From Publishing to Communicating Research Data

Armin Günther & Ina Dehnhard

Leibniz Institute for Psychology Information (ZPID)
Trier, Germany
Our Mantra

**Publishing data** is easy,

**Communicating data** is hard work!
The Pioneer Plaque: “to show scientifically educated inhabitants of some other star system, who might intercept it millions of years from now, when Pioneer was launched, from where, and by what kind of beings”
The Pioneer Plaque: “to show scientifically educated inhabitants of some other star system, who might intercept it millions of years from now, when Pioneer was launched, from where, and by what kind of beings”

Date of launch, encoded by the periods of 14 pulsars

Place of launch within the Solar System

Builders of the spacecraft and inhabitants of planet Earth

Trajectory of the spacecraft through the Solar System, encoded by an arrow
Publishing data

...is (not always) easy.
Communicating data... is hard work!
Our message

• We need not only technical solutions to *publish* data.
• We also need concepts and tools to effectively *communicate* these data.
Publishing research data “can [...] be as easy as spending a minute creating a figShare account and another minute uploading your file and adding a title and subject keywords. Click a button and your data has a permanent home and DOI.

Levelling up to Open Research Data
By Deborah Fitchett, Digital Access Coordinator at Lincoln University.

http://creativecommons.org.nz/2013/10/levelling-up-to-open-research-data/
Data Publication at Figshare

Data for “Power and Fairness in a Generalized Ultimatum Game”

- data.tsv
- prefs.annotated.csv

Published on 09 May 2014 - 14:21 (GMT)
File size in total is 38.49 KB

Categories
- Sociology
- Economics

Authors
- Giovanni Luca Ciampaglia

Tags
- fairness
- ultimatum game
- proposals
- preferences
- bargaining

URL: http://dx.doi.org/10.6084/m9.figshare.1021603 (Abruf 18.09.2014)
Data Publication at Figshare

- URL: http://dx.doi.org/10.6084/m9.figshare.1021603 (Abruf 18.09.2014)
“Intelligent openness” (Royal Society, 2012)

• **Accessible.** Data must be located in such a manner that it can readily be found.

• **Intelligible.** Data must provide an account of the results of scientific work that is intelligible to those wishing to understand or scrutinise them.

• **Assessable.** Recipients need to be able to make some judgment or assessment of what is communicated.

• **Usable.** Data should be able to be reused, often for different purposes.
Data Publication at Figshare

- Accessible?
- Intelligible?
- Assessable?
- Usable?

URL: http://dx.doi.org/10.6084/m9.figshare.1021603 (Abruf 18.09.2014)
Conclusion 1

• While more and more data are successfully and easily published, the communication of these data still may fail.

• But publishing data that are not intelligible does not increase transparency, it only adds noise to the scientific communication.
Communicating Data

From:
Researchers publish data

To:
Researchers communicate data to receivers

[Sender ➔ Message (Meaning) ➔ Receiver]
Communicating Data and Distance from Data-Origin

Data Origin

Individual Researcher

Research Team

near remote

Distance from data-origin
Communicating Data and Distance from Data-Origin

- Data Origin
- Individual Researcher
- Research Team
- Research Field (current)
- Distance from data-origin

near, remote
Communicating Data and Distance from Data-Origin

Data Origin

- Individual Researcher
- Research Team

Research Field (current)

Research Field (next generation/s)

Distance from data-origin

near
remote
Communicating Data and Distance from Data-Origin

Data Origin

- Individual Researcher
- Research Team
- Research Field (current)
- Research Field (next generation/s)

Distance from data-origin

Discipline (current)

Discipline (next generation/s)

near
remote
Communicating Data and Distance from Data-Origin

Data Origin

- Individual Researcher
- Research Team
- Research Field (current)
- Research Field (next generation/s)
- Distance from data-origin

Near/Remote

Other Disciplines, General Public

Discipline (current)

Discipline (next generation/s)
Conclusion 2

• To communicate research data we need metadata to ensure that the data are intelligible, assessable and usable.
• What set of metadata is needed depends on the kind of receivers/users of the data and their background knowledge.
How can I make my data submission as accessible and reusable as possible?

- Submit your data files in non-proprietary formats from which data can be easily extracted (e.g., CSV rather than PDF).
- Keep your file names short, informative, unique, and free of special characters.
- Consider submitting your data files in multiple formats if you think that will enhance their ability to be reanalyzed. View additional guidance and a list preferred Dryad file formats.
- Provide descriptive information within your data files (e.g., column headers in a spreadsheet).
- Provide a ReadMe file that provides contextual information about the data file so that it can be interpreted correctly.
- Provide titles, descriptions and keywords for your datafiles, to make the data more discoverable and to assist in understanding the relationship of the datafile to the publication.

https://datadryad.org/pages/faq
Disciplinary Metadata

While data curators, and increasingly researchers, know that good metadata is key for research data access and re-use, figuring out precisely what metadata to capture and how to capture it is a complex task. Fortunately, many academic disciplines have supported initiatives to formalise the metadata specifications the community deems to be required for data re-use. This page provides links to information about these disciplinary metadata standards, including profiles, tools to implement the standards, and use cases of data repositories currently implementing them.

For those disciplines that have not yet settled on a metadata standard, and for those repositories that work with data across disciplines, the General Research Data section links to information about broader metadata standards that have been adapted to suit the needs of research data.

Please note that a community-maintained version of this directory has been set up under the auspices of the Research Data Alliance.

Search by Discipline

Biology  Earth Science  General Research Data

Physical Science  Social Science & Humanities
Once you dig deeper into the metadata standards, there are links to websites that are completely indecipherable to the average joe (e.g., me). Jargon is abundant and navigation of these sites is, at best, based on an intimate knowledge of the metadata standard; at worst, erratic and inexplicable.

[...]

We have no easy way for researchers to start understanding and creating metadata.

Wanted: Better Tools and Websites for Data Management Help
Mar 06 2013, Carly Strasser

http://datapub.cdlib.org/2013/03/06/wanted-better-tools-and-websites-for-data-management-help/
Some Future Steps for Better Data Communication

• **Tools**: Development of data communication tools for researchers
• **Metadata standards**: Development of metadata for relevant data communication use cases
• (More) **Collaboration** between information scientists and research communities
Some Future Steps for Better Data Communication

• DataWiz:
  An Automated Assistant for the Management of Psychological Research Data
  Leibniz Institute for Psychology Information, Trier, Germany.

• Funding: German Research Foundation
Contact

Armin Günther: guenther@zpid.de
Ina Dehnhard: dehnhard@zpid.de

Psychology Information
ZPID - Leibniz Institute
ZPID - Leibniz-Gemeinschaft