PsychData – Experiences from 12 Years of Research Data Archiving

Peter Weiland, Ina Dehnhard
Leibniz Institute for Psychology Information (ZPID)
Trier, Germany
Who we are and what we’re doing

ZPID Services:
• **PSYNDEX**
  Reference Database for Psychology
• **PubPsych**
  European Search Portal for Psychology (includes about 11,000 records from NORART)
• **PsychAuthors**
  Database of psychologists from the German-speaking countries
• **PsychData**

ZPID Research:
• development of information technology systems,
• Information behaviour and information literacy
• scientometrics and the evaluation of science.

http://mapq.st/1kKeUJJ
Benefits of Research Data Archiving

- Possibility of Meta-Analyses
- Possibility of Re-Analyses
- Cost reduction
- Datasets could be used as a teaching tool
- Replication studies
- Fraud detection
- ... and many more

https://www.flickr.com/photos/127519682@N02/16189100439/
Obstacles for Research Data Sharing

- Time and effort for research data management
- Little or no recognition for documenting and archiving datasets
- Competition between researchers
- Detection of weaknesses in statistical analyses and datasets
The Situation in Psychology

- Small studies and small projects
- A bunch of different measures, non-standardized
- Privacy concerns
Data Request Response Rates

- **PsychData Project (Dehnhard et al, 2013):**
  - Between 2003 and 2011, we sent out 2302 information letters (mail or electronic mail) with a data request
    → 21 positive requests and 18 data deposits (0.8%)
  - Personal requests sent to 97 researchers between 2003 and 2010
    → 9 data deposits (9.3%)

- **Botella and Ortego (2010):**
  contacted the authors of 109 studies for a meta-analysis
  → got data from 13 studies (12%)

- **Wicherts et al (2006):**
  - 11% of contacted authors provided datasets
  - 27% after one reminder

https://www.flickr.com/photos/nathangibbs/98592171/
PsychData – About the project

An archive of primary research data in psychology

• Acquisition
• Documentation
• Preservation (long-term archiving)
• Access (distribution)
• Direct research support
**Timeline of PsychData**

- **2002 – 2015**

  - **Project Start**
    - Data Acquisition
    - Development of Metadata Schema
    - Raw data
  
  - **2004 – 2005**
    - PsychData was accredited by the German Data Forum (RatSWD)
  
  - **2005 – 2006**
    - integrating the documentation process in the research process (→ MyPsychData)
  
  - **2009 – 2010**
    - Wicherts: Only 27% of contacted authors made data available upon request
  
  - **2009**
    - „Stapel Affair“
  
  - **2014 – 2015**
    - OSC-Study: „Estimating the reproducibility of psychological science“
    - 127 datasets (44 studies) with 32,846,682 data points
    - DOI assignment (via da|ra, DataCite)

- **Database goes online**
  - Primary data

- **All National Institutes of Health funded research (> $500K) must have a plan to address the sharing and archiving of data.**
PsychData-Workflow

Data donator

- Research data
- Study description
- Documentation

PsychData

Standardized documentation

Data validation

Final revision

Long-term preservation

Data dissemination

PsychData
Example for Metadata in PsychData

Research Method Description
Questionnaire Data
Classification of Data Collection
Questionnaire Data: Fully Standardized Survey Instrument; Experimental Data: Repeated Measures, Laboratory Experiment
Research Instrument
Information forthcoming
Data Collection Method
Data collection in the presence of an interviewer - Individual Administration
Sample
105 subjects (52 females, 53 males; Mage = 76.15, SD=10.91, range 60-91 years)
Subject Recruitment
Personal contact to patients and subjects through psychological and psychiatric offices over the course of 20 years. Great pains were taken to include both sexes, various diagnostic groups, and age groups. The number of subjects was limited by the size of the study. Participation was voluntary, following the principles of the Declaration of Helsinki. Participants were provided with a detailed information sheet about the study. All participants agreed to participate in the study. All procedures were approved by the local ethics committee.

Data Collection Completion Date
2005
Data Collection Publication
2015
Publication
http://dx.doi.org/10.5160/psychdata.wtfz05lo22

Data Status
Current Data Set
Original Metadata
Questionnaire Data that can be either the subject or the instrument containing closed and/or open-ended questions
Transformation
Data from the subjects was scaled and then mathematically transformed into a machine-readable form.

Data Points
165+168 data points

Characteristics
Population
Experimental Period
Sample
Selection of 250 consecutive inpatients admitted to the Psychiatric and Psychotherapeutic Department of the University Hospital, Munich, Germany

Subject Recruitment
Personal contact to inpatients and subjects through psychological and psychiatric offices over the course of 20 years. Great pains were taken to include both sexes, various diagnostic groups, and age groups. The number of subjects was limited by the size of the study. Participation was voluntary, following the principles of the Declaration of Helsinki. Participants were provided with a detailed information sheet about the study. All participants agreed to participate in the study. All procedures were approved by the local ethics committee.

Sample Size
205 (90 females, 115 males; Mage = 58.44, SD=15.03, range 19-90 years)

Published Papers

citation

Used Test Methods

References

http://dx.doi.org/10.5160/psychdata.wtfz05lo22
The Problem of Domain Specific Knowledge in Psychology I

<table>
<thead>
<tr>
<th>Tag</th>
<th>MIN_TA200min</th>
<th>MAX_TA200max</th>
<th>AVG_TA200</th>
</tr>
</thead>
<tbody>
<tr>
<td>01.10.15</td>
<td>2,5</td>
<td>16,6</td>
<td>9,1</td>
</tr>
<tr>
<td>02.10.15</td>
<td>2,2</td>
<td>18,5</td>
<td>9,63</td>
</tr>
<tr>
<td>03.10.15</td>
<td>7,2</td>
<td>19,4</td>
<td>11,95</td>
</tr>
<tr>
<td>04.10.15</td>
<td>7,1</td>
<td>16,3</td>
<td>11,9</td>
</tr>
<tr>
<td>05.10.15</td>
<td>6,9</td>
<td>17,5</td>
<td>12,34</td>
</tr>
<tr>
<td>06.10.15</td>
<td>14,9</td>
<td>18,8</td>
<td>16,74</td>
</tr>
<tr>
<td>07.10.15</td>
<td>11,6</td>
<td>16,4</td>
<td>13,65</td>
</tr>
<tr>
<td>08.10.15</td>
<td>9,8</td>
<td>15,2</td>
<td>12,16</td>
</tr>
<tr>
<td>09.10.15</td>
<td>7,8</td>
<td>13,6</td>
<td>10,76</td>
</tr>
<tr>
<td>10.10.15</td>
<td>7,9</td>
<td>14,8</td>
<td>11,06</td>
</tr>
<tr>
<td>11.10.15</td>
<td>2,8</td>
<td>14,4</td>
<td>8,12</td>
</tr>
<tr>
<td>12.10.15</td>
<td>0,6</td>
<td>11,8</td>
<td>5,39</td>
</tr>
<tr>
<td>13.10.15</td>
<td>-1,5</td>
<td>5,7</td>
<td>2,71</td>
</tr>
<tr>
<td>14.10.15</td>
<td>1,9</td>
<td>4,4</td>
<td>3,29</td>
</tr>
<tr>
<td>15.10.15</td>
<td>2,7</td>
<td>5,7</td>
<td>4,15</td>
</tr>
<tr>
<td>16.10.15</td>
<td>3,9</td>
<td>6,2</td>
<td>5</td>
</tr>
<tr>
<td>17.10.15</td>
<td>4,1</td>
<td>6,9</td>
<td>5,43</td>
</tr>
<tr>
<td>18.10.15</td>
<td>2,2</td>
<td>9,6</td>
<td>5,88</td>
</tr>
<tr>
<td>19.10.15</td>
<td>3,4</td>
<td>9,7</td>
<td>6,65</td>
</tr>
<tr>
<td>20.10.15</td>
<td>6,9</td>
<td>10,6</td>
<td>8,66</td>
</tr>
<tr>
<td>21.10.15</td>
<td>6,7</td>
<td>12,2</td>
<td>8,98</td>
</tr>
<tr>
<td>22.10.15</td>
<td>8,5</td>
<td>11,2</td>
<td>9,9</td>
</tr>
<tr>
<td>23.10.15</td>
<td>9,1</td>
<td>14,3</td>
<td>11,19</td>
</tr>
<tr>
<td>24.10.15</td>
<td>8,8</td>
<td>15</td>
<td>11,19</td>
</tr>
<tr>
<td>25.10.15</td>
<td>7,6</td>
<td>14,2</td>
<td>11,55</td>
</tr>
<tr>
<td>26.10.15</td>
<td>5,2</td>
<td>15,6</td>
<td>9,23</td>
</tr>
<tr>
<td>27.10.15</td>
<td>4,8</td>
<td>9,9</td>
<td>6,76</td>
</tr>
<tr>
<td>28.10.15</td>
<td>6,6</td>
<td>11,7</td>
<td>8,89</td>
</tr>
<tr>
<td>29.10.15</td>
<td>4,3</td>
<td>12,5</td>
<td>9,46</td>
</tr>
<tr>
<td>30.10.15</td>
<td>3,1</td>
<td>13,1</td>
<td>7,63</td>
</tr>
<tr>
<td>31.10.15</td>
<td>2,1</td>
<td>11,1</td>
<td>6,94</td>
</tr>
</tbody>
</table>

Source:
http://www.am.rlp.de/Internet/AM/NotesAM.nsf/amweb/ff997cae2d6ac650c1257171002e8a47?OpenDocument&TableRow=2.1.0%2C2.8#2.1.
The Problem of Domain Specific Knowledge in Psychology II

Source:

Original Records
Questionnaire filled out by either the subject or the experimenter containing closed and/or open answers

Countries
Norway, England, Germany, Spain, Israel

Data Content
6106 subjects, 795 variables
DE3
Marital status (screening question)
"What is your(current) marital status?
Are you...?"
{1;2;3;4;5}
{6;7;8;9}
1 "Married“
2 "Unmarried partnership (living together)“
3 "Widowed“
4 "Divorced or separated“
5 "Never married"
6 "Missing value:Filtered“
7 "Missing value:Refused“
8 "Missing value:Do not know“
9 "Missing value"

<table>
<thead>
<tr>
<th>ID</th>
<th>DE3</th>
<th>HA03</th>
</tr>
</thead>
<tbody>
<tr>
<td>1000000171</td>
<td>2</td>
<td>3</td>
</tr>
<tr>
<td>1000000822</td>
<td>5</td>
<td>3</td>
</tr>
<tr>
<td>1000001003</td>
<td>2</td>
<td>1</td>
</tr>
<tr>
<td>1000001004</td>
<td>1</td>
<td>3</td>
</tr>
<tr>
<td>1000001005</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>1000001006</td>
<td>1</td>
<td>3</td>
</tr>
<tr>
<td>1000001007</td>
<td>4</td>
<td>3</td>
</tr>
<tr>
<td>1000001008</td>
<td>1</td>
<td>3</td>
</tr>
<tr>
<td>1000001009</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>1000001010</td>
<td>1</td>
<td>3</td>
</tr>
<tr>
<td>1000001011</td>
<td>2</td>
<td>3</td>
</tr>
<tr>
<td>1000001012</td>
<td>1</td>
<td>3</td>
</tr>
<tr>
<td>1000001013</td>
<td>1</td>
<td>4</td>
</tr>
<tr>
<td>1000001014</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>1000001015</td>
<td>4</td>
<td>3</td>
</tr>
<tr>
<td>1000001016</td>
<td>3</td>
<td>3</td>
</tr>
<tr>
<td>1000001017</td>
<td>3</td>
<td>3</td>
</tr>
<tr>
<td>1000001018</td>
<td>3</td>
<td>3</td>
</tr>
<tr>
<td>1000001019</td>
<td>3</td>
<td>3</td>
</tr>
<tr>
<td>1000001020</td>
<td>3</td>
<td>3</td>
</tr>
<tr>
<td>1000001021</td>
<td>1</td>
<td>4</td>
</tr>
<tr>
<td>1000001032</td>
<td>4</td>
<td>3</td>
</tr>
<tr>
<td>1000001033</td>
<td>2</td>
<td>2</td>
</tr>
<tr>
<td>1000001034</td>
<td>1</td>
<td>3</td>
</tr>
<tr>
<td>1000001035</td>
<td>1</td>
<td>4</td>
</tr>
<tr>
<td>1000001036</td>
<td>4</td>
<td>3</td>
</tr>
<tr>
<td>1000001037</td>
<td>1</td>
<td>4</td>
</tr>
<tr>
<td>1000001038</td>
<td>5</td>
<td>3</td>
</tr>
<tr>
<td>1000001039</td>
<td>1</td>
<td>3</td>
</tr>
<tr>
<td>1000001040</td>
<td>1</td>
<td>3</td>
</tr>
<tr>
<td>1000001041</td>
<td>1</td>
<td>4</td>
</tr>
</tbody>
</table>

HA03
SF 36 Health Survey instrument: Rate your health in general compared to one year ago
"Compared to one year ago, how would you rate your health in general now?"
{1;2;3;4;5}
{6;7;8;9}
1 "Much better“
2 "Somewhat better“
3 "About the same“
4 "Somewhat worse“
5 "Much worse"
6 "Missing value:Filtered“
7 "Missing value:Refused“
8 "Missing value:Do not know“
9 "Missing value"
Lessons Learned I

Retrospective Documentation is hard

- Undocumented variables and values
- Inconsistent handling of missing values
- Aggregated variables – formulae can not be reconstructed
- Personal Identifiers not removed
- Time consuming interaction with original researchers

https://www.flickr.com/photos/agrusoft/16357210343/
Lessons Learned II

Barriers for data sharing have to be low

→ Tools are needed that are integrated in the research process

https://www.flickr.com/photos/lostinfog/6317521866/
DataWiz

- Development of an Automated Assistant for the Management of Psychological Research Data
- Project started in October 2015
- Pilot study MyPsychData started in 2011 (39 users up to now)
- Project Goals:
  - Establishing domain-adequate research data management in psychology
  - Storing the research data sustainably
  - Providing incentives to share the data
- Can be implemented on-site or used as software as a service (SaaS) on a ZPID server
Takk, thank you!

For more information: http://www.psychdata.de/index_en.php
psychdata@zpid.de