Identifying psychological research data in the digital environment

Erich Weichselgartner
Leibniz Institute for Psychology Information (ZPID)
Trier, Germany

IDSC of IZA/GESIS/RatSWD Workshop:
Persistent Identifiers for the Social Sciences
University Club, Bonn – Feb. 1-2, 2011

PsychData

Walter Schneider

Name not unique identifier

Walter Schneider

PsychData

Make anonymous

↓

Sj2

PsychData

Psychologie Information
ZPID – Leibniz Institut

Major issues in Psychology (outline)

- It is not common practice in Psychology to share data
- Ethical principles: Confidentiality, privacy
- Lack of standardization of instruments: documentation very laborious, context important
- Change culture of the field
- Point out advantages for researcher, community, society
- PsychData: Discipline specific repository (archive)

PsychData

PsychData is an archive of primary research data in psychology. It was developed at the *Leibniz Institute for Psychology Information* (ZPID) in Trier, Germany, with partial funding by the German Research Foundation (DFG).

Goals

1. Acquisition
2. Documentation
3. Preservation (long-term archiving)
4. Access (distribution)
5. Direct research support *(limited resources!)*
    1. Tools to incorporate data sharing in the initial design of a study
    2. Direct deposit via web interface

PsychData

Psychologie Information
ZPID – Leibniz Institut

Very limited resources (2x0,5 staff).

Selection criteria based on quality:

- Large surveys
- Studies of unique populations
- Studies conducted at unique times
- Longitudinal studies
- Non-Replicability (data replication not feasible, excessively costly or prohibitive)
- Scientific value
  - Citation, research, and educational use as published in refereed scientific publications

Active solicitation required, hardly any volunteer donors!

PsychData

Psychologie Information
ZPID – Leibniz Institut

Volume

- ~ 60 Studies, annual growth ~ 10 studies
- ~ 80 Data sets
- ~ 40 Mio Data points

Data reuse

- Possible since June 2004
- ~ 10 requests per year

Psychologie Information
ZPID – Leibniz Institut

Exemplary Study

The Munich Twin Study (GOLD): Genetic Oriented Longitudinal Study of Differential Development (in preparation)

- Long term study, begin in 1937 with 180 monozygotic (identical) and dizygotic (fraternal) twins.
- Five waves
- Genetic vs. Environmental Determinants of Traits, Motives, Self-Referential Cognitions, and Volitional Control in Old Age
- http://www.mpipf-muenchen.mpg.de/BCD/PROJECTS/gold_g.htm

Psychologie Information
ZPID – Leibniz Institut

Benefits of data sharing (in Psychology)

- Provide incentives, rewards, and recognition for scientists who share and archive data.
  - Citation is a primary scholarly indicator of value (Lyon, 2007)
  - Sharing research data is associated with increased citation rate (Piwowar, Day & Fridsma, 2007)
- Make data sets citable as scholarly publications; establish citation standard
  - Long-Lived Digital Data Collections (NSF, USA)
    - „Strategies for location-independent identification of data objects, such as Digital Object Identifiers and permanent Universal Resource Locators (URLs) need to be developed and broadly applied to address this problem. “
  - Digital Repositories Programme (JISC, UK)
  - Project STD-DOI „Publication and Citation of Scientific Primary Data” (funded by the German Research Foundation, 2003-2005)

PsychData

Psychologie Information
ZPID – Leibniz Institut

Make data citeable (as a unique piece of work and not only a part of a publication); this requires

Persistent identification

- Long-term availability (resolver, data)
- Reliability

Possible solution: **The DOI System** (International DOI Foundation).

Components

- a specified numbering syntax
- a resolution service (based on the Handle System);
- a data model system (including the indecs Data Dictionary);
- policies and procedures for the implementation of DOI names through a federation of Registration Agencies.

PsychData

Psychologie Information
ZPID – Leibniz Institut

The Digital Object Identifier (DOI®) System

The DOI System provides a framework for

- persistent identification,
- managing intellectual content,
- managing metadata,
- linking customers with content suppliers,
- facilitating electronic commerce, and
- enabling automated management of media.

DOI names can be used for any form of management of any data, whether commercial or non-commercial.

Psychologie Information
ZPID – Leibniz Institut

The DOI System

- Examples
  - doi:10.1000/182
  - doi:10.1594/PANGAEA.484677
    - The prefix identifies the registrant of the name, and the suffix is chosen by the registrant and identifies the specific object associated with that DOI
  - http://dx.doi.org/10.1000/182
  - http://dx.doi.org/10.1594/PANGAEA.484677
    - DOIs can be incorporated into Web pages much like current links. But instead of pointing to a specific Web location, the DOI sends the browser off to a database, where it retrieves and displays whatever information the publisher chooses to offer.

PsychData

Psychologie Information
ZPID – Leibniz Institut

The DOI System

- **Registration agencies**
  - CrossRef, OPOCE, DataCite (GER: GESIS, TIB, ZB MED), etc.
    - On May 1st 2005 the TIB became the world's first DOI registration agency for scientific primary data

- **Publication agents** (data centers, e.g. PsychData)
  - Long-term archive

PsychData

The primary role of **Registration Agencies** (RAs) is to provide services to Registrants - allocating DOI® name prefixes, registering DOI names and providing the necessary infrastructure to allow Registrants to declare and maintain metadata and state data. This service is expected to encompass quality assurance measures, so that the integrity of the DOI® system as a whole is maintained at the highest possible level (delivering reliable and consistent results to users). This includes ensuring that state data is accurate and up-to-date and that metadata is consistent and complies with both DOI system Kernel and appropriate Application Profile standards.

PsychData

Psychologie Information
ZPID – Leibniz Institut

Registration Agency: **DataCite**

- DataCite is focused on improving the scholarly infrastructure around datasets. There will be a set of activities around establishing and sharing best-practices, identifying and solving some of the unique issues that arise with datasets.
- DataCite is focused on working with data centres and organisations that hold data. The details of their business models, workflows, and other requirements do not appear to be identical to those of publishers producing traditional journals.
- DataCite has a business model that meets the needs of non-commercial and sometimes smaller organisations; larger national-scale organisations (e.g., TIB, BL) carry the basic infrastructure costs and will reclaim where appropriate within their domain.

PsychData

Psychologie Information
ZPID – Leibniz Institut

Publication Agent

Data publications are processed by *publication agents*. Besides the publication tasks the agents are also responsible for long-term archiving of primary data ("data library"). Each agent covers its own thematic field.

Psychologie Information
ZPID – Leibniz Institut

Responsibilities of Publication Agent

Infrastructure and services

"Datasets should be easy to find, easy to access, easy to use."

- How to identify data set? Persistent identification.
  - „Strategies for location-independent identification of data objects, such as Digital Object Identifiers and permanent Universal Resource Locators (URLs) need to be developed and broadly applied to address this problem. " NSF, 2005
- Discovery
  - *Metadata* elements providing data history, authorship, and access information
  - Catalogues, Search engines
- Data access/Release policies
  - Legal restrictions
  - Property rights, confidentiality, privacy

PsychData

Psychologie Information
ZPID – Leibniz Institut

Requirements for PID infrastructure

- Trustworthy, secure, reliable, sustainable (e.g., defined sevice level agreements)
- Acceptance in community (PUB + COM)
- Standardized, interoperable; provides guidelines
- Added value
  - resource discovery (related works, derivatives)
  - track citation, usage logs

PsychData

Psychologie Information
ZPID – Leibniz Institut

Advantages of the DOI System

- Well established (e.g., all major publishers)
- Stable, redundant, no downtimes
- 1:1 relationship between metadata and identifier (use metadata to find identifier)
- „cited-by linking“

Psychologie Information
ZPID – Leibniz Institut

References

- A fair share. The concept of sharing primary data is generating unnecessary angst in the psychology community. (7 December 2006). *Nature, 444*, 653-654
- Breckler, S. (2009). Psychology needs to develop mechanisms for data sharing. *APA Monitor Online, 40 (2)*.
- Fienberg S. E., Martin M. E., Straf M. L. (1985). *Sharing research data*. Washington, D.C.: National Academy Press.
- Guilford, J. P. (1954). *Psychometric Methods. McGraw-Hill, New York, 2nd edition.*
- Piwowar H.A., Day R.S., Fridsma D.B. (2007). Sharing Detailed Research Data Is Associated with Increased Citation Rate. *PLoS ONE 2(3): e308. doi:10.1371/journal.pone.0000308*
- Roberts, F. S. (1979). Measurement theory: with applications to decisionmaking, utility, and the social sciences. Encyclopedia of Mathematics and its Applications, Vol. 7, Addison-Wesley, Reading, MA.
- Sobal, J. (1982). The Role of Secondary Data Analysis in Teaching the Social Sciences. *Library Trends, 30*, 479-488.
- Weichselgartner, E. (2008). PsychData: An archive for primary research data in Psychology. *Keeping the Records of Science Accessible: Can we afford it? High-level strategic conference organized by the Alliance for Permanent Access, the European Science Foundation (ESF) and the Hungarian Scientific Research Fund (OTKA) in Budapest, Hungary, November 4, 2008.*
- Wicherts, J. M., Borsboom, D., Kats, J., & Molenaar, D. (2006). The Poor Availability of Psychological Research Data for Reanalysis. *American Psychologist, 61,* 726-28.

Psychologie Information
ZPID – Leibniz Institut

Bibliography

Azar, B. (1999). Psychology needs to develop mechanisms for data sharing. *APA Monitor, 30 (8)*.
Dockser M. A.: My Data, Your Data, Our Data. Wall Street Journal, 2010/04/13.
Editorial: Data for eternity. Nature Geoscience 3, 219 (2010).
Klopp, T. (2010). OPEN DATA. Forscher sollen ihre Daten teilen. ZEIT ONLINE.
Mervis, J.: NSF to Ask Every Grant Applicant for Data Management Plan. Science Insider, 2010/05/05.
Procter, M. (1993). Analyzing other researchers´ data. In N. Gilbert (Ed.), Researching social life (pp. 255-269). London: Sage.
Sieber, J. E. (Ed.). (1991). Sharing social science data. Advantages and challenges. Thousand Oaks, CA: Sage.
Sieber, J. E. (1997). Credit allocation in psychology. *Science and Engineering Ethics*, *3,* 261-264.
Tucker, Jennifer (2009). Motivating Subjects: Data Sharing in Cancer Research. Falls Church, VA (Dissertation)

Events

CNR-ISTI (Italy) Workshop: GLOBAL SCIENTIFIC DATA INFRASTRUCTURES: THE BIG DATA CHALLENGES. Hotel La Palma, Island of Capri, Italy, 12-13 May 2011

Psychologie Information
ZPID – Leibniz Institut

Other Initiatives

Europe

- CESSDA: Council of European Social Science Data Archives (http://www.cessda.org/)
- UKDA: The UK Data Archive (http://www.data-archive.ac.uk/)

USA

- CHILDES: Child Language Data Exchange System (http://childes.psy.cmu.edu/)
- Henry A. Murray Research Archive at Harvard University (http://www.murray.harvard.edu/)
- Journal of Statistics Education Data Archive (http://www.amstat.org/publications/jse/jse_data_archive.htm)

Contact

http://www.psychdata.de/

Weichselgartner@zpid.de

PsychData team: Thomas Bäumer, Ina Dehnhard, Armin Günther, Günter Krampen, Jutta von Maurice, Leo Montada, Sebastian Mühlböck, Erich Weichselgartner

Member of

German Data Forum (RatSWD)

Partly funded by the German Research Foundation

Deutsche Forschungsgemeinschaft

DFG

About ZPID (http://www.zpid.de/index.php?lang=EN):

- ZPID's objective is to provide a comprehensive, sustainable, and professionally based documentation and communication of information in the field of psychology focusing on the German-speaking countries.
- Founded in 1977 at the University of Trier.
- Non-profit organization – co-funded by the Federal Republic of Germany and the German States.
- Member of the Leibniz Association (association of 86 scientific research institutions).
- Quality Assurance by External Evaluation, Scientific Advisory Board and Supervisory Board.
- Annual budget ~ US-$ 2.5 Mio (without competition-based grants).
- ~ 30 scientific and administrative staff.