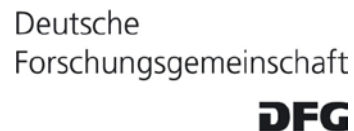


Forschungsdatenmanagement in der Psychologie: Anforderungen, Werkzeuge, Standards

Erich Weichselgartner, ZPID
Stv. Wiss. Leiter
Leitung FDZ PsychData

Gefördert von



Überblick

1. *Hintergrund*: Aktuelle Probleme der Wissenschaftskommunikation
2. *Anforderungen* an das Forschungsdatenmanagement gibt es zum einen von innen heraus, aus dem Bereich des wissenschaftlichen Arbeitens bzw. der guten wissenschaftlichen Praxis, und zum anderen von außen, zum Beispiel von Forschungsförderern, Fachgesellschaften und Verlagen.
3. *Werkzeuge* sollen das Management der Forschungsdaten unterstützen, einschließlich ihrer Vernetzung, ihrer dauerhaften Verfügbarkeit und ihres offenen Zugangs.
4. Von *Standards* im Sinne von Normen ist man in der Psychologie noch weit entfernt. In der Entwicklung sind Übereinkünfte, die in der Praxis Akzeptanz finden sollen. Dazu gehören Empfehlungen (Leitlinien, Richtlinien), Formate und Metadaten.

Aktuelle Probleme der Wissenschaftskommunikation:

1. Ignoranz & Verschwendung
2. Replizierbarkeitskrise
3. Mangelnde Nachnutzung
4. Tendenziösität
5. Neophilia

Konzeptueller Penis als Ursache für den Klimawandel

23. Mai 2017 – Peter Mühlbauer

Veröffentlicht in der Zeitschrift *Cogent Social Sciences*



Johann Heinrich Cohausen:
*„Der Atem von Jungfrauen
verlängert das Leben“ (1742)*
[sein Neffe wurde Professor an der
Universität Trier]

Medieninfo

Leibniz
Leibniz
Gemeinschaft

15.6.2017

10/2017

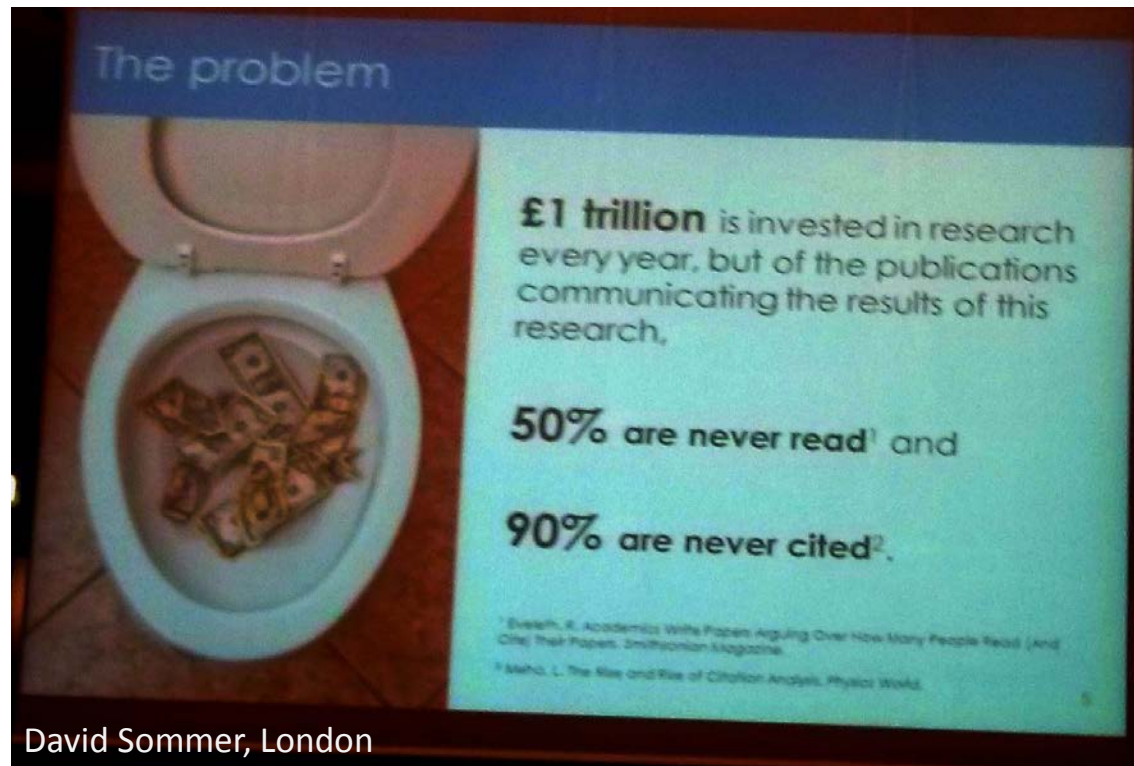
Präsidium beschließt Maßnahmen gegen Leibniz-Direktor
Wissenschaftliches Fehlverhalten: Datenfälschdarstellung, ungenügende
Datendokumentation und Verletzung der Aufsichtspflicht

Dieser Mann hat der Wissenschaft die Smarties geklaut

Der Forscher Diederik Stapel hat mehr als 50 Studien manipuliert, Kollegen
belogen und die Psychologie blamiert. Jetzt will er die Universitäten
verändern. Was geht in seinem Kopf vor?

Problem 1: Ignoranz & Verschwendung

- Doktorarbeiten werden kaum gelesen. Wichtig scheint nur noch, dass sie geschrieben werden. (Kaube, 2015)
- 85 Prozent aller Investitionen in Forschung sind verschwendet. Der größte Fehler: **Es wird nicht geprüft, ob vergleichbare Untersuchungen bereits existierten.** (Chalmers & Glasziou, 2009)
- Namhafte Forscher und Nobelpreisträger beklagen zu viel Forschungsmüll: Lancet-Special (fünf Analysen), Jan 2014 (<http://www.thelancet.com/series/research>)

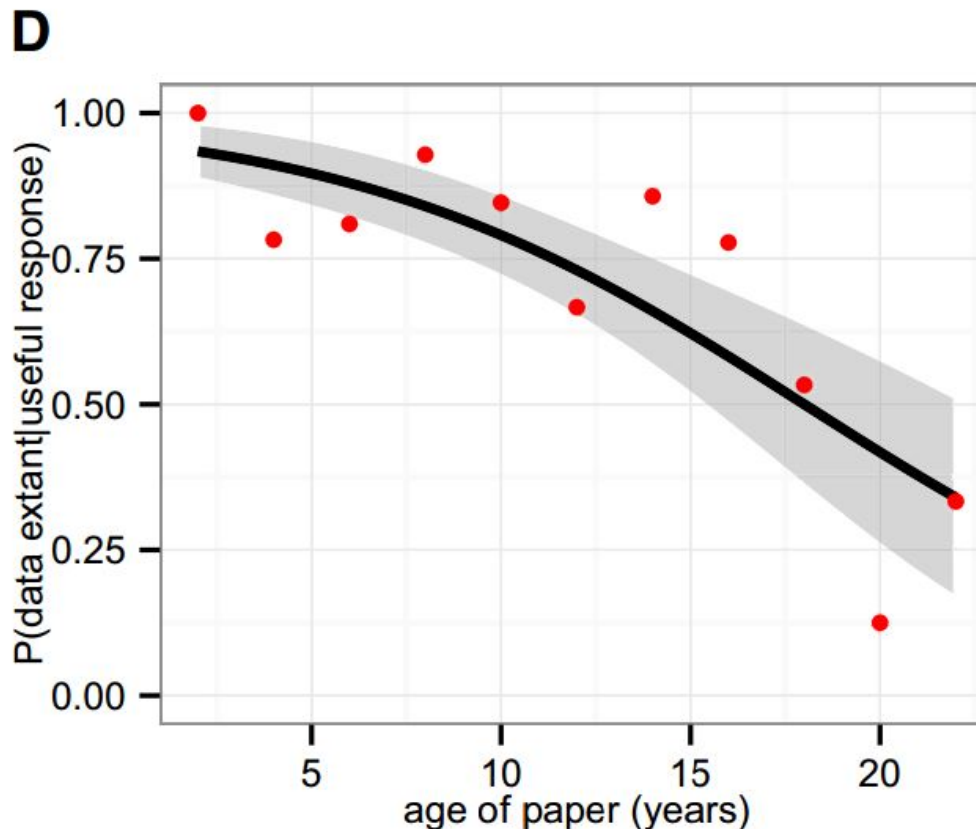


Problem 2: Replizierbarkeit



Problem 3: Nachnutzung

“The underlying data researchers analyze to come to their published conclusions ... **becomes less and less accessible to researchers over the years.**” (Vines et al, 2014; vgl. Dehnhard, Weichselgartner & Krampen, 2013; Wicherts et al, 2006)



(D) Predicted probability that the data were extant (either “shared” or “exist but unwilling to share”) given that we received a useful response.

Problem 4: Tendenziösität

Funding-induced Biases in Science (Wojick & Michaels, 2015)

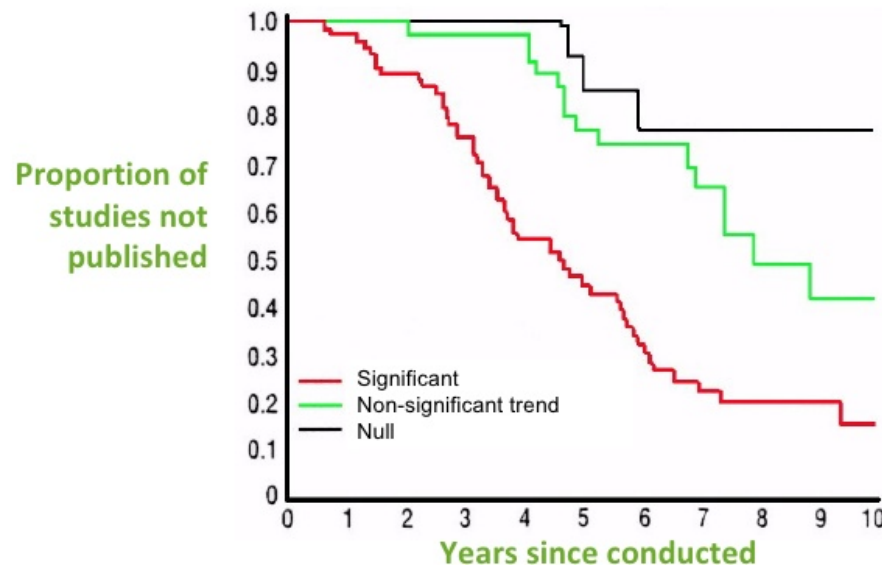
Reporting Biases¹

Publication Bias: Je nach Ergebnis wird publiziert oder nicht publiziert.

Outcome reporting bias: Selektive Auswahl von Resultaten.



Evidence for reporting bias



Source: Stern JM, Simes RJ. Publication bias: evidence of delayed publication in a cohort study of clinical research projects *BMJ* 1997;315:640-645.

cochrane training

¹ http://handbook.cochrane.org/chapter_10/table_10_1_a_definitions_of_some_types_of_reporting_biases.htm

Problem 5: *Neophilia*

Fokus auf neue Befunde: Die großen Fachzeitschriften lechzen nach "*sexy Themen und steilen Thesen*" (Randy Schekman, Nobelpreis für Medizin)

- Beschönigen (*massaging data*) und Fälschen
→ *false positives*
- Knapp 34% der Befragten gab fragwürdige Forschungspraktiken zu (Metaanalyse, Fanelli, 2009)

Beispiele:

- Medizin: Mori, Potti
- Stammzellenforschung: Woo-suk Hwang
- Physik: Schön

Überblick zu zurückgezogenen Aufsätzen bei

- PubMed retractions report (*live!*): <http://rpubs.com/neilfws/65778>
- Retraction Watch: <http://retractionwatch.com/>

Mangelnde Replizierbarkeit

- Reproducibility Project
 - **Psychologie:** 100 Studien aus großen Fachmagazinen: Replikationsrate je nach verwendetem Evaluationskriterium zwischen 36% und 47% (Estimating the reproducibility of psychological science. Open Science Collaboration. *Science*, 28 August 2015: 349 (6251); Renkewitz et al, 2015)

“Nicht-Kommunikation” negativer Ergebnisse

- <http://psychfiledrawer.org/>
- ‘Negative results’ now account for only 14% of published papers, down from 30% in 1990. ([How science goes wrong](#), The Economist, 19.10.2013)

[Casrai](#) standard dictionary of research administration information (2015)

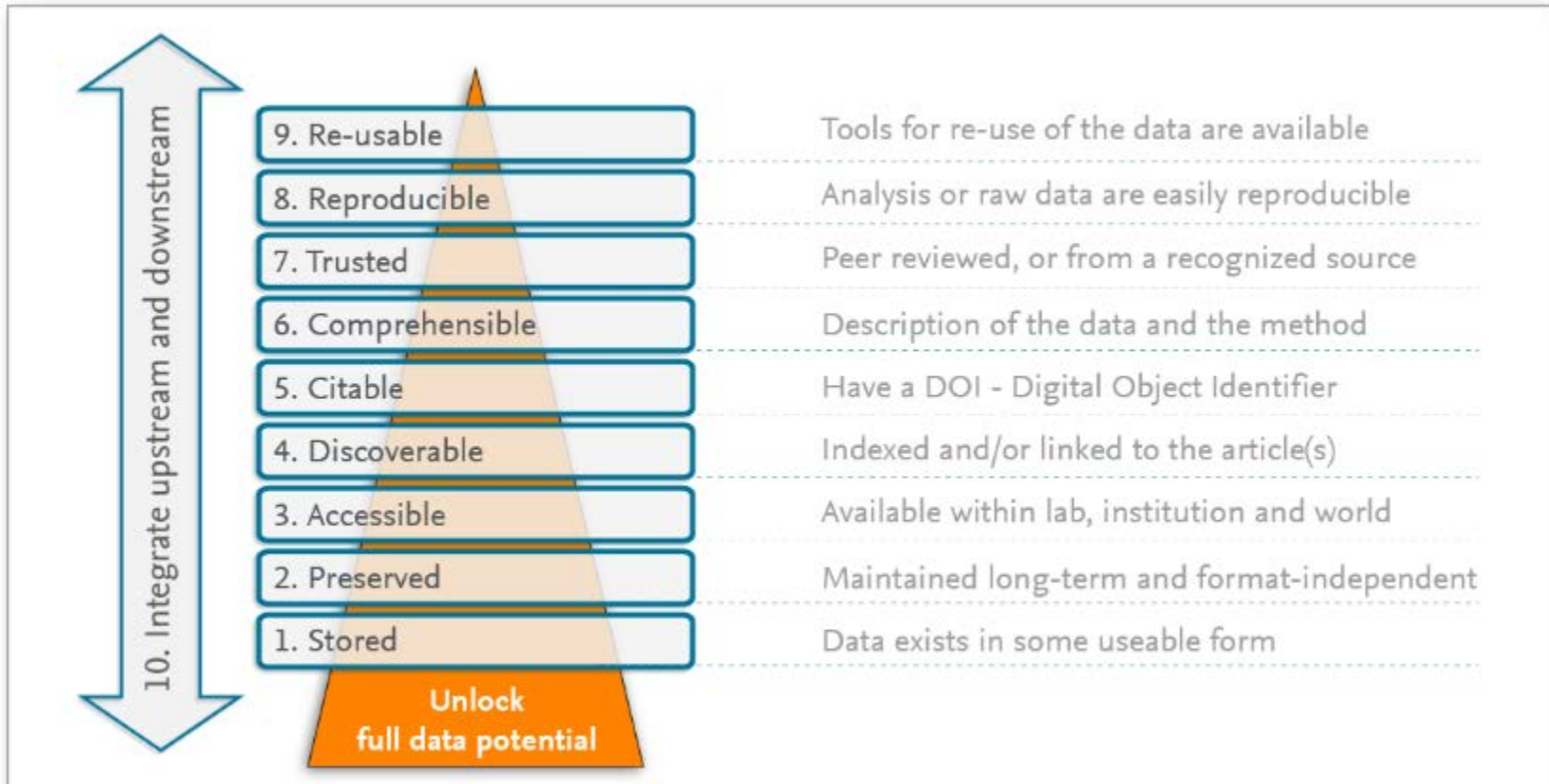
„Facts, measurements, recordings, records, or observations about the world collected by scientists and others, with a minimum of contextual interpretation. Data may be in any format or medium taking the form of writings, notes, numbers, symbols, text, images, films, video, sound recordings, pictorial reproductions, drawings, designs or other graphical representations, procedural manuals, forms, diagrams, work flow charts, equipment descriptions, data files, data processing algorithms, or statistical records.“

[Casrai](#) standard dictionary of research administration information (2015)

„Data that are used as primary sources to support technical or scientific enquiry, research, scholarship, or artistic activity, and that are used as evidence in the research process and/or are commonly accepted in the research community as necessary to validate research findings and results. All other digital and non-digital content have the potential of becoming research data. Research data may be experimental data, observational data, operational data, third party data, public sector data, monitoring data, processed data, or repurposed data.“

[University of Oxford](#) / Research Data Oxford (2013)

“The recorded information (regardless of the form or the media in which they may exist) necessary to support or validate a research project’s observations, findings or outputs“.



Quelle: [Elsevier](#)

Research Data Management helps researchers navigate the increasingly complex landscape of data planning, storage, and sharing:

- Data Management Plans
 - A data management plan is a document that describes how you will collect, organise, manage, store, secure, back up, preserve, and share your data.
- Active Data Management
 - Organising, naming, structuring and documenting
- Data Backup, Storage
- Data Sharing (Access)
 - Sharing, licensing, depositing, and citing your data
- Data Preservation and Archiving
- Data Security
 - Working with sensitive data

Vorbemerkung: Nach deutscher Nomenklatur wird die Psychologie den Geisteswissenschaften, den Sozialwissenschaften und den Naturwissenschaften zugeordnet und es kommt ein entsprechend vielfältiger Methodenkanon zum Einsatz.

Datenerhebung in der Psychologie (mit Beispielen)

- Beobachtung
 - Videoaufzeichnung von Unterricht
- Befragung, Umfrage, Interview
 - Auswertung eines standardisierten Fragebogens
- Fallstudie
 - Biographische Aufzeichnungen
- Experiment, Test
 - Resultate eines Intelligenztests
- Physiologische Methoden (EEG, GSR, fMRI, Eye Tracking)
 - Augenbewegungen bei statischem Bild (Fixationen, Sakkaden)

1. Anforderungen (Beispiele)

Forschungsförderer

- [DE] DFG-Denkschrift [Sicherung guter wissenschaftlicher Praxis](#) (2013): Sicherung und Aufbewahrung von Primärdaten
- [DE] Empfehlungen der Hochschulrektorenkonferenz [Gute wissenschaftliche Praxis an deutschen Hochschulen](#) (2013): Keine Datenmanipulation
- [UK] Wellcome Trust [Policy on data management and sharing](#) (2016): In cases where the proposed research is likely to generate data outputs that will hold significant value as a resource for the wider research community, applicants will be **required to submit a data management and sharing plan** to the Wellcome Trust **prior to an award** being made.
- [EU] [The European Code of Conduct for Research Integrity](#) (2017): Data Practices and Management
- [USA] [National Institutes of Health Plan](#) for Increasing Access to Scientific Publications and Digital Scientific Data from NIH Funded Scientific Research (2015): Digital Scientific Data (Expects investigators seeking more than \$500K in direct support in any given year to **submit a data sharing plan** with their application or to indicate why data sharing is not possible.)

1. Anforderungen (Beispiele)

Verlage (publisher requirements, publisher data policy)

- **Elsevier:** „We will encourage and support researchers and research institutions to share data where appropriate and at the earliest opportunity.”
 - [Mendeley Data](#) (repository)
 - [Data in Brief](#) (data publication)
- **Springer-Nature** (e.g. [Brain Imaging and Behavior](#)): Four different policy types.
Type 2: „Data sharing and evidence of data sharing encouraged.”
 - Springer Nature authors can now submit more of their research outputs via a new data submission portal powered by [Figshare](#). (repository)
 - [Scientific Data](#) (data publication) [Psychology: 4 in 2016, 5 in 2017]
- **AAAS** (e.g. [Science](#)): „ Datasets must be deposited in approved repositories with an accession number prior to publication. ... All materials that are necessary to understand and assess the research must be made available.”
- **APA**
 - [Archives of Scientific Psychology](#)): Seit 2013 kaum Beiträge!
[„ Users should contact the Editorial Office at the American Psychological Association for information on requesting data access.”]
 - “The [Journal of Experimental Psychology: Learning, Memory, and Cognition](#) may need to go back to preschool.” ([Oransky & Marcus](#), MARCH 10, 2017)
- **Ubiquity Press** ([Journal of Open Psychology Data](#)): Seit 2013 kaum Beiträge!

2. Werkzeuge

Werkzeuge sollen das Management der Forschungsdaten unterstützen, einschließlich ihrer Vernetzung, ihrer dauerhaften Verfügbarkeit und ihres offenen Zugangs.

- Datenmanagement-Pläne erstellen (disziplinübergreifend)
 - USA (NSF, NIH): DMPTool ([University of California Curation Center](#) of the [California Digital Library](#))
 - [Beispiel](#): Drug Use by High School Students in King County Schools
 - UK (EU): DMPonline ([Digital Curation Centre](#), Edinburgh, UK) (auch [DE](#) und [FR](#))
 - Musterpläne für [DFG](#) und [BMBF](#) (erstellt von HU Berlin)
- Datendokumentation
 - [PsychData-Dokumentationstool](#)
 - [Colectica for Excel](#) - Colectica for Microsoft Excel is a free tool to document your spreadsheet data using the open standard for data documentation.
- Datenkonvertierung
 - Von proprietären zu offen (standardisierten) Formaten
 - [SledgeHammer](#) (z.B. STATA → CSV)
 - [StatTransfer](#)
 - SPSS_IO (boelter@zpid.de)
- Speicherung, Archivierung, Zugang
 - Repositorien ([PsychData](#), [OSF Data Repository](#), [DataDryad](#)^{\$\$}, ...); Liste: [re3data.org](#)
 - Discovery (visibility, impact): Indexing, Linking, Browsing
 - Daten in den FDZ/DSZ des RatSWD [suchen](#)
 - [DataCite](#)

3. Standards

Von Standards im Sinne von Normen ist man in der Psychologie noch weit entfernt. In der Entwicklung sind Übereinkünfte, die in der Praxis Akzeptanz finden sollen. Dazu gehören Empfehlungen (Leitlinien, Richtlinien), Formate und Metadaten.

- Empfehlungen
 - DGPs: [Empfehlungen zu einem offenen und nachhaltigen Datenmanagement](#) (2016)
- Metadaten Standards und Ontologien: **Keine Metadatenstandards in der Psychologie**
 - [Research Data Alliance Metadata Directory](#) (verschiedene Disziplinen, keine Psychologie)
 - Sozialwissenschaften: [DDI](#)
 - [Linked Open Vocabularies](#) (verschiedene Disziplinen, keine Psychologie)
 - [BioPortal](#) (Psychology Ontology [APAONTO](#): APA thesaurus, flat ontology file, α)
 - Allgemein/Bibliographie
 - [Dublin Core](#)
- Formate (für langfristige Zugänglichkeit!)
 - American Standard Code for Information Interchange (ASCII): [ISO 8859](#)
 - XML, PDF (teilweise), CSV, R-Syntax
 - Unicode: [ISO 10646](#)
- Dokumentationsstandards
 - Reporting Standards for Research in Psychology ([JARS](#), APA, 2008)
 - Meta-Analysis Reporting Standards ([MARS](#), APA, 2008)
 - Consolidated Standards of Reporting Trials ([CONSORT](#), 2001)
 - Transparent Reporting of Evaluations with Nonexperimental Designs ([TREND](#))
 - Standards for Reporting on [Empirical Social Science Research](#) in AERA Publications
 - Standards for Reporting on [Humanities-Oriented Research](#) in AERA Publications

FILE FORMATS FOR LONG-TERM ACCESS

As technology changes, researchers should plan for both hardware and software obsolescence and consider the longevity of their file format choices to ensure long term readability and access.

File formats more likely to be accessible in the future have the following characteristics:

- Non-proprietary
- Open, documented standard
- Common usage by research community
- Standard representation (ASCII, Unicode)
- Unencrypted
- Uncompressed

Examples of preferred file format choices include:

- ODF, not Word
- ASCII, not Excel
- MPEG-4, not Quicktime
- TIFF or JPEG2000, not GIF or JPG
- XML or RDF, not RDBMS

Consider migrating your data into a format with the above characteristics, in addition to keeping a copy in the original software format. If you deposit your data in a [repository](#), your files may be migrated to newer formats, so that they're usable to future researchers.

[MIT Data Management](#)

Beispiel Metadaten für Forschungsdaten

Element	Value
Title	Name of the project of collection of datasets
Creator	Names and institutions of the people who created the data
Date	Key dates associated with the data, such as dates covered by the data or the date of creation
Description	Description of the resource
Keywords or Subjects	Keywords or subjects describing the content of the data
Identifier	Unique number of alphanumeric string used to identify the data
Coverage (if applicable)	Geographic coverage
Language	Language of the resource
Publisher	Entity responsible for making the dataset available
Funding Agencies	Organization or agency that funded the research
Access Restrictions	Where and how your data can be accessed by other researchers
Copyright	License associated with the resource
Format	Format of the data file

[University of Illinois at Urbana-Champaign](http://www.uiowa.edu)

Erich Weichselgartner (wga@zpid.de)

Leibniz-Zentrum für Psychologische Information und Dokumentation (ZPID), Trier

Love your data!

