



Research Data and Infrastructures

A Funder's Perspective

The German Research Foundation

Mission Statement:

The German Research Foundation (Deutsche Forschungsgemeinschaft, DFG) is the central, **self-governing** research funding organisation that promotes research at universities and other publicly financed research institutions in Germany.

The DFG serves **all branches** of science and the humanities by funding research projects and facilitating cooperation among researchers.

What is the issue?

Research Data ...

- ... are the basis of scientific findings.
- ... are inadequately used today.
- ... are accessible in a very limited manner.
- ... are not available on a long-term basis.

The Vision of the Research Councils

- Research data should be freely accessible, they should be easy to get and they should be professionally curated on a long-term basis.

Let's assume, ...

... secure storage and archiving, sharing and re-use of research data would be possible and it would be standard in science!

- We would have many more data in a specific discipline and would optimise the research process.
- We could compare very diverse data from other fields and would avoid duplicate work.
- Data sharing would ensure quality control and other data sets could support own results and interpretations, new interpretations would be possible.
- Own results could be acknowledged by other researchers.
- Unique, non-reproducible results would be documented.
- A free access to research data would ensure the standards of „good scientific practice“.

... and there is an incentive!

OPEN  ACCESS Freely available online

 PLoS one

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

Background. Sharing research data provides benefit to the general scientific community, but the benefit is less obvious for the investigator who makes his or her data available. **Principal Findings.** 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. **Significance.** This correlation between publicly available data and increased literature impact may further motivate investigators to share their detailed research data.

Citation: Piwowar HA, Day RS, Fridsma DB (2007) Sharing Detailed Research Data Is Associated with Increased Citation Rate. PLoS ONE 2(3): e308. doi:10.1371/journal.pone.0000308

But, ... my data - *my precious* !

For scientists in many disciplines exist “very good reasons” not to share data, which may mean “giving them away”. A selection of quotes:

- *I have spent xxx Euro of my budget and yyy years of my work time to produce these data.*
- *The data may be easily misunderstood.*
- *The data could be erroneous and anybody else may discover this.*
- *Somebody could find something interesting in my data.*
- *Anybody could publish my data before I do.*
- *It takes way too much time to understand and to format my data in the right context.*
- *Where to and how could I deliver my data?*
- *I fully loose control over my data.*

Many open issues

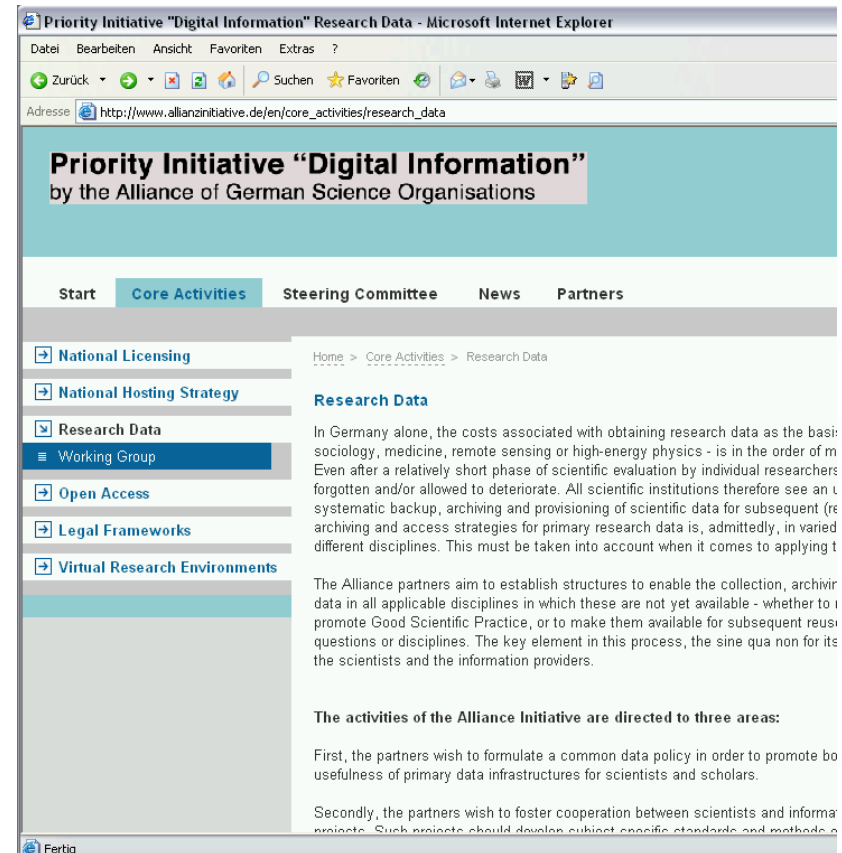
- Constitutional Freedom of Science: German Constitution § 5:
Does this include the freedom, NOT to share data?
- Who owns the data?
Funding organisation, host institution, funding applicant, scientist, publisher?
- How will data be provided?
What are the regulations for data re-use? Which licence model will be used? What are the incentives for scientists to share data?
(The availability of data not only is a technical issue ...)
- Quality control
Any data? What is possible, desirable and how? Who is responsible, which data where and how long should be archived?

The Alliance of German Science Organisations

- **The DFG together with:**
Max-Planck-Society
Helmholtz Association
Leibniz Association
Fraunhofer Association
and others

- **22 June 2008: Priority Initiative "Digital Information"**

- **The main topics are:**
National Licensing
Open Access
National Hosting Strategy
Research Data
Virtual Research Environments
Legal Frameworks



The German Research Foundation

- 1998: DFG „Proposals for Safeguarding Good Scientific Practice”
- 2003: Berlin Declaration on Open Access to Knowledge in the Sciences and Humanities
- since 2007: workshops and expert round table discussions
- 2008: National priority initiative „Digital Information "of the Alliance of German Science Organisations
- 2010: GWK/Joint Science Conference by the Federal and Länder Governments Commission on the Future of Information Infrastructure („KII“)

Recommendations for Secure Storage and Availability of Digital Primary Research Data

Committee on Scientific Library Services and Information Systems Subcommittee on Information Management

January 2009

- Definitions of research data, organisational concepts, meta data and standards
- Keeping the rights of scientists and their interests in access to data
- Free availability on a trans-regional level
- Quality control

www.dfg.de/lis/ at „Publications“

Research data management: approaches and strategies

- Shaping and moderation of an awareness building process aiming at using digital resources more efficient, develop and implement appropriate infrastructures and tools and to finally make research data available (**back up and archiving and re-use**).
- Close cooperation of scientists as data producers and providers to data repositories and also as users of those repositories with experts of the information management jointly shaping the process.
- Define discipline specific demands and specifications and include these into the process.
- Development of means for publication of research data (incl. peer-review).
- Identify best practise examples and use experience.
- Initiate pilot and exploration projects.

Current Activities

Supplement to the Guidelines for Proposal Preparation

(implemented at April 2010)

„ ... If during the proposed project scientific data will be gathered systematically, which may be feasible for later re-use, please describe the means how these data will be securely stored and made available for other scientists. Please regard the standards and the services of existing data repositories, if available in your specific scientific discipline ... “

Call for proposals: Information infrastructures for research data

(call closed at the 28. April 2010)

In order to improve the long-term preservation of research data, the German Research Foundation (DFG) had issued a call for proposals for projects aimed at developing and optimising information infrastructures that seek to achieve an efficient and sustained use of research data. This initiative is administered by the DFG's Scientific Library Services and Information Systems (LIS) division.

The German Research Foundation supports the systematic collaboration between scientists and information management experts.

Information management experts? E.g. libraries, archives ...

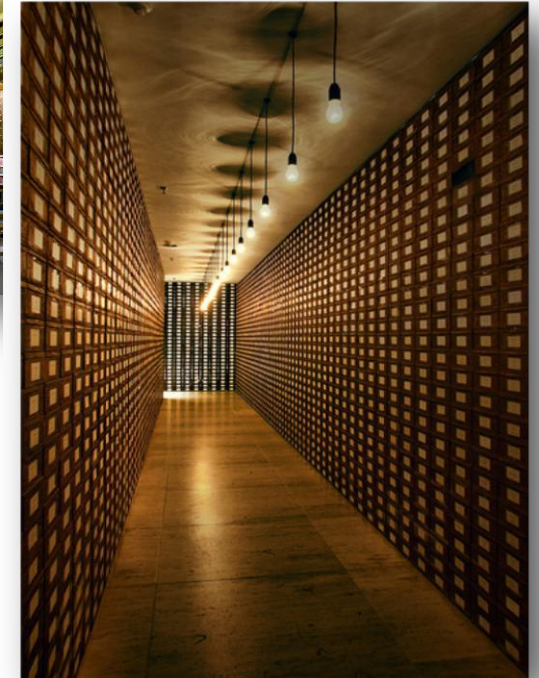


Old Reading Room of the British Library, London

http://en.wikipedia.org/wiki/File:British_Museum_Reading_Room_Panorama_Feb_2006.jpg cc: by-sa

"Archive of the Members of the German Bundestag", a work of art by
Christian Boltanski, in the basement of the Reichstag building

<http://www.flickr.com/photos/mr172/4451824642/>: cc: by-nc-sa



Data publications; e.g. ESSD



<http://www.earth-system-science-data.net/home.html>

Earth System Science Data (ESSD) is an international, interdisciplinary journal for the publication of articles on original research data(sets), furthering the reuse of high (reference) quality data of benefit to Earth System Sciences. The editors encourage submissions on original data or data collections which are of sufficient quality and potential impact to contribute to these aims.

Papers that pass a rapid access peer-review are immediately published on the Earth System Science Data Discussions (ESSDD) website. They are then subject to Interactive Public Discussion, during which the referees' comments (anonymous or attributed), additional short comments by other members of the scientific community (attributed) and the authors' replies are also published in ESSD.

What is on the agenda? A Summary

- Think disciplines
 - Define types of data, amount and characteristics, their specific life cycles and usage characteristics.
- Encouraging the publication of data sets
 - The willingness to provide access to data for scientific re-use should become standard in science and research
- Legal regulations
- Teaching and qualification
 - Scientists and data curators
- Working across sectors
- Digital collections of scientific objects
- Development and implementation of infrastructure
 - Collaboration between scientists and information management experts
 - Securing the interoperability of international and interdisciplinary networks



Thank you for your attention!

Stefan.Winkler-Nees@dfg.de