Influencing Data Culture in Psychology by Providing Data Management Support

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

Université d'été du GFII – INIST-CNRS – Nancy – 12 Sept 2013
Outline

• Research Culture in Psychology (and other fields)
• Benefits of data sharing
• Data sharing policies
• PsychData
• Lessons learned: Data management tools
• Discovery (DOI, PubPsych)
  ○ Bundle research data with the final paper
Data in Psychology

• Psychology is an **empirical science**. Psychologists collect, analyze and report **data**. Examples: Reaction times, fMRI scans, galvanic skin responses, judgments, essays, etc.

• Psychology deals with **unobservable phenomena** like knowledge, abilities, attitudes, and personality traits. The attempt to properly define and quantify such phenomena is extremely difficult; frequently there are no standardized instruments.

• The data generated in psychology can be very complex and there are **high costs** associated with their proper documentation.

• Data often relate to individual human research participants; the **confidentiality** and **privacy** of those participants needs to be protected.
Research Culture in Psychology versus Other Fields

“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)

Psychology has its own unique data culture:

1) In the physical sciences data are collected with shared instrumentation involving large groups of investigators. The data are shared among many scientists. (For example, high energy physics)

2) 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)
Research in the physical sciences: Grand collisionneur de hadrons, CERN

gravitycontrol.org
Research in psychology: Conditionnement classique

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 and other fields
(Weichselgartner, 2008)

1. Avoid unnecessary duplication of data collection
   • 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
   • Increase the amount of data available on any scientific question
8. Applying new theories to existing data
Benefits of data sharing in psychology and other fields

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 and research fraud

12. Sharing research data is associated with increased citation rate (Piwowar, Day & Fridsma, 2007)

Make data sets citable as scholarly publications; establish citation standard: *DataCite*, INIST member since 2010
There is a worldwide trend towards making use of the benefits of data sharing. Examples of policies:

German Research Foundation (1998)
*Primary data as the basis for publications shall be securely stored for ten years in a durable form in the institution of their origin.*

National Institutes of Health, USA (2003)
*Starting with the October 1, 2003 receipt date, investigators submitting an NIH application ... are expected to include a plan for data sharing.*

Wellcome Trust, UK (2010)
*All our funded researchers should maximise access to their research data with as few restrictions as possible.*

Publishers
*Many scientific journals require that authors make available the data included in their publications.*

More policies can be found at: [Site d'information sur les Données de la Recherche](#)
Change data culture in psychology: Provide infrastructure.

PsychData ([http://www.psychdata.de/](http://www.psychdata.de/)) is an institutional archive of primary research data in Psychology developed and maintained by ZPID and partially funded by the German Research Foundation.

Volume

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

Data reuse

- Possible since June 2004
- ~ 10 requests per year
Major lessons learned

1. Acquisition
The willingness to share data is very low, active solicitation required, hardly any volunteer donors!

<table>
<thead>
<tr>
<th>Year</th>
<th>Contacted</th>
<th>Received</th>
<th>%</th>
</tr>
</thead>
<tbody>
<tr>
<td>2001</td>
<td>811</td>
<td>18</td>
<td>2,2</td>
</tr>
<tr>
<td>2003</td>
<td>81</td>
<td>8</td>
<td>9,9</td>
</tr>
<tr>
<td>2009</td>
<td>11</td>
<td>1</td>
<td>9,1</td>
</tr>
</tbody>
</table>

2. Documentation
The documentation of data which were collected in the past (retrospective documentation) is extremely expensive. 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.
Develop **tools** to incorporate data sharing in the *initial design* of a study:

- **Documentation Manual** including rules how to write codebook
- **Interactive form** for metadata entry (plausibility checks)
- Direct deposit via **web interface**
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)
Science 2.0 – Networking scholarly materials – Bundle research data with the final paper

PubPsych is a free information retrieval system for psychological resources. It offers a comprehensive and balanced selection of resources from a growing number of international databases with a European focus, covering the needs of academic and professional psychologists. PubPsych includes 827,497 records (Sep 2013) and offers, where available, full-text linking, links to additional information and link resolving. ... PubPsych contains psychology relevant records from PASCAL.
PsychData contributes to a *gradual* cultural shift for psychology!

**Concerns**
- Ethical principles: Confidentiality and privacy of human participants
- Complexity of psychological data
- 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
Merci à tous pour votre attention.

http://www.psychdata.de/  weichselgartner@zpid.de

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

Member of

Partly funded by the German Research Foundation

Deutsche Forschungsgemeinschaft

Registrierungsagentur für Sozial- und Wirtschaftsdaten
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


