# Enabling Biographic Research in the 22nd Century: Archiving Academic Digital Footprints

Erich Weichselgartner

Leibniz Institute for Psychology Information (ZPID)

Trier, Germany
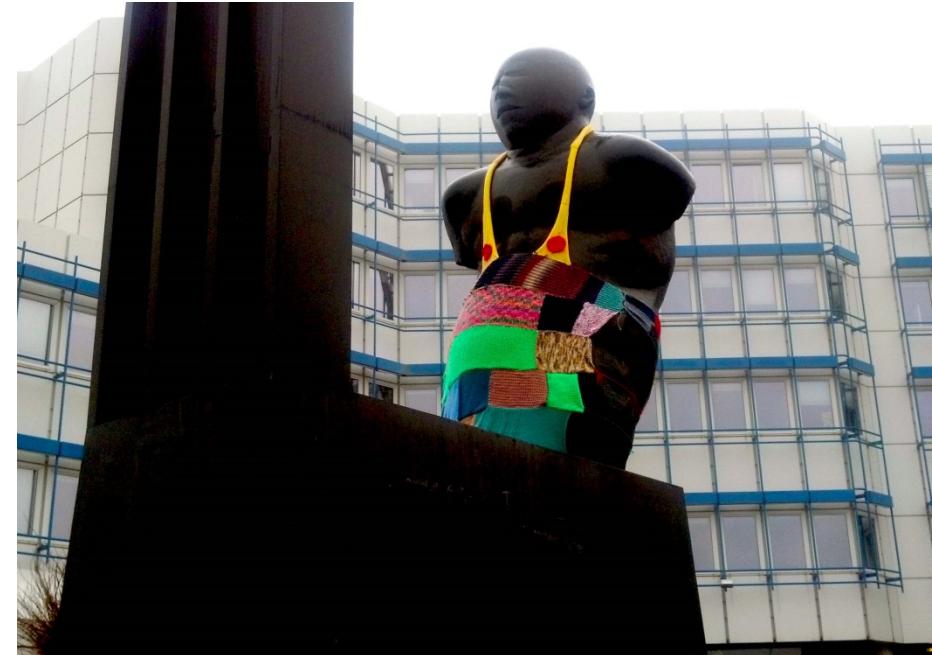
Web-Archiving Meeting 2013, Innsbruck, Austria – September 20th, 2013

Preface

About

1. I am not an archivist and not a historian
2. I will not give a technical presentation
3. Deputy scientific director @ research support organization
   1. Support scholarly communication in psychology
4. How to preserve digital heritage of academics (incl. *invisible college*)?
   1. Define relevant academic materials and processes (selection criteria)
5. New to the field; hope to learn

Overview

- Concerned with digital information relevant to scholarly communication
  - Enable future research on academics' lives
  - Ensure the survival of pertinent digital information
  - Preserve „digital trail" (prevent *Digital Dark Age*)
- Long term preservation of **scholar's heritage** and of **scholar's (informal) communication** (the *invisible college* – Price, 1986)
  - Intellectual genealogy
  - Intellectual heritage
- Records of scholars, expressions of scholarship
  - Research cycle
- Selection criteria: Traditional impact of scholarship vs altmetrics
- Problems (homographs, quality assessment, deceased scholars, legal issues)

Oldest ever found footprint of the human genus (1.5 million years old); Ileret, Kenya, 2009

Digital footprint

Biographical Research - The Study of Lives

- The collection and interpretation of personal or human documents (Allport 1942; Blumer 1969)

- Produce rich descriptions of persons

- Action logics: how persons and structures are interlinked; interaction with family, peers and institutions

- Using a broad range of (academic) material as e.g. personal artifacts, annals, chronicles, diaries, letters, conversations, photographs, memoranda, unpublished papers, autobiographies, interviews, lab data sheets, phone calls, e-mail, personal journals, conference proceedings, podcasts, videos

- Better preserve material in the lifetime of a scientist than to try to reconstruct it retrospectively (lesson learned from *PsychData*)

3,600 year-old Sky Disk of Nebra, Germany

3,400 year-old Phaistos Disk, Crete

http://longnow.org/

Analog world: Very long lasting Disks

5 year-old Rosetta Disk, Norsam Technologies, USA

Scholarly Communication

Scholarly Communication
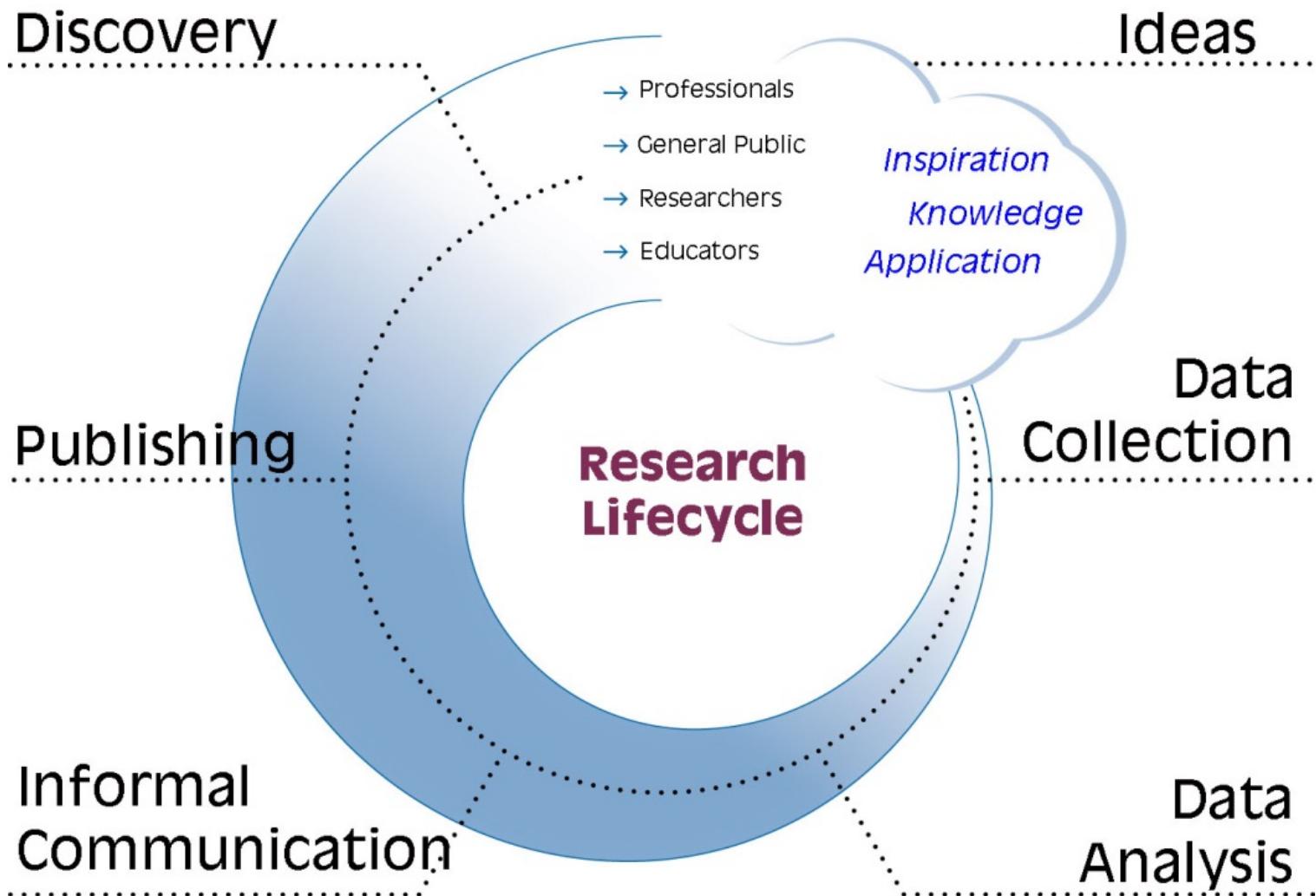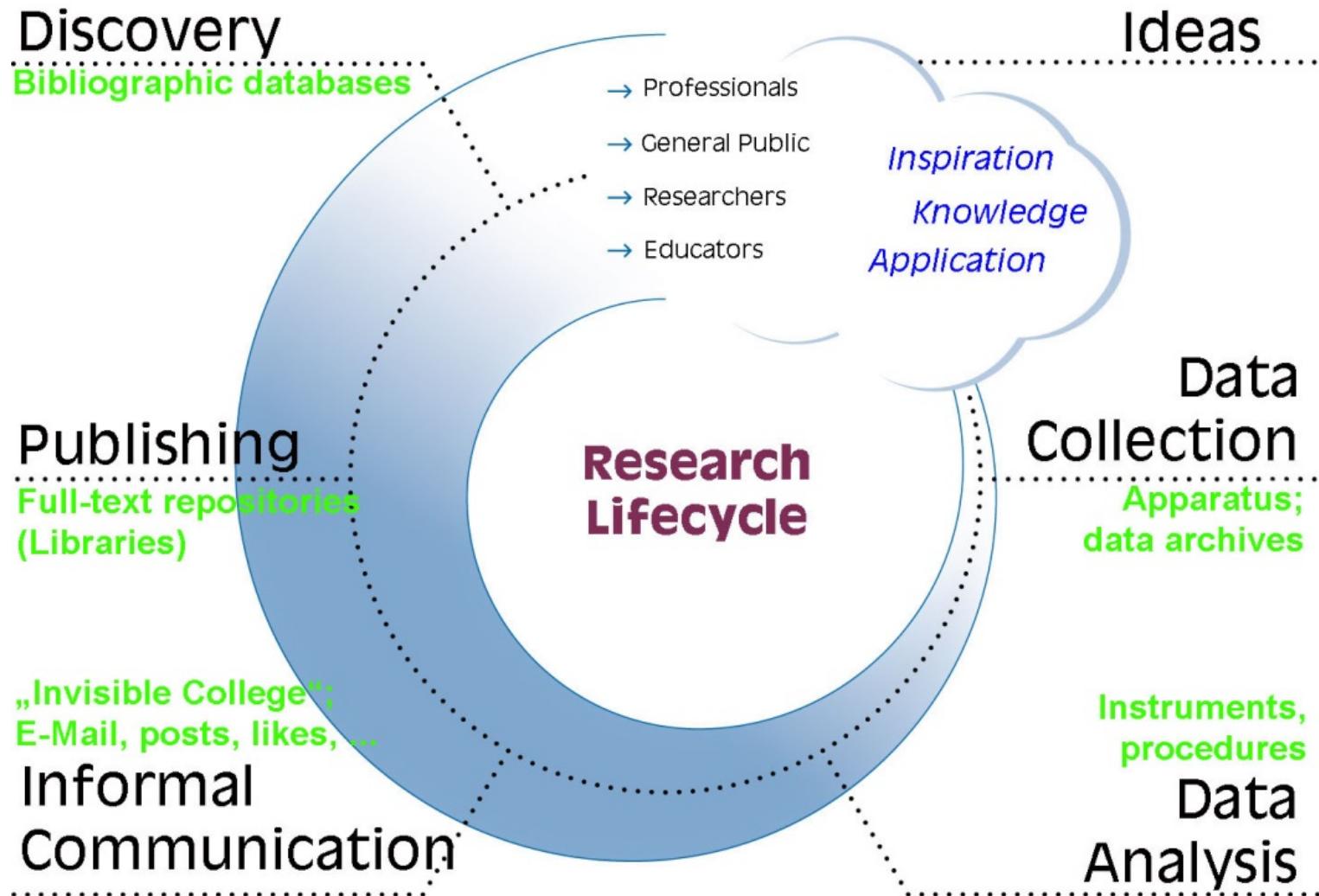
Expressions of scholarship: Criteria (quality and reputation)

- Degree, Position
  - Renown institution
- Publications and talks
  - Productivity (# of publications), impact (# of citations)
- Membership
  - Learned & professional societies, editorial boards, conference boards
- Reviewing
- Grants
  - e.g., DFG-Gepris
- Awards, honors, prizes
  - National Academy membership, the Nobel Prize
- Patents
- Experience

Academic Scholar Biographical Sketches

PsychAuthors: Author Profiles of Psychologists in the German Speaking Countries

Contents
- Address
- Education
- Experience (Employment History; Special Skills & Qualifications)
- Research Interests
- Teaching Experience
- Grants and Fellowships
- Awards and Honors
- Professional Affiliations
- Publications (incl. Presentations)

Example

PsychAuthors

Prof. Dr. Erich Schröger

akademische Qualifikationen:

- 1986 Diplom, LMU München
- 1991 Promotion, LMU München
- 1996 Habilitation, LMU München

Mitgliedschaften in Fachgesellschaften, Editorial Boards etc., Herausgeberschaften (national / international):

- Deutsche Gesellschaft für Psychologie (DGPs)
- Fachgruppe Allgemeine Psychologie der DGPs
- Fachgruppe Biologische Psychologie der DGPs
- Deutsche Gesellschaft für Psychophysiologie und ihre Anwendungen
- Gesellschaft für Kognitionswissenschaft

- Society for Psychophysiological Research (SPR)

Berti, S., Grunwald, M. & Schroeger, E. (2013). Age dependent changes of distractibility and reorienting of attention revisited: An event-related potential study. *Brain Research*, 156-166. `PSYNDEX`

Keitel, C., Maess, B., Schröger, E. & Müller, M. M. (2013). Early visual and auditory processing rely on modality-specific attentional resources. *NeuroImage*, 240-249. `PSYNDEX`

Knolle, F., Kotz, S. A. & Schroeger, E. (2013). Prediction errors in self- and externally-generated deviants. *Biological Psychology*, *92*(2), 410-416. `PSYNDEX`

Pieszek, M., Widmann, A., Gruber, T. & Schröger, E. (2013). The human brain maintains contradictory and redundant auditory sensory predictions. *PLoS ONE, 8*(1), No. e53634 `PSYNDEX`

PsychAuthors

Limitations

- Manual updates (annually)
  - (Benefit: Quality control)
  - Disadvantage: Dropouts
- Static view
- No versioning → No longitudinal research designs (e.g., change in interests)
- Covers only part of relevant information, content selection based on traditional metrics
- Only fraction of researchers

Status (August 2013)
- N ≈ 1.000 (800 public, 200 undisclosed)

Digital Footprints

The new invisible college

Digital Footprints are based on digital information you *actively create* about yourself. They provide data on what has been performed in the digital environment (e.g. what you clicked on, searched for, liked, where you went, your location, your IP address, what you said).
The data can be used in behavioural targeting, personalisation, digital **reputation**, social **influence** and other social media services.

Characteristics
of expertise

The measurement of expertise: Scientometric predictors of research performance

Metrics – analog world

- Publication counting (indicator for productivity, indicator for expertise)

- Circulation-counting (indicator for (potential) # of readers)

- Citation counting (indicator for impact)

  - Journal Impact Factor (JIF): Journals' average citations per article

    (OFTEN INCORRECTLY USED TO ASSESS THE IMPACT OF INDIVIDUAL ARTICLES)

  - H-Index: The h-index reflects both the number of publications and the number of citations per publication

The measurement of expertise: Scientometric predictors of research performance

Expressions of scholarship are becoming more ***diverse***

- Online reference management
  - How often is someone's paper bookmarked?
- The sharing of "raw science" like datasets, code, and experimental designs
- Semantic publishing or "nanopublication," where the citeable unit is an argument or passage rather than entire article
- Widespread self-publishing via blogging, microblogging, and comments or annotations on existing work.

"This diverse group of activities forms a composite trace of impact far richer than any available before. We call the elements of this trace **altmetrics**." (Priem & Hemminger, 2010)

The measurement of expertise: Scientometric predictors of research performance

Metrics – analog world

- Publication counting (indicator for productivity, indicator for expertise)
- Circulation-counting (indicator for # of readers)
- Citation counting (indicator for impact)

Metrics – digital world – altmetrics

- Downloads
- Views
- Links
- Bookmarks
- Conversations

Use metrics	**Attention metrics**	
↓	↓	**Preserve?**
Indicators for readers	Indicators for impact	

Citation of print resources

Information published in printed form in books, journals and newspapers is collected, catalogued and preserved by national and state libraries. The ISBN and ISSN provide ways to permanently find a particular item.

Citation of ephemeral Web resources

Internet resources tend to have a short life; their identification and persistent location pose complex problems that affect many technological and organizational issues involving the citation, retrieval and preservation of scientific resources. This is by no means a technical problem alone: persistent digital object identification, including texts, research data, images and the like, is still a major issue that prevents the use of today's Internet as a trustworthy platform for the research and dissemination of scientific content.   (Bellini et al, 2005)

Resolution service for scholarly materials (DOIs)

A trustworthy solution is to associate a persistent identifier (PI) with a digital resource that will remain the same regardless of where the resource is located. DOIs can be incorporated into Web pages much like current links. But instead of pointing to a specific Web location, the DOI sends the browser off to a database, where it retrieves and displays whatever information the publisher chooses to offer.

Major applications of the DOI system currently include:

- Persistent citations in scholarly materials (journal articles, books, etc.) through CrossRef, a consortium of around 3,000 publishers;
- Research datasets through DataCite, a consortium of leading research libraries, technical information providers, and scientific data centers.

Example

http://dx.doi.org/10.5160/psychdata.pdpr99ei20

        directory indicator   registrant code   suffix

Resolution service for researchers and scholars

As researchers and scholars, you face the ongoing challenge of distinguishing your research activities from those of others with similar names. You need to be able to *easily and uniquely attach your identity to research objects*

- How do you distinguish between articles belonging scholars with the same names (homographs)?

- How do you distinguish between articles belonging scholars with similar names?

- How can you be confident that you captured all results for a scholar when their name is recorded in different ways?

- And, can you be sure that names with unusual characters such as accents have been included?

ORCID (Open Researcher and Contributor ID) – "author ID", "author DOI"
An alphanumeric string that uniquely identifies an individual scientist.

Example

http://orcid.org/0137-1963-7688-2319

New possibilities due to PI

Provide each researcher with "a **constantly updated *digital curriculum vitae*** providing a picture of his or her contributions to science going far beyond the simple publication list". (Nature, 2009)

Examples of Altmetrics Tools

- Impact Calculator *ImpactStory*

- Researcher reputation graph PlumX

- Monitor, search and measure conversations about your publications
  The *Altmetric* Explorer

Alt-Metrics "storage"

Example: Mendeley

Why most published research findings are false.

John P A Ioannidis in *PLoS Medicine (2005)*

> probed in each **scientific** field. In this framework, a research finding is less likely to be true when

📄 Save PDF to library · **Related** research

3,755 readers

Why Most Published Research Findings Are False: Author's Reply to Goodman and Greenland

John P A Ioannidis in *PLoS Medicine (2007)*

> probed in each **scientific** field. In this framework, a research finding is less likely to be true when

📄 Save PDF to library · **Related** research

4,021 readers

Number of "readers" in Mendeley is the number of distinct users who have added the paper to their Library on Mendeley. It doesn't actually mean that they've read the paper.

Example (Google): ejop.psychopen.eu
- Google pagerank (pr): **5**   (0 ≤ pr ≤ 10)
- Link popularity (# of back links): **45**

Alt-Metrics "bookmarks"

The number of times an artifact has been *bookmarked*, the number of times the artifact has been *marked as a favorite*, the number of times a person or artifact has been *followed*, the number of times an artifact has been *placed in a library.*

Services

- AcademiaNET
- CiteULike
- Delicious
- Mendeley
- ResearchGATE
- Zotero

Alt-Metrics "conversations"

Conversations around the scientific articles in **tweets**, facebook, blog posts, and news. Discussion activity (comments, notes, ratings).
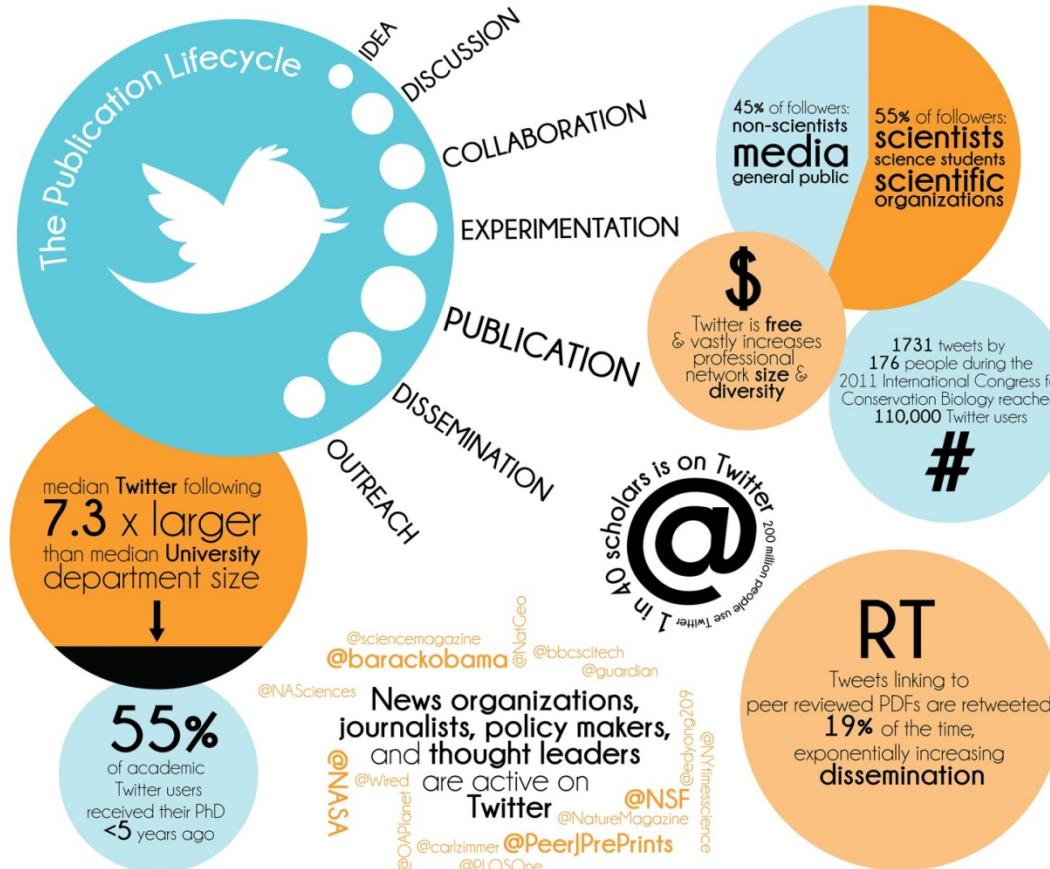
Darling et al (2013).

"The official mentor is listed on the dissertation, and university libraries keep these forever. After discovering the identity of a prior mentor, uncovering information about her or him can be more difficult, especially for more than one generation back. The individual and his or her compatriots may have died, and the written record may be poor. Many departments do not keep good records or even final vitas of emeritus (or deceased) professors, and histories of universities rarely record much about individual faculty members. For the more famous individuals, obituaries (and for some, even biographies) can help, although by their nature obituaries tend to present the deceased in a favorable light no matter what the individual's personality while living." (Roediger, 2005)

An academic, or scientific, genealogy, organizes a family tree of scientists and scholars according to dissertation supervision relationships.

PhDTree   Add New Profile   Search   Feedback

Hermann Ludwig Ferdinand von Helmholtz Edit | History | Comment

Professor
University of Berlin
Field: Physics
Website: http://www.helmholtz.de/ueber_uns/geschichte/he...
Profile Views:

Descendant:

- **PhD Students**
  - 1890 Arthur Gordon Webster
  - 1889 Michael Idvorsky Pupin
  - 1886 Philipp Eduard Anton von Lenard
  - 1886 Wilhelm Carl Werner Otto Fritz Franz Wien
  - 1884 Otto Richard Lummer
  - 1880 Heinrich Rudolph Hertz
  - 1879 Edward Leamington Nichols
  - 1875 Friedrich Hermann Schottky
  - 1870 Lórand Eötvös
  - 1869 William James
  - 1858 Friedrich Leopold Goltz
  - 1856 Wilhelm Max Wundt

Psychologie Information
ZPID – Leibniz Institut

Mathematics Genealogy Project

A service of the NDSU Department of Mathematics, in association with the American Mathematical Society.

Please email us with feedback.

Richard Courant

Biography MathSciNet

Ph.D. Georg-August-Universität Göttingen 1910

Dissertation: *Über die Anwendung des Dirichletschen Prinzipes auf die Probleme der konformen Abbildung*

Advisor: David Hilbert

Students:
Click here to see the students listed in chronological order.

| Name	School	Year	Descendants
Claire Adler	New York University	1942	
Leifur Asgeirsson	Georg-August-Universität Göttingen	1933	
Paul Bailyn	New York University	1964	
Isaac Battin	New York University	1949	
Herbert Busemann	Georg-August-Universität Göttingen	1931	20
Philip Cooperman	New York University	1951	
Norman David	New York University	1940	
Charles De Prima	New York University	1943	39
Avron Douglis	New York University	1949	3
William Feller	Georg-August-Universität Göttingen	1926	1015
Hans Fitting	Georg-August-Universität Göttingen	1931	

Example Prof. H. Irtel (1952 – 2008)

Looking for http://www.uni-mannheim.de/fakul/psycho/irtel

Obituary @ web adress

Last snapshot

Example Prof. H. Irtel (1952 – 2008)

INTERNET ARCHIVE
WayBackMachine
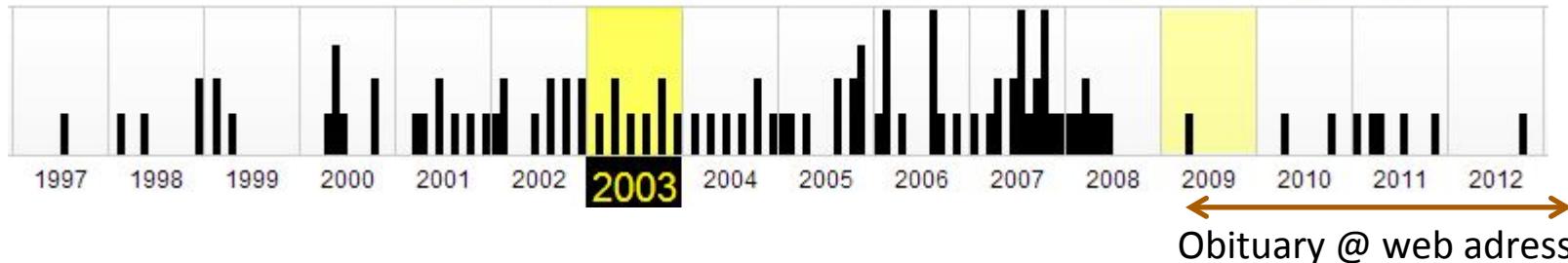
http://www.uni-mannheim.de/fakul/psycho/irtel/pxlab/lab_course/

12 captures
26 Sep 06 - 15 Mar 12

PXLab Lab Course Experiments

s on this page are described in the book Irtel, H. (1991). Experimentalps

s require a Java Version 5 Runtime Environment. Download and install the
sage appears in the buttons below.

Fehler. Klicken Sie hier, um weitere Informationen zu erhalten

ablog.pxd

Fehler. Klicken Sie hier, um weitere Informationen zu erhalten

anagram.pxd

Fehler. Klicken Sie hier, um weitere Informationen zu erhalten

bene.pxd

PXLab Lab Course Experiments

ribed in the book Irtel, H. (1991). Experimentalpsychologisches Praktikum. Hei

5 Runtime Environment. Download and install the Java Runtime Environment

Verification of Logical Statements (Baddeley, 1968)

ablog.pxd

Anagram Puzzles (in German)

anagram.pxd

Interocular Transfer of MAE

bene.pxd

Summary

❖ Preserve the memory of a scientific discipline in the digital age.

❖ Identify, evaluate, select and archive material relevant to intellectual genealogy and intellectual heritage.

❖ Are new forms of scientific communication and long-term preservation compatible? Emphasis on process vs. structure.

Thank You!

Erich Weichselgartner, Leibniz Institute for Psychology Information (ZPID), **Trier**, Germany
wga@zpid.de

Trier was founded in or before 16 BC.

A residence of the Western Roman Emperor until 395.

Emperor Constantine the Great, who first had expanded Trier's military base, moved the Roman capital to the banks of the Bosphorus.

Trier is the oldest seat of a Christian bishop north of the Alps.

The *Trier Cathedral* dates back to Roman times and is home to the Holy Tunic, a garment said to be the robe Jesus was wearing when he died.

Karl Marx was born in Trier in 1818.

References

Editorial (2009). "Credit where credit is due". *Nature*. 462: 825. doi:10.1038/462825a

Allport, G. W. (1942). The Use of Personal Documents in Psychological Science: Prepared for the Committee on Appraisal of Research. Bulletin #49 (New York: Social Science Research Council, 1942). This monograph was subsequently reprinted in 1951.

Bellini, E., Cirinnà, C., Lunghi, M. (2005). Persistent Identifiers for Cultural Heritage. Digital Preservation Europe Briefing Paper.

Blumer, Herbert (1969). *Symbolic Interactionism: Perspective and Method*. New Jersey: Prentice-Hall, Inc.

Darling E.S., Shiffman D., Côté I.M., Drew J.A. (2013) The role of Twitter in the life cycle of a scientific publication. *PeerJ PrePrints* 1:e16v1
http://dx.doi.org/10.7287/peerj.preprints.16v1

Joyce, Ch. (2009). Earliest Humanlike Footprints Found In Kenya. NPR Science Online.

Price, D. J. de Solla. (1986). *Little science, big science ... and beyond.* New York: Columbia University Press.

Priem, J., & Hemminger, B. H. (2010). Scientometrics 2.0: Toward new metrics of scholarly impact on the social Web. *First Monday, 15(7)*. Retrieved from
http://firstmonday.org/htbin/cgiwrap/bin/ojs/index.php/fm/article/view/2874/2570

Roediger, H.L. (2005). Intellectual Genealogy. *APS Observer, 18(4)*. Retrieved from
http://www.psychologicalscience.org/index.php/publications/observer/2005/april-05/intellectual-genealogy.html