# 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


http://mapq.st/1kKeUJJ

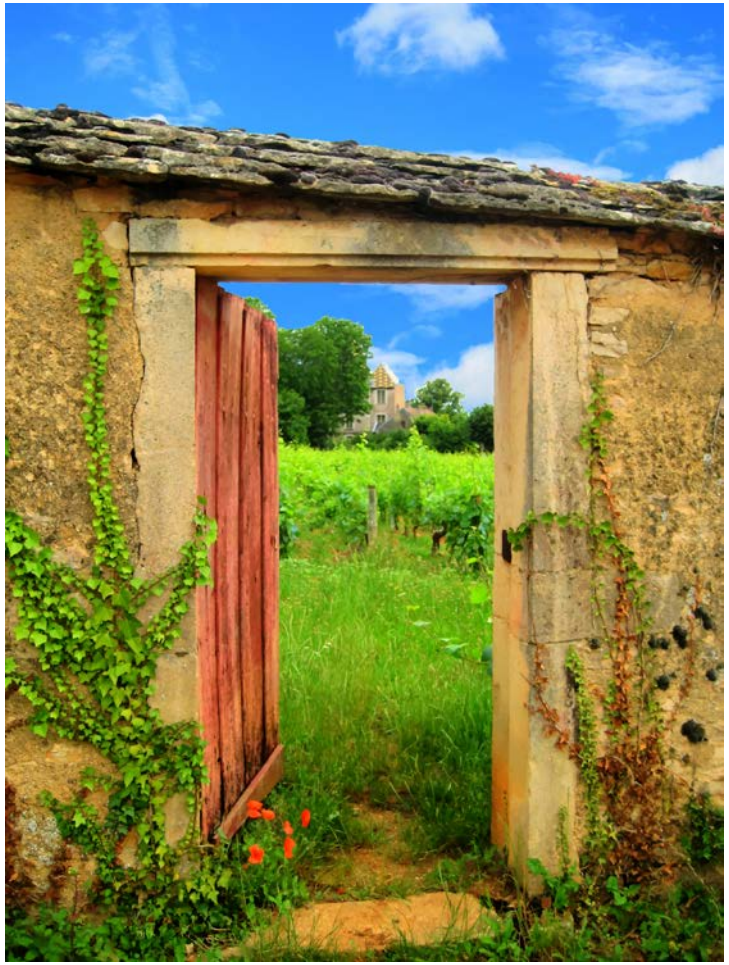
**ZPID Services:**

- *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.

# Benefits of Research Data Archiving



https://www.flickr.com/photos/127519682@N02/16189100439/

- 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

# 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

https://www.flickr.com/photos/ioelondon/6143379343/

# 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

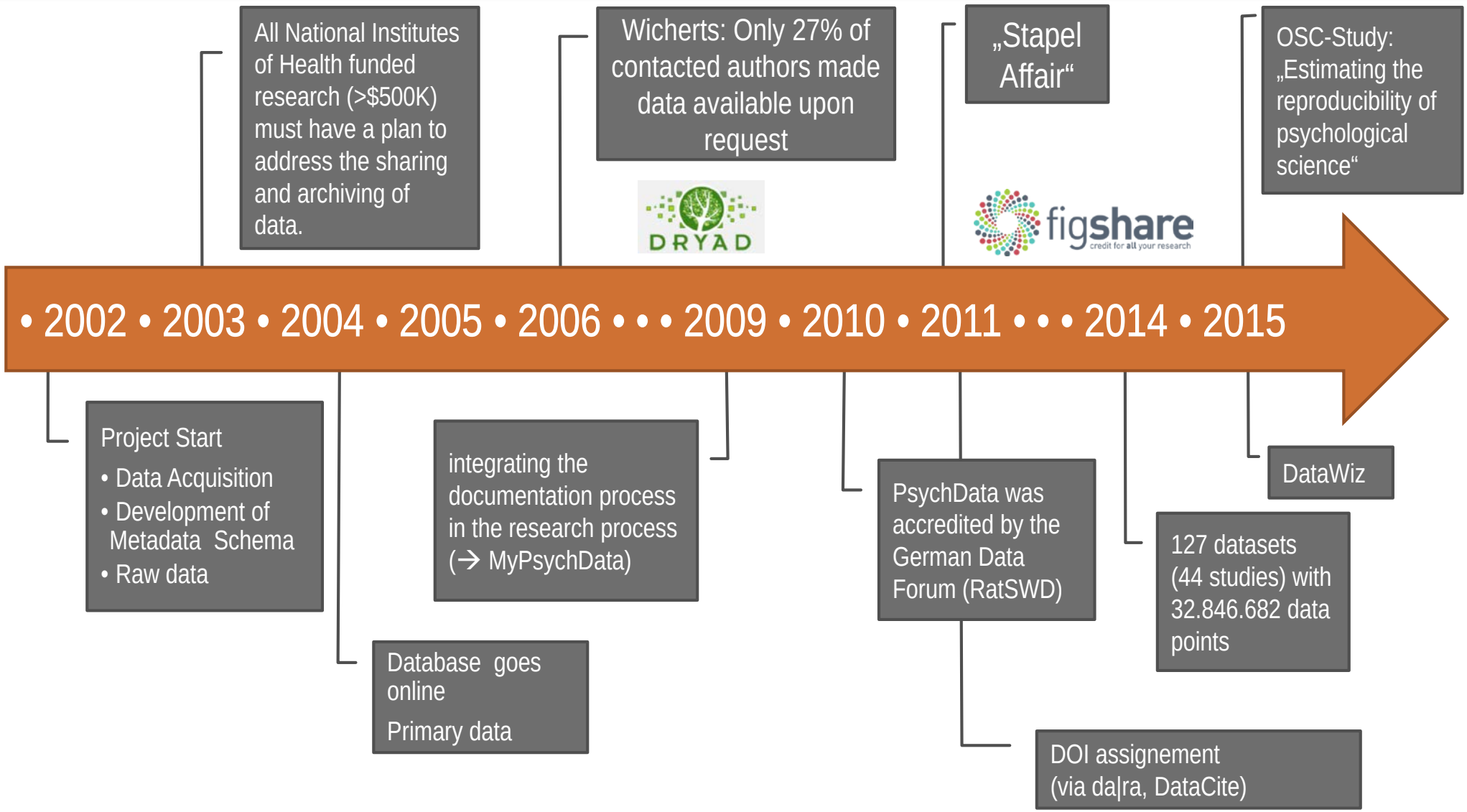
- **Documentation**

- **Preservation (long-term archiving)**

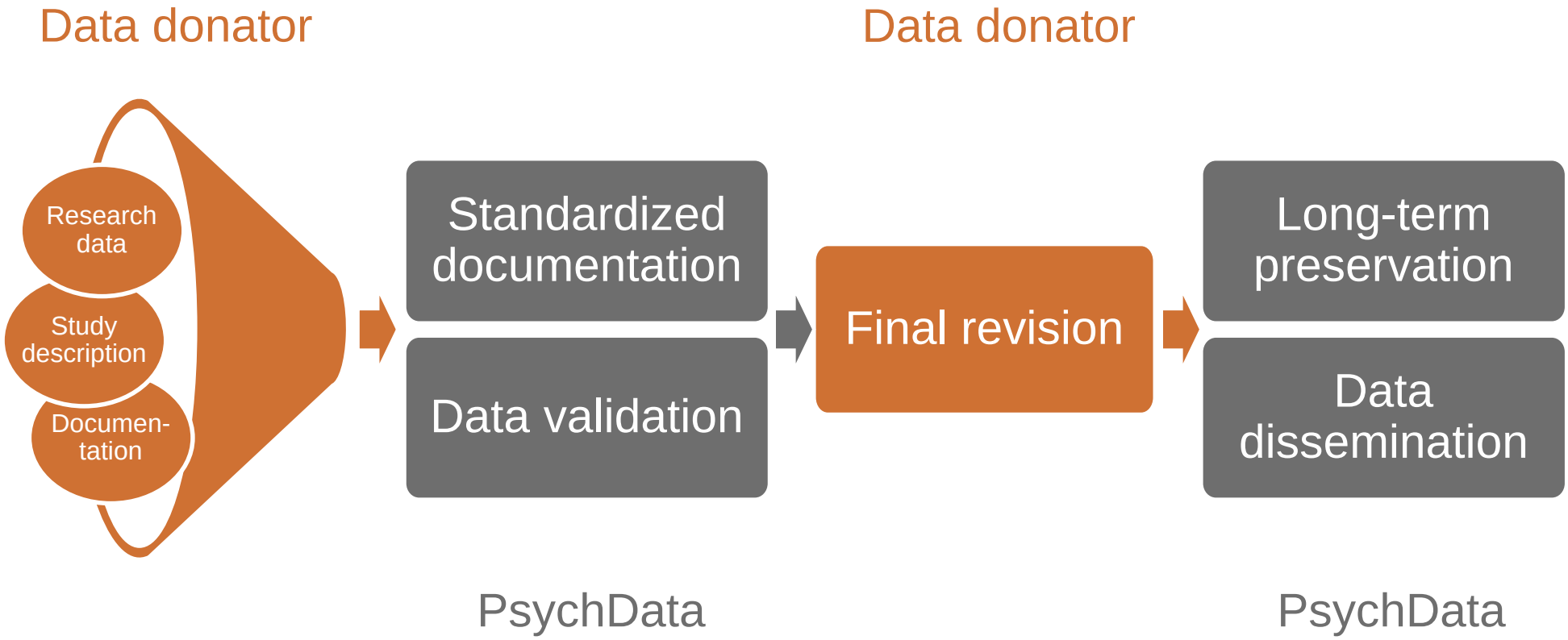
- **Access (distribution)**

- **Direct research support**

# Timeline of PsychData



# PsychData-Workflow



# Example for Metadata in PsychData

Psychologie Information
ZPID - Leibniz Institut
Mitglied der Leibniz-Gemeinschaft

**Scientific data of the Munich Longitudinal Study on the Genesis of Individual Competencies (LOGIC): Development of intelligence and thinking**

Print View

Index



**Researchers**

| Name |
|---|
| Weinert, Franz |
| Schneider, Wolfgang |
| Stefanek, Jan |

Top

**Dataset Information**

| | |
|---|---|
| **Title** | Scientific data of the Munich Longitudinal Study on the Genesis of Individual Competencies (LOGIC): Development of intelligence and thinking |
| **Title (German)** | Forschungsdaten der Münchner Longitudinalstudie zur Genese individueller Kompetenzen (LOGIK): Entwicklung der Intelligenz und des logischen Denkvermögens |
| **Citation** | Weinert, F., Schneider, W., & Stefanek, J. (2015). Scientific data of the Munich Longitudinal Study on the Genesis of Individual Competencies (LOGIC): Development of intelligence and thinking [Translated Title] (Version 1) [Files on CD-ROM]. Trier: Center for Research Data in Psychology: PsychData of the Leibniz Institute for Psychology Information ZPID. http://dx.doi.org/10.5160/psychdata.wtfz05lo22 |
| **Responsible for Data Collection** | Max-Planck-Institut für psychologische Forschung; Weinert, Franz; Schneider, Wolfgang |
| **Data Collection Completion Date** | 2005 |
| **Dataset Publication** | 2015 |
| **Dataset ID** | wtfz05lo22 |
| **Study Description** | The Munich "Longitudinal Study of the Genesis of Individual Competencies" (LOGIC) is a comprehensive examination of the differential description of developmental trajectories of cognitive skills and personality characteristics. It also describes individual differences in development due to the influence of varying school and classroom conditions. The changing state of the development of intelligence, |

**Research Method Description**

| | |
|---|---|
| **Research Method Description** | Questionnaire Data |
| **Classification of Data Collection** | Questionnaire Data: Fully Standardized Survey Instrument; Experimental Data: Repeated measures design, Laboratory Experiment |
| **Research Instrument** | information forthcoming |
| **Data Collection Method** | Data collection in the presence of an experimenter<br>- Individual Administration<br>- Paper and Pencil<br>- Reaction time<br>- Computer-supported |
| **Time Points** | repeated measurements |
| **Survey Time Period** | 1st wave: 1984 - 1985<br>2nd wave: 1985 - 1986<br>3rd wave: 1986 - 1987<br>4th wave: 1987 - 1988<br>5th wave: 1988 - 1989<br>6th wave: 1989 - 1990<br>7th wave: 1990 - 1991<br>8th wave: 1991 - 1992<br>9th wave: 1992 - 1993<br>10th wave: 1997 - 1998<br>11th wave: 2003 - 2005 |
| **Characteristics** | - |
| **Population** | Children |
| **Experimental Pool** | Individuals |
| **Sample** | Selection of 20 kindergartens in Greater Munich and in the Fürstenfeldbruck area (near Munich), that corresponded to socio-economic criteria of (West) Germany's general population in the year 1984. After information sessions at each of the facilities, 205 children were recruited. After the first assessment year (wave 1) approx. another 20 children were recruited. |
| **Subject Recruitment** | Personal contact to parents and subjects through psychological-technical assistants over the course of 20 years. Small gifts on special occasions. Yearly Christmas and birthday greetings with individually selected post cards. Individual aptitude tests for schooling and career paths as well as general counselling offered and performed by academic staff. Publication of a newsletter for the sample in the early assessment waves. Festive completion event in 1993 with gifts. Copy of the book "Development in Childhood" sent free of charge. During waves 10 and 11, subject compensation and offer of comparative performance evaluation. |
| **Sample Size** | 205 individuals (wave 1) |
| **Return/Drop Out** | After a drop out of 13 children in wave 2, a further 25 children were recruited, resulting in the following sample sizes for the further waves: 217 in wave 2; 213 in wave 3; 204 in wave 4; 200 in wave 5; 195 in wave 6; 194 in wave 7; 189 in wave 8; 186 in wave 9; 176 in wave 10 and still 153 subjects in wave 11.<br>The response rate was 74.6% in wave 11. The data on sample sizes refer to the entire LOGIC study. The number of subjects may vary between each of the single assessments. |
| **Gender Distribution** | 1st wave: 49% female subjects (n=100); 51% male subjects (n=105)<br>2nd wave: 48% female subjects (n=104); 52% male subjects (n=113)<br>3rd wave: 48% female subjects (n=102); 52% male subjects (n=111)<br>4th wave: 49% female subjects (n=98); 51% male subjects (n=105)<br>5th wave: 48% female subjects (n=96); 52% male subjects (n=104)<br>6th wave: 48% female subjects (n=93); 52% male subjects (n=100)<br>7th wave: 48% female subjects (n=93); 52% male subjects (n=101)<br>8th wave: 47% female subjects (n=89); 53% male subjects (n=100)<br>9th wave: 47% female subjects (n=87); 53% male |

**Data Status**

| | |
|---|---|
| **Data Status** | Complete Data Set |
| **Original Records** | Questionnaire filled out by either the subject or the experimenter containing closed and/or open answers, person-related data files |
| **Transformation** | Data from the subjects were coded and then immediately transferred into a machine-readable form. |

Top

**Description of the Provided Data**

| | |
|---|---|
| **Description** | Research data of the Hannover-Wechsler Intelligence Scale for Preschool Children (HAWIVA), 1st wave, 2nd measurement point |
| **File Name** | wtfz05lo22_w1_fd1.txt |
| **Data Content** | 185 subjects, 9 variables |
| **Data Points** | 185*9=1665 data points |
| **Variables** | - |
| **MD5 Hash** | e81911b11361ccbf130d9ae11b384292 |

| | |
|---|---|
| **Description** | Research data of Columbia Mental Maturity Scale (CMMS), 1st wave, 2nd measurement point |
| **File Name** | wtfz05lo22_w1_fd9.txt |
| **Data Content** | 185 subjects, 6 variables |
| **Data Points** | 185*6=1110 data points |
| **Variables** | subject ID (1), total scale raw scores (1), age level raw score (1), age level percentile rank (1), age deviation score (1), age level stanine (1) |
| **MD5 Hash** | 74ff0a942a010c8eef788f7d27bdba5b |

| | |
|---|---|
| **Description** | Research data of Columbia Mental Maturity Scale (CMMS), 3rd wave, 2nd measurement point |
| **File Name** | wtfz05lo22_w3_fd10.txt |
| **Data Content** | 211 subjects, 8 variables |
| **Data Points** | 211*8=1688 data points |
| **Variables** | subject ID (1), experimenter ID (1), age level raw score (1), age level percentile rank (1), age deviation score (1), age level stanine (1), total scale raw scores (1), 66 items scale raw score (1) |
| **MD5 Hash** | 54d43f2718682b2b836a752fb2c0f802 |

| | |
|---|---|
| **Description** | Research data of Columbia Mental Maturity Scale (CMMS), 4th wave, 1st measurement point |
| **File Name** | wtfz05lo22_w5_fd11.txt |
| **Data Content** | 194 subjects, 7 variables |
| **Data Points** | 194*7=1358 data points |
| **Variables** | subject ID (1), experimenter ID (1), age level raw score (1), age level percentile rank (1), age level stanine (1), age deviation score (1), total scale raw scores (1) |
| **MD5 Hash** | 769e15ef28296b53f7aa8162e29a92a6 |

| | |
|---|---|
| **Description** | Research data of Culture Fair Intelligence Test (CFT), 10th wave, 2nd measurement point |
| **File Name** | wtfz05lo22_w10_fd14.txt |
| **Data Content** | 174 subjects, 101 variables |
| **Data Points** | 174*101=17574 data points |
| **Variables** | information forthcoming |
| **MD5 Hash** | f64fb93cdab895bbafc3a6728862a5e1 |

**Publications Directly Related to the Dataset**


| Publications Directly Related to the Dataset |
|---|
| Kuhl, J., & Kraska, K. (1993). Self-regulation: Psychometric properties of a computer-aided instrument. German Journal of Psychology, 17 (1), 11-24. |
| Schneider, W. (1988). Identifying reciprocal causal effects in developmental patterns: An example from the Munich Longitudinal Study on the Genesis of Individual Competencies (LOGIC). Paper Nr. 3/1988. München: Max-Planck-Institut für Psychologische Forschung. |
| Schneider, W. (1990). Intelligenzentwicklung zwischen dem 4. und 8. Lebensjahr. In: W. Schneider, M. Knopf, E. Stern, A. Helmke, & J. Asendorpf (Hrsg.), Die Entwicklung kognitiver, motivationaler und sozialer Kompetenzen zwischen dem 4. und 8. Lebensjahr. München: Max-Planck-Institut für Psychologische Forschung. |
| Schneider, W. (2007). Entwicklung der Intelligenz im Kindesalter. In M. Hasselhorn & W. Schneider (Hrsg.), Handbuch der Entwicklungspsychologie (S. 277-288). Göttingen: Hogrefe. |
| Schneider, W. (2008). Entwicklung der Intelligenz und des Denkvermögens in Kindheit, Jugend und Erwachsenenalter. In: W. Schneider (Ed.), Entwicklung von der Kindheit bis zum Erwachsenenalter. Befunde der Münchner Längsschnittstudie LOGIK (S. 43-66). Weinheim: Beltz Psychologie Verlags Union. |
| Schneider, W., Bullock, M., & Sodian, B. (1998). Die Entwicklung des Denkens und der Intelligenzunterschiede zwischen Kindern. In: F.E. Weinert (Hrsg.), Entwicklung im Kindesalter (S. 53-74). Weinheim: Psychologie Verlags Union. |
| Schneider, W., Niklas, F., & Schmiedeler, S. (2014). Intellectual development from early childhood to early adulthood: The impact of early IQ differences on stability and change over time. Learning and Individual Differences, 32, 156-162. |
| Schneider, W., Perne, J., Bullock, M., Stefanek, J., & Ziegler, A. (1999). Development of intelligence and thinking. In: F.E. Weinert & W. Schneider (Eds.), Individual development from 3 to 12. Findings from the Munich Longitudinal Study (pp 9-28). Cambridge: Cambridge University Press. |
| Schneider, W., & Stefanek, J. (2004). Entwicklungsveränderungen allgemeiner kognitiver Fähigkeiten und schulbezogener Fertigkeiten im Kindes- und Jugendalter. Evidenz für einen Scherneffekt? Zeitschrift für Entwicklungspsychologie und Pädagogische Psychologie, 36 (3), 147-159. DOI: 10.1026/0049-8637.36.3.147 |
| Schneider, W., Stefanek, J., & Niklas, F. (2009). Development of intelligence and thinking. In: W. Schneider & M. Bullock (Eds.), Human development from early childhood to early adulthood. Findings from a 20 year longitudinal study (pp 7-33). New York: Psychology Press. |


Top

**Utilized Test Methods**

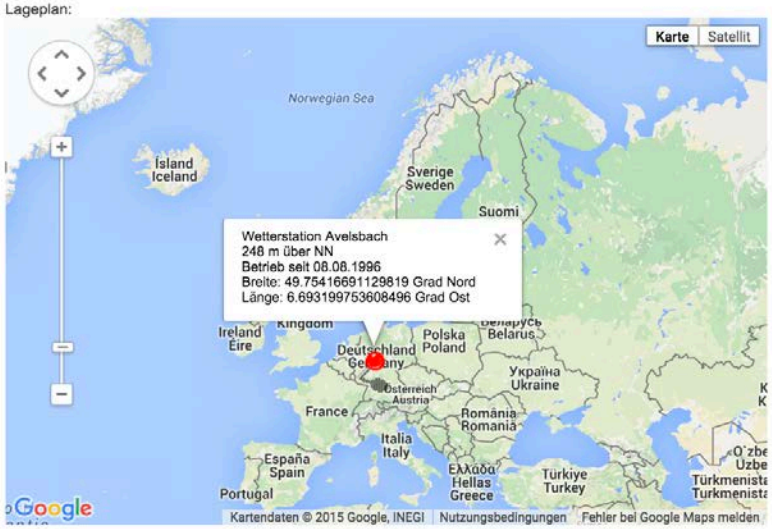

| Utilized Test Methods |
|---|
| Burgemeister, B., Blum, L., & Lorge, I. (1972). Columbia Mental Maturity Scale. New York: Hartcourt Brace. |
| Eggert, D. (1978). Hannover Wechsler Intelligenztest für das Vorschulalter (HAWIVA). Bern: Huber |
| Horn, W. (1969). PSB Prüfsystem für Schul- und Bildungsberatung. Göttingen: Hogrefe. |
| Kuhl, J., & Kraska, K. (1992). Selbstregulations- und Konzentrationstest für Kinder (SRKT-K). Göttingen: Hogrefe. |
| Schuck, K. D., & Eggert, D. (1975). Hannover-Wechsler-Intelligenztest für das Vorschulalter. Bern: Huber. |
| Tewes, U. (1985). Hamburg-Wechsler-Intelligenztest für Kinder (HAWIK-R). Bern: Huber. |
| Tewes, U. (1991). HAWIE-R Hamburg-Wechsler Intelligenztest für Erwachsene - Revision 1991. Bern: Huber. |
| Weiss, R. (1987). Grundintelligenztest CFT 20. Braunschweig: Westermann. |


Top

http://dx.doi.org/10.5160/psychdata.wtfz05lo22

# The Problem of Domain Specific Knowledge in Psychology I

Lageplan:




Bilder:

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

**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

| ID | DE3 | HA03 |
|---|---|---|
| 1000000171 | 2 | 3 |
| 1000000822 | 5 | 3 |
| 1000001003 | 2 | 1 |
| 1000001004 | 1 | 3 |
| 1000001005 | 1 | 1 |
| 1000001006 | 1 | 3 |
| 1000001007 | 4 | 3 |
| 1000001008 | 1 | 3 |
| 1000001009 | 1 | 1 |
| 1000001010 | 1 | 3 |
| 1000001011 | 2 | 3 |
| 1000001012 | 1 | 3 |
| 1000001013 | 1 | 4 |
| 1000001014 | 1 | 1 |
| 1000001015 | 4 | 3 |
| 1000001016 | 3 | 3 |
| 1000001017 | 3 | 3 |
| 1000001018 | 3 | 3 |
| 1000001019 | 3 | 3 |
| 1000001020 | 3 | 3 |
| 1000001021 | 1 | 4 |
| 1000001032 | 4 | 3 |
| 1000001033 | 2 | 2 |
| 1000001034 | 1 | 3 |
| 1000001035 | 1 | 4 |
| 1000001036 | 4 | 3 |
| 1000001037 | 1 | 4 |
| 1000001038 | 5 | 3 |
| 1000001039 | 1 | 3 |
| 1000001040 | 1 | 3 |
| 1000001041 | 1 | 4 |

**Source:**
Tesch-Römer, C., Motel-Klingebiel, A., & von Kondratowitz, H.-J. (2009). Old Age and Autonomy: The Role of Service Systems and Intergenerational Family Solidarity. Primary data of the european study OASIS [Translated Title] (Version 1) [Files on CD-ROM]. Trier: Center for Research Data in Psychology: PsychData of the Leibniz Institute for Psychology Information ZPID. http://dx.doi.org/10.5160/psychdata.thcs01ol03

**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

# The Problem of Domain Specific Knowledge in Psychology III

**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"

| ID | DE3 | HA03 |
|---|---|---|
| 1000000171 | 2 | 3 |
| 1000000822 | 5 | 3 |
| 1000001003 | 2 | 1 |
| 1000001004 | 1 | 3 |
| 1000001005 | 1 | 1 |
| 1000001006 | 1 | 3 |
| 1000001007 | 4 | 3 |
| 1000001008 | 1 | 3 |
| 1000001009 | 1 | 1 |
| 1000001010 | 1 | 3 |
| 1000001011 | 2 | 3 |
| 1000001012 | 1 | 3 |
| 1000001013 | 1 | 4 |
| 1000001014 | 1 | 1 |
| 1000001015 | 4 | 3 |
| 1000001016 | 3 | 3 |
| 1000001017 | 3 | 3 |
| 1000001018 | 3 | 3 |
| 1000001019 | 3 | 3 |
| 1000001020 | 3 | 3 |
| 1000001021 | 1 | 4 |
| 1000001032 | 4 | 3 |
| 1000001033 | 2 | 2 |
| 1000001034 | 1 | 3 |
| 1000001035 | 1 | 4 |
| 1000001036 | 4 | 3 |
| 1000001037 | 1 | 4 |
| 1000001038 | 5 | 3 |
| 1000001039 | 1 | 3 |
| 1000001040 | 1 | 3 |
| 1000001041 | 1 | 4 |

**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

https://www.flickr.com/photos/agrusoft/16357210343/

## 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

# Lessons Learned II

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

**Barriers for data sharing have to be low**

**→ Tools are needed that are integrated in the research process**

# 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 date
- 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

Member of the German Data Forum

Using DOIs to cite and link to research data

Partly funded by the German Research Foundation