

Fig. 1. Location of Georg-Forster-Station.

ESSDD
1, 1–13, 2008

Antarctic ozonesonde profiles

G. König-Langlo and H. Gernandt

Review Guidelines: Originality

- Are the data or methods new?

i.e., never measured or employed before or are they an improvement, e.g., in resolution or precision, of a known dataset or method?

In case of a sufficiently significant or (potentially) controversial datasets, an article about the data outcome of a replication of an experiment or observation may be accepted.

Review Guidelines: Significance

- Is there any potential of the data being useful?

(This is obviously the most difficult decision to take)

There are at least three sub-criteria to evaluate:

- **Uniqueness**
- **Usefulness**
- **Completeness**

To enable reuse and reinterpretation

Review Guidelines: Data Quality

In a Certified Repository

The data must be presented

- readily available and accessible to inspection and analysis to make the reviewer's task possible.

Its accuracy, the instrumentation employed and methods of processing should reflect the "state of the art" or "best practises".

Considering all conditions and influences presented in the article, these claims and factors must be mutually consistent.

The reviewer […] will make a judgement whether the claimed findings and its factors - individually and as a whole - are plausible and without detectable faults

First Experiences

Manuscript preparation:

- Difficult to give error estimates for „old data“ – the longer data are not documented in a „proper“ way the more difficult it gets to give a full documentation of the dataset

Reviewer:

„This is something I could do for my data as well!“ ➔ you are very welcome to do so!

Other:

- „This is really helpful for us. Now we can publish our data without having the fear that somebody else uses it withouth citing us properly. Moreover, we even get a publication for this.“

➔ Considered as very useful extra-service to the repository: full documentation to the dataset and rewards for the authors

Summary - Outlook

- Data for future reuse needs to be understandable:
  - Quality assured
  - What has been done to get this dataset?
- **Reward** for data publication, citable
- First article online as discussion paper

Outlook

- Special Issue with 18 papers to be published soon (Carboocean; CARINA)
- Development of new manuscript templates for other types of research data, e.g. modelling results (IPCC scenarios)

Peer reviewed data publication (= providing quality data for future reuse)

Sünje Dallmeier-Tiessen, Hans Pfeiffenberger

Alfred-Wegener-Institute, Helmholtz Association

Contact: suenje.dallmeier-tiessen@awi.de

http://www.earth-system-science-data.net/

References

- Deutsche Forschungsgemeinschaft (1998): Vorschläge zur Sicherung guter wissenschaftlicher Praxis, 1998. http://www.dfg.de/aktuelles_presse/reden_stellungnahmen/download/empfehlung_wiss_praxis_0198.pdf
- Earth System Science Data (Journal): http://www.earth-system-science-data.net/
- Empfehlung der Komission zum Umgang mit geistigem Eigentum bei Wissenstransfertätigkeiten und für einen Praxiskodex für Hochschulen und andere Forschungseinrichtungen, 10.04.2008. http://ec.europa.eu/invest-in-research/pdf/ip_recommendation_de.pdf
- Guardian – Free our Data: http://www.guardian.co.uk/technology/freeourdata
- International Council for Science (2004): Scientific Data and Information http://www.icsu.org/Gestion/img/ICSU_DOC_DOWNLOAD/551_DD_FILE_PAA_Data_and_Information.pdf
- NESTOR – Kompetenznetzwerk Langzeitarchivierung: http://www.langzeitarchivierung.de/
- Organisation for Economic Co-operation and Development (2007): Principles and Guidelines for Access to Research Data from Public Funding, 2007. http://www.oecd.org/dataoecd/9/61/38500813.pdf
- Murray-Rust, P. (2008): Open Data in Science. In: Nature Precedings, 2008. URL: http://hdl.handle.net/10101/npre.2008.1526.1