

Data Sharing in Psychology: A German Initiative

Erich Weichselgartner

Leibniz Institute for Psychology
Information (ZPID)

Trier, Germany



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.

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.

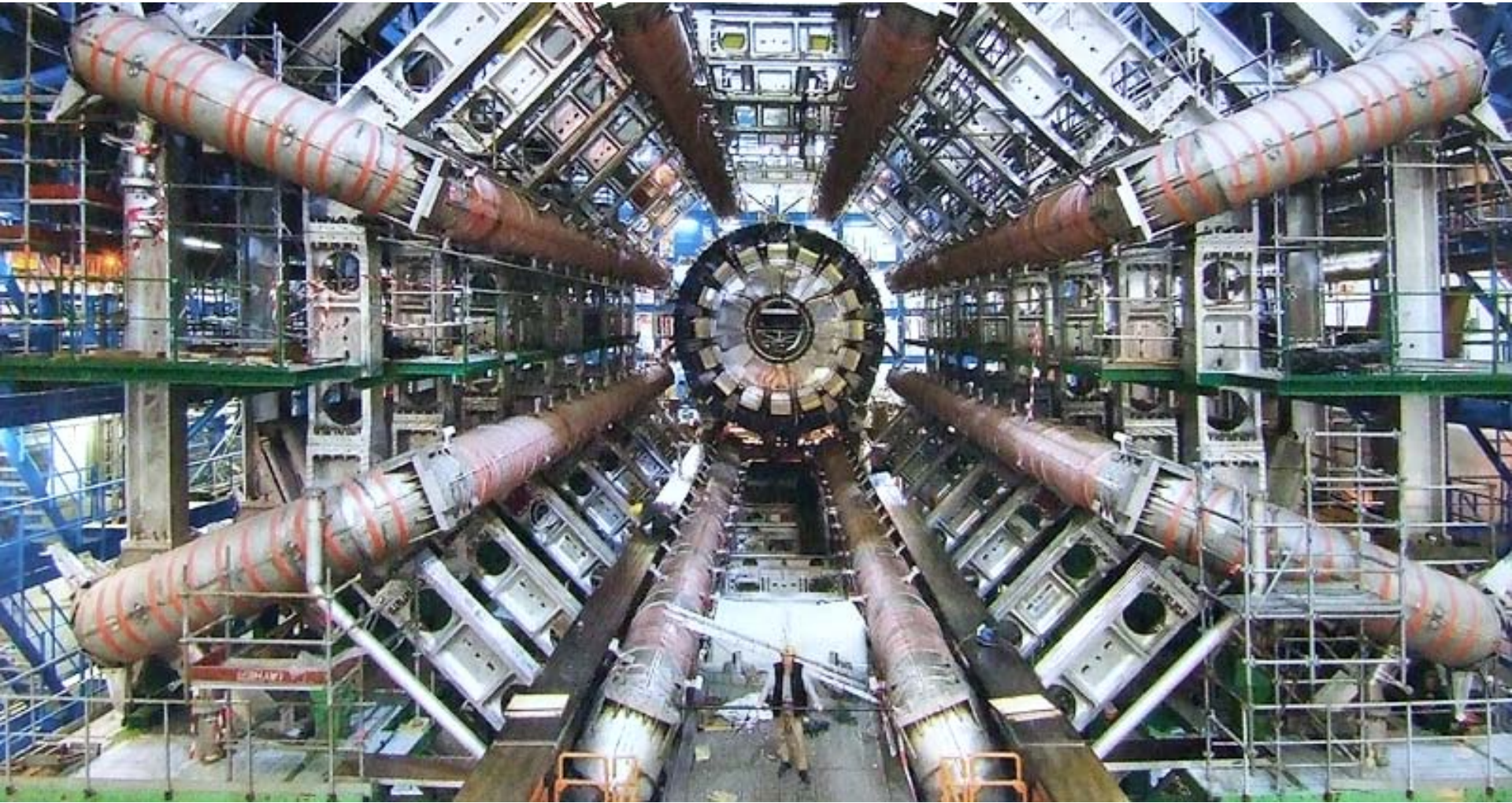
Research Culture in Psychology

- Psychology deals with unobservable phenomena like knowledge, abilities, attitudes, and personality traits. The attempt to properly define and quantify such phenomena is extremely difficult.
 - Measurement theory (Roberts, 1979)
- Psychology is an empirical science. Psychologists collect, analyze and report data. Examples:
 - Reaction times
 - fMRI scans
 - Galvanic skin responses
 - Judgments
 - Essays
 - ...

Research Culture in psychology and other fields

- “**Psychology** has its own unique data culture. Because our data often relate to individual human research participants, we spend a great deal of effort to protect the **confidentiality** and **privacy** of those participants. This is one of the reasons why psychologists are reluctant to place their data in publicly accessible archives or to share their data with anyone other than other psychologists who are working in closely related areas.” (Breckler, 2009)
- In the **physical sciences** data are collected with shared instrumentation involving large groups of investigators. The data are shared among many scientists.

Large Hadron Collider at CERN



Research Culture in psychology and other fields

- In the **social sciences**, such as economics, sociology or political science, coordinated data collection efforts are common. Scientists participate in the design of those activities, and base much of their own research on them. The data are often archived in publicly accessible facilities and available for anyone to scrutinize.
- For example, SOcio-Economic Panel (SOEP)

Data Sharing Policies in Psychology

- American Psychological Association Ethical Principles (2003).
After research results are published, psychologists do not withhold the data on which their conclusions are based from other competent professionals who seek to verify the substantive claims through reanalysis. ([8.14](#))
- Journal policies (example)
When a manuscript is accepted for publication, the author will be asked to provide assurance that the data in unanalyzed form and the exact details of the procedures will be available to other investigators for at least five years after publication. (Fienberg et al, 1985)

Need for more data sharing in Psychology

➤ “Psychologists need to rethink their reluctance to share data. Their discipline is 'softer' than some others: rarely do data on issues such as playground bullying or the usefulness of psychotherapy reveal really clear-cut answers. This makes the rigor with which the data are handled fundamental to research outcomes — and increases the desirability of having them open to examination by peers.” (Nature Editorial, 2006)

➤ Wicherts et al. (University of Amsterdam) selected the November and December 2004 issues of four top journals published by the American Psychological Association, which requires its authors to agree to share their data with other researchers after publication. In June 2005, Wicherts wrote to each corresponding author requesting data for simple reanalysis. Six months and several hundred e-mails later, he abandoned the mission, having received only a quarter of the data sets. (Wicherts et al, 2006)

Benefits of data sharing in Psychology

1. Avoid unnecessary duplication of data collection
 1. Save time and money of respondents and of researchers
2. Reanalysis: Verification (same problem, same data)
3. Secondary analysis (different problem, same data)
4. Meta-analysis (same problem, several independent data sets)
5. Refinement (alternative analyses)
6. Testing the generality of research findings
7. Create new enlarged data bases
 1. Increase the amount of data available on any scientific question
8. Applying new theories to existing data

Benefits of data sharing in Psychology

9. Provision of resources for training
 1. The reanalysis of previously collected data is one of the best ways of teaching research methodology.
 2. Secondary data are models for collecting one's own data (Sobal, 1982)
10. Monitor historical changes
11. Protection against faulty data
12. Sharing research data is associated with increased citation rate (Piwowar, Day & Fridsma, 2007)
13. Make data sets citable as scholarly publications; establish citation standard (→ „Publication and Citation of Scientific Primary Data“; [project](#) funded by the German Research Foundation)

PsychData is an institutional archive of primary research data in Psychology developed and maintained by ZPID and funded by the German Research Foundation.

Goals

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

1. Acquisition

Active solicitation required, hardly any volunteer donors.

Selection criteria

- 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

2. Documentation

Facilitate future discovery, access, and use. Retain the meaning of the data over time; allow different research groups to analyze the data.

- **Metadata**

Metadata (data about data) is descriptive information about an object or resource whether it be physical or electronic. It constitutes the information that enables the effective, efficient, and accurate use of those resources. It prevents misuse, misinterpretation, or confusion.

- **Codebook**

A document which lists the variables in a dataset, possible values for each variable, and the definitions of codes that have been assigned to these values.

3. Preservation

Long-term preservation of electronic materials.

Stable and reliable archival system for primary data in Psychology:

- Use open standards in storing the data instead of proprietary file formats and software applications.
- Migration and emulation from one generation to the next.
- Security (protecting against piracy; preservation against catastrophe; preservation of integrity - what is the authoritative data set?).

4. Access (concern: privacy of research subjects)

- Metadata: Freely accessible on the Web; OAI-PMH support in preparation.
- Primary data:
 - Protecting the rights and privacy of human subjects: Identifiers are removed from data.
 - Data sharing agreements are used to restrict the transfer of data to others and to require that data be used only for research purposes.
 - Distribution on CD-ROM (write once read many optical medium). Once written, the data cannot be altered.
- Technical assistance for requestors.

5. Tools

Tools to incorporate data sharing in the *initial design of a study*.

- Manual including rules how to write codebook
- Form for metadata entry (plausibility checks)

Major lesson learned

The documentation of data which were collected in the past (*retrospective documentation*) is much more expensive than using the PsychData documentation tools during the course of a study. Problems:

- Undocumented variables and values;
- Inconsistent handling of missing values;
- Personal identifiers not removed;
- Lost data (example: diskettes thrown out by beadle).

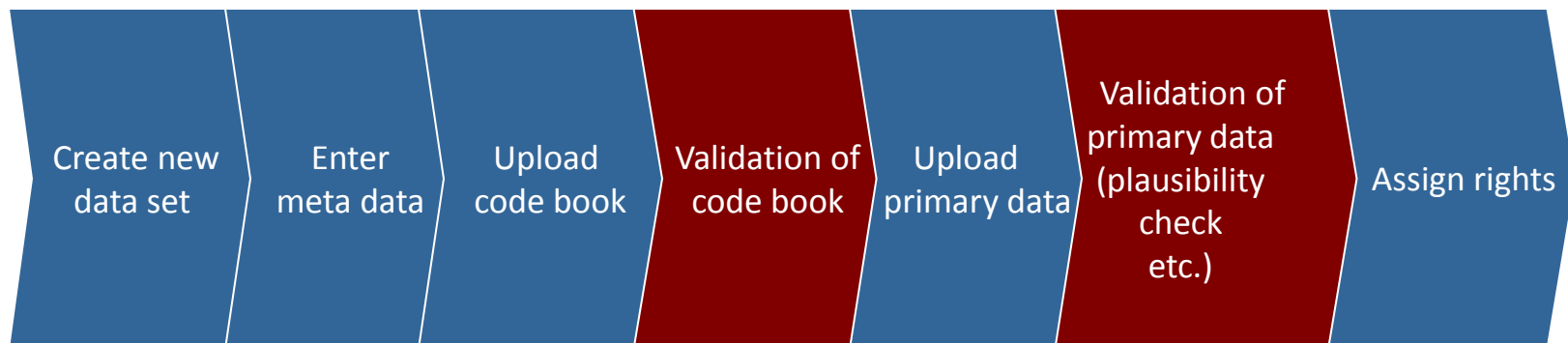
Time consuming checking and processing. Laborious interaction with original researcher(s) required.

6. Web interface

- Web-based data entry and upload, no ZPID staff needed
 - Beta-Version released August 2009, being tested by University of Graz (Austria)
 - Separate data spaces for institutions
 - Secure transmission (VeriSign Secure Site SSL Certificates, \$100,000 warranty)



Reuse of data



Exemplary Study

Kurz, D., Brinkhoff, K.-P., Tietjens, M. & Endrikat, K. (1995/2004). *Youth Sport Study*.

Relation of physical activity to psychosocial and personality development

- 3426 participants, 11-20 years old (52% female, 48% male)
- Questionnaire, 110 items, 647 variables

Reuse: Examples

Youth sport study (1995)

- Testing alternative hypothesis (Dissertation, University of Trier)
- Training (Statistics course, University of Bamberg)
- Training (Statistics course, University of Hamburg)
- Comparative analysis (Dissertation, University of Potsdam)
- Secondary analysis (Master thesis, University of Bielefeld)

Exemplary Study

Silbereisen, R. K., & Eyferth, K. (1982/2004). *Berlin Youth Longitudinal Study (BYLS) "Youth development and substance use"*.

Adolescent development and drug abuse, drug abuse and peer relations

- Seven waves from 1982 to 1988
- 1434 participants, 9-17 years old (50% female, 50% male)
- Questionnaire, 379 variables

Reuse: Examples

Berlin Youth Longitudinal Study : "Youth development and substance use" (1982).

- Training (Research methods, University of Jena)
- Reanalysis {International postdoctoral program with the participation of:

[Institute of Education](#), London, UK

[Center for Applied Developmental Science](#), Jena, Germany

[Center for Human Growth and Development](#), Michigan, USA

[Max-Planck-Institut für Bildungsforschung](#), Berlin, Germany

[Collegium for Advanced Studies](#), Jyväskylä, Finland

[University of Stockholm](#), Stockholm, Sweden}

Exemplary Study

Tesch-Römer, C., Motel-Klingebiel, A., & von Kondratowitz, H.-J. (2001/2009). *Old Age and Autonomy: The Role of Service Systems and Intergenerational Family Solidarity. Primary data of the European study OASIS.*

The relation between intergenerational family help and welfare state support.

- Data collected 2001
- 6106 participants from England, Germany, Israel, Norway and Spain
- Questionnaire, 795 variables

Reuse: Examples

Old Age and Autonomy: The Role of Service Systems and Intergenerational Family Solidarity. Primary data of the European study OASIS.

- Secondary analysis (Research project on the mobility of senior citizens funded by the Federal Ministry of Education and Research, University of Applied Sciences Bonn-Rhein-Sieg)
- Secondary analysis (Master thesis, University of Oldenburg)
- Refinement (Research project, University of Heidelberg)

Summary

Small initiative, unique in Psychology

Concerns

- Ethical principles: Confidentiality and privacy of human participants
- **Language issue (problems of translation): (Inter) National use**

Acceptance

- Usage is steadily growing
- Interdisciplinary usage: Education, Linguistics, Robotics, Sports Science
- Invitations to meetings and training courses for young researchers at the beginning of large projects
- Multiple uses of same data set: training, graduate work, research projects

PsychData contributes to a *gradual* cultural shift for psychology!

You need a long breath. Change happens slowly!

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.

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).
- 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

- DataCite Summer Meeting 2010 - [Making datasets visible and accessible](#). June 7-8th 2010, Hannover, Germany
- 22nd CODATA International Conference, 24-27th October 2010, Cape Town, South Africa

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, Peter Weiland

Member of



German Data Forum
(RatSWD)

Partly funded by the German Research
Foundation

Deutsche
Forschungsgemeinschaft

DFG