



DFG

# Access to Digital Research Data: DFG Perspectives

# Agenda

1. What is the issue?
2. Status quo and major challenges
3. DFG approaches
4. Perspectives

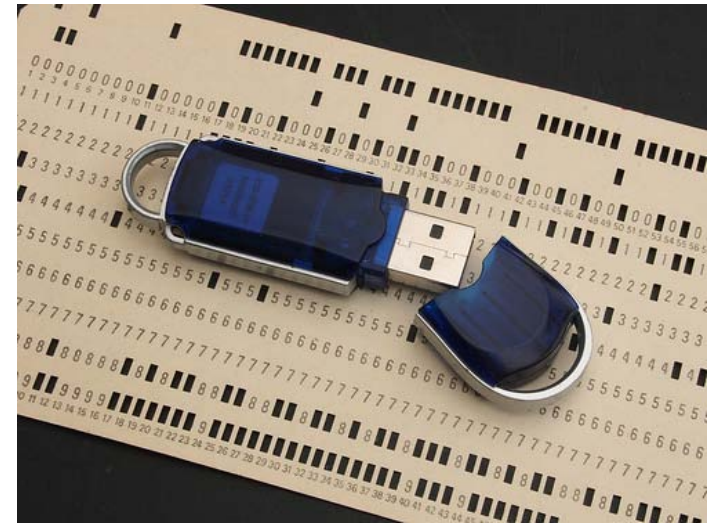
# Agenda

- 1. What is the issue?**
2. Status quo and major challenges
3. DFG approaches
4. Perspectives

# Research Data

## A definition

“A reinterpretable representation of information in a formalized manner suitable for communication, interpretation, or processing. Examples of data include a sequence of bits, a table of numbers, the characters on a page, the recording of sounds made by a person speaking, or a moon rock specimen.” (OAIS Reference Model)



Ian-S; <http://www.flickr.com/photos/ian-s/2152798588>; cc: by-nc-nd

# Berlin Declaration on OA

“Open access contributions include original scientific research results, raw data and metadata, source materials, digital representations of pictorial and graphical materials and scholarly multimedia material.”

(Berlin Declaration on Open Access to Knowledge in the Sciences and Humanities, 2003)

Vision of the Scientific Organisations:

Research data should be freely accessible, re-usable and professionally curated on a long-term basis.

# Data Sharing: Central Elements

- Secure Storage
- Standardised metadata
- Accessibility
- Long-term preservation
- Appropriate organisational frameworks and services

# Agenda

1. What is the issue?
- 2. Status quo and major challenges**
3. DFG approaches
4. Perspectives

## Research data ...

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



# Challenges and open questions

- How will data be provided?
- Quality control
- Legal aspects

# Why to share research data?

- Publicly funded research is a public good.
- The re-use of research data enables efficient and innovative science.
- Open access to data is an integral part of good scientific practice.
- Data-driven science relies on standardized and re-usable data.

# Policy Papers and Reports

- [“OECD Principles and Guidelines for Access to Research Data from Public Funding”](#) (2007)
- [“The EUROHORCs and ESF Vision on a Globally Competitive ERA and their Road Map for Actions to Help Build It”](#) (2008)
- [“Final report of the Blue Ribbon Task Force on Sustainable Digital Preservation and Access”](#) (2010)
- [“Riding the Wave: How Europe can gain from the rising tide of scientific data”](#) (2010)

# Incentive: Increased citation rates

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

# Data archiving as a good investment

<http://researchremix.wordpress.com/2011/05/19/nature-letter/>

## Research Remix

May 19, 2011

[Full Text And Details For Nature Letter "Data Archiving Is A Good Investment"](#)

Filed under: Uncategorized — Heather Piwowar @ 12:24 pm

I'm happy to say that we've just had a Letter to the Editor published in *Nature*:



**Piwowar, HA, Vision, TJ, & Whitlock, MC (2011). Data archiving is a good investment *Nature*, 473 (7347), 285-285 DOI: [10.1038/473285a](https://doi.org/10.1038/473285a)**

We hope publishing the argument in this high-visibility venue will inspire hallway conversations amongst scientists and influence how they view long-term data archive funding. Particularly those scientists who also wear hats in funding agencies!

**Research Blogging** The letter is currently behind a paywall. As is [permitted by Nature's preprint policies](#), I include the text we initially submitted below. It is very similar to what appears in the final article (linked above).

The published letter is also very short. The original article-length draft is at the bottom of this post. Needless to say, it includes nuances lost in the shorter versions. The [README associated with the data](#) has additional information about methods.

While doing this research I wrote a few blog posts about my methods and early results. Here are the links:

- [Studying Reuse Of GEO Datasets In The Published Literature](#)
- [Early Results: Public Data Archiving Increases Scientific Contribution By More Than A Third](#)
- [Rough Estimate Of Papers Per Dollar](#)

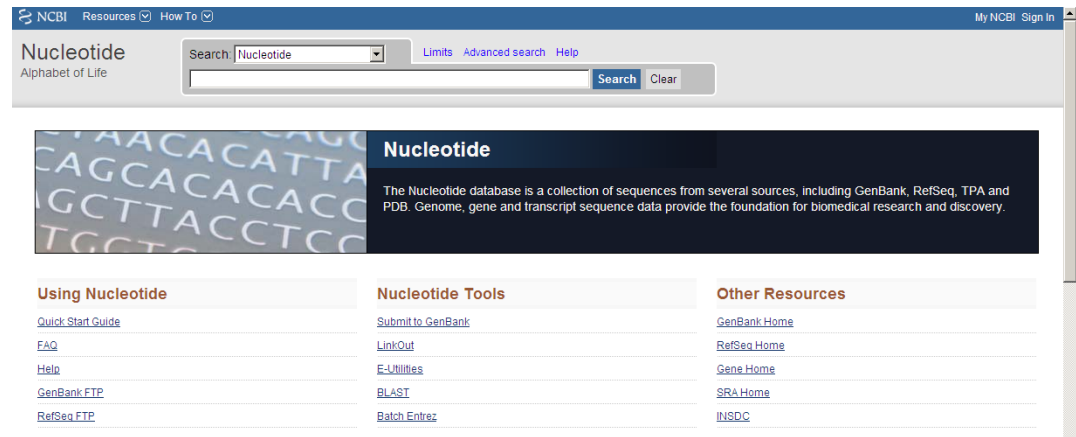
The data behind the analysis is openly available in the Dryad data archive. Please reuse it, and feel free to contact me if you have questions! Data citation:

**Piwowar HA, Vision TJ, & Whitlock MC (2011). Data from: Data archiving is a good investment *Dryad Digital Repository* : [10.5061/dryad.j1fd7](https://doi.org/10.5061/dryad.j1fd7)**

# Infrastructures for Research Data

## Some examples -1-

- The GenBank database is designed to provide and encourage access to the most up to date and comprehensive DNA sequence information.  
(<http://www.ncbi.nlm.nih.gov/sites/entrez?db=nucleotide>)
- World Data Center for Climate: Data for climate research are collected, stored and disseminated.  
(<http://mud.dkrz.de/wdc-for-climate/>)



The screenshot shows the NCBI Nucleotide database homepage. At the top, there is a navigation bar with 'NCBI Resources' and 'How To' links. Below this is a search bar with a dropdown menu set to 'Nucleotide' and a search button. The main content area features a large image of DNA sequence letters (A, C, G, T) and a dark box with the text: 'Nucleotide. The Nucleotide database is a collection of sequences from several sources, including GenBank, RefSeq, TPA and PDB. Genome, gene and transcript sequence data provide the foundation for biomedical research and discovery.' Below this, there are three columns of links: 'Using Nucleotide' (Quick Start Guide, FAQ, Help, GenBank FTP, RefSeq FTP), 'Nucleotide Tools' (Submit to GenBank, LinkOut, E-Utilities, BLAST, Batch Entrez), and 'Other Resources' (GenBank Home, RefSeq Home, Gene Home, SRA Home, INSDC).



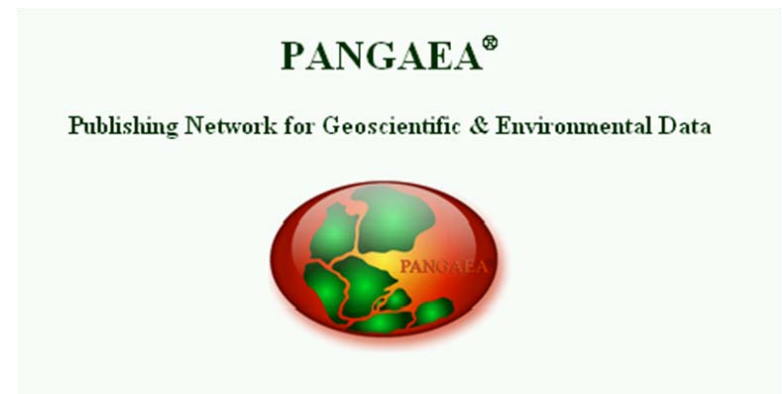
# Infrastructures for Research Data

## Some examples -2-

- DNA-Bank Network: DNA bank databases of all partners are linked and are accessible via a central web portal, providing DNA samples of complementary collections. (<http://www.dnabank-network.org/>)



- PANGAEA: geo-referenced data from earth system research. (<http://www.pangaea.de/>)



# Agenda

1. What is the issue?
2. Status quo and major challenges
- 3. DFG approaches**
4. Perspectives



# DFG: approaches and strategies

- Shape and moderate an **awareness building process (back-up and archiving and re-use)**
- Support **close cooperation** between scientists and information management experts
- Define **discipline specific demands and specifications**
- Develop mechanisms for the **publication of research data** (including peer-review)
- Identify **best-practise examples**
- Initiate **pilot and exploration projects**

# DFG Activities

## Policy Level

- 1998: DFG „Proposals for Safeguarding Good Scientific Practice”
- 2003: Berlin Declaration on Open Access to Knowledge in the Sciences and Humanities
- 2006: DFG Position Paper: Scientific Library Services and Information Systems – Funding Priorities through 2015
- 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 State Governments: Commission on the Future of Information Infrastructure („KII“)

# Recommendations for Secure Storage and Availability of Digital Research Data

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

January 2009

- Definitions of research data, organisational concepts, metadata and standards
- Encourage the commitment of scientists to share their research data
- Free availability on a trans-regional level
- Quality control

# Supplement to the DFG Guidelines and Proposal Preparation Instructions

Implemented in April 2010

## 3.7 Data handling

Improving the handling of research data is a priority both for national and international research organisations and for science in general. In order to enhance the long-term preservation of research data, the DFG funds projects that seek to achieve an efficient and sustained use of research data.

If research data will be systematically produced using DFG project funds, describe what measures will be implemented to ensure their management, curation and long-term preservation for future reuse. Please regard existing standards and data repositories in your discipline where appropriate.

# DFG Funding Level

## Pilot Projects

- 2010: Call for proposals „Information infrastructures for research data”
- April 2011: 27 projects were granted funding by the DFG Joint Committee
- 9.9 million Euros will be provided

For more information see: [Information für die Wissenschaft Nr. 18 \(2011\)](#)

## Information Management in Collaborative Research Centres

Main objective: to systematically manage the data collected within the framework of the Collaborative Research Centre in the long term

# International Cooperation



Knowledge Exchange is a co-operative effort that supports the use and development of Information and Communications Technologies (ICT) infrastructure for higher education and research.

The Knowledge Exchange partners are:

- [Denmark's Electronic Research Library](#) (DEFF) in Denmark
- [German Research Foundation](#) (DFG) in Germany
- [Joint Information Systems Committee](#) (JISC) in the United Kingdom
- [SURFfoundation](#) in the Netherlands

# Knowledge Exchange



Knowledge Exchange Vision:

To make a layer of scholarly and scientific content openly available on the Internet.

# Knowledge Exchange Working Group on Primary Research Data: Main Objectives

- Establishing a network of peers, informed and aware of policy implications and strategic developments of the partner organisations addressing the challenges of managing research data.
- Informing the partner organisations, affiliated institutions, the European Commission and other stakeholders of the Knowledge Exchange partners' views on preferred infrastructural developments and supporting the open data movement.



# Agenda

1. What is the issue?
2. Status quo and major challenges
3. DFG approaches
- 4. Perspectives**

# Perspectives

## What is on the agenda?

- Think disciplines
- Encourage the publication of data sets
- Legal frameworks
- Teaching and qualification
- Digital collections of non-textual materials
- Explore the international dimension of data management



DFG

**Thank you!**

**Contact:**

**DFG – Scientific Library Services and Information Systems:** <http://www.dfg.de/lis>

- Dr. Stefan Winkler-Nees
- Dr. Franziska Regner (Knowledge Exchange)