

Do psychologists in Germany now search for relevant research publications on the Internet in the same way they did at the end of the 1990s? Does the use of social media in teaching really produce better learning outcomes because of the informal learning potential inherent in social media? Do university students participate more actively as text contributors in bottom-up wikis initiated by their fellow students than they do in top-down wikis initiated by their instructors? What can we do about ethical beliefs pertaining to the use of the Internet in academia that are not in accordance with the law (e.g., a belief that forwarding a digital copy of an article to a colleague is a good thing even if the publishing contract explicitly grants this right to the publisher only)? What are the main motives behind scholarly blogging? Which groups of scientists are more likely to self-cite their peer-reviewed publications in their blog posts? Can citations in Web 2.0 such as, for example, retweeting on Twitter and social bookmarking on Delicious serve as indicators of the academic impact of a particular scholar? Which bibliometric tools can a scholar rely on in order to keep track of all citations of his or her publications? Will scientific texts be soon written by computers instead of human beings? These are some of the most important issues addressed in 21 papers of the present interdisciplinary volume, which is concerned with the influence of the Internet on various scholarly practices in Germany and worldwide.

The editors of this volume are members of the Interdisciplinary Junior Researchers Group "Science and the Internet" at the Heinrich Heine University Düsseldorf, Germany.



d|u|p düsseldorf university press

d|u|p

d|u|p

science and the internet

# science and the internet

Edited by:

Alexander Tokar / Michael Beurskens / Susanne Keuneke / Merja Mahrt  
Isabella Peters / Cornelius Puschmann / Timo van Treeck / Katrin Weller

Science and the Internet

**d|u|p**



# Science and the Internet

Edited by:

Alexander Tokar / Michael Beurskens /  
Susanne Keuneke / Merja Mahrt /  
Isabella Peters / Cornelius Puschmann /  
Timo van Treeck / Katrin Weller

**d|u|p**

**Bibliographic Information published by the Deutsche  
Nationalbibliothek**

The Deutsche Nationalbibliothek lists this publication in  
the Deutsche Nationalbibliografie; detailed bibliographic  
data is available in the Internet at <http://dnb.dnb.de>.

© düsseldorf university press, Düsseldorf 2012

<http://www.dupress.de>

Cover Design: Juliane Ezold

Layout and Typesetting: Alexander Tokar

Printed and bound in Germany by

Druckerei Frick, Krumbach

ISBN 978-3-943460-16-2

# Acknowledgements

This volume would not exist without the generous financial support of our two main sponsors: the Düsseldorf University Strategic Research Fund and the German Research Foundation (DFG). The editors also wish to express their sincere gratitude to Professor Hans Süssmuth from Düsseldorf University Press for agreeing to publish this volume and answering all our questions concerning the preparation of the manuscript. Our thanks are also due to all the (numerous!) people who contributed to the success of the International Conference “Science and the Internet” (#cosci12), which underlies this volume. Special gratitude goes to Ralph Schroeder, René König, Isa Jahnke, Sascha Friesike, Ulrich Herb, Jean Burgess, Stefanie Hausteine, and Daniel Mietchen for serving as discussants in conference sessions. Their critical comments on all papers presented at #cosci12 have helped the authors finalize their contributions to this volume. Our thanks also go to our student assistant, Magdalena Amann, who (together with Doreen Pomsel, Ulrike Ezold, and Ramona Canu, whom we thank as well!) ensured the participants’ good mood during the conference and, after the conference, proofread the German contributions in this volume and designed the Table of Contents in this book. Last but not least, we would like to thank Juliane Ezold for designing the cover, and various editors at Papercheck.com for proofreading all English contributions in the volume. Needless to say, the editors take full responsibility for any remaining errors and shortcomings.



# Table of Contents

Science and the Internet: Introduction	1
<i>Alexander Tokar, Michael Beurskens, Susanne Keuneke, Merja Mahrt, Isabella Peters, Cornelius Puschmann, Timo van Treeck, and Katrin Weller</i>	

## **PART ONE**

### **The Internet and Scholarly Practices from a Cross-Cultural Perspective**

Online Trends from the First German Trend Study on Science Communication	13
<i>Alexander Gerber</i>	

From Analog to Digital Psychology: Results from Surveys on Information Behavior among German Psychologists between 1997 and 2010	19
<i>Hans Bauer, Gabriel Schui, and Günter Krampen</i>	

How and Why Do Turkish Scholars Use Social Networking Platforms?	31
<i>Selva Ersöz Karakulakoğlu and Övünç Meriç</i>	

Digital History in Portugal: A Survey	43
<i>Maria Cristina Guardado and Maria Manuel Borges</i>	

## **PART TWO**

### **The Internet and Teaching**

Informal Learning via Social Media—Preparing for Didactical Designs	59
<i>Isa Jahnke</i>	

Collaborative and Cooperative Text Production in Wikis <i>Claudia Bremer</i>	73
Fostering Crossmedia Literacy in Formal Educational Contexts: Conceptual Considerations and Case-Specific Results <i>Sandra Hofhues, Christian Geier, and Lena Griebhammer</i>	87
Belief(s) in eLearning – Zusammenhänge zwischen eLearning und Lehr-/Lern-Überzeugungen in Lehrportfolios <i>Timo van Treeck</i>	99

### **PART THREE**

#### **The Internet and Legal Issues**

Law: Friend or Foe in Scientific Internet Use? <i>Michael Beurskens</i>	115
Do Easily Copied Internet Media in the Library Lead to Plagiarism? <i>Gabriel Gomez</i>	131
Scientists and Librarians Create an Environmental Toxicology Data Repository <i>Deborah Keil and Kenning Arlitsch</i>	145

### **PART FOUR**

#### **The Internet and New Modes of Scholarly Communication**

Notes towards the Scientific Study of Public Communication on Twitter <i>Axel Bruns and Jean Burgess</i>	159
Scholarly Blogging: A New Form of Publishing or Science Journalism 2.0? <i>Cornelius Puschmann and Merja Mahrt</i>	171

Self-Citation of Bloggers in the Science Blogosphere 183  
*Hadas Shema, Judit Bar-Ilan, and Mike Thelwall*

Semantic Change of the Publication-Concept? 193  
*Alexander Tokar*

**PART FIVE**  
**The Internet and Scholarly Impact**

Citations in Web 2.0 209  
*Katrin Weller and Isabella Peters*

Google Scholar versus Google Scholar:  
 Among Publish or Perish, Scholarometer, and My Citations,  
 Which Citation Count Tool Is Telling Which Truth? 223  
*Ulrich "Tibaut" Houzanme*

The H-Index: What Is It, How Do We Determine It, and  
 How Can We Keep Up With It? 237  
*Timothy Ireland, Kathy MacDonald, and Peter Stirling*

**PART SIX**  
**The Internet and the Future of Science**

Publishing against the Machine: A New Format of  
 Academic Expression for the New Scientist 251  
*Adam Sofronijevic*

Developing Scientific Software: The Role of the Internet 263  
*Aleksandra Pawlik, Judith Segal, Helen Sharp, and Marian Petre*

Öffentlichkeit und Neue Medien: das Projekt „InsideScience“ 275  
*Thorsten Greiner, Jesús Muñoz Morcillo,  
 Caroline Y. Robertson-von Trotha, und Klaus Rümmele*



# Science and the Internet: Introduction

**Alexander Tokar, Michael Beurskens, Susanne Keuneke,  
Merja Mahrt, Isabella Peters, Cornelius Puschmann, Timo  
van Treeck, and Katrin Weller**

Interdisciplinary Junior Researchers Group “Science and the Internet,” Heinrich Heine University Düsseldorf

The present volume represents a selection of English and German papers from the International Conference “Science and the Internet,” which was held in August 2012 in Düsseldorf (Germany). The conference was organized by the Interdisciplinary Junior Researchers Group “Science and the Internet” (<http://nfgwin.uni-duesseldorf.de/>), which is comprised of the above-named authors of this introductory chapter and the editors of this volume.

The central theme of the conference was the influence of the Internet on various scholarly practices, such as gathering research data, finding relevant literature, presenting research results, communicating with colleagues and students, etc. (see <http://nfgwin.uni-duesseldorf.de/de/cosci12>). It should thus come as no surprise to readers that the volume begins with PART ONE, “**The Internet and Scholarly Practices from a Cross-Cultural Perspective**,” which addresses the changes in scholarly practices that have been brought about by the Internet thus far in a very straightforward manner. The first contribution in PART ONE is concerned with scholarly communication in Germany. In his article “Online Trends from the First German Trend Study on Science Communication,” Alexander Gerber demonstrates that the use of online social media remains a fairly marginal phenomenon among academics in Germany. Reasons for this circumstance, according to Gerber, range from a simple unawareness of such useful research-related Web sites as SlideShare to a fairly negative attitude toward such well-known social media services as Twitter. A somewhat similar conclusion (i.e., that no significant changes have been made so far) is also arrived at in the second contribution of PART

ONE, the article entitled “From Analog to Digital Psychology: Results from Surveys on Information Behavior among German Psychologists between 1997 and 2010” by Hans Bauer, Gabriel Schui, and Günter Krampen. Among other things, the authors assert that in 2010, established subscription-based journals remained the main source of information for German psychologists, whereas open access journals as well as preprint and document servers were rarely used for the purpose of finding relevant research literature. In contrast to the papers by Gerber and Bauer et al., the other two contributions in PART ONE have a narrower focus. The article “How and Why Do Turkish Scholars Use Social Networking Platforms?” by Selva Ersöz Karakulakoğlu and Övünç Meriç compares the attitudes of a group of Turkish natural and social science scholars towards the use of social media in research and teaching. In stark contrast to their expectations, the authors discovered social science scholars in Turkey to be much more positive about social media than their counterparts in the natural sciences. While the latter are of the opinion that social networking Web sites do not support traditional education methods, the former regard them as a natural tool for both teaching and research. The article “Digital History in Portugal: A Survey” by Maria Cristina Guardado and Maria Manuel Borges raises the question as to the extent to which digital tools, such as online prosopographical databases, are utilized in history research projects in Portugal. After analyzing the Web sites of 13 history research units in the country, the authors conclude that the use of digital tools is currently on the rise among historians in Portugal.

PART TWO, “**The Internet and Teaching**,” deals with the opportunities that the Internet offers for teaching in academia. Contrary to the view held by the aforementioned natural scientists from Turkey, Isa Jahnke contends in her article “Informal Learning via Social Media—Preparing for Didactical Designs” that the use of social media in teaching can produce better learning outcomes due to the informal learning potential inherent to social media (as well as the Internet in general). That is, social media utilized in a teaching context can trigger unplanned learning, in the course of which students can find solutions to a number of problems (related to what they are doing in class) outside of formal instruction given by their teachers. The article contains a description of three recent teaching projects that aimed to achieve better learning outcomes through informal learning via social media. The question of whether the use of Web 2.0 tools can improve teaching in academia is also (somewhat implicitly) addressed in the second contribution in PART TWO, Claudia Bremer’s article “Collaborative and Cooperative Text Production in Wikis.” Her study examined several university-related wikis launched and maintained by students of Frankfurt am Main University. The

main finding is that similar to top-down wikis initiated by university instructors, bottom-up wikis initiated by students are also characterized by a fairly slow rate of voluntary text contribution. In other words, students contribute to wikis only when they are required to do so, such as to earn credit points for a course whose instructor insists on using a wiki. A more optimistic conclusion can be found in the article “Fostering Crossmedia Literacy in Formal Educational Contexts: Conceptual Considerations and Case-Specific Results” by Sandra Hofhues, Christian Geier, and Lena Griebßhammer. This contribution focuses on the use of crossmedia (i.e., several media formats: text, audio, and video) in teaching at the university level as a means of promoting media literacy among students. After outlining barriers impeding the use of crossmedia in teaching, the authors proceed to describe their own teaching experiences with crossmedia in the seminar “Crossmedia Ethnological Communication,” taught at the University of Augsburg in Winter 2011/2012, whose participants were required to work with three media formats: text (blog), audio (podcast), and video. Despite the usual students’ criticism that too much work was required for a relatively small number of credit points, the course was by and large positively evaluated by its participants. Finally, the last contribution in PART TWO, Timo van Treeck’s article “Belief(s) in eLearning – Zusammenhänge zwischen eLearning und Lehr-/Lern-Überzeugungen in Lehrportfolios,” asks if there is a connection between university instructors’ educational beliefs and e-learning scenarios practiced by them in their lectures and seminars. In other words, is a university instructor who conceptualizes teaching as “giving” knowledge to students likely to solely upload his or her course materials—such as lecture notes—on the Internet? In contrast, are more interactive e-learning scenarios (e.g., those that involve the use of collaborative text production) more likely to be implemented by university instructors with more collaborative educational beliefs, i.e., those that emphasize social interaction in class and, in particular, students’ independent acquisition of knowledge? On the basis of his analysis of 31 teaching portfolios created by participants of didactical workshops conducted at various German universities between 2004 and 2011, van Treeck argues that e-learning scenarios do indeed to a very large extent depend on university instructors’ conceptions of teaching, especially of their own role as educators.

PART THREE, “**The Internet and Legal Issues,**” begins with the article “Law: Friend or Foe in Scientific Internet Use?” by Michael Beurskens. This contribution demonstrates that ethical beliefs in academia and legal issues very often do not go hand in hand. For example, while in academia, plagiarism is regarded as one of the most severe violations of the rules of good

scientific practice, the law, as pointed out by Beurskens, largely ignores this issue. For instance, the concept of self-plagiarism is unknown to the law. Also, the practice of sending digital copies of an article to colleagues via e-mail is considered acceptable, ethical behavior in academia, despite the fact that publishing contracts typically assign this right to the publisher only, thereby making this practice illegal. An interesting point that Beurskens makes is that instructors in the Faculty of Law at Düsseldorf University believe that their illegal teaching-related actions—such as providing their students with a broad selection of teaching materials (usually in the form of online readers)—are covered by certain exceptions to copyright law and thus are not illegal. While the paper by Beurskens focuses on several Internet-related issues (plagiarism, reuse of content in teaching, open access, and privacy), the second contribution in PART THREE is devoted exclusively to plagiarism. As the title “Do Easily Copied Internet Media in the Library Lead to Plagiarism?” makes clear, Gabriel Gomez’s article attempts to determine whether readily accessible information on the Internet is one of the causes of plagiarism among students in U.S. colleges and universities. As Gomez conjectures, a student who is accustomed to casual copying practices on the Internet (e.g., in the context of an e-mail message or a Facebook post) may easily include copied information in an assignment, thereby committing plagiarism. In addition, according to Gomez, plagiarism can arise from certain misconceptions among students about the nature of information on the Internet and the reasons for citing other people’s work in academic publications. With regard to the former, some students believe that any information that is freely available on the Internet (including information that is only accessible because the university library has paid for it) does not belong to anybody and hence does not need to be cited. As for the latter, some students are simply unaware of the fact that citing in academic publications is a requirement for more than just ethical reasons. The list of references at the end of an article also has a very important practical function, as it serves to familiarize readers with other relevant publications in the same research field. The last contribution in PART THREE, the article “Scientists and Librarians Create an Environmental Toxicology Data Repository” by Deborah Keil and Kenning Arlitsch, is not a classic research paper, but rather an outline of a research proposal aimed at developing an environmental toxicology data repository for the Wasatch Front, a metropolitan region in the state of Utah in the U.S. The article begins with a description of both the geographical and industrial peculiarities of this region, necessitating the creation of such a repository. Thus, according to the authors, a number of factors peculiar to the Wasatch Front (e.g., the presence of major extractive industries and military installa-

tions) contribute to environmental pollution in this area, which, in turn, is very likely to account for significant increases in lung and bronchial cancer, as well as above-average incidences of asthma and autism in the Wasatch Front. As Keil and Arlitsch point out, the major shortcoming of most American environmental toxicology data repositories is that they cover the entire U.S. and thus do not take into account regional peculiarities, such as those of the Wasatch Front. The authors suggest that the new regionally focused data repository that they propose for the Wasatch Front area will help to address many of the health issues listed above and serve as a model for other regional data repositories in the U.S.

It is a well-known fact that the Internet has given rise to new modes of academic communication. One prominent example is Twitter, which is the topic of the first article in PART FOUR, “**The Internet and New Modes of Scholarly Communication.**” The article “Notes towards the Scientific Study of Public Communication on Twitter” by Axel Bruns and Jean Burgess is a methodological paper that describes how public communication on Twitter can be studied. One important approach is hashtag analysis, which has become popular in recent years. As Bruns and Burgess point out, the hashtag-based approach of categorizing tweets along with some other important metrics provided by Twitter (e.g., the status of a tweet: an original tweet, a retweet, a reply, or a tweet containing a URL) can offer important information about interaction patterns during a particular event (such as the most active users discussing the event and the peak of interactive activity during the event). In addition, a researcher can resort to user metrics, which, among other things, can help systematize the typology of Twitter users on the basis of their preferred interactive patterns (i.e., e.g., whether they mainly post original tweets or retweet other users’ tweets). The next two contributions in PART FOUR focus on blogging in academia. In an attempt to establish the main motives behind scholarly blogging, Cornelius Puschmann and Merja Mahrt, the authors of the second paper “Scholarly Blogging: A New Form of Publishing or Science Journalism 2.0?,” conducted a Web-based survey of scholarly bloggers active on the platform SciLogs (<http://www.scilogs.de/>), which contains over 60 blogs. The majority of respondents were found to regard blogging as a means of presenting their fields of research to a general public, rather than as a platform for debating a specific aspect of their research with a relatively small network of peers. This perspective probably arises from the fact that the respondents do not regard their blogs as appropriate outlets for original research. Thus, Puschmann and Mahrt conclude that scholarly blogging has thus far failed to replace traditional scholarly publishing, and it is far from being clear what role, if any, blogging will play in

academia in the future. An answer to this question raised by Puschmann and Mahrt can be found in the following contribution in PART FOUR, the article “Self-Citation of Bloggers in the Science Blogosphere” by Hadas Shema, Judit Bar-Ilan, and Mike Thelwall. The authors demonstrate that blogging about one’s publications (i.e., self-citing them in a scholarly blog) often results in a noticeable increase in the number of downloads of the full texts of these publications (provided that they are freely available on the Internet). The primary objective of the study by Shema et al. was, however, to identify the defining characteristics of self-citers in scientific blogs. For this purpose, the authors analyzed blog posts from the blog aggregator Research Blogging (<http://researchblogging.org/>) and found the average self-citer to be a male who has earned a doctorate and is affiliated with a university or a research institute. With regard to academic disciplines, the largest number of self-citers stems from the field of computer science, although mathematics is credited with the largest number of self-citing blog posts. PART FOUR ends with the article “Semantic Change of the Publication-Concept?” by Alexander Tokar. This paper focuses on academics’ publications lists, which, in contrast to public communication on Twitter and scholarly blogging, cannot be regarded as a new mode of scholarly communication, as they have existed in academia long before the Internet era. Tokar argues, however, that online-based publications lists differ from their pre-Internet counterparts in that the former do not only list scholars’ publications, but also provide digital access (usually in the form of downloadable PDF files) to various publications that they list. Another important difference is that online-based publications lists sometimes include unpublished manuscripts, which, unlike articles in refereed journals and edited volumes, do not qualify as academic publications in the traditional meaning of the term “publication.” Tokar hypothesizes that the concept of academic publication is currently undergoing a semantic change: While a traditional scholarly publication is defined by the semantic feature [quality control through other experts in the appropriate research field] (hence the term “peer review”), an unpublished manuscript is defined by the semantic feature [quality control through the author of the manuscript only].

The point of departure of PART FIVE, “**The Internet and Scholarly Impact**,” is the article “Citations in Web 2.0” by Katrin Weller and Isabella Peters. Its central claim is that scholarly communication in Web 2.0 exhibits several activities that resemble citing in traditional academic publications and thus may also be worthy of consideration as an indicator of the impact of a particular scholar. Examples analyzed by Weller and Peters include retweeting on Twitter, which can be analogized to quoting in traditional publications; bookmarking on social bookmarking services, such as Delicious,

which can be considered an indicator of the interest of a research community in a particular publication / the work of a particular scholar; and citing in the context of a scholarly blog. One of the most interesting findings of this study is that social bookmarking systems cover 28 percent more articles than Scopus, a well-known database of citations for articles in academic journals. According to Weller and Peters, this fact suggests that users of social bookmarking services create a more holistic view of academic authors by means of bookmarking. It is also worth noting that Internet users clearly reward more products of scholarly practice than traditional citations databases, such as Scopus, do. Thus, in addition to academic articles, numerous scholarly blogs (which are, of course, missing from Scopus) are bookmarked. The next contribution in PART FIVE, the article “Google Scholar versus Google Scholar: Among Publish or Perish, Scholarometer, and My Citations, Which Citation Count Tool is Telling Which Truth?” by Ulrich “Tibaut” Houzanme, focuses on measuring scholarly impact with the help of the three tools listed in the title of the article, all of which make use of Google Scholar data. In an attempt to determine which of these tools is “telling which truth,” the author conducted a citation analysis of the thirty most influential information scientists (fifteen from the U.S. and fifteen from the UK). According to Houzanme’s research, Scholarometer proved to be the best tool of the three tools under investigation, followed by My Citations and Publish or Perish. Noteworthy is the fact that Publish or Perish, according to Houzanme, lags considerably behind both Scholarometer and My Citations, as it often requires disambiguation with regard to disciplines and authors’ names. The last contribution in PART FIVE, the article “The H-Index: What Is It, How Do We Determine It, and How Can We Keep Up With It?” by Timothy Ireland, Kathy MacDonald, and Peter Stirling, deals with a well-recognized aspect of information science—the h-index. In particular, the authors suggest an approach to determining the h-index of an individual scholar on the basis of the information contained in different citations databases—Web of Science, Scopus, and Google Scholar. It is well known that the h-index has become a crucial factor in determining tenure and promotion in many academic disciplines. Accordingly, junior researchers in untenured positions in particular may wish to keep track of how often their publications are cited by other researchers in order to be able to update their h-index. The authors of the article propose a citations tracking system meant to enable academics to successfully cope with this task.

The last part of the volume, “**The Internet and the Future of Science,**” dares to explore potential scientific practices of the future. The first contribution in PART SIX, the article “Publishing against the Machine: A New Format

of Academic Expression for the New Scientist” by Adam Sofronijevic, is essentially a plea for scientists to start cooperating with machines (more actively!) in such areas as text production and text comprehension. The paper begins with the interesting observation that the continuous doubling of the corpus of world knowledge has not yet given rise to dramatic changes in scientific communication. In this regard, the article reminds readers of similar claims made in the articles by Gerber and Bauer et al. in PART ONE of this volume. However, what Sofronijevic focuses on here has far broader implications for scholarly practices in the future. Sofronijevic argues that despite the presence of computers and the Internet, human beings are still the ones to produce scientific texts, which, according to Sofronijevic, is deplorable, given that machines are already capable of assisting us with text production. Indeed, conference abstracts and even texts such as the present introductory chapter consist to a large extent of formulaic expressions, such as *this paper argues that...*, *the author raises the question as to whether...*, *the results obtained corroborate our hypothesis that...*, *in stark contrast to the findings of...*, which can easily be added to the database of a machine. It is very likely that in the future, such scientific texts will be produced, at least in part, by machines rather than human beings. The topic “Machines as Scientists’ Practical Assistants” is continued in the next contribution in PART SIX. As reported in the article “Developing Scientific Software: The Role of the Internet” by Aleksandra Pawlik, Judith Segal, Helen Sharp, and Marian Petre, the authors interviewed 27 scientists developing software for various research-related purposes. (These scientists are not professional software developers!) Of 27 interviewed scientists, 24 acquired software-developing skills nearly exclusively by teaching themselves on the Internet. A related finding of the qualitative interviews conducted by the authors is that the Google search engine serves as the most important source of information of which the interviewed scientists make use when they need to fill in gaps in their software development knowledge (such as when dealing with a software bug). Linking these results with the main tenor of Sofronijevic’s paper, one might exclaim: *Yes, we have computers and the Internet, but human beings are still the ones to develop scientific software.* The final contribution in PART SIX addresses the use of videos in science. The article “Öffentlichkeit und Neue Medien: das Projekt „InsideScience“” by Thorsten Greiner, Jesús Muñoz Morcillo, Caroline Y. Robertson-von Trotha, and Klaus Rümmele describes the project “InsideScience” at the Karlsruhe Institute of Technology. The aim of the project is to familiarize the general public with the work of two research groups (also affiliated with the Karlsruhe Institute of Technology)—“Computational Particle Physics” and “Humanoid Robots: Learning and

Cooperating Multimodal Robots”—which, like the project “InsideScience,” are both funded by the German Research Foundation (<http://www.dfg.de>). The essence of “InsideScience” is that junior researchers from the two aforementioned research groups make videos in which they attempt to explain to laypeople (in an accessible manner) what the projects “Computational Particle Physics” and “Humanoid Robots” are actually about. These videos are then posted to popular video hosting platforms, such as YouTube and Vimeo. This openness of “InsideScience” is indicative of an approach to presenting research to the general public that is likely to gain greater popularity in the near future.

Nearly all of the papers in this volume end with a brief summary of the remaining research questions that arise in connection with what has been discovered by the authors. The editors of the present volume thus hope that our book will instigate further important contributions to what will hopefully soon become a new established interdisciplinary field of study—“Science and the Internet.”



## **PART ONE**

---

# **THE INTERNET AND SCHOLARLY PRACTICES FROM A CROSS- CULTURAL PERSPECTIVE**



# Online Trends from the First German Trend Study on Science Communication

**Alexander Gerber**

German Research Centre for Science & Innovation Communication  
(innokomm)

*Communication both among scientists and between science and society is undergoing a paradigm shift that is fundamentally redefining the communicative requirements. This change is driven by the horizontalization of information, with the Internet as the enabling technology. However, technology only sets the framework, whereas the real challenges and solutions are deeply rooted in the culture and in the system of knowledge creation itself.*

## Introduction

If we divide the development of science communication into four phases, each with an approximately 15-year span, we are today witnessing the beginning of a fifth phase: After the Age of Utopia and Science Fiction in the 1950s and early 1960s, the Age of Enlightenment in the 1970s, the Era of Acceptance Building until the mid-1990s, and recently, the PUSH paradigm, policy-makers today advocate for a participatory and truly dialogic approach toward communication, seeking greater transparency and even the ideal of a “scientific citizenship.” The challenges emerging from this new paradigm were addressed in the first German trend study on science communication—a comprehensive survey conducted between September 2009 and October 2011 among 326 science journalists and Public Relations (PR) managers, scientists, and communication researchers, leading to an adjacent two-stage Delphi

study with 30 renowned experts, researchers, and practitioners who aimed to find answers to the challenges expressed by the community.

The findings (Gerber, 2011a) show trends in decreasing salaries, revenues, and media coverage on the one hand and increasing PR resources and direct online communication by scientists on the other. The four main trends identified and investigated within the study were as follows:

- 1) Influences of social media on science PR, science journalism, and scientific communication.
  - i) Science journalism is moving from classic storytelling and translating to moderating and investigating.
  - ii) The increasing demand in social media consulting and training in science PR is a clear sign of an upcoming second wave of professionalization among communicators.
  - iii) Scientists witness paradigmatic shifts both in expectations as well as in self-perception. Public science and open science will increasingly flatten scientific hierarchies (“sharing” vs. “status”).
  
- 2) Transparency through open science and citizen science create promising opportunities for renewed “cultures of communication,” which then could foster scientific citizenship.
  - i) If practitioners in science communication are serious about involving the entire spectrum of civil society in terms of a scientific citizenship, they will need to abandon their main focus on research results. New formats will switch more authentically between results- and process-oriented perspectives.
  - ii) Science PR / science media / science education have to make the relativity of the knowledge they create more transparent. New media can serve as an enabling technology for this approach.
  
- 3) Fundamental changes both in the self-perception of science communicators as well as the required qualification for professional science communicators in their new roles as “mediators.”
  - i) Sustainable science communication in the fifth development stage will exceed the previous approaches of the “packaging industry.”
  - ii) The communicator’s performance will increasingly be measured by whether he or she succeeds in assembling the pieces of the transdisciplinary puzzle within the institution instead of merely disseminating project results or cultivating a brand.

- iii) In lieu of squeezing news through narrowing information channels, PR will need to “agenda-surf” on socio-political waves. As a consequence of PR professionalization, it is becoming more and more important to communicate cooperatively.
- 4) Obstacles of a change-resistant scientific system that hardly incentivizes “real” outreach at all, leading to challenges such as measuring or norming communication impact.
  - i) Today’s criteria for research funding, evaluation, and the appointment of senior positions mostly disregard communication achievements. New systems of incentivizing outreach to the lay public will be considered. Such catalysts will foster a broad change in communication culture.
  - ii) Hereby, one of the major challenges and success factors lies in the ability to measure and compare communication performance, especially regarding social media activities. This will lead to a debate about norming and standardizing these new measurements. For the short term, specific communication audits will be sufficient.

By investigating these trends, the study could answer several of the challenges that were expressed beforehand by the community and were based on further statistical analyses, e.g., the decreasing sold circulation of popular science magazines in Germany by an average of 30 percent within the last 10 years (Gerber, 2011b).

## **Challenges in the Fifth Phase of Science Communication**

As high as the change potential of Internet technologies might be in academia, the vast majority of both scientists and science communicators mostly react hesitantly, at best, to the new opportunities. In compliance with the findings of Bader et al. (2012), Dernbach and Schreiber (2009), Kleinmann et al. (2008), and Procter et al. (2010), the yet unpublished results of a comprehensive Web technology use and needs analysis conducted between February and June 2012 in nine German research institutes (Gerber, 2012) leave no doubt that the use of interactive and social media in academia is still a niche phenomenon. (The study combined an online survey among 617 participants with 30 in-depth interviews with decision-makers and standardized Web analyses of nine scientific institutions.)

Among the 14 most common online tools surveyed, ten were widely unknown (ranging from 35.8% publicity in the case of ResearchGate to 16.1% in the case of SlideShare). The best-known tools, however, were also those that were rejected by the majority of researchers, e.g., Twitter, which was rejected by 80.5% of the respondents.

In contrast to the assumptions of the aforementioned analysis, the willingness of researchers to use certain Web technologies does not automatically increase with the sheer publicity of these technologies. Whereas scientists are similarly familiar with ResearchGate und iTunes University (35.8% and 37.2%, respectively), the ratio between opponents and users of these platforms is more than twice as high in the case of the Apple platform (4.08) as in the case of ResearchGate (1.92); i.e., the degree of refusal is significantly more distinct. The only conclusion from this appears to be that sheer education or information about the new possibilities has not solved and will not solve the problem of slow adoption in academia. Nonetheless, the lack of orientation becomes obvious in the fact that 73.7% of the researchers interviewed explicitly expressed a demand for training in social media.

In compliance with these figures, the aforementioned Delphi study (Gerber, 2011a) showed clearly that a diffusion of Web 2.0 tools in the academic world is only partly a question of technology acceptance. For instance, online communication is still not taken into account in most cases of evaluations or allocations of research funding. The Delphi experts, therefore, demanded a critical discourse about possible incentives for scientists to reach out into the online world.

If online outreach, however, became a relevant criterion for academic careers, we would have to find more empirically sound ways to measure, compare, or even standardize and audit the impact of such outreach.

There is a widening gap between the masses of scientists communicating “ordinarily” on the one hand and the very few cutting-edge researchers and (mostly large and renowned) institutions experimenting extensively with the new online opportunities on the other. Therefore, it will be a major challenge for science communication scholars to provide appropriate advice and practical orientation.

Furthermore, the study investigated how an increase in transparency through open science could lead to a new culture of science and scientific communication. Such a dialogue would have to complement classic approaches of dissemination to engage with a much wider spectrum of societal groups than the dissatisfying spectrum that has been reached up to now. For instance, accordant studies over the last decade have unmistakably shown that civic scientific literacy within the European Union reaches no more than

an average of 13% (Miller, 2012). The obvious conclusion from this is that the supposed “social contract” between science and society has not been fulfilled through the PUSH paradigm.

Scientific institutions, more than ever, have difficulties in convincing lay people of the importance of their role in society. To avoid becoming easy prey for governments that might want to reduce state deficits, the aim of widespread scientific citizenship is far more than an academic endeavor.

One approach suggested by the Delphi experts is trying to distort the process of knowledge creation much less than has been done both in science PR and science journalism. Eighty percent of the experts see a major deficit in science being reduced to research results, whereas the process and mechanisms of innovation and scientific achievement remained unclear. This “science in the making” seems to be the hardest to convey by the media, even though scientists and science PR experts regard this aspect as the most important in inspiring young people to enter a career in research and development. Science blogging was mentioned by many of the experts as a possible solution due to its authenticity.

Taking all of this into account, the main challenges seem to be much less technological than they are cultural and systemic. Possible solutions, therefore, do not lie primarily in the tools or in the strategies but in the systemic frameworks of knowledge creation and dissemination as we have practiced them for decades, if not centuries. Permeating “open science” under closed paradigms can succeed only if the embedding communication frameworks are adapted. This will include just as many new forms of impact measurement, recognition, and qualification, as it will include obvious solutions from the archaic toolbox of the Enlightenment phase. The scientific community, however, still has a long way to go to properly understand the causes, effects, and solutions for this cultural change.

## References

- BADER, A., FRITZ, G., & GLONING, T. (2012). *Digitale Wissenschaftskommunikation 2010-2011—Eine Online-Befragung*. Gießen: Gießener Elektronische Bibliothek. Retrieved November 10, 2012 from <http://tinyurl.com/ajz8z7t>
- DERNBACH, B., & SCHREIBER, P. (2009). *Wissenschaft evaluiert*. Unpublished paper presented at Forum Wissenschaftskommunikation, Berlin.
- GERBER, A. (2011a). *Trendstudie Wissenschaftskommunikation: Vorhang auf für Phase 5. Chancen, Risiken und Forderungen für die nächste Entwicklungsstufe*

- der Wissenschaftskommunikation*. Berlin: edition innovare / innokomm Forschungszentrum.
- GERBER, A. (2011b). Auflagenentwicklung in Wissens- und Wissenschaftsmedien zwischen 2000 und 2011. Presentation. Retrieved November 21, 2012 from <http://tinyurl.com/3bwfl7t>
- GERBER, A. (2012). Social Media in Science. Nutzungs- und Potenzialanalyse sozialer Medien in der Wissenschaft (Research project). Berlin: innokomm Forschungszentrum. Retrieved November 21, 2012 from <http://tinyurl.com/agr6xsm>
- KLEINMANN, B., ÖZKILIC, M., & GÖCKS, M. (2008). Studieren im Web 2.0 (Project report). Hannover: HIS GmbH. Retrieved September 21, 2012 from <https://hisbus.his.de/hisbus/docs/hisbus21.pdf>
- MILLER, J. D. (2012). The sources and impact of civic scientific literacy. In M. BAUER, R. SHUKLA, & N. ALLUM (Eds.), *The culture of science* (pp. 217-240). New York: Routledge.
- PROCTER, R., WILLIAMS, R., & STEWART, J. (2010). If you build it, will they come? How researchers perceive and use Web 2.0 (A Research Information Network report). London: Research Information Network. Retrieved November 10, 2012 from <http://tinyurl.com/ae6a8uv>

# **From Analog to Digital Psychology: Results from Surveys on Information Behavior among German Psychologists between 1997 and 2010**

**Hans Bauer, Gabriel Schui, and Günter Krampen**

Leibniz Institute for Psychology Information (ZPID), Trier

*Data from four surveys on information behavior of German psychologists conducted between 1997 and 2010 are compared. Results indicate that information behavior of researchers does not seem to have changed fundamentally, with the main focus being on efficiently accessing all pertinent publications. Differences in information needs and behavior between researchers and practitioners point out the importance of considering differing outlooks and available resources regarding scholarly information.*

## **Introduction**

The use of electronic communication media has become virtually universal in modern societies over the last decades. Scholarship is of course no exception to this. For example, digitalized communication holds the promise of increasing quantity and quality of research results by speeding up and extending access to resources and by facilitating cooperative work (Dutton & Jeffreys, 2010).

At the same time, it is equally apparent that use of electronic media differs among academic disciplines. Roughly speaking, adoption has so far been much more pervasive in the natural sciences than in the arts and humanities, with social sciences roughly in between (Education for Change, 2002). Even

within those broad categories, there are differences in the ways electronic media are used (Kling & McKim, 2000).

Epistemic and social foundations of disciplines have been identified as the reasons for such divergences (e.g., Kling & McKim, 2000). These works generally follow the notion that researchers in the natural sciences have a more unified understanding of their discipline and cooperative work is much more commonplace than in more “individualist” disciplines (Becher & Trowler, 2001).

It has thus been cautioned that to better understand the big picture, discipline-specific developments in electronic scholarly communication have to be considered (Cronin, 2003). Psychology is a special case in that it encompasses both mechanistic and subject-oriented ways of theorizing; and even though mainstream psychology is now strongly oriented toward the natural science paradigm, it is far from being a unified field, with a multitude of micro-theories coexisting and competing (Lück et al., 2011).

So far, the question of whether and how psychologists have adapted their information behavior in the digital age does not seem to have been investigated. The Leibniz Institute for Psychology Information (ZPID) provides information for the psychology community in German-speaking countries. To adapt its services appropriately, it has been conducting surveys about information needs and behaviors of the community. In the present contribution, results of four surveys sampled between 1997 and 2010 are compared to draw inferences about the transition to digital scholarship in psychology. However, it must be cautioned in advance that the surveys were not designed for time comparison. Therefore, the present study can only serve as a first exploration and basis for further investigation.

## Methods

### Survey Samples and Their Comparability

Core features of survey samples and prior publications referring to the four surveys analyzed are tabulated in Table 1. Only surveys #1.1 and #1.2 were explicitly designed for comparison purposes (i.e., items are identical). Because sampling dates are reasonably close, the corresponding samples can be meaningfully pooled into the composite survey sample #1 (middle column). Its target population (researchers) and composition of academic positions roughly resemble those of survey #3. By contrasting surveys #1 and #3, changes in researchers' information behavior from around the end of the millennium to 2010 can thus be evaluated.

Survey #2, which was administered in 2003-2004, differs starkly from the others in terms of the target group and gender composition. Therefore, no straightforward comparison with regard to changes over time can be made. However, it offers the interesting possibility of contrasting information behavior among psychologists engaged in research versus (mostly clinical) practice. Therefore, the sample was retained, even though interpretation must remain tentative because survey date and target population are confounded.

### **Comparability of Survey Items**

Surveys #1.1 and #1.2 contained the same set of 25 items pertaining to the frequency of use of different information sources. Scale endpoints were labeled “never” versus “very often.” Surveys #2 and #3 each contained a semantically similar set of items on the use of information sources, with endpoints “never” versus “weekly” and “never” versus “all the time,” respectively. These can be regarded as sufficiently similar to allow comparisons to the items in #1.

However, between the surveys, the particular items contained in the information source item sets only partially overlapped. For example, survey #1 asked about the use of bookstores as a means of procuring information, but surveys #2 and #3 did not. Conversely, survey #3 asked about social networking platforms, while surveys #1 and #2 did not. As the last example shows, the change in information sources that were included in the surveys is, to a good degree, due to the increasing importance of digital communication.

Surveys #2 and #3 each included a set of items asking for the subjective importance of each of several properties of an “information service” (#2) or a “database or other information source” (#3) such as up-to-dateness and ease of use, both with scale endpoints “unimportant” versus “very important.” Table 2 lists categories that served as a conceptual guideline for comparing the frequency of use and importance of information service property items. The actual items falling under these categories are reported in the results section along with their respective scores. (A full list of items in all surveys, including exact wordings and additional descriptive statistics, can be obtained from the corresponding author [hans.bauer@zpid.de](mailto:hans.bauer@zpid.de).)

All scales for frequency of use and information service property items were transformed to a percentile scale to compare means and to facilitate interpretation. (Percentile values reported in the results section thus represent means on a scale interpretable as “percentage of the original scale maximum.”) Also, because partially different sets of information sources and information service properties were asked about in the different surveys, rank

Table 1

Core features of survey samples and prior publications referring to surveys						
	Survey #1.1	Survey #1.2	Survey #1	Survey #2	Survey #3	
Sampling date	1997	1999	–	2003-2004	2010	
Target population	Senior members in German universities' psychology depts.	Junior members in German, Austrian, and Swiss universities' psychology depts.	–	Members of German psychology practitioners' association (BDP)	Members of German psychology researchers' association (DGPs)	
Sampling method	Full sampling	Austria, Switzerland: Full population; Germany: every third person in list	–	Random sampling of 1,500 persons from BDP members' directory	Full sampling of 10 of 15 DGPs sections (representing all 15 sections due to multiple memberships)	
Survey format	Postal	Postal	–	Postal	E-mail invitation to Web questionnaire	
Valid responses (response rate)	265 (48%)	221 (36%)	486 (41%)	324 (22%)	298 (17%)	
Academic titles / positions <sup>a</sup>	18% PD, 82% FP	45% Postgrad, 48% Postdoc, 7% PD	21% Postgrad, 22% Postdoc, 13% PD, 44% FP	Not inquired	12% Postgrad, 26% Postdoc, 17% PD, 39% FP, 7% missing	
% female <sup>b</sup>	16	32	23	67	38	
Mean age	missing	36	missing	44.0	42.9	
Used in <sup>c</sup>	A, B	C	–	D, E	F, G, H	

Notes. <sup>a</sup> Abbreviations used: PD – Privatdozent (roughly comparable to assistant professor); FP – Full Professor. <sup>b</sup> Gender missing < 3 % in all samples.

<sup>c</sup> A – Montada, Krampen, & Burkard (1999); B – Krampen, & Montada (2000); C – Neppi, Wiesenhütter, Krampen, & Montada (2001); D – Becker (2004);

E – Krampen, Becker, Labouvie, & Montada (2004); F, G, H – Krampen, Fell, & Schui (2011, 2012a, 2012b).

Table 2. List of comparison categories for frequency of use of information sources and subjective importance of information service properties

Information sources	Library, literature databases, publication contents, WWW
Information services	Up-to-dateness, quality assurance, internationality, cost, ease of use, search speed, workflow integration, full text access

values are provided as a rough indicator of item score relative to other items in the respective survey.

Finally, surveys #1.1 and #3 contained open-ended questions about “desired improvements in PSYNDEX” (i.e., a psychology literature database produced by the ZPID; see <http://www.zpid.de>) (#1.1) and “typical difficulties while searching for information” (#3). Responses to these were crudely evaluated to identify potential shortcomings in information resources offered to the respondents at the time they were surveyed.

## Results

### Use of Information Sources

In the composite survey #1, libraries as a means of finding information scored on average 71% (on a scale from “never” to “very often”), ranking second in frequency of use among all 25 sources that were inquired about. Survey #3 did not ask about overall use of libraries, but instead about online library catalogs, which also ranked quite high (third of 32 sources) with a score of 61%. Direct access to the library scored 84% on a scale from “unimportant” to “very important” in survey #3. In contrast, mean use of either online library catalogs or local libraries was only 40% (ranking 14th / 15th among 31 sources) in the practitioner survey #2, on a scale from “never” to “weekly.” (For the frequency of use data, width of 95% confidence intervals for means ranged from 3.7% to 6.6% in survey #1, from 2.8% to 7.9% in survey #2, and from 3.2% to 8.7% in survey #3.)

All three surveys asked about the use of two psychology-focused literature databases: PsycINFO (with a strong focus on English language publications) and PSYNDEX (indexing publications with at least one coauthor from the German-speaking countries). In the late 1990s researchers’ survey #1, the most commonly used format was CD-ROM, whereas PSYNDEX and PsycINFO scored 63% (seventh) and 67% (fourth), respectively. Online

versions were, at this time, used less frequently: 33% for both databases. Surveys #2 and #3 did not ask about database format.

In the 2010 researchers' survey #3, as compared to the late 1990s researchers' survey #1, use of PSYNDEX had somewhat declined (57%, ranking sixth), while PsycINFO was used somewhat more often (73%, second). In the 2003 practitioner survey #2, database use in general was markedly lower, but PSYNDEX was being used more often: 25% (17<sup>th</sup>) for PSYNDEX and 11% (21<sup>st</sup>) for PsycINFO. Finally, use of unspecified "other" databases (not inquired about in survey #2) increased from survey #1 to #3: 38% (CD-ROM; 15<sup>th</sup>) to 45% (12<sup>th</sup>).

In both researchers' surveys, journals ranked first among the inquired information sources: In survey #1, which asked about browsing publications, browsing of "the top journals in your field" was rated 75%. In survey #3, use of online journals (not specifying in what way, e.g., browsing or chaining) scored 88%. Unspecified journal use was also fairly common among practitioners (79%), but only ranking sixth. Instead, books (87%) were rated as the most heavily used source of information. In survey #1, browsing of "many books" as a means of getting information scored 58% (ninth).

Only survey #3 differentiated between electronic and print versions of publications. With 49% (10<sup>th</sup>), print journals were used markedly less frequently than online journals by researchers in 2010. Use of open access online journals, which was asked for separately, was rated 56% (seventh). Thus, use of open access journals was less frequent than of journals in general.

The change in survey focus is especially apparent in questions related to use of WWW resources as information sources. Whereas survey #1 only inquired about unspecified "search on the Internet (WWW)," which was rated 48% (10<sup>th</sup>), survey #3 included 14 WWW-related sources. Most commonly used among them were "conventional search engines" (59%, fifth) and Google Scholar (56%, seventh). With 41% (15<sup>th</sup>), Google Books was used less frequently. Usage score was 44% for authors' Web sites (13<sup>th</sup>), 39% for unspecified discipline-specialized Web sites (16<sup>th</sup>), and 25% for document / preprint servers (24<sup>th</sup>). Research-oriented social networking platforms and (micro-)blogs were hardly used at all (8% / 30<sup>th</sup> and 7% / 31<sup>st</sup>, respectively).

Survey #2 included five WWW-related items. For practitioners in 2003-2004, use of unspecified "Internet search engines" was fairly common (73%, seventh), which was even more true for Google (82%, fourth). Web sites pertinent to professional activity scored 56% (eighth).

## Subjective Importance of Information Service Properties

Surveys #2 and #3 asked about perceived importance of several properties of information services (effectively referring to literature databases). Both practitioners in 2003-2004 and researchers in 2010 considered “up-to-dateness of contents” very important: 92% (ranking second among 12 properties inquired about) for practitioner survey #2 and 97% (second among 31 properties) for researcher survey #3 on a scale from “unimportant” to “very important.” (For the importance of properties data, width of 95% confidence intervals for means ranged from 2.2% to 7.0% in survey #2 and from 2.1% to 7.6% in survey #3.)

Quality assurance of contents was also deemed very important by both groups: Practitioners rated “service professionalism” 95% (first) and “certified information” 87% (fourth). Researchers rated “correctness of information” 94% (sixth) and “professional quality assurance” 85% (10<sup>th</sup>). However, as can be seen from the ranks, even though researchers considered these aspects important, there were several others they deemed at least equally significant, taking random error into account (keep in mind that researchers were asked for a total of 31 properties, whereas practitioners were only asked for 12).

Rated importance was also fairly similar with respect to interdisciplinarity of contents (70% and ninth rank in survey #2, 67% and 18<sup>th</sup> rank in survey #3) and search speed (#2: 78%, sixth rank; #3: 80%, 12<sup>th</sup> rank). Integration of services into workflow was considered similarly important: Possibilities for “subsequent processing of information, e.g., in personal databases” were rated 67% (10<sup>th</sup>) by practitioners. Researchers rated “seamless connection of resources” 72% (15<sup>th</sup>) and “dataset exporting capabilities” 68% (17<sup>th</sup>).

Even though “low user fees” were rated as quite important by practitioners (79%, fifth), researchers seemed to place even more emphasis on this aspect: 88% (seventh) for “open access (no charge).” However, the difference may also be due to the wording of the items.

Notable differences between practitioners and researchers emerged with regard to the “internationality of contents”: While practitioners rated this aspect 71% (eighth), researchers considered it one of the most important properties, with a rating of 96% (fourth). There also appears to be a difference regarding preferred search style: Practitioners considered “straightforward, uncomplicated search technology” one of the most important aspects (91%, third). Researchers, in contrast, gave fairly low ratings to ““intelligent” search engines” (52%, 20<sup>th</sup>) and “recommender systems” (38%, 26<sup>th</sup>), and higher ratings to “powerful search syntax” (71%, 15<sup>th</sup>) and “many searchable database fields” (75%, 13<sup>th</sup>).

The property considered most important by researchers was “direct access to full text” (98%). Practitioners were not asked about this in the context of information service features. However, when asked on a scale from “very low” to “very high” how important they thought online full text services would become in the future, they gave a rating of 68%. On a scale from “never” to “weekly,” they rated their current frequency of use of online full text services at 25%.

### **Open-Format Answers Regarding Problems in Information Search**

Survey #1.2 (see Table 1) asked about “desired improvements” regarding the literature database PSYINDEX. Online access to the database was the most commonly given single answer. Integration of the database into literature management workflow, linking to full texts, and integration with other literature databases were also often mentioned. Another class of frequently given answers referred to improved up-to-dateness and search features.

Survey #3 more generally asked researchers for “typical difficulties encountered while searching for information.” Most often, problems with access to the full text of publications were mentioned. Difficulties concerning search strategy (e.g., choosing proper keywords, filtering results, identifying all pertinent literature) were also quite common. Other notable issues included insufficient coverage of literature by databases (such as confinement to a single discipline, to certain languages, or to certain publication types) and flaws in usability (especially heterogeneity in user interfaces of databases).

### **Discussion**

Finding relevant scholarly information is pertinent to all research-related contexts, and the widespread digitalization of communication has affected the ways in which such information can be obtained. In the present work, information behavior and needs of German psychologists were examined by comparing the results of four surveys conducted between 1997 and 2010, a critical period in terms of the transition to various forms of digital communication.

The surveys appear to be the only ones regarding psychologists in Germany. However, they varied considerably with regard to target population characteristics and survey items. To obtain more valid data, prospectively designed investigations need to be carried out. In addition, because of the known weaknesses of the survey method, such as social desirability or differ-

ing interpretation of items (cf. Lietz, 2010), it is necessary to complement self-report data with observation of actual information behavior. The ZPID aims to implement both approaches, and the purpose of the present work is to serve as a starting point.

In general, information behavior of researchers in psychology does not seem to have changed fundamentally from the late 1990s to 2010. At both points in time, journals are rated as the most frequently used source of information. Notably, in 2010, use of open access journals is still rather infrequent compared to established subscription-based journals. Also, preprint and document servers are only rarely used as information sources in the psychology researchers' community. In terms of disciplinary culture, credit seems to be mainly allocated through publication in well-established, reputable journals, in contrast to the often-cited "preprint culture" in high-energy physics (Kling & McKim, 2000). However, recent data also indicate a trend toward increased publishing in open access journals (Krampen et al., 2012).

Researchers' database use appears to have remained stable in the preceding decade, although there does seem to be a slight shift toward more "international" (i.e., English-language) contents. This is consistent with an ongoing increase in the quota of English-language publications by psychologists from German-speaking countries (Krampen et al., 2012). Use of general-purpose search engines in searching for scholarly information has, as might be expected, increased from the late 1990s to 2010. However, it is important to point out that when searching for such information, psychology researchers still report using specialized databases more often than these search engines, and also more frequently than Google Scholar.

There has been some debate about the role of research libraries in the course of digitalization (e.g., Applegate, 2008). Library Online Public Access Catalogs (OPACs), the primary digital medium of searching library holdings, are one of the most frequently used means of searching information reported by psychology researchers in 2010, even though frequency appears to be somewhat lower than library use in the late 1990s. Despite virtualization, researchers still consider direct access to the library very important.

In general, and consistently over time, psychology researchers' main concern in information search was efficiently identifying and accessing all publications pertinent to their research topic, as evidenced by their comments regarding difficulties and desired improvements in scholarly information searches. Therefore, literature databases like PSYINDEX need to be integrated into more comprehensive systems.

Besides efficiency in obtaining information, the notion of trustworthiness or "approvedness" of sources (usually by peer review) figures prominently in

researchers' information needs. These two motives are probably the main reason for psychologists' near non-existent use of new communication forms like blogs and social networking sites as research information sources. However, most recently, fueled by increased awareness of questionable research practices in psychology (John et al., 2012), increased discontent has been voiced concerning the current publishing regime (e.g., Chambers, 2012), and eventually, innovations in communication may be embraced. This is particularly likely in the case of open access publishing, data sharing, and similar arrangements aimed at increased transparency and removal of barriers to access.

Finally, in the case of psychology in particular, information service providers should be aware of practitioners' perspectives on scholarly information, as evidenced by the practitioners' survey. Comparing this population with the researchers, some remarkable differences in information needs and behavior emerged. Practitioners used general purpose search engines more, but specialized databases less, than researchers did. When using databases, they placed more emphasis on ease of use and less on coverage of international publications. Practitioners also rated books as a more frequently used source of information than journals and made less use of library catalogs.

Clearly, these differences can largely be explained in terms of the availability of information services (e.g., journals or databases via the university library). Making information resources available to practitioners is a crucial challenge if psychological practice is to be grounded in empirical foundations. The results also reinforce the importance of including books and native-language publications in resources such as literature databases and library collections.

## References

- APPLEGATE, R. (2008). Whose decline? *Reference & User Services Quarterly*, 48, pp. 176-189.
- BECKER, R. (2004). Informationsbedarf und Informationsverhalten in der psychologischen Praxis (Report). Retrieved September 21, 2012 from <http://psydok.sulb.uni-saarland.de/volltexte/2011/2744/pdf/praktikerbefragung2004.pdf>
- BECHER, T., & TROWLER, P. R. (2001). *Academic tribes and territories* (2nd ed.). Buckingham: Open University Press.
- CHAMBERS, C. (2012, July 18). *The dirty dozen: A wish list for psychology and cognitive neuroscience*. Retrieved November 10, 2012 from <http://tinyurl.com/azba23n>
- CRONIN, B. (2003). Scholarly communication and epistemic cultures. *New Review of Academic Librarianship*, 9, pp. 1-24.

- DUTTON, W. H., & JEFFREYS, P. W. (2010). World wide research: An introduction. In W. H. DUTTON, & P. W. JEFFREYS (Eds.), *World wide research. Reshaping the sciences and humanities* (pp. 1-18). Cambridge, MA: MIT Press.
- EDUCATION FOR CHANGE (2002). Researchers' use of libraries and other information sources: Current patterns and future trends. Retrieved October 6, 2012 from <http://www.rslg.ac.uk/research/libuse/>
- JOHN, L. K., LOEWENSTEIN, G., & PRELEC, D. (2012). Measuring the prevalence of questionable research practices with incentives for truth telling. *Psychological Science, 23*, pp. 524-532.
- KLING, R., & MCKIM, G. (2000). Not just a matter of time: Field differences and the shaping of electronic media in supporting scientific communication. *Journal of the American Society for Information Science, 51*, pp. 1306-1320.
- KRAMPEN, G., BECKER, R., LABOUVIE, Y., & MONTADA, L. (2004). Internet-Ressourcen für die Psychologie sowie Informationsbedarf und Informationsverhalten von BDP-Mitgliedern. *Report Psychologie, 29*, pp. 588-600.
- KRAMPEN, G., FELL, C., & SCHUI, G. (2011). Psychologists' research activities and professional information-seeking behavior. *Journal of Information Science, 37*, pp. 439-450.
- KRAMPEN, G., FELL, C., & SCHUI, G. (2012a). Professionelles Informationsverhalten von Psychologen im Arbeitsfeld „Forschung und Lehre“. *B.I.T.online, 15*, pp. 117-126.
- KRAMPEN, G., FELL, C. B., & SCHUI, G. (2012b). Professionelle Publikationspräferenzen von Mitgliedern der Deutschen Gesellschaft für Psychologie (DGPs). *Psychologische Rundschau, 63*, pp. 175-178.
- KRAMPEN, G., & MONTADA, L. (2000). Nutzung alter und neuer Medien bei Literaturrecherchen von Experten. In G. KRAMPEN, & H. ZAYER (Eds.), *Psychologiedidaktik und Evaluation II* (pp. 89-99). Bonn: Dt. Psychologen-Verlag.
- KRAMPEN, G., SCHUI, G., & BAUER, H. (2012). ZPID-Monitor 2010 zur Internationalität der Psychologie aus dem deutschsprachigen Bereich: Der ausführliche Bericht (Report). Retrieved November 10, 2012 from <http://tinyurl.com/betmubo>
- LIETZ, P. (2010). Research into questionnaire design: A summary of the literature. *International Journal of Market Research, 52*, pp. 249-272.
- LÜCK, H. E., LEPLow, B., & SELG, H. (2011). *Geschichte der Psychologie* (5th ed.). Stuttgart: Kohlhammer.
- MONTADA, L., KRAMPEN, G., & BURKARD, P. (1999). Persönliche und soziale Orientierungslagen von Hochschullehrern/innen der Psychologie zu Evaluationskriterien über eigene berufliche Leistungen. *Psychologische Rundschau, 50*, pp. 69-89.
- NEPPL, R., WIESENHÜTTER, J., KRAMPEN, G., & MONTADA, L. (2001). Literaturrecherche-Strategien wissenschaftlicher Mitarbeiter/innen in psychologischen Instituten. In G. KRAMPEN, & H. ZAYER (Eds.), *Psychologiedidaktik und Evaluation III* (pp. 125-134). Bonn: Dt. Psychologen-Verlag.



# How and Why Do Turkish Scholars Use Social Networking Platforms?

**Selva Ersöz Karakulakođlu and Övünç Meriç**

Faculty of Communication, Maltepe University, Turkey

*To establish the differences between academic disciplines with regard to the use of social networking platforms in teaching and research, we conducted a focus group study among Maltepe University scholars in the fields of social and natural sciences. Our central finding is that while natural scientists in Turkey do not think that social networking platforms can support traditional education methods, Turkish social scientists regard them as a natural tool for both research and teaching.*

## Introduction

With the emergence of Web 2.0 technologies, many dimensions of academic life have undergone changes that have had an effect on the way scholars work. Online social media sites such as blogs, microblogging sites, and wikis are increasingly being used by academics for disseminating information, informal communication, and interaction with peers (Gruzd & Staves, 2011).

In this article, we will show how representatives of different academic disciplines—in our case, Turkish social and natural science scholars—make use of social networking platforms in teaching and research. The focus is on differences in attitudes towards the use of the Web 2.0 services under investigation among representatives of the aforementioned disciplines. Before proceeding to the methodology and findings of our study, we will briefly summarize the main academia-related advantages and disadvantages of some of

the most popular social media sites. Also, we will provide some important statistical information concerning the Internet use in Turkey.

## **Scholars' Use of Social Media Sites**

Blogging and microblogging enable a real-time interaction between users with the help of different devices, technologies, and applications (Ebner & Schiefner, 2008). In addition to Twitter, which seems to be the most popular microblogging service nowadays, many educational institutions also use a special microblogging platform designed for education: Edmodo.com. They use this platform to send notes, links, files, alerts, assignments, and events to each other (Holotescu & Grosseck, 2010).

Microblogging sites are mainly used for dissemination of teachers' publications and materials, and for creating user groups in which students and teachers can work collaboratively. These sites are capable of facilitating virtual classroom discussions, as well as constructing a viable platform for metacognition (the practice of thinking about the learning), which can benefit comprehension and retention of the material (Grosseck & Holotescu, 2008). One of the major disadvantages of blogging and microblogging is that these activities demand time. Additionally, the lack of courtesy on the part of students or the loss of meaning in an online environment can constitute a serious problem. Finally, it must be mentioned that blog entries currently do not count as academic publications (see, e.g., Lovink, 2008).

Facebook, a very popular social networking Web site, enables scholars to exchange ideas (with colleagues and students), to disseminate publications, to foster teamwork, and more (Cobbs, 2008). While Facebook is widely used by students, many instructors avoid it for several reasons:

The reasons for instructors' nonparticipation in online social networking sites tended to center on three issues: privacy and surveillance, teacher identity, and time. Many instructors cited concerns about their online privacy, Web presence, adware, spyware, or spam as their main reasons for nonparticipation. (Vie, 2008, p. 18)

## **Internet Use in Turkey**

According to research conducted by Mediascope Europe in 2012, there were approximately thirty-four million Internet users in Turkey ("Türkler internete zaman geçiriyor," 2012, July 7). There were approximately thirty-one million Facebook users as of March 2012, according to Internet World Stats.

Among the social networking sites' users, the young population comes first, and Turkey has a very young population: The majority of the Turkish population is represented by young people, about forty percent of whom are under the age of fifteen. Seventy-two percent are below the age of thirty-five ("Population of Turkey," 2003, February 2). These statistics are also reflected in the social media use trends in Turkey:

Young males are typically the earliest adopters in emerging internet markets and digital platforms. The strong presence of these early adopters in Turkey—71% of users are aged 15-34 and 58% are male—implies that there is still room for further advancement. (Flanagan, 2009)

## Research Questions and Method

While the youth population in Turkey has been surrounded by social networking sites, Turkish scholars have also become involved with online social media, although there still remain problems like the digital divide and a lack of Internet literacy.

The most important research questions concerning the use of social networking Web sites in academic institutions in Turkey are as follows: To what extent have Turkish scholars adopted social media platforms? Do university instructors regard social media tools as course supports? Do they believe that the interaction created by these Web sites is helpful for the regulation of their courses?

To answer these questions, we conducted focus group interviews with scholars from different disciplines at the Maltepe University in Istanbul. We sought to identify scholarly trends in social and natural sciences. Some recent studies have indicated that discipline differences might account for dissimilarities in the use of social media: Scholars in the sciences make use of social media technologies earlier and more frequently than their counterparts in the humanities (Gardiner et al., 2006; Maron & Smith, 2008; Dubini et al., 2010).

For the data collection, we have chosen two small groups composed of six social science faculty members representing communication studies and six natural science faculty members representing computer engineering. Each group consisted of instructors with different academic titles, including assistant professors and lecturers. The participants were mostly between the ages of thirty-five and forty-five.

We determined as keywords "participation," "reflection," "interaction," "support," and "interpretation." These keywords were introduced to the groups and we allowed the participants to speak freely about the subjects.

These key concepts were taken from Taylor and Maor's (2000) study of online learning. We analyzed the discussions using audio files of the interviews. Apart from the determined concepts, emergent themes coming out of the responses were also determined and analyzed.

The participants were asked to talk freely on the determined keywords without much interference on our part. In the beginning of the interviews, we acquainted the participants with the subject of our study and the aforementioned keywords. Then we asked the interviewees whether they had accounts on social networking Web sites and whether they used them for educational purposes.

## Findings

### The Use of Social Networking Web Sites

**Natural Sciences:** Of the six participants, only two had accounts on social networking sites. All participants reported that they used Microsoft Network (MSN) and e-mail. While they used MSN only in their private lives, they used e-mail to communicate with students. The participants who used social networking sites reported to have both Facebook and Twitter accounts. They did not use blogs, wikis, or video sharing sites. All of them used File Transfer Protocol (FTP).

**Social Sciences:** All of the participants had accounts on social networking sites. While all of them had Facebook accounts, only four had Twitter accounts and only three used blogs. None of them used wikis or video sharing sites. Five of the six used MSN. All participants stated that they used e-mail for communication. All of them used FTP.

### Participation

The participants were asked to talk freely about this keyword. To trigger a discussion, we asked them whether the communication with students by means of social networking Web sites increased students' participation in the classroom.

**Natural Sciences:** Of the six participants, five argued that communicating with students via online social networks did not increase their participation in the classroom. Professor 5 stated that in engineering sciences, it was very

difficult to educate students by means of social networking sites. Professor 1 drew attention to another negative dimension:

In particular, there is a problem with students using Facebook for educational purposes. They give copies to each other. Facebook is becoming a media for providing copies. I do not approve of that. (our translation)

As was pointed out by Professor 2, Erasmus students were more willing to communicate with their professors via social networking sites.

**Social Sciences:** All participants stated that the use of social networking sites in teaching positively affected lecture attendance and participation in the classroom. Professor 1 confirmed this phenomenon:

Social media are very fast. Sharing information about the lecture on these platforms increases the attendance. (our translation)

Professor 3 agreed with this assessment:

Working with my shared files on Facebook and Twitter in class increases the attendance. Students show more interest. (our translation)

Professor 4 also recognized the usefulness of these sites:

Especially in application classes, social networking sites are a very effective and productive way to increase attendance. I realized that there is a considerable increase in attendance compared to before and after using social networking sites. I can see that students who make use of my shared lecture notes on my blog are more active in class. (our translation)

## Reflection

On social networking sites, teachers and students can create a virtual identity and, thus, be connected to each other in a huge Internet platform that enables them to reach and share all kinds of information and knowledge. Given this, it appears that the use of online social networks as a medium for reflection might be a new channel for teaching methods. All kinds of audio, visual, and textual documents that are shared by the teachers and students might promote critical thinking and reflexivity.

**Natural Sciences:** The idea that social networking sites do not support reflection and even affect it in a negative way seems to be widely spread among Turkish professors of natural sciences. Except for one professor, the participants agreed that social networking sites were not platforms suitable for natural sciences. Professor 2 expressed the following concerns:

Social networking sites cannot be used to manage projects and develop ideas. It seems like it makes science easier. No work and no effort. Students popularize everything. All work done is undervalued. (our translation)

Only one professor had moderate views regarding online social networks:

The aim is important. I establish a target group on Facebook according to my aim and share the documents. I share information through mobile technology. It expedites the students' work to reach information, but I do not use it for scientific sharing. (our translation)

**Social Sciences:** The general opinion of all the professors interviewed was that the use of social networking sites increased their students' ability to access important information and their critical thinking skills. This belief was shared by Professor 5:

In application classes, in terms of visual design and video analysis, the students can both develop and criticize themselves. Narration in classes supported by social media applications increases the ability to interpret information. (our translation)

Professor 6 agreed:

Social networking sites prepare students for the field. They increase creativity. Students follow up their final projects and project results through Facebook, Twitter, and blogs and get feedback. I find this very important. (our translation)

The interviewees, thus, shared the view that social networking sites can affect reflection in a positive way.

## Interaction

As was already mentioned above, social networking sites can be used for the dissemination of teachers' publications and course materials, and for creating user groups in which students and teachers can work collaboratively. In this way, social networking sites can help build a connection between teachers and students.

**Natural Sciences:** The general trend was not to interact with students by means of social networking sites. Of the four professors, only two had accounts on social networking platforms, and only one of them reported interacting with students on these platforms. According to this participant, social networking sites can facilitate teacher-student communication:

Through Facebook it is very easy to create an activity, to make announcements and get feedback instantly with the group I established. With mobile technologies

students are immediately informed about all messages I relayed and make comments. This makes my job easier. (our translation)

Other professors said that they were against social networking sites because they did not want to expose their privacy.

**Social Sciences:** Professors in this group all had positive views about interacting with students by means of social networking sites. Professor 3's response was representative of this group's optimism:

When you get in touch with a student on Facebook, your contact in everyday life improves. It becomes a double-sided interaction. You get to know the student better. (our translation)

## Support

**Natural Sciences:** The participants did not practice the use of social networking sites for educational purposes in their courses. What is more, they were against the students' use of social networks for this purpose. According to the participants, face-to-face interaction was more important. They accentuated discredit for students who followed the courses in social networks and shared documents via these. The quite common belief was that the content posted on social networking platforms did not support education in natural sciences. According to one of the participants, these sites are simply not compatible with the educational efforts of learning institutions:

The use of social networking sites is causing a loss of time and misunderstanding. I do not think that the use of these platforms suffices to master a subject, no matter if the subject is related to natural sciences or social sciences. (our translation)

**Social Sciences:** In this group, social networking sites were regarded as a supportive element in teaching. The participants did not oppose the use of social networking platforms by students. Moreover, they considered these platforms to be part of the social sciences. One of the participants reiterated the following benefits of these sites:

Social networks are always part of communication sciences. Social networking platforms themselves are course material. Thus, teachers and students can be able to work on the same material. (our translation)

## Interpretation

This term refers to students and educators' co-construction of meaning in a congruent and connected manner. This co-construction of meaning (of what

is going on in class) can affect, in an important way, academic success, motivation, communication, interaction, and more.

**Natural Sciences:** The participants had negative attitudes towards the use of social networking platforms, believing that these prevented their dialogue with students. The participants did not regard the use of these platforms as a process of sharing and interpretation of information and knowledge. One of the participants shared the following concerns:

In the field of social sciences, the relationship between students and instructors might be different, but in natural sciences, this process operates in a different way. Communication in social networking sites does not permit a dialog. (our translation)

**Social Sciences:** All of the participants argued that the higher degree of interaction between instructors and students had affected course productivity in a positive way. Students' online presence affected in an important way the content production and their interpretation of the course material. It was useful not only for the course content but also for the teaching rules and social behavior.

During our interviews, it was observed that two notions, other than those five categories we had previously determined, became prominent: "privacy" and "event management and mobility." These were mentioned by ten (out of twelve) participants during their free talk on the use of social networking platforms. In the following sections, we will enlarge on what was stated in connection with these notions by our participants.

## Privacy

Social networking sites are public and free, and thus readily available for everyone. In general, students often disregard the rules of grammar and formal precepts when posting on online social networks. Furthermore, social networking sites do not operate according to normative notions of power as well as sexual and educational hierarchies between teachers and students.

**Natural Sciences:** All participants were concerned about the possible loss of privacy when utilizing online social networks. This fear also created anxiety about the general loss of control regarding what is going on with social media. One participant expressed the following concerns:

The process could get out of control. Common use of personal information could be a disadvantage. I cannot risk myself by sharing my life. I don't want my personal life to become students' material. (our translation)

**Social Sciences:** The interviewees in this group also had anxieties about the secrecy of their private lives. They pointed out that providing information about their private lives in their profiles on social networking sites would result in sharing this information with their students. They feared that this would result in an uncontrollable flow of information. Because of this, the participants limited the access to their profiles on online social networks by means of not adding their students to their “friends” lists and by sharing information only through user groups.

### **Event Management and Mobility**

Students use services like Facebook and Twitter, not only to stay in touch with existing friends and to make new ones, but also to exchange information about classes, lectures, and all kinds of events or whatever else might interest them. The most prevalent use of social networking sites in the university community was creating profiles and groups to communicate with others.

**Natural Sciences:** Even the participants who did not use social networking platforms had reached an agreement concerning the use of these platforms for event management and relaying various notifications. The common view was that events organized with the help of social media could increase participation and mobility.

**Social Sciences:** All participants in this group organized events by means of launching educational groups. According to them, social networking sites facilitated event organization. Participants emphasized the idea that it increased the class participation and mobilized students.

### **Conclusion**

The main objective of our interviews was to determine the differences in the use of social networking sites by natural and social science scholars in Turkey. Our point of departure was the expectation (based on the previous research on this matter) that natural science scholars would be more enthusiastic about the use of social media in teaching and research than their counterparts in social sciences. But our interviews did not confirm this expectation: On the contrary, whereas the interviewees from the natural sciences think that social networking sites do not support traditional education methods and, thus, cannot be used as an appropriate tool for teaching, the interviewees

from the social sciences regard social media as a natural tool for both teaching and research.

Social networking sites are constructing milestones in every field of life, including education. It seems that the future of education will develop around them. The use of social networking platforms will also develop in the Turkish higher education system. An important aspect to remember is that social networking platforms should be integrated to course content and should not be used only for better and effective communication with students but also for a more valuable use in combination with university courses:

The problem is not so much providing access for students surrounded by technology but rather effectively integrating technological literacy instruction into the composition classroom in meaningful ways. (Vie, 2008, p. 9)

Some limitations of this study and some open research questions for the future work should be added to our conclusion. In addition to a very small number of our interviewees (12), it must be also mentioned that in this study we took only the scholars' point of view into consideration. Furthermore, the technical dimensions related to the environment where this research was carried out—the Maltepe University—undeniably constitutes another limitation: The participants were all members of staff of this university, so the university's technical infrastructure might account for their use of social networking sites. In the future, the effectiveness of social networking sites in education can be analyzed both from the teachers' and students' perspectives to better understand the weaknesses and strengths of these platforms.

## References

- COBBS, J. T. (2008, February 1). Learning 2.0 for associations. Message posted to [http://blog.missiontolearn.com/files/Learning\\_20\\_for\\_Associations\\_eBook\\_v1.pdf](http://blog.missiontolearn.com/files/Learning_20_for_Associations_eBook_v1.pdf)
- DUBINI, P., GALIMBERTI, P., & MICHELI, M. R. (2010). Authors publication strategies in scholarly. Unpublished paper presented at the 14th International Conference on Electronic Publishing, Helsinki.
- EBNER, M., & SCHIEFNER, M. (2008). Microblogging—more than fun? In I. A. SÁNCHEZ, & P. ISAÍAS (Eds.), *Proceedings of IADIS Mobile Learning Conference 2008* (pp. 155-158). Retrieved September 22, 2012 from [http://lamp.tu-graz.ac.at/~i203/ebner/publication/08\\_mobilllearn.pdf](http://lamp.tu-graz.ac.at/~i203/ebner/publication/08_mobilllearn.pdf)
- FLANAGAN, J. (2009, November 17). State of the Internet: Global perspectives and the role of Turkey in the digital world. Retrieved September 22, 2012 from <http://www.slideshare.net/mytayf/comscore-turkey-presentation>

- GARDINER, D., MCMENEMY, D., & CHOWDHURY, G. (2006). A snapshot of information use: Patterns of academics in British universities. *Online Information Review*, 30(4), 341-359.
- GROSSECK, G., & HOLOTESCU, C. (2008). Can we use Twitter for educational activities? Unpublished paper presented at the 4th International Scientific Conference eLearning and Software for Education, Bucharest.
- GRUZD, A., & STAVES, K. (2011). Trends in scholarly use of online social media. Retrieved September 22, 2012 from <http://conway.isri.cmu.edu/hicss2011-sci-workshop/gruzd-hicss-trends.pdf>
- HOLOTESCU, C., & GROSSECK, G. (2010). Using microblogging in education. Case study: Cirip.ro. Retrieved November 10, 2012 from <http://tinyurl.com/8yunoa>
- Internet world stats. Usage and population statistics. Europe. (2012, April 11). Retrieved September 22, 2012 from <http://www.internetworldstats.com/europa2.htm>
- LOVINK, G. (2008). *Zero comments. Blogging and critical Internet culture*. London: Routledge.
- MARON, N. L., & SMITH, K. K. (2008). Current models of digital scholarly communication: Results of an investigation conducted by Ithaka for the association of research libraries. Washington: Association of Research Libraries. Retrieved September 22, 2012 from <http://www.arl.org/bm~doc/current-models-report.pdf>
- Population of Turkey. (2003, February 2). Retrieved September 22, 2012 from <http://www.turkishodyssey.com/turkey/pop/pop.htm>
- TAYLOR, P., & MAOR, D. (2000). Assessing the efficacy of online teaching with the constructivist on-line learning environment survey. In A. HERRMANN, & M. M. KULSKI (Eds.), *Flexible futures in tertiary teaching. Proceedings of the 9th annual teaching learning forum*. Retrieved September 22, 2012 from <http://lsn.curtin.edu.au/tlf/tlf2000/taylor.html>
- Türkler internette zaman geçiriyor. (2012, July 11). Retrieved September 22, 2012 from <http://www.teknokulis.com/Haberler/Internet/2012/07/11/turkler-internette-zaman-geciriyor>
- VIE, S. (2008). Digital divide 2.0: “Generation M” and online social networking sites in the composition classroom. *Computers and Composition*, 25(1), 9-23.



# Digital History in Portugal: A Survey

Maria Cristina Guardado<sup>I</sup> and Maria Manuel Borges<sup>II</sup>

<sup>I</sup> School of Technology and Management Águeda, University of Aveiro, Portugal, <sup>II</sup> Faculty of Arts and Humanities, University of Coimbra, Portugal, and CETAC-MEDIA

*The emergence of concepts like “digital humanities” and “digital history” shows how digital technology is becoming increasingly important to the humanities and corresponding disciplines. This study is a first look into digital history in Portugal. Using information available on the Web, it is possible to see that historians are using more and more digital tools to complete and communicate their work as well as to collaborate with other researchers.*

## Introduction

Over the last decades, a new technological infrastructure appeared, and access to information has become increasingly important and fast, giving rise to new services, practices, and requirements in several areas, including research and development (R&D). However, the adoption of new technologies to support and to communicate science is not homogeneous in different scientific communities, and the process is influenced by the epistemic culture of each community (Knorr-Cetina, 1999). If, in the field of “hard sciences,” the use of digital tools is considered natural, arts and humanities have been less keen on taking advantage of the benefits offered by the technological infrastructure to support research. Nevertheless, humanists are using more and

more digital technologies in their work (Hayles, 2012), making “digital humanities” a topic that has been attracting interest from a growing number of people (Little, 2011; Gold, 2012).

Although the association between humanities and computers has been established for quite some time, especially in linguistics and text analysis, the changes arising from the development of the Internet and the World Wide Web (WWW) have led to a terminological shift from “humanities computing” to “digital humanities” (Svensson, 2009; Hayles, 2012), the latter emerging not as an academic discipline, but as “a hybrid domain, crossing disciplinary boundaries and also traditional barriers between theory and practice, technological implementation and scholarly reflection” (Flanders et al., 2007).

Early on, the establishment of technological infrastructure and large-scale digitization projects were the main focus of digital humanities. Presner (2010) found that in this “first wave,” digital humanities tended to be more quantitative, centered on the development of databases, text encoding, and other issues concerning text analysis. In this current second wave, digital humanities have become more “qualitative, interpretative, experimental, emotive, generative in character” (Schnapp & Presner, 2009). However, for some authors, a new wave may be in progress due to some epistemic transformations operated by the use of digital (Berry, 2011), which might lead us to a new discipline, distinct from the traditional humanities (Evans & Rees, 2012).

Because the humanities are a large field of knowledge, comprising many distinct disciplines and tribes (Becher & Trowler, 2001), which have their own traditions and cultures (Knorr-Cetina, 1999), these transformations are not all homogeneous. Thus, we chose to center our attention on one of the humanistic disciplines: history.

In the humanities in general, “new media and new technologies have challenged historians to rethink the ways that they research, write, present and teach about the past” (Cohen & Rosenzweig, 2006), leading to digital history. Although this term has been used for more than a decade—it was first used in 1997 by Ayers and Thomas (Cohen et al., 2008)—few people have tried to define it. Thus, there is no established definition. According to Andersen (2002), digital history has to do with the work developed by historians who privilege the use of digital technologies in their activities and not specifically with the use of new sources in the construction of history. Cohen et al. (2008) reiterated this through questions regarding the collaboration and interaction between historians, students, and other publics. These two dimensions—the use of digital tools and “particular ways of working” (which in-

volve collaboration as well as openness)—are referenced by Murphy (2012) when characterizing digital history and are used in this study.

However, even though digital media was seen relatively early as an opportunity to reconfigure historical science and scientific practices to attract new audiences (O'Malley & Rosenzweig, 1997; Ayer, 1999), some authors claim that this has not occurred with the expected intensity, despite the presence of its significant history in the WWW (Cohen, 2004; Cohen & Rosenzweig, 2006; Turnbull, 2010). Such behavior has many causes and is associated with the practices adopted to construct and disseminate historical knowledge as well as the entire context in which historians do their work.

Founded in the 1990s, the Virginia Center for Digital History at the University of Virginia and the Center for History and New Media at George Mason University (CHNM) were two of the earliest centers devoted to digital history in the U.S. The information available about their research projects reflects clearly the importance of information and communication technologies (ICT). Online databases enabling access to digitized primary sources (texts, images, videos, or podcasts), bibliographic, or geographical information are the most frequent mark of that significance. If in some cases the collections are built from archive material, in others the primary sources are provided by the public in general. Projects like “The September 11 Digital Archive” or “Hurricane Digital Memory Bank,” developed by CHNM in partnership with other institutions, reveal the importance of collaboration outside academia, taking advantage of “crowdsourcing” by engaging with the public to help their projects.

But this engagement is not limited to the building of collections. “Transcribe Bentham” is a project launched in 2010 by the University College London (UCL) “with the aim of recruiting volunteers from around the world, whatever their background, to help transcribe the unpublished manuscripts of Jeremy Bentham” (Causer, 2012), a British philosopher who lived from 1748 to 1832. This project is one of the best-known examples of “crowdsourcing” in the academic context.

Another project that includes a collaboration component is “London Lives 1690 to 1800,” developed by the Universities of Hertfordshire and Sheffield. It makes available, in a fully digitized and searchable form, a wide range of primary sources about eighteenth-century London. By providing access to a large number of archive documents and datasets created by several other projects, it is possible for users “to link together records relating to the same individual, and to compile biographies of the best documented individuals” using a wiki available on the project Web site. To make digitized documents searchable, encoding schemes are used, and the text is marked up in XML.

Encoding text is also used in many other projects, such as “1641 Depositions,” a collaborative project between Trinity College, Dublin, the University of Aberdeen, and the University of Cambridge, which aims to conserve, digitize, transcribe, and make available online witness testimonies by Protestants and some Catholics concerning their experiences of the 1641 Irish rebellion.

Building prosopographical databases is very common in historical research since these databases allow historians to obtain information specific to individuals by analyzing common characteristics of a group and learn about patterns of relationships and activities. Various prosopographical databases are available on the Web.

Historical Geographical Information Systems (historical GIS) associate spatial data with other data included in standard databases (attribute data). By combining them with visualization and analytic tools, they allow historians to georeference historical events, identify changes in the geographic space over time, or respond to specific research problems such as the relationship between some diseases and water sources. “Mapping Medieval Chester,” by Swansea University, Queen’s University Belfast, and the Centre for Computing in the Humanities at King’s College London, and “A Vision of Britain between 1801 and 2001,” by the University of Portsmouth, are two examples of projects with those characteristics.

With the development of ICT, visualization techniques became more and more used in history research projects. King’s Visualization Lab (KVL), based at the Department of Digital Humanities at King’s College London, is a lab that specializes in visual representations of archaeology, historic buildings, cultural heritage organizations, and academic research and that uses 3D modeling, among other techniques, in several projects.

## **Research Design and Preliminary Findings**

In Portugal, the available data shows that almost the entire national scientific community has access to the national research and education network, which supports numerous applications and services related to collaboration, knowledge, and safety (UMIC, 2011). Nonetheless, studies about the adoption of technologies by the scientific community are scarce.

The study presented here is part of ongoing research that aims to understand the impact of digital media in the mechanisms of production and dissemination of scientific knowledge within a specific epistemic community: historians. This paper will look at whether and how digital technologies are being used in history research projects.

Considering that scientific research in Portugal is mainly developed in Research & Development Units (R&D Units), funded and evaluated by the Foundation for Science and Technology (FCT), this was the setting chosen for our study. Thirteen out of the 17 history R&D Units currently funded by FCT (Mattoso et al., 2011) were chosen (see Appendix, Table 2). These units integrate around a thousand researchers. The four remaining R&D Units were excluded for two reasons: two for being specialized in archaeology; the other two because they are specialized in history and philosophy of science and integrate mainly researchers from other scientific fields, with epistemic cultures and scientific practices that differ from those of historians.

To identify potential digital history projects, the 13 R&D Units were contacted and asked to provide information about the projects in development. Because the responses were few, our attention fell on the Web sites of the R&D Units. Some Web pages had not been updated (in some cases the information about the research projects was two or more years out of date), making it difficult to obtain a clear idea about ongoing projects, so it was necessary to undertake some complementary research to confirm the information retrieved. During this process, carried out between May and July 2012, we considered numerous research projects, but only ongoing projects (except those for which it was not possible to find more information than the designation)—those finished within the last two years or those that reveal the use of ICT and are hosted by the R&D Units.

Data analysis shows that the use of digital tools in history research projects is prevalent even though the information available is frequently vague and the tools used are often not specified. Of the 38 projects considered (see Table 1), only three are not associated with the use of ICT. However, as ongoing projects, they were included in the research.

As stated before, a database is a digital tool frequently used in historical research and is associated with 58% of the projects. There are several cases in which prosopographical databases were, or are being, built. For example, “Corpo do Estado Maior do Exército Português,” carried out by researchers from Centro de Estudos de História Contemporânea Portuguesa (CEHCP), focuses on the elite staff of the Portuguese Army from the 1940s to 1974 and identifies the most influential officers. Another project that gave rise to a prosopographical database was “Fasti Ecclesiae Portugaliae” (Project 4). This project, developed by the Centro de Estudos de História Religiosa (CEHR) between 2002 and 2006, gave historians the opportunity to identify the bishops, dignitaries, and canons who served in Portuguese cathedrals during the Middle Ages. Findings were published in some books, although the database has not yet been finished. The project’s Web page states that when complet-

ed, the prosopographical database will be disclosed. Nevertheless, there is no information stating whether it will be available online.

The use of prosopographical information is frequently associated with social network analysis and representation (using graph models), which occurs in three of the nine projects that use that type of data.

Besides prosopographical databases, other databases are mentioned in projects such as “Engenho e Obra” (Project 35) and “Marconi em Lisboa” (Project 38) from the Instituto de História Contemporânea (IHC) or “Grupos intermédios em Portugal e no Império Português” (Project 22), from the Centro Interdisciplinar de História, Culturas e Sociedades (CIDEHUS). However, no details about those databases or the tools used are displayed in either case. “Nobreza Medieval Hispânica” (Project 28), developed by the Centro de Estudos da População Economia e Sociedade (CEPESE), produces a database with prosopographical and bibliographic information to publish a biographical dictionary on medieval Galician-Portuguese nobility.

There are, however, some databases available online. For example, the “Project LITTERA” (Project 33), at Instituto de Estudos Medievais (IEM), created a database freely accessible on the WWW; the *cantigas* (songs) contained in the medieval Galician-Portuguese *Cancioneiros* (songbooks) can be searched for by author, title, *cancioneiro*, genre, type, or subject. Also available is bibliographic information about the troubadours and the persons mentioned in the songs as well as digitizations of the manuscripts, musical scores, and audio files.

The transcription and publication of historical sources is frequent in Portuguese research projects. For instance, in a CEHR project entitled “Portugaliae Monumenta Misericordiarum” (Project 9), developed between 2001 and 2011, ten volumes were published. These volumes gathered not only the most significant documents in the study of the Portuguese *miseriórdias* (assistance institutions existing in Portugal since the fifteenth century), but also some thematic bibliographies, guides from the archives of those institutions, and some unpublished studies (Paiva, 2010). Moreover, the Centro de Estudos Históricos (CEH) is planning to publish several textual primary historical sources (Project 11). However, they do not mention whether they intend to use the print or the electronic format.

The data available shows that the results of the research projects are frequently presented in books, conference proceedings, and other traditional formats. However, a Centro de História de Além-Mar (CHAM) project intends to create a *Virtual Encyclopedia of Portuguese Expansion* from the fifteenth to the eighteenth century (Project 13); a preliminary version is already available. In the same R&D Unit, another project is being developed

for the virtual environment. “De todas as Partes do Mundo” (Project 14), based on the inventory of the estate of a Portuguese noble from the sixteenth century, creates a virtual exhibition and a 3D virtual rebuilding of the Palace of Vila Viçosa, built by the Duke.

Another project that has exploited the potential of digital technologies is “DynCoopNet<sup>PT</sup>” (Project 17). In fact, this was the Portuguese part of a larger project entitled “DynCoopNet: Dynamic Complexity of Cooperation-Based Self-Organizing Networks in the First Global Age,” founded by the European Science Foundation, and involved several institutions, including Centro de Investigação Transdisciplinar “Cultura, Espaço e Memória” (CITCEM). This was a truly multidisciplinary project that used several digital technologies, such as structured databases, historical GIS, and mathematical modeling. The main tool used was Time Link, a software “specially developed to support micro-historical research with a strong emphasis on network analysis and prosopography” at the University of Coimbra (DynCoopNet<sup>PT</sup>, 2009).

As stated above, one of the characteristics of digital history is collaboration. An analysis of the R&D Unit Web sites shows that in more than 61% of the projects, people from different disciplines, units, universities, institutions, and countries integrate research teams. Nevertheless, this collaboration does not seem to involve the general public. When using collaborative tools, such as Google Docs, data analysis does not allow us to conclude whether they are being used. The few units that responded to the e-mailed questionnaire reported using Dropbox to share files.

## Conclusions and Further Research

The adoption of ICT to support and to communicate science is a reality in all scientific fields, including the humanities, which are less keen on taking advantage of the technological infrastructure.

As a result of this new reality, the concepts of digital humanities and digital history emerged, indicating an increased use of technologies by humanists in general and by historians in particular. In spite of this new trend, the use of new technologies by the scientific community is still quite unknown in Portugal.

In the case of historians, data shows that researchers are increasingly using digital tools to do their work. When compared to international history research projects, the Portuguese projects make use of the same tools and techniques. The major difference is related to patterns of collaboration. In a number of non-Portuguese projects, public participation is one of the main

characteristics, while in Portugal the collaboration is mainly inside the scientific community, involving historians and other researchers.

In the future, we aim to cross the data obtained in this study with data collected through semi-structured interviews conducted with the directors of the selected R&D Units and through a survey applied to individual researchers. This will allow us to better understand how Portuguese historians perceive digital history and what strategies and tools are being used in the construction of historical knowledge in Portugal in the digital age.

## References

- 1641 Depositions. Retrieved November 10, 2012 from <http://tinyurl.com/af5qq9z>
- ANDERSEN, D. L. (2002). Defining digital history. *Journal of the Association for History and Computing*, 5(1).
- AYERS, E. L. (1999). The pasts and futures of digital history. Retrieved September 23, 2012 from <http://www.vcdh.virginia.edu/PastsFutures.html>
- BECHER, T., & TROWLER, P. R. (2001). *Academic tribes and territories: Intellectual enquiry and the culture of disciplines* (2nd ed.). Buckingham: SRHE, Open University Press.
- BERRY, D. M. (2011). The computational turn: Thinking about the digital humanities. *Culture Machine*, 12, 1-22.
- CAUSER, T. (2012). Two years on, and progress update, 1 to 7 August 2012. Transcribe Bentham: A participatory initiative. Retrieved September 23, 2012 from <http://www.ucl.ac.uk/transcribe-bentham/>
- COHEN, D. J. (2004). History and the second decade of the Web. *Rethinking History: The Journal of Theory and Practice*, 8(2), 293-301.
- COHEN, D. J., FRISCH, M., GALLAGHER, P., MINTZ, S., SWORD, K., MURRELL, A., THOMAS III, W. G., & TURKEL, W. J. (2008). The promise of digital history. *Journal of American History*, 95(2), 442-451.
- COHEN, D., & ROSENZWEIG, R. (2006). *Digital history: A guide to gathering, preserving, and presenting the past on the Web*. Fairfax: CHNM.
- DynCoopNet<sup>PT</sup>: Dynamic complexity of cooperation-based self-organizing networks in the first global age. About the Portuguese project. (2009). Retrieved September 23, 2012 from <http://www.dyncoopnet-pt.org/about-portuguese-project>
- EVANS, L., & REES, S. (2012). An interpretation of digital humanities. In D. M. BERRY (Ed.), *Understanding digital humanities* (pp. 21-41). London: Palgrave Macmillan.
- FLANDERS, J., PIEZ, W., & TERRAS, M. (2007). Welcome to digital humanities quarterly. *DHQ: Digital Humanities Quarterly*, 1(1). Retrieved September 23, 2012 from <http://digitalhumanities.org/dhq/vol/1/1/000007/000007.html>

- GOLD, M. K. (2012). The digital humanities moment. In M. K. GOLD (Ed.), *Debates in the digital humanities* (pp. IX-XVI). Minneapolis: University of Minnesota Press.
- HAYLES, N. K. (2012). How we think: transforming power and digital technologies. In D. M. BERRY (Ed.), *Understanding digital humanities* (pp. 42-66). London: Palgrave Macmillan.
- Hurricane Digital Memory Bank. Retrieved November 10, 2012 from <http://tinyurl.com/b4zqdw1>
- KLV: King's Visualization Lab. Retrieved November 10, 2012 from <http://tinyurl.com/bz4zcdy>
- KNORR-CETINA, K. (1999). *Epistemic cultures: How the sciences make knowledge*. Cambridge: Harvard University Press.
- LITTLE, G. (2011). We are all digital humanists now. *The Journal of Academic Librarianship*, 37(4), 352-354.
- London lives 1690 to 1800: Crime, poverty and social policy in the metropolis. (2012, April). Retrieved September 23, 2012 from <http://www.londonlives.org/>
- Mapping medieval Chester. Retrieved September 23, 2012 from <http://www.medievalchester.ac.uk/index.html>
- MATTOSO, J., ALVELOS, H., DUARTE, I., FERRÃO, J., AMARAL, J. F., LIMA, L. P., MESQUITA, P. E., PEREZ, R. M., & KOULALDIS, V (2011). *Ciências Sociais e Humanidades: internacionalização, pluralismo, pluridisciplinaridade, avaliação, disseminação e relação entre as políticas científicas nacional e comunitária*. Lisboa: CCCSH. Retrieved November 23, 2012 from <http://tinyurl.com/by7raun>
- MURPHY, R. (2012). Digital history: Some tools and techniques. In *Digital arts and humanities: scholarly reflections* (pp. 12-16). Cork: Digital Academy at University College Cork.
- O'MALLEY, M., & ROSENZWEIG, R. (1997). Brave new world or blind alley? American History on the World Wide Web. Retrieved September 23, 2012 from <http://chnm.gmu.edu/essays-on-history-new-%20media/essays/?essayid=13>
- PAIVA, J. P. (2010). Projecto *Portugaliae Monumenta Misericordiarum*. Retrieved November 10, 2012 from <http://tinyurl.com/b4epffp>
- PRESNER, T. (2010). Digital humanities 2.0: A report on knowledge. Retrieved September 23, 2012 from <http://cnx.org/content/m34246/1.6/>
- SCHNAPP, J., & PRESNER, T. (2009). Digital humanities manifesto 2.0. Retrieved September 23, 2012 from [http://humanitiesblast.com/manifesto/Manifesto\\_V2.pdf](http://humanitiesblast.com/manifesto/Manifesto_V2.pdf)
- SVENSSON, P. (2009). Humanities computing as digital humanities. *DHQ: Digital Humanities Quarterly*, 3(3). Retrieved September 23, 2012 from <http://digitalhumanities.org/dhq/vol/3/3/000065/000065.html>
- SVENSSON, P. (2010). The landscape of digital humanities. *DHQ: Digital Humanities Quarterly*, 4(1). Retrieved November 10, 2012 from <http://tinyurl.com/7dhflz5>
- The September 11 digital archive: Saving the histories of September 11, 2001. (2011). Retrieved September 23, 2012 from <http://911digitalarchive.org/>

TURNBULL, P. (2010). Historians, computing and the World-Wide-Web. *Australian Historical Studies*, 41(2), 131-148.

UMIC–AGÊNCIA PARA A SOCIEDADE E CONHECIMENTO. (2011). *A Sociedade da Informação em Portugal 2010*. Retrieved September 23, 2012 from [http://www.unic.pt/images/stories/publicacoes4/SIP2010\\_completo.pdf](http://www.unic.pt/images/stories/publicacoes4/SIP2010_completo.pdf)

A vision of Britain between 1801 and 2001. (2012). Retrieved September 23, 2012 from [www.visionofbritain.org.uk](http://www.visionofbritain.org.uk)

## Appendix

Table 1. Projects considered

Project	Host R&D Unit	Digital tools used						Project outputs																	
		In collaboration	databases	historical GIS	3 D visualization	network graph	Monographs / book chapters			Dictionaries / encyclopedias			Other publications			Exhibition			Database						
							print	electronic	not specified	print	electronic	not specified	print	electronic	not specified	virtual	traditional	online	offline	not specified	audiovisual materials	conferences / scientific meetings / workshops			
1.	CEHCP	*	*											*					*						
2.	CEHCP	*	*					*						*										*	
3.	CEHR	*								*														*	
4.	CEHR	*						*															*		
5.	CEHR	*												*										*	
6.	CEHR	*								*														*	
7.	CEHR	*						*																*	
8.	CEHR	*																						*	
9.	CEHR	*	*					*																*	
10.	CEHR	*						*	*															*	
11.	CEH	*								*														*	
12.	CEIS 20	*												*										*	
13.	CHAM	*	*	*						*														*	*
14.	CHAM	*	*	*	*			*					*	*		*	*						*	*	*
15.	CHAM	*	*	*				*						*									*	*	*
16.	CHAM	*	*	*	*			*					*	*							*			*	*
17.	CITCEM	*	*	*	*			*					*	*						*				*	*
18.	CITCEM	*	*	*	*			*					*	*						*				*	*
19.	CIDEHUS	*	*					*					*	*						*				*	*
20.	CIDEHUS	*	*					*					*	*						*				*	*
21.	CIDEHUS	*	*					*					*	*			*			*				*	*
22.	CIDEHUS	*	*	*	*			*					*	*					*				*	*	*
23.	CIDEHUS	*	*	*	*			*					*	*					*				*	*	*
24.	CIDEHUS	*	*	*	*			*					*	*				*		*			*	*	*
25.	CIDEHUS	*	*	*	*			*					*	*				*		*			*	*	*
26.	CIDEHUS	*	*	*	*			*					*	*				*		*			*	*	*
27.	CIDEHUS	*	*	*	*			*					*	*				*		*			*	*	*
28.	CEPESE	*	*					*					*	*						*			*	*	*
29.	GHEH	*	*					*					*	*					*				*	*	*
30.	GHEH	*	*					*					*	*					*				*	*	*
31.	GHEH	*	*					*					*	*					*				*	*	*
32.	HEM	*	*					*					*	*					*				*	*	*
33.	HEM	*	*					*					*	*					*				*	*	*
34.	HEM	*	*					*					*	*					*				*	*	*
35.	IHC	*	*					*					*	*		*	*		*				*	*	*
36.	IHC	*	*					*					*	*				*		*			*	*	*
37.	IHC	*	*					*					*	*		*	*		*				*	*	*
38.	IHC	*	*					*					*	*		*	*		*				*	*	*

Table 2. R&amp;D Units considered

<b>Abbreviation</b>	<b>Unit designation</b>	<b>Institutional affiliation</b>
CEHCP	Centro de Estudos de História Contemporânea Portuguesa	ISCTE – Instituto Universitário de Lisboa
CEHR	Centro de Estudos de História Religiosa	Universidade Católica Portuguesa
CEH	Centro de Estudos Históricos	Universidade Nova de Lisboa. Faculdade de Ciências Sociais e Humanas
CEIS 20	Centro de Estudos Interdisciplinares do Século XX – CEIS 20	Universidade de Coimbra
CHC	Centro de História da Cultura	Universidade Nova de Lisboa. Faculdade de Ciências Sociais e Humanas
CHSC	Centro de História da Sociedade e da Cultura	Universidade de Coimbra. Faculdade de Letras
CHAM	Centro de História de Além-Mar	Universidade Nova de Lisboa. Faculdade de Ciências Sociais e Humanas and Universidade dos Açores
CITCEM	Centro de Investigação Transdisciplinar “Cultura, Espaço e Memória”	Universidade do Porto and Universidade do Minho
CIDEUS	Centro de Investigação Transdisciplinar de História, Culturas e Sociedades da Universidade de Évora	Universidade de Évora
CEPESE	Centro de Estudos da População, Economia e Sociedade	Universidade do Porto
GHEs	Gabinete de História Económica e Social	Universidade Técnica de Lisboa. Instituto Superior de Economia e Gestão
IEM	Instituto de Estudos Medievais	Universidade Nova de Lisboa. Faculdade de Ciências Sociais e Humanas
IHC	Instituto de História Contemporânea	Universidade Nova de Lisboa. Faculdade de Ciências Sociais e Humanas

Table 3. Project titles

Nr.	Project title
1	O Corpo do estado maior do exército português: apogeu e queda
2	Terras Além dos Mares: Direitos de Propriedade no Império Português Moderno
3	Centenário da República e da Lei de Separação
4	Fasti Ecclesiae Portugaliae: prosopografia do clero catedralício português (1071- 1325)
5	História e Memória Local
6	A Igreja Católica e o Estado Português no século XX: os Cardeais Mendes Belo (1907-1929), Gonçalves Cerejeira (1929-1971), António Ribeiro (1971-1998) e a República Portuguesa
7	Memória e História de Fátima
8	A Participação da Igreja Portuguesa no Concílio Vaticano II
9	Portugaliae Monumenta Misericordiarum
10	Religião, Sociedade e Laicidade nos Países Ibero-Americanos
11	Publication of historical primary sources
12	Pró-Memória Portugal Vinte
13	Enciclopédia Virtual da Expansão Portuguesa
14	De Todas as Partes do Mundo, O Património do 5º Duque de Bragança, D. Teodósio I
15	A Nobreza e Estado da Índia no século XVI
16	Portugal e o Sul de Marrocos: contactos e confrontos (séculos XVI-XVIII)
17	DynCoopNet <sup>PT</sup>
18	Espaços Urbanos: dinâmicas demográficas e sociais (séculos XVII-XX)
19	Decisão Política, necessidades colectivas e afirmação profissional: o Hospital de Todos os Santos em perspectiva
20	Engenharia, Tecnologia, Cultura Material e Património
21	Espaços urbanos: história e património cultural
22	Grupos intermédios em Portugal e no Império Português: as familiaturas do Santo Ofício (c. 1570-1773)
23	História do Alentejo, séculos XII-XX
24	Inquirir da honra: comissários do Santo Ofício e das Ordens Militares em Portugal (1570-1773)
25	O sistema de caridade e assistência: apropriação social, trajectórias sociais, discursos institucionais. O caso da região de Évora (séculos XV-XVIII)
26	Os utentes do sistema de assistência entre a população de Évora: a importância das bases de dados demográficas (1535-1800)
27	Viagens, Turismo, Lazer e Património no Sul em perspectiva histórica. Dos finais do século XVII à primeira metade do século XX
28	Nobreza Medieval Hispânica: séculos VIII-XVI
29	Corporate Responses to European Integration, 1957-2007
30	História do Sistema Bancário (séculos XIX e XX)
31	O ouro do Brasil – agentes, remessas e conjunturas (1720-1807)
32	Projecto Imago
33	Littera, edição, atualização e preservação do património literário medieval português
34	Regnum Regis - As inquirições de 1220 e a génese da memória documental do reino medieval português
35	Engenho e Obra. História da Engenharia em Portugal no Século XX
36	Guia Bibliográfico da I República
37	História da Emigração e das Comunidades Portuguesas
38	Marconi em Lisboa





## **PART TWO**

---

# **THE INTERNET AND TEACHING**



# Informal Learning via Social Media—Preparing for Didactical Designs

**Isa Jahnke**

Department of Applied Educational Science, Umeå University, Sweden

*The paper argues that didactical designers, teachers, and researchers can learn from informal learning situations to build a meaningful learning experience in formal education. By illustrating three research projects, five theses will be outlined in order to describe the shift from traditional teaching to “learning to be creative.”*

## Introduction

Informal learning usually takes place when a learner has unsolved issues outside of the formal instruction that is provided by a teacher. Sometimes such informal, unsolved issues are clear problems of which an individual is aware, yet sometimes they are less clear. Imagine that a person who wants to know something starts to search for an answer. Such “unsolved problems” are, for instance, improving a swim style by watching YouTube videos, verifying that information provided by others is correct, observing how a speaker reacts to difficult questions, and so forth. Currently, we also see a change in discussion cultures of daily-life groups: When facts are discussed, at least one person takes her smartphone and googles the information—hence, unplanned learning takes place.

Informal learning makes it clearly visible that the learners are their own pace-makers; they choose the topic and the context, in particular, “learning

what” (content), “learning how” (pace and style), “learning when” (time), “learning where” (online or a physical location), and “learning with whom” (Shurville et al., 2008). Why is it important to discuss informal learning? Because informal learning can lead to a deeper understanding and a different quality of a learning outcome; it enables the learner to expand her thinking beyond a receptive behavior within formal schooling and beyond the traditional reproduction of existing knowledge.

A combination of both informal learning and formal education might be a win-win situation for learners. The research question is: To what extent is informal learning in formal education designable? Can we create didactical designs for technology-embraced informal-*in*-formal learning? To answer this question, we want to know what informal learning is and what we can learn from it in order to inform and form didactical designs for teaching.

In this article, I will describe informal learning in connection with online groups and social media, and with regard to the understanding of didactical designs. Then I will highlight results from three research projects—InPUD, CSCL@Work, and iPad-Didactics—to illustrate some theses on the way to “learning to be creative.”

## **Emerging Forms of Learning, “Cultures of Participation”**

In the past few years, new forms of online communities and new forms of learning have emerged. John S. Brown makes it clear:

Whatever your particular interest is, there is some niche community, already formed on the network you can join. (...) These resources not only provide facts. They are also tools you can use to build things to tinker with, to play with, to reflect on, and to share with others. And most importantly, you will learn from other people’s comments and from what they do with your creations. (Brown, 2009)

The pressure to rethink schooling today is increasing more than ever because the innovation is coming from outside and into the classrooms (Collins & Halverson, 2009). We currently do not know if formal schooling will be replaced or not, but new forms of both formal and informal learning have emerged around the edge of formal schooling (Brown, 2009). Social media affect the relationship between formal education, informal learning out of schools, and at the workplace (Goggins et al., in press). Studies illustrate a transformation in education through innovation in computing (Mørch & Skaanes, 2010; Tuomi-Gröhn & Engeström, 2003). New “cultures of participation” have emerged (Fischer, 2011; Jahnke & Haertel, 2010).

## **Informal, Non-Formal, Formal Learning**

Informal learning is related to incidental learning, “learning en passant” (Reischmann, 1986) and “experiential learning” (Kolb, 1984). According to Kolb’s learning cycle, learning occurs in four steps: A learner has a) a concrete experience, b) she observes and reflects, c) she draws conclusions and forms abstract concepts, and d) she tests new concepts in new situations. A person conducts the four steps by contrasting her experiences with the experiences of the others (Schön, 1983; Daudelin, 1996).

Incidental and experiential learning can occur in both planned and unplanned learning situations which are designated as formal, non-formal, and informal learning. The concepts differ in a) the degree of organization; b) formal certificates, credits or a degree; and c) the criterion of “who triggers learning” (Ainsworth & Eaton, 2010). Formal learning is triggered by instructors / teachers and organized by an educational institution. The learner receives credits or a formal degree. Non-formal learning is also a form of planned learning that is organized by an external person, but it occurs outside educational institutions. Informal learning is a self-directed learning situation (or not organized at all) that is triggered by the learner instead of an external teacher. It does not culminate in the conferring of any degree.

The primary difference between the forms of learning is the external organizer. Formal and non-formal learning are related to a teacher and a tutor, who give instructions and rules; informal learning is related to an inspiring environment and supporting structures (Watkins & Marsick, 1992). Online forums, weblogs, and social networking sites (e.g., LinkedIn or Facebook) are just a few examples where informal learning can take place.

## **Social Media in Academia**

In their special issue entitled “Web 2.0 Goes Academia,” Jahnke and Komers (2009) show eight different scenarios for educational workers regarding the ways in which social media can be used in teaching and learning, as well as in research.

Online groups have been studied by many research teams, such as Stahl (2006), Preece and her colleagues (2004), and Wenger et al. (2001). According to Preece’s studies, such groups differ in four areas: a) the group size (e.g., groups with 25 members or less to groups with 1,000 or more), b) the primary content (e.g., discussion boards about a stock exchange, online communities on marathon training, and political topics), c) the lifespan (e.g., several years or only for organizing one event), and d) the degree of presence (e.g., online communication, face-to-face communication, or a mixture of

both forms). Online groups (van de Sande, 2010) and social media applications differ in their concepts towards private identity and public accessibility. Private identity means that the users have anonymous nicknames and keep their identity private, whereas the information that they share is public; for instance, in online discussion boards, the content has public access that differs from social networking sites, where a special degree of the private identity is shown.

### **Didactical Designs for Teaching and Learning**

Learning is knowledge construction, and collaborative learning is defined as a form of co-creation of new knowledge among a group of people, that is, “an active process of constructing rather than acquiring knowledge” (Duffy & Cunningham, 1996, p. 171). This represents a shift from teaching, where learning has been seen as information consumption, to a focus on learning as knowledge construction (Barr & Tagg, 1995). The learning paradigm refers to the need for learners to become active agents within the learning process (i.e., prosumers ← producers and consumers). Studies point out the positive relationship between being active and a deeper learning outcome (Chapman, 2003).

Regarding formal schooling, the following questions remain: How to design “active knowledge co-construction,” how to design situations where learners are prosumers; what are appropriate didactical designs? The term “didactical design” follows from the German concept of “Didaktik” by Klafki (1963, 1997) and is inspired by Hudson (2008), Fink (2003), and Lund and Hauge (2011), who stress the differences between teaching concepts and learning activities and call them designs for teaching and designs for learning. From this point of view, a didactical design includes teaching objectives (see Figure 1). It also incorporates the plan to achieve those objectives in such a way that the learners are able to develop competencies and skills which the teachers have in mind (design for learning activities). It seeks to transform the teaching aims into learning activities and, finally, it includes different forms of feedback and assessment to assess the learning progress. Process-based, formative assessment seems to be the most effective method to foster learning (Bergström, 2012). The difference from an instructional design is that didactical designs also include the strategy that incorporates the design for social relations (i.e., teacher-student and student-student).

The digital didactical design approach is the advanced model that integrates educational technology. To each of the four design levels, the design-question is: How can information and communication technology, social media or, for instance, iPads support the activity? The benefit from social

media is that it makes learning visible. Mårell-Olsson and Hudson (2008) illustrate in two case studies different ways of compiling digital portfolios in which students develop the ability to “collect, organize, interpret and reflect on their own individual learning and practice and become more active and creative in the development of knowledge” (p. 73).

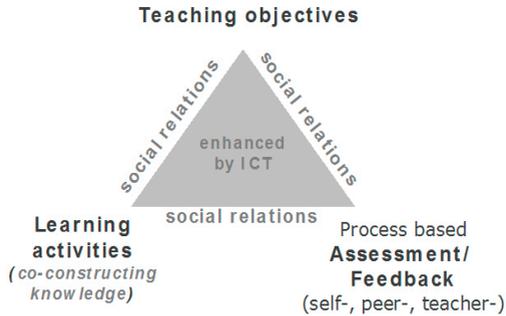


Figure 1. Framework for a digital didactical design

## Examples

In the following sections, three projects and derived theses will be described.

### InPUD—Informal Learning in Higher Education

InPUD is an example of an informal learning community of approximately 1,500 students that is embedded into higher education. In 2002, an online forum at a computer science program was launched (Jahnke, 2010a). The free and open online forum has been offered to support students in conducting their computer science studies (i.e., the pursuit of B.A. / M.A. degrees). Learning is defined as the co-construction of knowledge among new and senior students, study advisors, and faculty members. The sub-boards exist for a) courses such as lectures and seminars (e.g., to discuss exercises or content of lectures) and b) study organization where, for example, users share knowledge about the computer science study. The decision about the topics mainly depends on what the students want to discuss. InPUD is characterized by a large size and an extended lifespan; it commenced in 2002 and continues today by providing a space for interactions, usually asynchronously. InPUD is a PHP-based technical system. Users need only an Internet access in order to read the contributions. To post, registration with a free chosen username is

required. InPUD supports public communication that is based on the anonymity of its users. This is different from traditional learning management systems, which often require the real names of the users.

The data collection, analysis, and redesign were conducted in iterative cycles of research and development from 2002 to 2009. The data-gathering process included mixed methods such as open-ended interviews, standardized questionnaires, user statistics, content analysis, and log files. The detailed results about InPUD are illustrated in Jahnke (2012, 2010b, 2006).

One result is that the InPUD community indicates a special feeling of a membership. This ambience is expressed in terms such as “we help each other,” “that’s the sense of a community” (interviewees). InPUD activates a) the user’s perception of having a specific form of social proximity, which is triggered by technology and b) the conative level of learning. The term “conation” refers to a concrete action that is conducted by a learner; s/he does not only know, but s/he really acts, s/he is willing to do something and really does (Kolbe, 1990). The concept of conation stresses what a learning outcome really is; the learning outcome is seen by a changed behavior of the learner.

The conative level of learning is often neglected in formal schooling, where the cognitive learning of “what” and textbook knowledge is focused on without supporting the learners’ ability to practice such knowledge in action. Traditional teaching neglects the designs for learning as an active process that includes reflective action (i.e., students as prosumers), but also neglects to create designs for social relations among students and between the teacher and the students. To make this gap smaller, solutions such as InPUD can be useful. An online board can be a differentiator that supports the individual needs of the users.

**Thesis 1:** The addition of informal learning by means of social media expands formal education and leads to an all-embracing learning experience that activates learners on all levels such as the cognitive, affective, and conative levels; this is what we call designing for technology-embraced informal-*in*-formal learning (Jahnke, 2012).

## **CSCL@Work—Informal Work-Based Learning**

Research in the field of organizational learning emerged in 1978 and gained further attention in the 1990s, when challenges centered on the creation of organizational cultures to support existing knowledge sharing. However, the challenge in contemporary companies is the creation of new knowledge, and is driven by a primary question: How do organizations create new knowledge

when the answer to a particular problem is not available within the company; how does one design learning when the answer is not known (Fischer, 2011)?

Within a first workshop at ACM Group 2010 ([www.csclatwork.org](http://www.csclatwork.org)), it became clear that social media affects collaborative learning at work. The studies by Elizabeth King (2010) and Gurzick and White (2010) pointed out that Facebook, as well as World of Warcraft, play important roles in developing competencies for being successful at work. The resulting book entitled “CSCL@Work, Computer-Supported Collaborative Learning at the Workplace,” edited by Goggins et al. (in press), includes thirteen case studies about collaborative learning within the workplace and how it is enhanced by social media.

The results from our meta-analysis indicate that CSCL@Work is a timely challenge for researchers to develop a new, integrated understanding of working and learning as they are embraced by social media. The cases illustrate the following:

- a) Learning occurred in unexpected and unusual online learning places through social media: The cases demonstrated unstructured connections to the employee’s workplaces in social media.
- b) Learning activities by the employees incorporated feedback from diverse people: The successful cases enabled a change of feedback partners and established learning loops.
- c) Learning took place across established organizational boundaries: Communication took place with people who were not available within the traditional organizational boundaries.

**Thesis 2:** Social media enable informal learning at the workplace in unexpected, unusual places and across established boundaries.

### **iPad-Didactics**

The Odder Project began in 2012 in Denmark, where an entire municipality provided iPads to approximately 2,000 students and 180 teachers in seven schools. A qualitative approach, as part of a larger study, was used to explore how teachers used the iPads in the classrooms in Denmark. In April 2012, thirteen classroom observations and ten interviews were conducted in five schools in Odder. The teaching subjects ranged from languages (Danish and English) and arts to mathematics and physics. The classes ranged from pre-school (grade 0) to ninth grade. The classroom observations were based on the didactical triangle that included a) the design of “teaching aims,” b) “learning activities,” and c) “different forms of process-based feedback /

assessment.” These three elements are connected by the design of “social relations” and the role of the iPads (Figure 1).

Data from the observations and interviews were analyzed according to each classroom before being open coded (Bryman, 2008). The results from the pilot study show five examples named transformative learning (math, second grade), complex learning (language, grade 0), peer-reflective learning (language, seventh grade), collocated collaborative learning (language and arts combined, eighth grade), and personalized learning (physics, ninth grade), which are illustrated in detail by Jahnke and Kumar (in press). From this study in Denmark, we learned how simple it can be to foster “learning to be creative” by using iPads. The studied classrooms represent active learning that was focused on action where students obtained the assignment to produce something, and while doing so, they reflected and learned. Creativity, which signifies the creation of something new, plays a central role (Jahnke & Haertel, 2010). The teachers did not solely focus on outcomes or exams, nor did they expect students to reproduce the facts. The teachers’ designs for teaching and learning included active student participation and student engagement. The teachers had a learner-centered approach. They scaffolded the learning process by providing feedback and personalizing the learning experience for students who experienced difficulties. The iPad served as a “booster” to foster learning as a process.

**Thesis 3:** When using social media, there is a shift in teaching practices from learning as consumption to a focus on action and a focus on relationships.

What we learned from Odder is that there is a shift from “textbook learning” to a focus on action—when ICT has been implemented in classrooms. Besides the traditional teaching objectives, learning “what” (i.e., information from textbooks), the Odder teachers applied new designs for “learning to be creative” (Jahnke, 2011) adopting the iPads that activated the students to create solutions to problems where no answer was available (i.e., the problem can be part of a task or assignment that is given by the teachers, or the students must find the problem and create tasks.)

**Thesis 4:** The iPad makes a difference. It is not seen as technology.

When we started our research on iPads in schools and universities, people asked us if there was a difference from laptops. Now, after a first pilot study, we know that iPads differ in many aspects. The most important difference is that the teachers we interviewed said “the iPad works,” “you open an iPad

and it works,” “you don’t waste time like with the laptops where the batteries are out of energy or the software wasn’t installed.” One teacher’s response to our question regarding why there was such hype around the iPads was: “There is no technology in there!” and she pointed her finger to the iPad. Of course, an iPad is made of technical elements and it is a pure technical device. However, with that quote, we understood the teacher’s point of view: They perceive the iPad as a device that is easy to use rather than a complex, complicated device. That is one major difference from the laptop.

### **Informal Learning via Social Media: The Ways in Which It Affects Science**

With the emergence of social media applications, online communities, and forums (i.e., Wikipedia, which was launched in 2001; Facebook, which was launched in 2004; YouTube, which was launched in 2005; and Twitter, which was launched in 2006), science has been affected. While collecting and storing significant data each day such as texts, words, videos, numbers, and so forth, the term “big data analytics” has been introduced. The term describes the huge amount of data and the possibilities of analyzing what people do online every second. The concept of big data analytics has been transformed into “learning analytics.” The goal of this “booming domain” (Siemens, 2012) is to find appropriate methodologies to analyze big data for specific contexts; for instance, to improve knowledge management in large organizations and to improve learning within the learning sciences and in higher education. Duval and Verbert describe the research area in the following way:

Learning analytics focuses on collecting traces that learners leave behind and using those traces to improve learning. In this domain, there are two major approaches: 1. Educational Data Mining can process the traces algorithmically and point out patterns or compute indicators. 2. Information visualization can present the traces in ways that help learners or teachers to steer the learning process. (Duval & Verbert, 2012)

A relatively new approach is that of “teaching analytics,” which was explored within the first workshop entitled “Towards Theory and Practice of Teaching Analytics,” which took place at the Seventh European Conference on Technology Enhanced Learning, Saarbrücken (<http://ec-tel.eu/>). The motivation for such a new research field is based on high-performance classrooms, which are characterized by 1:1 computers that generate big data. The organizers argued that teachers needed different information in order to make decisions about their teaching plans and practices, and to adjust the decisions

“in a meaningful and actionable format” (Vatrapu et al., 2012). To support the teachers, the researchers want to discuss how to use and analyze the big data using visual analytics methods for reflecting on teaching and learning (Vatrapu et al., 2012).

**Thesis 5:** Teaching analytics becomes the most important research methodology that complements traditional methods for studying teaching and learning.

However, the question contemplates the purposes for which the data will be used in the future: Is the purpose to improve learning or to control learners? The responsibility and the need for ethical standards in educational work and social sciences are on the increase.

This research is only a cursory example. It shows that traditional research methods such as interviews and surveys, which are mainly based on subjective perception by the respondents (i.e., what they say), will be complemented by more “objective” data (i.e., what they do) by analyzing their logging data. Together, both methods allow the situation—the whole big picture—to become visible.

## Conclusion

Berger and Luckmann (1967) explained the social construction of reality. Within the age of social media, this concept has changed into the socio-technically constructed reality; there is no objective reality, but we construct it in complex and often hidden social and socio-technical mechanisms. In this age, emerging forms of informal learning by social media have affected the concepts of learning in companies, in formal teaching, and in higher education. It also affected the socio-technical construction of digital didactical designs.

Is there a gap between the Homo Interneticus (Krotoski, 2011) and the Homo Didacticus? What kind of digital didactical designs for teaching and learning are useful to support meaningful, challenge-based learning in higher education and in schools? What we learned from our studies is that new digital didactical designs include designs for the active co-construction of new knowledge; they have a focus on action and they design the relationships. The approach of “learning to be creative” includes: a) the design of technology-embraced informal-*in*-formal learning that supports the conative level of learning, b) the design for learning in a way that learning at unexpected unusual (online) places across established boundaries will be fostered, and c) the

utilization of social media as a “booster” to intensify learning as a process strengthened through the design for creating social relations.

The approach “learning to be creative” is shifting to a “focus on action,” where the social relations among the peers, as well as the teacher-student, are integrated into the didactical designs for a situation that is unknown.

## Acknowledgements

I thank Andreas Olsson, Lars Norqvist, Krister Lindwall, and Peter Vinnervik at Umeå University for their work in the Odder Project, as well as their wonderful hints, questions, comments, and ideas. I want to thank Sean Goggins for the CSCL@Work joint project and exciting discussions, as well as Swapna Kumar for her reflections on iPad-Didactics. I am grateful to Timo van Treeck for providing me with the opportunity to present a keynote lecture at the International Conference “Science and the Internet” (<http://nfgwin.uni-duesseldorf.de/de/cosci12>).

## References

- AINSWORTH, H. L., & EATON, S. (2010). *Formal, non-formal and informal learning in the sciences*. Calgary: Onate Press.
- BARR, R., & TAGG, J. (1995). From teaching to learning. A new paradigm for undergraduate education. *Change*, 27(6), 198-200.
- BERGER, P., & LUCKMANN, T. (1967). *The social construction of reality*. Middlesex: Penguin.
- BERGSTRÖM, P. (2012). *Designing for the unknown. Didactical design for process-based assessment in technology-rich learning environments*. (Doctoral dissertation, Umeå University). Retrieved September 23, 2012 from <http://umu.diva-portal.org/smash/record.jsf?pid=diva2:516108>
- BROWN, J. S. (2009). Foreword. In A. COLLINS, & R. HALVERSON, *Rethinking education in the age of technology* (pp. ix-x). New York: Teachers College Press.
- BRYMAN, A. (2008). *Social research methods* (3rd ed.). New York: Oxford University Press.
- CHAPMAN, E. (2003). Alternative approaches to assessing student engagement rates. *Practical Assessment, Research & Evaluation*, 8(13). Retrieved September 23, 2012 from <http://pareonline.net/getvn.asp?v=8&n=13>
- COLLINS, A., & HALVERSON, R. (2009). *Rethinking education in the age of technology: The digital revolution and schooling in America*. New York: Teachers College Press.

- DAUDELIN, M. (1996). Learning from experience through reflection. *Organizational Dynamics*, 24(3), 36-48.
- DUFFY, T. M., & CUNNINGHAM, D. J. (1996). Constructivism: Implications for the design and delivery of instruction. In D. H. JONASSEN (Ed.), *Handbook of research for educational communications and technology* (pp. 170-198). New York: Simon and Schuster Macmillan.
- DUVAL, E., & VERBERT, K. (2012). Learning analytics. Retrieved September 23, 2012 from <http://eeced.campussource.de/archive/8/3336>
- FINK, D. L. (2003). *Integrated course design*. Kansas City: Idea Center.
- FISCHER, G. (2011). Understanding, fostering, and supporting cultures of participation. *Interactions*, 18(3), 42-53.
- GOGGINS, S., JAHNKE, I., & WULF, V. (in press). *CSCL@Work, computer-supported collaborative learning at the workplace*. New York: Springer.
- GURZICK, D., & WHITE, K. (2010). Transforming CSCL in the workplace by supporting online personal networks. Unpublished paper presented at Work Workshop at the ACM Conference on Supporting Groupwork (GROUP), Florida.
- HUDSON, B. (2008). A didactical design perspective on teacher presence in an international online learning community. *Journal of Research in Teacher Education*, 15(3-4). Retrieved September 23, 2012 from <http://umu.diva-portal.org/smash/get/diva2:158554/FULLTEXT01>
- JAHNKE, I. (2006). *Dynamik sozialer Rollen beim Wissensmanagement. Soziotechnische Anforderungen an Communities und Organisationen*. Wiesbaden: DUV.
- JAHNKE, I. (2010a). A way out of the information jungle—a longitudinal study about a socio-technical community and informal learning in higher education. *International Journal of Socio-technology and Knowledge Development*, 4, 18-38.
- JAHNKE, I. (2010b). Dynamics of social roles in a knowledge management community. *Computers in Human Behavior*, 26, 533-546.
- JAHNKE, I. (2011). How to foster creativity in technology enhanced learning. In B. WHITE, I. KING, & P. TSANG (Eds.), *Social media tools and platforms in learning environments* (pp. 95-116). Berlin: Springer.
- JAHNKE, I. (2012). Technology-embraced informal-in-formal learning. In A. RAVENSCROFT, S. LINDSTAEDT, C. D. KLOOS, & D. HERNÁNDEZ-LEO (Eds.), *21st century learning for 21st century skills. 7th European conference of technology enhanced learning, EC-TEL 2012, Saarbrücken, Germany, September 18-21, 2012. Proceedings* (pp. 395-400). Heidelberg: Springer.
- JAHNKE, I., BERGSTRÖM, P., LINDWALL, K., MÄRELL-OLSSON, E., OLSSON, A., PAULSEN, F., & VINNERVIK, P. (2012). Understanding, reflecting and designing learning spaces of tomorrow. In A. SÁNCHEZ, & P. ISAÍAS (Eds.), *Proceedings of IADIS mobile learning 2012* (pp. 147-156). Berlin: IADIS Press.
- JAHNKE, I., & HAERTEL, T. (2010). Kreativitätsförderung in der Hochschule—ein Rahmenkonzept. *Hochschulwesen*, 3, 88-96.

- JAHNKE, I., & KOMMERS, P. (Eds.). (2009). *International Journal of Web Based Communities*, 5(4), 477-483.
- JAHNKE, I., & KUMAR, S. (in press). iPad-didactics. Didactical designs for iPad-classrooms. In C. MILLER, & A. DOERING (Eds.), *The new landscape of mobile learning: Redesigning education in an app-based world*. London: Routledge.
- KING, E. M. (2010). Digital media and gaming spaces as models of CSCL and CSCW in practice. Unpublished paper presented at the workshop on Computer-Supported Collaborative Learning (CSCL) at GROUP2010, Sanibel.
- KLAFKI, W. (1963). *Studien zur Bildungstheorie und Didaktik*. Weinheim: Beltz.
- KLAFKI, W. (1997). Critical-constructive didactics. In M. ULJENS (Ed.), *Didaktik* (pp. 215-228). Lund: Studentlitteratur.
- KOLB, D. (1984). *Experiential learning: Experience as the source of learning and development*. Englewood Cliffs: Prentice-Hall, Inc.
- KOLBE, K. (1990). *The conative connection*. Boston: Addison-Wesley Publishing.
- LUND, A., & HAUGE, T. E. (2011). Designs for teaching and learning in technology rich learning environments. *Nordic Journal of Digital Literacy*, 4. Retrieved September 23, 2012 from <http://www.idunn.no/ts/dk/2011/04/art04>
- MÄRELL-OLSSON, E., & HUDSON, A., (2008). To make learning visible: In what way can ICT and multimedia contribute? *Tidskrift för lärarutbildning och forskning*, 3-4, 73-90.
- MØRCH, A. I., & SKAANES, M. A. (2010). Design and use of an integrated work and learning system: Information seeking as critical function. In S. LUDVIGSEN, A. LUND, I. RASMUSSEN, & R. SÄLJÖ (Eds.). *Learning across sites: New tools, infrastructures and practices* (pp. 138-155). London: Routledge.
- PREECE, J., ABRAS, C., & MALONEY-KRICHMAR, D. (2004). Designing and evaluating online communities. *International Journal of Web based Communities*, 1, 2-18.
- REISCHMANN, J. (1986, October). Learning “en passant”: The forgotten dimension. Presentation at the Conference of the American Association of Adult and Continuing Education, Florida. Retrieved September 23, 2012 from <http://www.unibamberg.de/fileadmin/andragogik/08/andragogik/aktuelles/86AAACE-Hollywood.pdf>
- SCHÖN, D. (1983). *The reflective practitioner*. New York: Basic Books.
- SHURVILLE, S., O’GRADY, T., & MAYALL, P. (2008). Educational and institutional flexibility of Australian educational software. *Journal of Campus-Wide Information Systems*, 25(2), 74-84.
- SIEMENS, G. (2012). Learning analytics: Envisioning a research discipline and a domain of practice. Presentation at the Second International Conference on Learning Analytics and Knowledge, Vancouver.
- STAHL, G. (2006). *Group cognition: Computer support for building collaborative knowledge*. Cambridge, MA: MIT Press.

- TUOMI-GRÖHN, T., & ENGESTRÖM, Y. (2003). *Between school and work: New perspectives on transfer and boundary crossing*. Amsterdam: Pergamon Press.
- VAN DE SANDE, C. (2010). Free, open, online, help forums: Convenience, connection, control, comfort, and communication. *International Journal of Sociotechnology and Knowledge Development*, 2(4), 1-17.
- VATRAPU, R., HALB, W., & BULL, S. (2012). TaPTA Workshop at EC-TEL 2012: Towards Theory and Practice of Teaching Analytics.
- WATKINS, K., & MARSICK, V. (1992). Towards a theory of informal and incidental learning in organization. *International Journal of Lifelong Education*, 11(4), 287-300.
- WENGER, E., MCDERMOTT, R., & SNYDER, W. M. (2002). *Cultivating communities of practice*. Boston: Harvard Business School.

# Collaborative and Cooperative Text Production in Wikis

**Claudia Bremer**

studiumdigitale, Goethe University Frankfurt

*In the context of the pedagogical discussion around learner orientation, active learning, learner participation, and the widespread use of Web 2.0 tools, wikis become more and more attractive for e-learning settings. But while many lecturers expect wikis to increase learner participation, they often have to acknowledge that the tool itself does not necessarily have any effect on the learners' involvement. Based on this observation, a study was conducted that examined the effects of a more bottom-up usage of wikis in projects initiated by students in comparison to the usage of wikis in seminar settings.*

## Introduction

Due to their technical features, wikis can be used for the provision of online material as well as for collaborative and cooperative production of texts in e-learning settings. In wikis, all participants can edit the content either with or without registration—depending on the system's settings. Even the wiki's structure can be developed by the participants while they edit pages and links (Ebersbach et al., 2008). Meanwhile, wikis, as well as blogs and discussion boards, are integrated into most learning management systems—so wikis can be easily integrated into e-learning settings. Often the choice between wikis, blogs, or other tools is not an easy one. In this context, it is helpful to look at Hippner's (2006) approach towards social Web tools, by which he differentiated along the criteria of information, relation, and communication. He tried

to describe how each single tool is appropriate or supports each of these aspects. Ebersbach et al. (2008) picked up on the Hippner's idea but differentiated between the aspects "information," "relation," and "collaboration." The result of their categorization is shown in Figure 1.

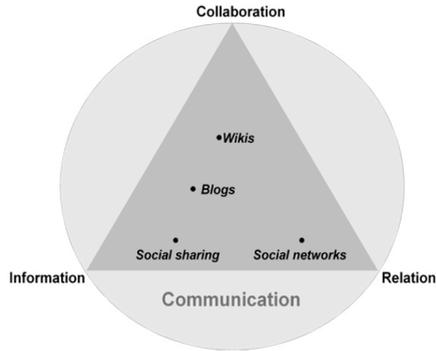


Figure 1. Tools categorized by Eberwald et al. (2008, p. 35)

In this context, the question can be raised as to whether wikis are a social network tool at all. Baumgartner (2006) noted that, in fact, wikis are not a social networking tool but they might still have the effects of social networks:

[...] If wikis are used in a closed group settings to collaboratively work on a certain topic, then this is not designed to make new social networks since the group members were selected by other methods before and registered on this server. If a wiki is used in an open collective setting such as Wikipedia, then it might be true that a wiki can have an effect as a social network. All users who work on specific topics seem to have similar interests. They did meet on a website in order to work on the same topic. (p. 4)

Some major differences between wikis and other Web 2.0 tools also concern the chronology of contributions. While blog posts are sorted chronologically and focus more on an information function, wikis are designed for collaboration (Godwin-Jones, 2003). Although the contribution of a single author can be seen, the visibility of the authors is not the main intention of wikis, whereas in blogs and especially in social networking tools the author of a contribution is more visible. In comparison, wikis are more output-oriented. This means that the authors are visible only in order to reconstruct and show previous versions of a page. This aspect can affect the motivation of the learners if they want their individual contributions to be visible.

Instead of formal hierarchies as in classroom settings or in traditional learning management systems, where the teacher's role is clearly defined, wiki systems lack any hierarchies. Unless other settings are applied, all users are equal (Ebersbach et al., 2008; Konieczny, 2007; Raitman et al., 2005; Schwartz et al., 2004). Nevertheless, differences in participation do appear. On this issue, Nielsen's (2006) well-known 90-9-1-rule seems to apply to social networks: 90% of the users behave more or less as lurkers, 9% contribute a few statements, and 1% contribute more. This participation rate was confirmed by Stegbauer (2009), who did some major studies on Wikipedia, as well as by Ebersbach et al. (2008), and by Thelen and Gruber (2005). It also seems to be valid for online settings in general, if no other external motivation factors such as assignments in formal e-learning settings are applied.

One major advantage of wikis is their ease of use. Studies by Augur et al. (2005) observed that students had almost no problems using a wiki in an e-learning setting. This result was confirmed by Beißwenger and Storrer (2010), who applied wikis at university and at schools. In this regard, Farmer (2004) emphasized that Wikipedia is used by millions of users who might not all be well trained in terms of media competencies, so it must have some ease of use. On the other hand, Panke and Thillosen (2008) expressed concerns regarding the application of more sophisticated functions such as the uploading and integrating of images. Despite these varying opinions upon their ease of use, wikis seem to have a growing share in e-learning settings at universities, schools, and companies.

## **Wikis in E-learning Scenarios**

Looking at the usage of wikis in e-learning scenarios, we find applications in nearly all educational fields: Wikis are used in schools (Beißwenger & Storrer, 2010; Döbeli Honegger, 2005a, 2005b, 2006; Jonietz, 2005; Klampfer, 2005); most intensively in universities (Augar et al., 2005; Bristow, 2005; Edington et al., 2005; Ferris & Wilder, 2006; Gaiser & Thillosen, 2009; Lamb, 2004; Schoderet et al., 2006; Schwartz et al., 2004; Xu, 2007), and in continuing adult education programs, or as knowledge management tools in companies (Bartel, 2006; Brahm et al., 2007; Ebersbach et al., 2008; Majchzak et al., 2006; Robes, 2006).

The range of settings in which wikis are used is widespread. Some institutions use wikis mainly for information purposes as a substitute for Web pages in which they provide relevant general study information to students, who choose wikis because they are so easily edited (Konieczny, 2007; Kleimann, 2007). Some wikis cover specific disciplines such as the StudiGer, i.e., an

information platform of the Technical University of Dortmund German studies (Beißwenger & Storrer, 2008, 2010) and the Pflegewiki, a portal for health-care issues (Panke & Thilloßen, 2008). By now, the Pflegewiki is used mainly by experts since the involvement of more experienced users in the field of health care has reduced the students' involvement (Panke & Thilloßen, 2008)—a phenomenon often observed in online and offline communication processes when participants of different status groups interact (Kerr, 1983; Piezon & Donaldson, 2005). A wiki project where students developed final exams preparation learning materials in an online community is described by Buchem and Hagenhofer (2009).

### **Accompanying Lectures with Wikis**

Wikis are also used to accompany lectures as a tool to either provide material or to have the lectures documented by students. Kleimann (2007) described the usage of a wiki to document lectures through the production of scripts by students, while Hermann and Janzen (2009) described a setting in which students also produce exercises. In this context, O'Neill (2005) asked the question of whether lecture scripts should be provided to students at all or whether students should produce them on their own. But when Klauer et al. (2006) experimented with this kind of setting, they observed that students in a competitive field such as medicine are not willing to take notes for other students unless they are rewarded in some way.

### **Collaborative and Cooperative Text Production in Seminars**

As stated above, wikis are especially appropriate for cooperative and collaborative content production. This process can be applied to gain a deeper and more profound understanding of the learning material (Konieczny, 2007) through active learning processes, especially when learners research material and collect and select it (Dewald et al., 2000; Konieczny, 2007; Kuh, 1996; Longworth, 2003; Ruhl et al., 1987). Collaborative text production is mainly applied in small course settings such as seminars with up to 25 or 40 participants. Examples are described by Beißwenger and Storrer (2008) for students of German philology, by Egloffstein and Städtler (2006) for a pure online setting, by Thelen et al. (2005) for a cooperative setting between students of two universities, by Hodel and Haber (2007) for a planned history course, and by Schorderet (2006) for courses in literature. In a later section of this article, I will give two examples in geographic sciences and theology.

## **Using Public Wikis with Students or Making Student Wikis Public**

Some authors promote the idea that university students should use public wikis such as Wikiversity in order to interact with other groups such as teachers and school students so that they can connect to practical issues and real discussions (Spannagel & Schimpf, 2009). Konieczny (2007), Grauerholz (1999), and Bruns and Humphreys (2005) also supported the idea because of the similarity of Wikipedia articles to certain types of students' paper assignments. Students might also get feedback from experts and learn to interact in a public online discussion. On the other hand, this might be an obstacle for students' participation since they might feel reluctant to write in public wikis if they perceive their own writing capabilities as insufficient. In this case, it helps to provide a separate wiki environment in which texts are prepared and there are internal feedback loops for tutor and / or peer reviews for quality assurance.

Especially in projects where students create wiki-based portals that later become available to the public, wikis seem to have some important advantages. Effects of communal constructivism might unfold when students provide material to a larger community (Holmes et al., 2001). Examples for this kind of wikis are the M/Cyclopedia of New Media at the Queensland University of Technology (Bruns & Humphreys, 2005); the JuraWiki at the Ruhr University Bochum, eLIB, a portal for instructional design at the Universities of Dortmund and Duisburg-Essen (Beißwenger & Storrer, 2008); the Pflegewiki (Panke & Thilloßen, 2008), and the later described projects Podcast-Wiki Physics, ExkursionenWiki, and BasisReliPaed at the University of Frankfurt. Providing information to a larger public could also be considered to be a form of "community work" (Hollis, 2002) or "service learning" (Weigert, 1998) with satisfying and motivating effects for students. Also, the feeling to be part of a larger community might have motivating effects for students (Moskaliuk & Kimmerle, 2008; Ryan & Deci, 2000).

Still, the problem exists that students might reduce their participation when wikis are opened up to a larger public due to the involvement of experts, as had happened in the case of the Pflegewiki (Panke & Thilloßen, 2008). One solution to this problem is that in contrast to public wikis, these projects provide information to the public, but only students can edit pages and contribute material. It depends on the objectives of the projects as to whether this is an appropriate solution. Anyway, if a wiki is planned to be opened to the public, it could be a solution that the access to the wiki is limited to the students during the development phase and the opening is announced in advance, a procedure that was applied in the Pflegewiki and will

be applied for the ExkursionenWiki. In case of BasisReliPaed, the material was prepared in a wiki and published in another platform on the educational server of the State of Hesse.

## **Study of Wikis at the University of Frankfurt**

Based on the observation that in wikis initiated by teachers, the students made almost no contributions beyond the most necessary ones (e.g., assignments) and did not use the wiki to produce texts collaboratively, the question was raised as to whether this behavior might change if wikis are initiated by students. At that time, the university promoted some student projects in which a number of wikis were started by students in order to improve the study conditions for their fellows. In this context, the question was raised as to whether these bottom-up wikis might foster more voluntary participation than top-down wikis initiated by teachers. For the study, several projects were examined concerning aspects such as 1) the purpose of the use of a wiki, 2) the existence of an “editorial team,” 3) the number of people involved in the team, 4) the number of members within an inner circle of writers, and 5) the number of readers and target groups.

The research questions were as follows: Do people overwrite each others’ articles? How is the collaborative text production process organized? Are there any rules? If so, how were these rules arrived at? Are the articles reviewed and / or approved? Are the objectives achieved? Is the target group reached? What is the motivation of the writers to participate and to get involved? Are there any incentives?

The projects’ participants were interviewed and examined in two rounds: In the first one, at least one team member of each project was interviewed on the telephone based on the questions listed above. For the second round, the same project participants were invited for a discussion round of three hours at which one team member from each of the three student projects (OHEF, Blended Learning, and BioKemika) and from both L-News and OKAPI participated. The teachers who initiated ExkursionenWiki and BasisReliPaed were interviewed separately.

The project Podcast-Wiki Physics had the intention to provide information about the research fields in physics in order to help students to decide where to write their bachelor or master thesis in this discipline. This idea was adopted by the project BioKemika, which provides learning material for the usage of databases in chemistry to fellow students. The project Blended Learning is intended to produce materials for psychology students on methods and statistics, and the project OHEF had the objective to help students

with general study relevant material in history and philosophy. KA-Wiki is a project that covers a part of a larger wiki project for culture and science students.

Two projects were examined that were initiated by a Center or a Department: L-Wiki of the Center for Prospective Teachers and OHEF by the Department of History and Philosophy: Both projects have the objective to provide information to students. Furthermore, two projects were included where wikis were used in a classic course setting: ExkursionenWiki is used in the seminar “Geography of Differences.” Here, prospective teachers describe excursions in and around Frankfurt in a wiki that will be opened up to teachers later, teachers in training, and prospective teachers who also want to conduct excursions. In the seminar “BasisReliPaed in theology,” prospective teachers produced learning material for schools that were used later in religion classes in the ninth grade.

## Results

For the results, the answers were analyzed in the different categories and compared to each other. One major result of the study is that in the bottom-up wikis initiated by students more or less the same mechanisms occur as in wikis initiated by teachers. Most students—except for some members of the inner circle—needed to be motivated extrinsically in order to contribute material to the wiki. All the student projects had problems in gaining active writers, and only a small number of students were actively involved in the writing process. Often, extrinsic motivational aspects had to be applied in order to motivate fellow students to get involved into the projects: In the case of BioKemika, students were either rewarded by credit points they received in an introductory seminar for their text contributions or they received gifts sponsored by companies. In all other student projects, contributions were mainly made by the inner circle of the team, ranging from five to 20 members. While collaborative writing in the wiki successfully took place in the wikis used in the seminar settings when demanded by the teachers, the writing processes in the student projects were mainly individual.

In the case of BioKemika, the members of the editorial team produce articles, send them to other team members by e-mail, and give their fellow students one week to react. The team also reported that students outside of the editorial team preferred to provide their text contribution per e-mail, which then was edited and published by an editorial team member. The editors assume that the reason for this procedure lies in the insecurity of contributors concerning quality of text style or a lack of technical competencies. This exact procedure was intentionally used in the L-Wiki project, run by the

university's Center for Prospective Teachers. Here, students were not allowed to contribute text into the wiki directly but had to hand in their contributions via e-mail. Accordingly, participation was low and text production was mainly limited to the editor of the platform, a person employed at the Center. The same situation occurred in OKAPI, a wiki run by the Department of Philosophy that has the objective to provide learning material covering scientific techniques to students. As soon as one person became officially responsible for the wiki, others withdrew, relying on that one person to take care of the whole wiki.

The Ka-Wiki provides Web pages and serves as the content management system and the learning management system of the Department of Cultural Anthropology. Although every teacher and student, as well as guest teachers can edit and add all pages, only a small group of people makes use of this opportunity, mainly the ones who have initiated the project. Teachers also use the wiki to document their courses. Although every page up to the main page is open to every registered user, no vandalism was ever observed.

## Summary

Being doubtful about the quality of their contributions seems to prevent students from participating, especially if experts and people from higher status groups are involved (Ebersbach et al., 2008; on the loss of motivation due to a low subjective ranking of the own contribution, see Kerr & Bruun 1983). This uncertainty can be reduced by the provision of examples and training, consulting, and review processes. BioKemika applied creative and effective mechanisms to encourage students to participate and became part of the study program in the Chemistry Department. Despite some disappointments and unmet expectations, overall almost all of the projects provided satisfying results according to the interviewed team members and teachers.

But real collaborative writing process among two or more students only occurred in projects where the teachers demanded this from their participants and where teams were set up in class (BasisReliPaed, ExkursionenWiki, and during the courses of the Blended Learning project). But as soon as the obligatory examination fell away due to the new Bachelor study program, the Blended Learning tutor groups vanished and so did the wiki. The results also show that if the wiki can be accessed by the public, writing and quality assurance processes often are conducted by e-mail. Only if the wiki is used as a secure working environment, then the collaborative writing process might happen in the wiki itself (BasisReliPaed, ExkursionenWiki).

Almost in all cases, the aspects of communal constructivism and service learning applied, which means that the production of material for a larger group, maybe even the public (which in the case of the ExkursionenWiki and BasisReliPaed are not only fellow students but also teachers and schools), motivated the students to raise the quality of the material they produced.

## References

- AUGAR, N., RAITMAN, R., & ZHOU, W. (2004). Teaching and learning online with wikis. In R. ATKINSON, C. McBEATH, D. JONAS-DWYER, & R. PHILLIPS (Eds.), *Beyond the comfort zone: Proceedings of the 21st ASCILITE Conference* (pp. 95-104). Perth: ASCILITE. Retrieved September 25, 2012 from <http://www.ascilite.org.au/conferences/perth04/procs/pdf/augar.pdf>
- BARTEL, T. (2006). Nutzung von Wikis in Unternehmen—Kollaboratives Arbeiten mit Wikis im Unternehmensumfeld. In 8. KnowTech “Mit WM besser im Wettbewerb!” (pp. 161-168). München: CMP-WEKA Verlag.
- BAUMGARTNER, P. (2006). Web 2.0: Social software & e-Learning. *Computer + Personal (CoPers), Schwerpunktheft E-Learning und Social Software*, 14 (8), 20-24 and 34. Retrieved November 10, 2012 from <http://tinyurl.com/a4stabh>
- BEIBWENGER, M., & STORRER, A. (2008). Wiki-Einsatz in universitären Blended Learning-Szenarien: Konzepte und Erfahrungen aus der Dortmunder Germanistischen Linguistik. In S. HAMBACH, A. MARTENS, & B. URBAN (Eds.), *e-Learning Baltics 2008. Proceedings of the 1st International eLBA Science Conference* (pp. 129-138). Stuttgart: Fraunhofer IRB Verlag.
- BEIBWENGER, M., & STORRER, A. (2010). Kollaborative Hypertextproduktion mit Wiki-Technologie. Beispiele und Erfahrungen im Bereich Schule und Hochschule. In E.-M. JAKOBS, K. LEHNEN, & K. SCHINDLER (Eds.), *Schreiben und Medien. Schule, Hochschule, Beruf* (pp. 13-36). Frankfurt am Main: Peter Lang.
- BioKemika. (2012, March 5). Retrieved September 22, 2012 from <http://biokemika.uni-frankfurt.de/wiki/BioKemika:Hauptseite>
- BRAHM, T., SELINA I., & BRUNO W. (2007). Pädagogische Einsatzszenarien von Wikis unter besonderer Berücksichtigung der Nutzung an der FHS St. Gallen—Hochschule für Angewandte Wissenschaften. In S. SEUFERT, & T. BRAHM (Eds.), *“Ne(x)t Generation Learning”: Wikis, Blogs, Mediacasts & Co.—Social Software und Personal Broadcasting auf der Spur* (pp. 54-68). St. Gallen: Swiss Center for Innovations in Learning. Retrieved September 25, 2012 from <http://www.scil.ch/fileadmin/Container/Leistungen/Veroeffentlichungen/2007-02-euler-seufert-next-generation-learning.pdf>
- BREMER, C. (2011). Studentische E-Learning Projekte in der Hochschule. In H. DÜRNBERGER, S. HOFHUES, & T. SPORER (Eds.), *Offene Bildungsinitiativen:*

- Fallbeispiele, Erfahrungen und Zukunftsszenarien* (pp. 41-55). Münster: Waxmann.
- BRISTOW, R. (2005). Beyond email: Wikis, blogs and other strange beats. *Ariadne*, 42. Retrieved September 25, 2012 from <http://www.ariadne.ac.uk/issue42/beyond-email-rpt>
- BRUNS, A., & HUMPHREYS, S. (2005). Wikis in teaching and assessment: The M/cyclopedia project. In *Proceedings of the 2005 international symposium on wikis* (pp. 25-32) New York: ACM Press. Retrieved September 25, 2012 from <http://www.wikisym.org/ws2005/proceedings/paper-03.pdf>
- BUCHEM, I., & HAGENHOFER, T. (2009). Didaktische Entscheidungen bei der Konzeption einer Web 2.0-basierten Lernercommunity. In A. SCHWILL, & N. APOSTOLOPOULOS (Eds.), *Lernen im digitalen Zeitalter* (pp. 26-28). Berlin: Logos.
- DEWALD, N., SCHOLZ-CRANE, A., BOOTH, A., & LEVINE, C. (2000). Information at a distance: Instructional design issues. *The Journal of Academic Librarianship*, 26, 33-44.
- DÖBELI HONNEGGER, B. (2005a). Wiki und die starken Lehrerinnen. In F. STEFFEN (Ed.), *Unterrichtskonzepte für informatische Bildung. Fachtagung Informatik und Schule der Gesellschaft für Informatik* (pp. 173-183). Berlin: Logos Verlag.
- DÖBELI HONNEGGER, B. (2005b). Wikis: a rapidly growing phenomenon in the German-speaking school community. In *Proceedings of the 2005 international symposium on wikis* (pp. 113-116). New York: ACM Press. Retrieved September 25, 2012 from <http://www.wikisym.org/ws2005/proceedings/paper-10.pdf>
- DÖBELI HONEGGER, B. (2006). Wiki und die starken Texte. Schreibprojekte mit Wikis. *Deutschmagazin*, 1(06), 15-19. Retrieved September 25, 2012 from <http://beat.doebe.li/publications/wiki-und-die-starken-texte-doebeli-honegger-deutschmagazin.pdf>
- EBERSBACH, A., GLASER, M., & RICHARD, H. (2008). *Social Web*. Berlin: Springer.
- EDINGTON, J., FUNK, J., THORPE, R., & WARRINGTON, J. (2005). Professional applications of wikis and weblogs. Retrieved September 25, 2012 from <http://www.docstoc.com/docs/28679464/h3The-Professional-Applications-of-Wikis-and-Weblogsh3#>
- EGLOFFSTEIN, M., & STÄDTLER, H. (2006). Didaktisch fokussierte Umsetzung des Wiki-Konzepts in einem virtuellen Seminar zum Selbstorganisierten Lernen. In C. RENSING (Ed.), *Proceedings der Pre-Conference Workshops der 4. e-Learning Fachtagung Informatik DeLFI* (pp. 115-123). Berlin: Logos Verlag.
- eLIB. (n.d.). Retrieved July 13, 2012 from <http://www.literature.at/elib>
- FARMER, S. J. (2004, June 10). The wide world of wiki: Choosing a wiki for an element of a fully online undergraduate course. Message posted to <http://radio.weblogs.com/0120501/2004/06/10.html>
- FERRIS, S. P., & WILDER, H. (2006). Uses and potentials of wikis in the classroom. *Innovate*, 2(5). Retrieved November 10, 2012 from <http://tinyurl.com/abqb7cw>

- GAISER, B., & THILLOSEN, A. (2009). Hochschullehre 2.0 zwischen Wunsch und Wirklichkeit. In N. APOSTOPOULOS, H. HOFFMANN, V. MANSMANN, & A. SCHWILL (Eds.), *E-Learning 2009. Lernen im digitalen Zeitalter* (pp. 185-196). Münster: Waxmann.
- GODWIN-JONES, R. (2003). Emerging technologies. Blogs and wikis: Environments for online collaboration. *Language Learning & Technology*, 7(2). 12-16. Retrieved September 25, 2012 from <http://lt.msu.edu/vol7num2/pdf/emerging.pdf>
- GRAUERHOLZ, L. (1999). Creating and teaching writing intensive courses. *Teaching Sociology*, 27, 310-323.
- HERMANN, C., & JANZEN, A. (2009). E-Lectures-Wiki—Aktive Nutzung von Vorlesungsaufzeichnungen. In N. APOSTOPOULOS, H. HOFFMANN, V. MANSMANN, & A. SCHWILL (Eds.), *E-Learning 2009. Lernen im digitalen Zeitalter* (pp. 127-138). Münster: Waxmann.
- HIPPNER, H. (2006). Bedeutung, Anwendung und Einsatzpotentiale von Social Software. In K. HILDEBRANDT, & J. HOFFMANN (Eds.), *Social Software* (pp. 6-16). Heidelberg: dpunkt.
- HODEL, J., & HABER, P. (2007). Das kollaborative Schreiben von Geschichte als Lernprozess. Eigenheiten und Potenzial von Wiki-Systemen. In M. MERKT, K. MAYBERGER, R. SCHULMEISTER, A. SOMMER, & I. VAN DEN BERK (Eds.), *Studieren neu erfinden—Hochschule neu denken* (pp. 43-53). Münster: Waxmann.
- HOLLIS, S. (2002). Capturing experience: transforming community service into service learning. *Teaching Sociology*, 30, 200-213.
- HOLMES, B., TANGNEY, B., FITZGIBBON, A., & SAVAGE, T. (2001). Communal constructivism: Students constructing learning *for* as well as *with* others. Retrieved November 10, 2012 from <http://tinyurl.com/azbamas>
- JONIETZ, D. (2005). Ein Wiki als Lernumgebung? Überlegungen und Erfahrungen aus schulischer Sicht. In J. M. HAAKE, U. LUCKE, & D. TAVANGARIAN (Eds.), *DeLFI 2005: 3. Deutsche e-Learning Fachtagung Informatik* (pp. 35-44). Berlin: Logos Verlag.
- JuraWiki. (2012, June 10). Retrieved September 25, 2012 from <http://www.jurawiki.de/StartSeite>
- KERR, N. L. (1983). Motivation losses in small groups: A social dilemma analysis. *Journal of Personality and Social Psychology*, 45, 819-828.
- KERR, N. L., & BRUUN, E. (1983). Dispensability of member effort and group motivation losses: Free rider effects. *Journal of Personality and Social Psychology*, 44, 78-94.
- KLAMPFER, A. (2005). Wikis in der Schule—Eine Analyse der Potentiale im Lehr-/Lernprozess. (Bachelor thesis, University of Hagen). Retrieved November 10, 2012 from <http://tinyurl.com/aspg98d>
- KLAUER, G., MELAMED, R., TILLMANN, A., REINHOLD, S., & KANDSPERGER, L. (2006). MediaWiki als Werkzeug zur kooperativen Erstellung einer

- Vorlesungsmitschrift in der Humananatomie. In C. RENSING (Ed.), *Proceedings der Pre-Conference Workshops der 4 e-Learning Fachtagung Informatik DeLFI* (pp. 107-104). Berlin: Logos Verlag.
- KLEIMANN, B. (2007). eLearning 2.0 an deutschen Hochschulen. In M. MERKT, K. MAYBERGER, R. SCHULMEISTER, A. SOMMER, & I. VAN DEN BERK (Eds.), *Studieren neu erfinden—Hochschule neu denken* (pp. 149-201). Münster: Waxmann.
- KONIECZNY, P. (2007). Wikis and Wikipedia as a teaching tool. *International Journal of Instructional Design and Distance Learning*, 4(1), 15-30. Retrieved September 25, 2012 from [http://www.itdl.org/journal/jan\\_07/article02.htm](http://www.itdl.org/journal/jan_07/article02.htm)
- KUH, G. (1996). Guiding principles for creating seamless learning environments for undergraduates. *Journal of College Student Development*, 37(2), 135-148.
- LAMB, B. (2004). Wide open spaces: Wiki ready or not. *Educause Review*, 39(5), 36-48.
- LONGWORTH, N. (2003). *Lifelong learning in action: Transforming education of the 21st century*. London: Kogan Page.
- L-Wiki. (2011, February 28). Retrieved September 25, 2012 from <http://www.l-wiki.uni-frankfurt.de>
- MAJCHZAK, A., WAGNER, C., & YATES, D. (2006). Corporate wiki users: results of a survey. In *Proceedings of the 2006 international symposium on wikis* (pp. 99-104). New York: ACM Press. Retrieved September 25, 2012 from [http://www.cs.helsinki.fi/u/vahakota/stuff/corporate\\_wiki\\_users.pdf](http://www.cs.helsinki.fi/u/vahakota/stuff/corporate_wiki_users.pdf)
- MOSKALIUK, J., & KIMMERLE, J. (2008). Wikis in der Hochschule—Faktoren für den erfolgreichen Einsatz. Retrieved November 11, 2012 from <http://bit.ly/RNzEOx>
- NIELSEN, J. (2006, October 6). Participation inequality: Encouraging more users to contribute. Retrieved November 10, 2012 from <http://tinyurl.com/vvyzf>
- OKAPI. (2012, September 5). Retrieved September 25, 2012 from <http://okapi.uni-frankfurt.de/index.php?title=Hauptseite>
- O'NEILL, M. E. (2005). Automated use of a wiki for collaborative lecture notes. In *Proceedings of the 36th SIGCSE technical symposium in computer science education* (pp. 267-271). New York: ACM Press.
- PANKE, S., & THILLOSEN, A. (2008). Unterwegs auf dem Wiki-Way. Wikis in Lehr- und Lernsettings. Retrieved November 11, 2012 from <http://bit.ly/SOM5vX>
- Pflegewiki. (2012, September 21). Retrieved September 25, 2012 from <http://www.pflegewiki.de/wiki/Hauptseite>
- PIEZON, S. L., & DONALDSON, R. L. (2005). Online groups and social loafing: Understanding student-group interactions. *Online Journal of Distance Learning Administration*, VIII(IV). Retrieved November 11, 2012 from <http://bit.ly/H0Vp3y>
- Podcast-Wiki Physik Online. (n. d.). Retrieved September 22, 2012 from <http://podcast-wiki.physik.uni-frankfurt.de/wiki/Hauptseite>

- RAITMAN, R., AUGAR, N., & ZHOU, W. (2005). Employing wikis for online collaboration in the e-Learning environment: Case study. In X. HE, T. HINTZ, M. PICCARDI, Q. WU, M. HUANG, & D. TEIN (Eds.), *Proceedings of the 3rd international conference on Information Technology and Applications* (pp. 142-146). Los Alamitos: IEEE Computer Society.
- ROBES, J. (2006). e-Learning 2.0: Neue Perspektiven für die Weiterbildung. Retrieved November 10, 2012 from <http://tinyurl.com/ba7qxov>
- RUHL, K. L., HUGHES, C. A., & SCHLOSS, P. J. (1987). Using the pause procedure to enhance lecture recall. *Teacher Education and Special Education, 10*, 14-18.
- RYAN, R. M., & DECI, E. L. (2000). Self-determination theory and the facilitation of intrinsic motivation, social development, and well-being. *American Psychologist, 55*(1), 68-78.
- SCHODERET, A. (2006). E-Learning über Online-Edition literarischer Texte mit Wiki. In E. SEILER SCHIEDT, S. KÄLIN, & C. SENGSTAG (Eds.), *E-Learning—alltagstaugliche Innovation?* (pp. 183-194). Münster: Waxmann.
- SCHWARTZ, L., CLARK, S., COSSARIN, M., & RUDOLPH, J. (2004). Educational wikis: Features and selection criteria. *The International Review of Research in Open and Distance Learning, 5*. Retrieved November 11, 2012 from <http://bit.ly/Panz0g>
- SPANNAGEL, C., & SCHIMPF, F. (2009). Öffentliche Seminare im Web 2.0. In A. SCHWILL, & N. APOSTOLOPOULOS (Eds.), *Lernen im Digitalen Zeitalter* (pp. 13-20). Berlin: Logos.
- STEGBAUER, C. (2009). *Wikipedia. Das Rätsel der Kooperation*. Wiesbaden: Verlag für Sozialwissenschaften.
- StudiGer. (2012, September 21). Retrieved September 25, 2012 from <http://www.studiger.tu-dortmund.de>
- THELEN, T., & GRUBER, C. (2005). Textproduktions- und Kommunikationsprozesse in WikiWikiWebs. In H.-W. HUNEKE (Ed.), *Geschriebene Sprache: Strukturen, Erwerb, Modellbildung* (pp. 183-202). Weinheim: Deutscher Studienverlag.
- WEIGERT, K. M. (1998). Academic service learning: Its meaning and relevance. *New Directions for Teaching and Learning, 73*, 3-10.
- XU, L. (2007). Project the wiki way: using wiki for computer science course project management. *Journal of Computing Sciences in Colleges, 22*(6), 109-116.



# Fostering Crossmedia Literacy in Formal Educational Contexts: Conceptual Considerations and Case-Specific Results

Sandra Hofhues<sup>I</sup>, Christian Geier<sup>II</sup>, and Lena Griebhammer<sup>II</sup>

<sup>I</sup> Hamburg University of Applied Sciences, <sup>II</sup> University of Augsburg

*First, this contribution will explain what crossmedia means in the face of the omnipresence of computers, the Internet, and digital media, and it will define the term and its underlying concept in comparison to other commonly used terms. Then the constraints that (might) affect the fostering of crossmedia concepts in university teaching will be examined. Referring to basic assumptions, an example from the University of Augsburg will show how crossmedia concepts can be integrated on a micro-pedagogical level in an individual academic course. Finally, questions will be asked regarding personal requirements for teaching and learning in crossmedia conditions as well as the challenges involved in embedding them on a structural level.*

## Introduction

Publicly used and media-facilitated communication that does not represent mass communication but many-to-many communication is still at a very early stage, and it cannot be established yet how sustainable it will prove to be in terms of use and occurrence. However, because of or despite the fact that this is a most recent phenomenon, it contributes to academic learning

through media use on three levels: media as a source of (scientific) information, media as a tool for coping with new challenges, and media as a way of creating realistic learning and teaching environments (Marotzki & Jörissen, 2008). As a consequence, the consideration of media concepts in university teaching also needs to go beyond focusing on just one particular medium; instead, it needs to encompass and make available a “variety of a number of media areas” (Sesink, 2008, p. 15) and to facilitate systematic analysis of these media areas in a self-evaluating process.

Following Jakubetz (2011), the subject of crossmedia, as referred to in this article, is expected to provide some form of “salvation” in journalism because it combines technical, strategic, and conceptual aspects in a sophisticated way, adds further options for circulation to the existing way of media distribution, and brings producers and recipients closer together than before. Whether such a concept should be identified as crossmedia or media convergence is of secondary interest insofar as it refers to two sides of the same coin: Crossmedia specifies the links between various media formats, mainly from the perspective of content producers and with regard to economic factors. Media convergence describes different media formats joining together from the viewpoint of recipients (Schorb, 2007). In relation to this, for example, the effect of media formats is examined because, “with the growing convergence of media, the boundaries between ‘information’ and other media have become increasingly blurred” (Buckingham, 2010, p. 59). There is also the question of which new challenges have to be met by recipients or users when dealing with interconnected media formats. In this respect, the term “crossmedia” places more emphasis on the planning and design of media production, whereas the term “media convergence” highlights changes in media from an observant-analytical angle.

The following will be mainly referred to as *crossmedia* while encompassing both aspects: The objective is to broaden the subject of media convergence (which is mainly influenced by communication science and media pedagogics) and to include economic perspectives of crossmedia planning and design. Thus, production areas, ways of distribution, and crossmedia products will be considered together.

Broadening perspectives in this way can be put into the context of the discourse on media competence in general and in relation to the academic environment. Furthermore, it can be included in the discourse on comprehensive media education; this highlights its significance for lifelong learning in comparison to obtaining isolated skills in the short term (see Hofhues & Schiefner-Rohs, in print) and significantly intensifies the awareness of new (not yet tangible) developments within the media, thus paving the way to

flexible responses to such developments. At the same time, crossmedia is, necessarily, only one area of media-related, competency-oriented support that nevertheless has its own significance (Sesink, 2008) given the omnipresence of computers, the Internet, and digital media and the ongoing process of breaking boundaries in science and society (Derrida, 2001; Baecker, 2007).

## Crossmedia in University Teaching

The consideration alone of what crossmedia is or could be in the face of changing learning habits, increasingly vague learning environments, and media shifts implies that engaging with crossmedia in academic learning and teaching is essential. A closer look at the curricula of media-related study courses at German universities reveals, however, that the analysis of this phenomenon is almost neglected. Accordingly, the normative premise for theoretical-conceptual considerations has to be a deficit-oriented one. In the context of university education, however, such deficits can be explained because comprehensive considerations question, at least in part, the traditional curricula of the media and communication sciences and enable interdisciplinary access to the issue. In line with this, certain constraints can be identified that potentially can obstruct the fostering of crossmedia in university education. University teachers as well as students are affected by this; the reasons may vary regarding relevance, the attribution of competence, and implementation (either curricular or extra-curricular), as will be shown below.

**Lack of relevance.** It may be commonplace that crossmedia content is only made the subject of academic courses if the phenomenon is seen as a challenge for planning and designing media products or for researching media effects. If, as it is often the case, teachers do not show a lot of interest in interdisciplinary engagement with certain media, the situation concerning crossmedia is hardly any better. Crossmedia is too deeply associated with the commitment to disseminate all content through all channels at all times. Although this idea has already been abandoned in everyday journalism (Jakubetz, 2011, p. 32), the presumption of total media networking currently still prevails in academic research and teaching. In addition, the perceived relevance of the content (marginal topic vs. subject of genuine research and teaching) differs insofar that even those who are at least familiar with the term assume that excessive expectations, as described above, loom large over them. Furthermore, teachers and students often shy away from interdisciplinary analysis. At universities, orientation toward design is under-represented: analyzing the status quo of phenomena is much better established than de-

sign-oriented analysis and appropriately structured academic courses. Even in the field of media pedagogics, there is disagreement on how relevant action-based teaching and learning is for the scientific analysis of crossmedia or for other areas of media education. In this respect, the outcome presents two constraints: regarding crossmedia as relevant content or as a significant subject and regarding subject-based perspectives.

**Self-attribution of competence.** Taking the current subject-related discussion as a reference point, neither students nor teachers will be attributed with “media literacy” as such—media literacies are acquired and developed too individually, and with regard to the intended development of media literacies, there are too many different perspectives on the academic learning environment with its core subject areas. This cautious and also critical assessment is at odds with the public debate on the increasing presence of media in normal life and, figuratively, also in everyday life at university. It is also at odds with some hopes that are often publicly discussed in connection with the ubiquity of the media: for example, those of democratic participation, the utilization of digital media in academic learning and teaching, and competent media usage in general. Instead, the omnipresence of computers, the Internet, and digital media often leads to students’ and teachers’ being overwhelmed by the implementation of the media in academic learning and teaching. As another consequence, they hardly use digital media (Grosch & Gideon, 2011; Kleimann, 2007). Accordingly, in the context of the university, the net generation is “de-mystified” (Schulmeister, 2009), and teachers are increasingly required to build academic media literacies (Mayrberger, 2010; Wedekind, 2008), which encompass knowledge, skills, and attitudes in the field of the media in a context-appropriate way. With regard to crossmedia, this means, for example, knowledge of media convergence, skills for conceptualizing and analyzing various ways of content distribution through the media, and a certain attitude toward crossmedia in studies or work (Sesink, 2008). For university teachers, this also includes teaching skills, so that crossmedia-inspired academic courses could convey not only sound theoretical knowledge but also offer real-life learning situations and, at times, certain qualifications. At this point, the university faces a conflict that is difficult to resolve: In particular, the aspects of professional qualification in crossmedia learning scenarios are often rejected, and the use of more action-based learning is frequently misinterpreted as “bad” practice (i.e., not scientific enough).

**Lack of implementation.** Teachers often neglect to encourage students’ active participation, which can be stimulated by means of action-based learning. Implementation often fails in terms of time, structure, and content because of staff discontinuity and a lack of suitable ways of communication

within the university. It also needs to be established that it is often a priority for students to do well in a (subject-related) project and that making progress in crossmedia takes second place (in study courses with no focus on media). In this respect, the lack of curricular implementation impacts the engagement with crossmedia in two ways: It affects students, who are overstrained by the demands of self-organized learning in complex media scenarios (Reinmann, 2009) and also by the increasing complexity of content. Also, it affects teachers, who, in consideration of their own routine, resort to seemingly proven teaching methods when facing the demands imposed on them and a structural framework that is either vague or missing.

The consideration of these constraints reveals implications regarding the potential of fostering crossmedia concepts and skills that correlate with concepts and structures of micro-teaching scenarios (e.g., academic courses, projects), with the individual person and his / her media literacies, and with the implementation of an adequate framework (e.g., the option of building credits or co-curricular organization).

### **Example of Implementing Crossmedia**

There are different ways to deal with the constraints outlined above. A pragmatic approach would be to accept them as a part of the university framework and make the best of the situation in day-to-day teaching. In the following, an example from the University of Augsburg will show how the challenges identified above—relevance, individual media competence, and structural implementation—can be met in a constructive way. A learning scenario will be described that incorporates these aspects through content planning and the implementation of teaching methods. This example will clarify how crossmedia concepts and practice enable a form of learning and teaching that includes a variety of participants in the learning process and makes formal learning more accessible (Baecker, 2007).

### **The Course “Crossmedia Ethnological Communication”**

The media laboratory of the Institute of Media and Education Technology and the Chair of European Ethnology (both at the University of Augsburg) took steps to incorporate crossmedia in the study course “Crossmedia Ethnological Communication” during the winter semester of 2011/12. The course aimed at examining ethnological questions, presenting the results with the help of crossmedia, and providing media tools for research to find answers to the questions and to enable networking among the students. A broad definition of the ethnological research includes a variety of questions relating to

cultural and traditional aspects of everyday life, festivities, and celebrations (e.g., Brednich, 1994; Kaschuba, 1999). This approach makes ethnological research accessible for students of other disciplines. In addition, scientific empirical methods of ethnological research (Götttsch, 2001) can be used together with crossmedia, i.e., interviews (podcasts), observation (videos), and fieldwork diaries (weblog).

The course, planned and conducted in team-teaching by Christian Geier and Lena Griebhammer, was organized as an interdisciplinary course and made available for 20 participants from study courses of two different faculties: the Faculty of Philosophy and Social Sciences and the Faculty of Philology and History. It emerged in the first session that a considerable effort regarding coordination and communication was needed: The adjustment of the various requirements of the different study courses alone—including the history of arts and culture, media and communication studies (MuK), sociology, and teaching degrees—took considerably more time and attention than expected. Despite these difficulties, the structure of the course could be maintained, although, upon the request of the participants, minor adjustments to the schedule had to be made during the course.

To not over-challenge the participating students, the use of media was confined to three different formats: text (weblog), audio (podcast), and video. The intention was to encourage the students to produce more than one medium and become aware of the fact that the media were connected in a potential way (Buckingham, 2010, p. 69). The three media were prepared in such a way that the research of the ethnological question led from one type of media to the next. Two weblogs were continued by students beyond the study course, but the crossmedia approach of the course was not pursued any further.

All students had some everyday experience with media but only a little experience in media production and design (according to the information provided by the students at the beginning of the course). Only some of the students had more comprehensive media knowledge—they were majoring in MuK. The study course requires the theoretical analysis of media and of past and recent empirical studies of media use.

A particular focal point of teaching at the start of the course was the handling of text when using the available media. In line with Buckingham (2010, pp. 62-63), working with text as a common element of all three media and, along with this, developing reading and writing skills can be understood as a target dimension in the area of media literacy. Following Schorb (2007)—and with action-oriented learning in mind—it can also be said that “media activity begins with understanding and judging the sign language of media”

(p. 27). Basic technical training provided support for producing and designing podcasts and videos and their integration in a weblog. As was repeatedly mentioned during the plenary sessions, one particular session on crossmedia and corporate design was dedicated to focusing on core aspects of strategy and linking content. The course was organized in a project-oriented way so that students were able to set their own milestones along which they could develop, present, and alter their media products. Due to the media experience of the course instructors, it was also possible to provide feedback on content and to reflect critically on product design and ideas for distribution and discuss them during the plenary sessions. The podcast and the weblog were completed during the semester to ensure a continuous production process and to give the participants the opportunity to work on the technically more demanding video production during the semester break. It was an important teaching and learning objective to build media knowledge and the ability to analyze media critically before starting with practical media work. At the same time, the technical complexity of the project was taken into account: Based on text as a common denominator, conceptual aspects, commonalities, and differences of the individual media as well as technical aspects were implemented (step by step and increasing in complexity). A lot of attention was also paid to mutual critical reflection. The essential tasks of the course were solved to the satisfaction of the lecturers, given the instruction they provided to the students. The results are documented in weblogs, and most of them are available; only a few videos are password-protected to respect the privacy of some fieldwork participants and interviewees.

Table 1. Crossmedia outcomes of the course (examples)

Name	...wordpress.com	Topic	Media
FazinationKiosk	zeitungskiosk	Phenomenon of kiosk	Blog, Audio
snapshotme	snapshotme	Photography as profession and for daily life	Blog, Audio, Video
Halloween	hallowin	Halloween	Blog, Audio, Video
Es weihnachtet sehr...	facebookxmas	Christmas	Blog, Audio, Video
wgmalanders	wgmalanders	WG-life	Blog, Audio, Video
Bahnekdoten	bahnstories	Rail journeys in Tweets	Blog, Audio, Video
Volksmusik in Bulgarien	bgvolksmusik	Bulgarian folk music	Blog, Audio, Video
Augsburg-Lechhausen	domazone	Portrait of the urban quarter Lechhausen	Blog, Audio

According to the self-assessment of the lecturers, the structure of the course proved to be successful, which is confirmed by the evaluation results, which

indicate the feasibility of crossmedia teaching and learning scenarios in principle and confirm that students are not over-challenged by them. However, it must be noted that the students rated the amount of work that had to be invested in the course as significantly higher than in other seminars.

### **Implications for Planning Learning Scenarios**

The case shows that crossmedia planning and realization in education is a challenge that exceeds the usual planning needs for an academic course: The course was team-taught by the two lecturers who both have some knowledge of each other's field of expertise. Ethnological topics and media education content were equally considered. At the same time, the course was open to students from different study courses so that there were opportunities for peer coaching in the sense of an exchange between students who were more experienced in the media and those who were more advanced in ethnological studies. The course was also a regular part of each subject curriculum, with the result that queries regarding the applicability of the course were limited to the number of achievable credit points; at least during the winter semester of 2011/12, the course was, in principle, embedded in the respective studies' courses. Regarding the constraints outlined above, the following implications can be identified.

Our use of crossmedia has proved to represent a cross-sectional dimension: While ethnologic issues and media issues were equally considered, the reception of media-related content had stronger leaning toward digital literacies and production / design skills (Schorb, 2007). However, through the example of an academic course, the potential of crossmedia analysis becomes clear: Crossmedia seems to be particularly suitable for documenting an awareness of the continuous shift in the media and for meeting the challenges of individual subjects by studying a topic while making constructive use of modified methods and approaches. At the same time, crossmedia as a phenomenon lends itself to planning for interdisciplinary teaching. Access was certainly made easier by the fact that the course was embedded in the subject curricula. Without this, individual teachers and students would have been left to make their own assessments of the relevance of crossmedia in the course, which would not necessarily have been without merit, but it would have turned any comprehensive thematization of the course into a personal and organizational challenge.

The study course concept described above shows that competence-related demands on teachers and students change again if they conceptualize courses in terms of crossmedia and act accordingly in the appropriate scenarios. Consequently, the growing demands on both target groups and the already long-

standing debate on fostering media literacy and information skills at universities need to be put in context. That is to say, if academic courses are supposed to be planned and realized in a crossmedia-oriented way, basic media competence will be essential as a disposition for media practice with a clear emphasis on media literacy, not information literacy (see Schiefner-Rohs, 2012). After all, “media literacy [is] the prerequisite for being master, not servant of the media” (Schorb, 2007, p. 23). At the same time, media literacies in the context described here need to be understood in a broader sense than merely in relation to creating a crossmedia product. To avoid widely heterogeneous learning outcomes and to teach media-related cultural skills in a critical, self-reflective way (Sesink, 2008), technical training is indispensable and as essential as issues of critical assessment of media, which have been discussed constantly by teachers and students during courses (Schorb, 2007, p. 26).

While the first aspect of crossmedia as a cross-section dimension follows an educational idea, the aspect of the individual development of competence focuses on practical media literacies that both students and teachers obtain during the course. In contrast to an academic course where the presence of lecturers ensures media pedagogical support as well as technical help, this kind of support needs to be provided within an environment where formal learning conflicts with an open setting; this concerns the concept of media-supported teaching and learning environments, but especially support for students and teachers in an actual teaching and learning situation (e.g., Seufert & Euler, 2003). Therefore, implementation in curricula would not only place emphasis on the subject-related relevance but, above all, it would lift the burden of developing media literacies as a singular task and point toward long-term competence development. Both perspectives also indicate the significance of increased cooperation and networking given the changes in how knowledge is generated at universities. However, not all approaches to crossmedia concepts and practice (Mahrtdt, 2009, p. 7, 17) are useful in the field of academic work, yet there would be scope for orientation. Opportunities for self-study or co-curricular projects relating to crossmedia might also emerge, which could be located between independent study and tutoring through teachers (see Hofhues & Heudorfer, in print). In any case, the analysis of crossmedia requires open-mindedness on the part of teachers and students, readiness to engage with the subject, and motivation to develop away from trodden paths and to get involved in interdisciplinary examination of (research) questions while facing all the challenges that usually come with such changes in reality.

## Conclusion

The analysis of the issue of crossmedia has shown that, for the time being, there is a need for new terms and concepts with everyday relevance that would meet the changes happening in learning habits, as well as in places for generating knowledge and in academic study and learning. Thus, engaging with crossmedia is by no means part of day-to-day teaching at universities. One option—apart from others—would be to plan academic courses with a crossmedia context that would not only be characterized by their particular focus on topic and content but also view the phenomenon from a distinctive design perspective and organize teaching in a new way (e.g., by means of a project-oriented approach). Insofar as the considerations exhibit similarities to projects based on action-orientated media pedagogics, there is a vital difference in discussions in school and informal education: This difference is rooted in the university context itself, which largely neglects the systematic development of media literacies. The chance, therefore, lies in interconnecting the university context: in crossmedia as an up-to-date topic area as well as in everyday life.

## References

- BAECKER, D. (2007). *Studien zur nächsten Gesellschaft*. Frankfurt: Suhrkamp.
- BREDNICH, R. W. (Ed.). (1994). *Grundriss der Volkskunde. Einführung in die Forschungsfelder der Europäischen Ethnologie*. Berlin: Reimer.
- BUCKINGHAM, D. (2010). Defining digital literacy. What young people need to know about digital media. In B. BACHMAIR (Ed.), *Medienbildung in neuen Kulturräumen. Die deutschsprachige und britische Diskussion* (pp. 59-71). Wiesbaden: VS.
- DERRIDA, J. (2001). *Die unbedingte Universität*. Frankfurt: Suhrkamp.
- GÖTTSCHE, S. (Ed.). (2001). *Methoden der Volkskunde. Positionen, Quellen, Arbeitsweisen der Europäischen Ethnologie*. Berlin: Reimer.
- GROSCHE, M., & GIDEON, G. (2011). *Mediennutzungsgewohnheiten im Wandel. Ergebnisse einer Befragung zur studiumsbezogenen Mediennutzung*. Karlsruhe: KIT Scientific Publishing.
- HOFHUES, S., & HEUDORFER, A. (in print). Students' publishing projects and their impact on teaching and learning. In M. PATRUT, & B. PATRUT (Eds.), *Social media in higher education: Teaching in Web 2.0*. New York: IGI Global.
- HOFHUES, S., & SCHIEFNER-ROHS, M. (in print). Crossing Boundaries: Handeln in Medienprojekten zwischen Medienbildung und ökonomischer Bildung. *merz Wissenschaft 2012*.

- JAKUBETZ, C. (2011). *Crossmedia*. Konstanz: UVK.
- KASCHUBA, W. (1999). *Einführung in die Europäische Ethnologie*. München: Beck.
- KLEIMANN, B. (2007). E-Learning 2.0 an deutschen Hochschulen. In M. MERKT, K. MAYRBERGER, R. SCHULMEISTER, A. SOMMER, & I. VAN DEN BERK (Eds.), *Studieren neu erfinden—Hochschule neu denken* (pp. 149-158). Münster: Waxmann.
- LAUGHEY, D. (2011). Media studies 1.0: Back to basics. *Media Education Research Journal*, 2(2), 57-64.
- MAHRDT, N. (2009). *Crossmedia. Werbekampagnen erfolgreich planen und umsetzen*. Wiesbaden: Gabler.
- MAROTZKI, W., & JÖRISSEN, B. (2008). Medienbildung. In U. SANDER, F. VON GROSS, & K. U. HUGGER (Eds.), *Handbuch Medienpädagogik* (pp. 100-109). Wiesbaden: VS.
- MAYRBERGER, K. (2010). Web 2.0 in der Hochschullehre—Überlegungen zu einer (akademischen) Medienbildung für „E-Learning 2.0“. In B. HERZIG, D. M. MEISTER, H. MOSER, & H. NIESYTO (Eds.), *Jahrbuch Medienpädagogik 8. Medienkompetenz und Web 2.0* (pp. 309-328). Wiesbaden: VS.
- REINMANN, G. (2009). Selbstorganisation auf dem Prüfstand: Das Web 2.0 und seine Grenzen(losigkeit). In B. BACHMAIR (Ed.), *Medienbildung in neuen Kulturräumen* (pp. 75-89). Wiesbaden: VS.
- SALMON, G. (2008). The future for podcasting. In G. SALMON, & P. EDIRISINGHA (Eds.), *Podcasting for learning in universities* (pp. 169-177). Maidenhead: Open University Press.
- SCHIEFNER-ROHS, M. (2012). *Kritische Informations- und Medienkompetenz. Theoretisch-konzeptionelle Herleitung und empirische Betrachtungen am Beispiel der Lehrerausbildung*. Münster: Waxmann.
- SCHORB, B. (2007). Zur Bedeutung und Realisierung von Medienkompetenz. In B. SCHORB, N. BRÜGGEN, & A. DOMMASCHK (Eds.), *Mit eLearning zu Medienkompetenz. Modelle für die Curriculumgestaltung, Didaktik und Kooperation* (pp. 15-34). München: kopaed.
- SCHULMEISTER, R. (2009). Gibt es eine „Net Generation“? Erweiterte Version 3.0. Retrieved November 10, 2012 from <http://tinyurl.com/yepm251>
- SESINK, W. (2008). Bildungstheorie und Medienpädagogik—Versuch eines Brückenschlags. In J. FROMME, & W. SESINK (Eds.), *Pädagogische Medientheorie* (pp. 13-35). Wiesbaden: VS.
- SEUFERT, S., & EULER, D. (Eds.). (2003). *Nachhaltigkeit von eLearning-Innovationen*. St. Gallen: Swiss Centre for Innovations in Learning.
- WEDEKIND, J. (2008). Medienkompetenz für (Hochschul-) Lehrende. *Zeitschrift für E-Learning*, 3(2), 24-37.



# **Belief(s) in eLearning – Zusammenhänge zwischen eLearning und Lehr-/Lern-Überzeugungen in Lehrportfolios**

**Timo van Treeck**

Abteilung für Bildungsforschung und Bildungsmanagement, Heinrich-Heine-Universität Düsseldorf, Kompetenzteam Hochschuldidaktik, Fachhochschule Köln

*Die Gestaltung von Lernumgebungen ist stark abhängig von den Lehr-/Lern-Überzeugungen der Lehrenden. Dies gilt insbesondere dann, wenn Lehrende neue Lehrformen einsetzen wollen oder sollen. Lehr-/Lernüberzeugungen (beliefs) werden deshalb anhand von Leistungsdarstellungen zur Lehre in Lehrportfolios untersucht und in Beziehung zum Einsatz von eLearning-Formen gebracht.*

## **Einleitung**

Die Einflussgrößen auf die Gestaltung guter Lehre sind vielfältig. Neben Rahmenbedingungen, Diversität der Zielgruppe, Methodeneinsatz oder Assessmentformat wird vor allem die Lehrkompetenz als Schlüsselfaktor gesehen. Dessen entscheidende Rolle ist darauf zurückzuführen, dass Lehrende je nach Lehrkompetenz in der asymmetrischen Lehr-/Lernbeziehung die Lernumgebung maßgeblich gestalten können. Kompetenz wird hier verstanden als Disposition, in unsicheren, herausfordernden Situationen erfolgreich zu handeln. Gerade in dem wegen seiner Komplexität und interpersonaler Variablen durch hohe Misserfolgswahrscheinlichkeit geprägten Bereich des Lehrens und Lernens, sind deshalb Kompetenz, die Reflexion von deren Entwicklung sowie die ständige Reflexion der Lehr-/Lernsituationen von

Tokar, A., Beurskens, M., Keuneke, S., Mahrt, M., Peters, I., Puschmann, C., van Treeck, T., & Weller, K. (Hrsg.). (2012). *Science and the Internet* (S. 99-111). Düsseldorf: Düsseldorf University Press

entscheidender Bedeutung. Wie kann damit angemessen umgegangen werden?

Ein im angloamerikanischen Raum schon seit Anfang der 1990er Jahre in Bewerbungen eingesetztes und seit Mitte der 1990er Jahre in der hochschuldidaktischen Weiterbildung hierzulande genutztes Instrument stellt das Lehrportfolio dar (Auferkorte-Michaelis & Szczyrba, 2004). Ein Portfolio ist eine Sammlung verschiedener Artefakte (Handouts, Medien, Bescheinigungen, Arbeitsdokumente, usw.), entweder um die weitere Planung, die Reflexion oder die Präsentation der Leistungen, z.B. für Bewerbungen, zu unterstützen. Das Lehrportfolio fokussiert aus Lehrendensicht das Thema Lehre und enthält in der Regel verschiedene strukturierende Elemente, wie Lehrphilosophie, Lehrmethoden, Zielgruppe, usw. (Auferkorte-Michaelis & Szczyrba, 2004; von Queis, 1994; Szczyrba & Gotzen, 2012). Welche Bedeutungen haben Lehrportfolios konkret für die Entwicklung und Darstellung von Lehrkompetenz? (Zu der Rolle von Lehrkompetenz und Lehrportfolios in Berufungsverfahren vgl. Szczyrba, 2010; Trautwein, 2012.)

So genannte educational beliefs von Lehrenden (Trautwein, 2012) lassen sich entweder direkt den Lehrkompetenzen zuordnen, oder sie sind zu verstehen als Einstellungen, welche die Kompetenzentwicklung unterstützen, behindern oder in bestimmte Richtungen lenken. Eine besondere Bedeutung kommt den beliefs vor allem dann zu, wenn Veränderungen in der Lehre angestrebt werden, da für deren Erfolg elementar ist, wie Interventionen von den Lehrenden angenommen und gegebenenfalls transformiert werden. So behauptet Kember (1997), dass Lehrende je nach ihren Lehr-/Lern-Überzeugungen (bzw. conceptions) Tutorien so umfunktionieren, dass sie entgegen ihrem eigentlichen Ziel als Mini-Vorlesungen durchgeführt werden. Um Lehre zu ändern, ist also Wissen darüber notwendig, wie Lehrende sich selbst und ihre Lehre sehen:

An important part of what may be needed to change actual teaching and perceptions of teaching is knowledge of what the teachers themselves see as their own approaches to teaching, and how these approaches are experienced by students. (Trigwell & Prosser, 2004)

Da educational beliefs jedoch häufig nicht bewusst vorliegen und früh im Lebensverlauf gebildet werden, wird von einigen Forschern die Sichtbarmachung der Beliefs über „storytelling“ empfohlen (vgl. zum digital storytelling auch Bäcker et al., 2011), eine Methode, die auch in Workshops zur Portfolio-Erstellung häufig eingesetzt wird. Die Erfassung von beliefs erscheint über Lehrportfolios möglich, ebenso sollen Veränderungen im Lebensverlauf sichtbar werden (Trautwein, 2012; van Treeck, 2012).

Untersucht wurden deshalb Lehrportfolios, um die mögliche Struktur, Ausprägungen und Inhalte von educational beliefs in Lehrportfolios darzustellen und vor allem deren Bezüge zum Einsatz von eLearning auszuarbeiten, das hier als ein Beispiel für einen Change-Prozess innerhalb der Lehre herausgegriffen wird.

## eLearning und beliefs – Thesen

Mit dem Einsatz von eLearning werden ganz unterschiedliche Erwartungshaltungen, aber auch Vorurteile verknüpft. Immer noch stark scheinen die Vorstellungen zu sein, schon allein der Einsatz von Medien könne Probleme lösen oder Zeit sparen (Kritik dazu bei Reinmann, 2012). Verbunden damit mag auch sein, dass immer wieder propagiert wird, es sei vor allem wichtig, Inhalte zur Verfügung zu stellen und diese dann, nach einmaliger Herstellung, leicht zwischen verschiedenen Lehrenden oder Hochschulen auszutauschen und (gemeinsam) wiederverwenden zu können. Dass dieser Prozess der gemeinsamen Nutzung von Content immer noch eine Herausforderung darstellt, illustriert die eLearning-Redeweise, dass „ProfessorInnen eher ihre Zahnbürste teilen als ihren Kursinhalt“ (Quelle unbekannt). Betont sei hier zusätzlich, dass die alleinige Contentbereitstellung kein gutes eLearning ausmacht, da Materialbereitstellung weder Lehre noch Lernen darstellt – dazu bedarf es einer Einbindung in weitergehende didaktische Konzepte.

Betrachtet man diese und andere Erwartungshaltungen an eLearning-Realisationen, lassen sich diese sehr gut zu verschiedenen educational beliefs und verwandten Konzepten in Beziehung setzen, zu verschiedenen Konzepten darüber also, welche Grundlagen Lehren und Lernen haben und wie sie ablaufen sollten:

Tabelle 1. Beliefs in Bezug zu eLearning-Umsetzungen

<b>beliefs und verwandte Konzepte</b>	<b>eLearning-Umsetzung</b>
Inhaltsorientierte / dozentenorientierte Lehransätze liegen vor bzw. serviceorientierte Ausrichtungen der Lehre	Materialien aus Lehrveranstaltungen sollen bereitgestellt werden
Strukturierte Informationsvermittlung (als eine conception der Lehre)	Content-Erstellung durch Lehrende
Soziale Beziehungen und Eigenaktivität wer-	Online-Kollaboration soll

den als wichtig für Lernprozesse angesehen	durchgeführt werden
Soziale Beziehungen werden als wichtig für Lernprozesse angesehen	Ablehnung von eLearning (ohne sozialen Austausch)
Kontrollverlust befürchtet	Ablehnung von eLearning / Social Software
Hohe Toleranz gegenüber Fehlern und Kontrollverlust vorhanden	Internet als Experimentierraum

Es ist zum Beispiel möglich, dass sich viele Lehrende mit ihren Erwartungshaltungen an traditionellen Konzepten zu computer-based-learning orientieren. Diese jedoch waren lange Zeit vor allem instruktionsorientiert (Jahnke & Koch, 2009). Diese Erwartungshaltungen würden beispielweise beliefs widersprechen, dass Lernen ein sozialer Prozess ist. Lehrende mit diesem educational belief dürften einen eLearning-Einsatz mit der Begründung ablehnen, dass es für Lernen notwendige soziale Prozesse nicht unterstützt und gegebenenfalls sogar eine Vereinsamung der Lehrenden fördere.

## Forschungsdesign

Zur Erhebung impliziter und expliziter Ausprägungen der educational beliefs in den Lehrportfolios wurde das Konzept der Lehr-/Lern-Überzeugungen bzw. beliefs gegenüber den ähnlichen Konzepten „approaches to teaching“ und „conceptions of teaching“ abgegrenzt. Approaches werden in der Regel durch Fragebögen erhoben und sind stärker handlungsorientiert ausgerichtet, wohingegen die conceptions sich vor allem durch eine Strukturierung in fünf verschiedene Ausprägungen auszeichnen und eher ohne Bezug auf konkrete Lehrsituationen formuliert werden. Teilweise werden die Begriffe „beliefs“ und „conceptions“ in der Forschungsliteratur synonym verwendet. Es sei jedoch betont, dass sich beliefs sehr frühzeitig über den Lebensverlauf aus bilden und zudem schwer veränderlich sind. (Weitere Details hierzu finden sich bei van Treeck, 2012.)

Grundlage für die Analyse der educational beliefs und der abgegrenzten Konzepte waren 31 Lehrportfolios, die im Rahmen von Workshops und Coachings, die von 2004 bis 2011 durchgeführt wurden, entstanden sind. Zielgruppen der Workshops waren der wissenschaftliche Mittelbau sowie jüngere und erfahrene ProfessorInnen verschiedener Hochschulen. Die Lehrportfolios wurden zum Feedback durch die Workshop-Leitung und für weitere Studien freigegeben. Das Lehrportfoliomaterial wurde anonymisiert.

Die Portfolios wurden inhaltsanalytisch nach Mayring (2008) interpretiert (van Treeck, 2012). Dabei erfolgte eine Paraphrasierung und Reduktion der Daten.

## **Approaches und eLearning zur Abgrenzung von beliefs**

Wenn man die verschiedenen verwandten Konzepte, mit denen Einstellungen zur Lehre untersucht werden, in Bezug auf die Lehrportfolios betrachtet, kann festgestellt werden, dass vor allem die so genannten approaches mit ihren inhärenten Widersprüchen so abgebildet werden, wie zu vermuten war. Bei den approaches handelt es sich um Lehransätze oder Lehrorientierungen, die neben einer Einstellung zur Lehre auch direkt eine damit verbundene Handlungsorientierung beinhalten. Für den deutschsprachigen Raum hat beispielsweise Lübeck (2009) diesen Ansatz mittels einer angepassten Version des ATI (approaches to teaching inventory) erhoben. Die approaches werden dabei vor allem zwischen den beiden Polen lehrendenzentriert / inhaltsorientiert und studierendenzentriert / lernorientiert gesehen. Untersucht hat Lübeck einerseits die Ausprägungen dieser beiden Pole bei verschiedenen Lehrenden und andererseits die fachkulturellen Unterschiede bei den Lehransätzen (Lübeck, 2010). Die Autorin ist zu dem Ergebnis gekommen, dass eine Person auf beiden Polen stark ausgeprägte Ansätze vertreten kann und diese dabei je nach Veranstaltungsformat (Vorlesung, Übung, Seminar, Labor) variieren können. Dies kann als Bestätigung für die Konstruktion des ATI verstanden werden, soll dieser doch auf eine konkrete Lehrveranstaltung bezogene Einstellungen messen, also den Lehrkontext berücksichtigen. Trotzdem hat Lübeck auf dieser Basis auch fachkulturelle Unterschiede in den Ausprägungen lehrendenzentrierter / inhaltsorientierter gegen studierendenzentrierter / lernorientierter Ansätze feststellen können.

Versucht man, Aussagen zu den approaches in den Lehrportfolios ausfindig zu machen, um diese gegenüber den educational beliefs abzugrenzen, lässt sich feststellen, dass die approaches dort so formuliert werden, dass deren beiden Pole zusammen auftreten. Dies geschieht in der Regel an Stellen, an denen die approaches nicht explizit mit einer konkreten Lehrveranstaltung in Bezug gesetzt, sondern in den Kapiteln Lehrphilosophie, Lehransatz oder Lehrkonzeption der Lehrportfolios ausgeführt werden. Hier finden sich dann Aussagen, die sich den Fragen aus dem ATI zuordnen lassen und nach der Rolle des / der Lehrenden fragen: 1) ÜbermittlerIn von Wissen (Wissen wird weitergegeben / übertragen), 2) Stoff strukturierend (Wissen muss für die Lehre strukturiert werden), 3) ErmöglicherIn (Lernumgebung gestalten, um Entwicklungen zu ermöglichen).

Dabei zeigte die Analyse der Portfolios in diesem Bereich, dass das Verständnis der Lehrendenrolle als ÜbermittlerIn von Wissen häufig zusammenfällt mit dem Selbstverständnis als ErmöglicherIn. So wird beispielsweise ein Zusammenhang gesehen zwischen der Rolle als ÜbermittlerIn von Lerninhalten und der als ErmöglicherIn durch das Einfordern von eigenständigem Lernen:

In meinen Lehrveranstaltungen möchte ich gezielt und didaktisch ansprechend Lehrinhalte vermitteln und gestalten, gleichzeitig jedoch durch das Einfordern von eigenständigem Lernen ein anregendes und effektives Arbeitsklima erreichen, in dem die Studierenden Selbstverantwortung für den eigenen Lernerfolg und einen kritischen und frischen Geist kultivieren. (LP 27, 121)

Teilweise scheinen diese Rollen als Gegensätze verstanden zu werden, wie es vordergründig gesehen das Paradigma des „shift from teaching to learning“ (Barr & Tagg, 1995) postuliert. In den Lehrportfolios dürfte aber vor allem eine starke Betonung des studierendenzentrierten Ansatzes, als Gegengewicht zu dem eher traditionell weiter verbreiteten lehrendenzentrierten / inhaltsorientierten Ansatz wichtig sein. Es handelt sich also möglicherweise um eine Überbetonung einer Ausrichtung, um eine günstigere Balance herzustellen:

Als Lehrende an der Hochschule bin ich die Person, die die Studierenden zum eigentätigen Denken anregt. Die Vermittlung von Wissen stellt dabei nur einen kleinen Teil dar. Vielmehr sollen die Studierenden dazu ermutigt werden, selbst Fragen zu stellen, selbst Dinge wissen zu wollen und diesen dann auf den Grund zu gehen. (LP 3, 18)

Betrachtet man die approaches in Bezug auf die in den Lehrportfolios thematisierten eLearning-Umsetzungen, dann lassen sich die Befunde folgendermaßen zusammenfassen:

- Lehrende, welche in den approaches eine ausschließliche Wissensübermittlungs-Rolle thematisieren, machen keinerlei Angaben zu von Ihnen durchgeführten eLearning-Maßnahmen.
- Lehrende, welche in den approaches eher eine ErmöglicherInnen-Rolle thematisieren und Gruppenarbeit für wichtig halten, setzen vereinzelt im eLearning auf reine Materialbereitstellung.
- Lehrende mit Selbstverständnis als ErmöglicherInnen oder mit einer Doppelrolle ErmöglicherInnen / ÜbermittlerInnen berichten davon, sowohl Online-Kommunikation als auch die umfangreichere Online-Aktivität der Kollaboration umzusetzen.

Diese Ergebnisse sind insoweit überraschend, als vermutet werden konnte, dass die reine Wissensübermittlung häufig durch die Bereitstellung von Online-Materialien begleitet würde. Möglicherweise war den Lehrenden dieses für die Darstellung in ihren Portfolios nicht bedeutsam genug. Sollte dieser Zusammenhang aber darüber hinausgehen, so kann man vermuten, dass entweder die Möglichkeiten zur Online-Bereitstellung nicht ausreichend bekannt sind, oder dass die Wissensübermittlung so stark mit der Person des Lehrenden verbunden wird, dass eine mediengestützte Vermittlung bewusst nicht in Betracht gezogen wird.

Dass andererseits vereinzelt Lehrende trotz ihrer Tendenz zur Rolle des / der ErmöglicherIn online nur Materialien zur Verfügung stellen, mag ebenfalls an Unkenntnis der Möglichkeiten liegen oder an einer Ablehnung der Unterstützung solch persönlicher Prozesse durch eLearning-Maßnahmen. Eine Abwehr gegenüber vermeintlich unpersönlichem Kontakt in internetgestützten Formaten ist als Grund auch nicht unwahrscheinlich.

Besonders für den Einsatz von Web 2.0 und Social Software haben beispielsweise Jahnke und Koch (2009) sowie Mayrberger (2008) mögliche theoretische Bezüge zwischen studierenden- gegen dozierendenzentrierten Ansätzen zur Lehre bereits hergestellt. (Zu den Besonderheiten bei der Implementierung von Social Software vgl. auch van Treeck, 2010.)

## Conceptions und eLearning zur Abgrenzung von beliefs

Die Lehransätze wurden hier vor allem über die Zuordnung von Aussagen zu den fünf verschiedenen conceptions erhoben, wie sie Kember (1997) strukturiert:

Tabelle 2. Conceptions of teaching nach Kember (1997)

Informations- vermittlung	Strukturiere Informations- vermittlung	Interaktiver Prozess	Verstehen Ermöglichen	Sichtweisen ändern Entwicklung fördern
------------------------------	--	-------------------------	--------------------------	---

Für die Informationsvermittlung lässt sich festhalten, dass die Lehrenden diese Konzeption der Lehre in der Regel mit einem weiter gefassten Konzept verbinden, z.B. mit der Rolle des Kooperationsexperten oder mit dem Ziel, die Studierenden auf unerwartete Ereignisse vorzubereiten:

Mein für meine Lehrphilosophie bestimmendes Rollenverständnis als Lehrender kombiniert das Bild des Wissens-Vermittlers und das des Experten im Kooperati-

onsprozess, der in einen Denkstil einführt und eigenes Arbeiten ermöglicht. (LP 24, 5)

Die Ausführungen zur strukturierten Informationsvermittlung erfolgen dagegen meist so, dass sie den approaches zugeordnet werden können. Als allgemeiner interaktiver Prozess werden dagegen Gruppenphasen zur Erstellung von Produkten beschrieben oder es werden strukturelle Notwendigkeiten für eine gelingende Interaktion geschildert, wie beispielsweise die folgende:

Klarheit und Transparenz der Lernziele, deren Kontrolle und Ergebnissicherung in gemeinsamen Plenum-Diskussionen sind dabei von entscheidender Bedeutung (LP 27, 30)

Das Ziel, Verstehensprozesse anzuregen, formulieren Lehrende vor allem in Bezug auf die wissenschaftliche Vorgehens- und Denkweise sowie Methodenkompetenz:

Des Weiteren sollen Studierende aber auch ein Verständnis entwickeln, z.B. lernen, wie man bestimmte Marktforschungsmethoden anwendet (LP 13, 497-645)

Die fünfte conception, die Förderung von Entwicklung und Änderung von Sichtweisen, visieren die Lehrenden in einem ähnlichen Bereich an: In der Ausbildung einer wissenschaftlichen Grundhaltung, wobei auch explizit deren affektive Seite betont wird:

Auch universitäre Bildung bleibt an der Oberfläche und verfehlt den entscheidenden Schritt in Richtung Verhaltensänderung, wenn sie auf den rein kognitiven Bereich beschränkt bleibt und nicht auch die übrigen Bereiche menschlichen Erkennens (Emotion, Handeln, Ästhetik) berührt. (LP 22, 21)

Die Entwicklung eines Problembewusstseins bei den Studierenden wird ebenso angestrebt, wie das Verlassen einer konstatierten passiven Grundhaltung auf Seiten der Studierenden.

Besonders auffallend in Zusammenhang mit den Aussagen zum eLearning-Einsatz ist, dass diese nur im Umfeld der fünften conception auftreten, welche mit dem Ideal der Änderung von Einstellung und der Begleitung der Entwicklung der Studierenden zusammenhängt. Möglicherweise findet sich hierin eine anspruchsvolle didaktische Ausrichtung der Lehrenden wieder, die auch in den Workshops, in deren Umfeld die Portfolios entstanden sind, häufig thematisiert worden sein dürfte. Die Differenz zwischen conceptions und approaches mag hier für den eLearning-Bereich auf eine Differenz zwischen Umsetzungswunsch und -möglichkeiten hinweisen, also gegebenenfalls auf einen Unterschied zwischen espoused theories und theories in use (Trautwein, 2012).

## Beliefs und eLearning

Die über die Portfolios erfassten educational beliefs oder Lehr-/Lern-Überzeugungen bilden einen ersten Ansatz für eine Kategorisierung und Typisierung von educational beliefs im deutschsprachigen Raum. Hier lassen sich vor allem allgemeine Überzeugungen zu Struktur und Zielen von Wissenschaft (Forschung und Lehre) und zu den Bedingungen guten Lehrens und Lernens unterscheiden.

In den ausgewerteten Datensätzen waren als allgemeine Ziele vor allem die Befähigung der Studierenden zu einer Lebens- und Weltbewältigung genannt worden, ebenso die Ausbildung einer positiven Haltung zur selbständigen Erkenntnissuche:

Zum universitären Lernen gehört für mich das selbständige und dialogische Erschließen von Gegenständen, das Entwickeln von eigenen Fragen, die Fähigkeit, diesen Fragen methodisch fundiert und sachlich angemessen nachzugehen, sowie die Fähigkeit, eigene Untersuchungsergebnisse so zu präsentieren, dass sie für andere verständlich und wissenschaftlich nachvollziehbar sind. (LP 14, 7)

Forschung und Lehre werden in den Portfolios in der Regel als eine Einheit beschrieben. Besonders häufig ist die eigene Forschung ein Ansatzpunkt für die eigene Lehrtätigkeit, darüber hinaus wird diese Verbindung aber auch als vorteilhaft für die Überprüfung von wissenschaftlichen Erkenntnissen auf ihre Validität sowie für die Ausbildung von Kernkompetenzen angesehen. Vereinzelt lassen die Portfolio-Darstellungen darauf schließen, dass die Verbindung von Forschung und Lehre als ein dem Bologna-Prozess entgegengesetztes Ideal gesehen wird, das aber nichtsdestotrotz vertreten wird.

Besonders stark wird die Bedeutung von sozialen Beziehungen für das Lehren und Lernen betont. So wird der Aufbau einer lernförderlichen Atmosphäre von den Lehrenden unterstützt, indem Regeln gemeinsam erarbeitet und befolgt werden oder eine vertrauensvolle, respektvolle Grundhaltung eingenommen wird:

Dabei ist es mir auch ein Anliegen, dass ausnahmslos alle Studierenden eine Atmosphäre vorfinden, in der sie sich äußern können. (LP 15, 22)

Andererseits werden auch spezifische Vorteile ausgeführt, die durch eine Beachtung der sozialen Aspekte des Lernens unterstützt werden, wie beispielsweise Lernen durch Feedback der KommilitonInnen, Herstellen eines persönlichen Bezugs zum Thema oder Ausbildung sozialer Kompetenzen.

Lehre wird in den meisten Portfolios als praxisorientiert angesehen, d.h. die spätere Berufstätigkeit der Studierenden wird von den Lehrenden als Ziel der Lehre verstanden und in den meisten Fällen werden auch konkrete Akti-

vitäten innerhalb der Lehre daraufhin ausgerichtet. Letzteres erfolgt auf vielfältige Weise: durch Praxisexkursionen, Einladung von KollegInnen, Thematisierung von Stellenausschreibungen, realitätsnahe Praxissimulationen oder Praxisprojekte. Der lernförderliche Aspekt der Verbindung von Theorie und Praxis wird vereinzelt ausdrücklich thematisiert.

Auffallend häufig wird die Bedeutung von Feedback für einen erfolgreichen Lehr-Lern-Prozess angeführt:

Als sehr gut wurde auch beurteilt, dass Feedback- und Diskussionsrunden bezüglich der anderen [Produkte] geübt wurden und wir im Seminar vorher Feedbackkriterien erarbeitet hatten. So entstand ein als sehr angenehm und entspannt empfundenes Lernklima. (LP 12,21)

Es wird Wert auf die Strukturierung eines methodischen Feedbacks gelegt, die positive Aufnahme des Feedbacks seitens der Studierenden wird thematisiert und Feedbackmaßnahmen sowie Reflexionsanlässe bewusst in verschiedene Phasen der Lehre eingebunden, z.B. für die Planung und Durchführung von Abschlussarbeiten oder im Rahmen einer Vorgehensreflexion oder Zwischenbilanz:

Ich lasse Studierende z.B. Zwischenbilanzen zum Stand ihres Lernprozesses anfertigen, damit sie auf einer Metaebene ihre Lernfortschritte reflektieren. Das steigert die Motivation und das zielorientierte Lernen. (LP 12, 9)

Eigenaktivität von Studierenden wird aus verschiedenen Gründen als lernförderlich eingeschätzt:

- Studierende sollen aktiv werden, um Lehrveranstaltung auch für andere lernförderlich zu gestalten (Arbeitsklima, Lernen durch Lehren).
- Eigenaktivität ist lernförderlich und wichtig für Kompetenzerwerb bzw. Lernen ist immer ein aktiver Prozess.

Möglichkeiten für Eigenaktivitäten der Studierenden bieten die Lehrenden dabei unter anderem über die Mitgestaltung der Lehrsettings durch die Studierenden. Elementarer Bestandteil und explizites Ziel von Lehrportfolios ist die Reflexion der eigenen Lehre. Die Einschätzung der Bedeutung dieser Reflexion für die eigene Lehre dürfte ebenfalls ein Aspekt der educational beliefs sein. Als Anlässe für diese Selbstreflexion werden von den Lehrenden dabei sowohl Evaluationsergebnisse, als auch die Teilnahme an hochschuldidaktischen Weiterbildungen angeführt:

Evaluationen meiner Lehre sind für mich als junger Dozent ein zentrales Element meiner professionellen und persönlichen Weiterentwicklung und dienen meiner

Auffassung nach der Optimierung der von mir zu verantwortenden Lehr- und Lernprozesse. (LP 11,16)

Mit der Teilnahme am Hochschuldidaktischen Qualifizierungsprogramm kann ich in Fortbildungen zu unterschiedlichen Themenbereichen (Planung von Lehrveranstaltungen, schwierige Situationen in der Lehre) mein Selbstbild als Lehrende kritisch reflektieren und Impulse von außen in meine Lehrveranstaltungen integrieren. (LP 20, 24)

Betrachtet man die verschiedenen eLearning-Umsetzungen, sind diese mit den angeführten beliefs gut in Einklang zu bringen. So wird eine starke Ausprägung der beliefs, in Bezug auf die Bedeutung der sozialen Beziehungen sowie der Eigenaktivität, besonders in denjenigen Lehrportfolios genannt, in welchen auch kollaborative und kommunikative eLearning-Umsetzungen dargestellt werden. Auch die stark auf Praxis und Reflexion ausgerichteten beliefs-Aussagen geschehen im Umfeld von kollaborativen eLearning-Umsetzungen insofern sinnvoll und passend, als die Praxisphasen hier möglicherweise gemeinsam vorbereitet und reflektiert werden können. Die Verbindung von Blended-Learning und Praxisphasen erscheint vor allem sinnvoll, weil die Mischung von Online- und Offlinephasen besonders gut Praxisphasen begleiten kann, indem vor- und nachbereitende Offline-Phasen als Präsenzveranstaltung durchgeführt werden und die Begleitung der Tätigkeit im Praxisfeld durch Online-Aktivitäten gestützt wird (vgl. für Lehrportfolios in Praxisphasen auch van Treeck & Hannemann, 2012).

## **Fazit**

Lehrportfolios als Reflexionsobjekte und Leistungsdarstellungen über die Lehre eignen sich gut, um einen Einblick in die Lehr-/Lern-Überzeugungen bzw. educational beliefs von Lehrenden zu erhalten. Zur weiteren Ausgestaltung und Ausdifferenzierung der im englischsprachigen Bereich stark diskutierten Ansätze für den deutschsprachigen Raum, sind jedoch noch weitere Forschungsaktivitäten angeraten. Für den Einsatz von eLearning zeigte sich hier, dass sich die diskutierten beliefs mit den eLearning-Umsetzungen (differenziert nach Online-Materialbereitstellung und Online-Kommunikation sowie Kollaboration) gut in Deckung bringen ließen. Auf soziale Beziehungen und Eigenaktivität ausgerichtete beliefs traten nicht im Zusammenhang mit reiner eLearning-Materialbereitstellung auf. Die approaches to teaching und der eLearning-Ansatz stellten sich dagegen widersprüchlich dar, weil hier z.B. eine eher auf reine Wissensübertragung ausgerichtete Materialbereitstellung mit approaches zu Gruppenorientierung und dem Rollenver-

ständnis als ErmöglicherIn kollidierten. Für eine genauere Untersuchung von speziellen eLearning-Formaten sowie einer Ausdifferenzierung und Typisierung der beliefs werden weitere Erhebungen kombiniert werden müssen. Unklar bleibt beispielsweise noch, welche Bezüge zwischen Vorstellungen über notwendige und mögliche Kontrolle eines Lernprozesses und dem Einsatz von offenen Online-Lernformen (unter anderem Einsatz von Social Software) auffindbar sind.

## Danksagung

Diese Arbeit wurde gefördert durch den Strategischen Forschungsfonds der Heinrich-Heine-Universität Düsseldorf.

## Referenzen

- AUFERKORTE-MICHAELIS, N., & SZCZYRBA, B. (2004). Das Lehrportfolio in der Reflexions- und Schreibwerkstatt. In B. BEHRENDT., H.-P. VOSS, & J. WILDT (Hrsg.), *Neues Handbuch Hochschullehre* (E 2.6). Berlin: Raabe.
- BÄCKER, E.-M., CENDON, E., & MÖRTH, A. (2011). Das E-Portfolio für Professionals. Zwischen Lerntagebuch und Kompetenzfeststellung. In P. ARNOLD (Hrsg.), *E-Portfolios* (S. 37-50). Innsbruck: StudienVerlag.
- BARR, R. B., & TAGG, J. (1995). From teaching to learning: A new paradigm for undergraduate education. *Change*, 6(27), 12-25.
- JAHNKE, I., & KOCH, M. (2009). Web 2.0 goes Academia—does Web 2.0 make a difference? *International Journal of Web Based Communities*, 5(4), 484-500.
- KEMBER, D. (1997). A reconceptualisation of the research into university academics' conceptions of teaching. *Learning and Instruction*, 7(3), 255-275.
- LÜBECK, D. (2010). Wird fachspezifisch unterschiedlich gelehrt? Empirische Befunde zu hochschulischen Lehransätzen in verschiedenen Fachdisziplinen. *Zeitschrift für Hochschulentwicklung*, 5(2), 7-24.
- LÜBECK, D. (2009). Lehransätze in der Hochschullehre. (Dissertation zur Erlangung des akademischen Grades Doktor der Philosophie). Berlin: Freie Universität Berlin. Heruntergeladen am September 27, 2012 von [http://www.diss.fu-berlin.de/diss/receive/FUDISS\\_thesis\\_000000011078](http://www.diss.fu-berlin.de/diss/receive/FUDISS_thesis_000000011078)
- MAYRBERGER, K. (2008). Fachkulturen als Herausforderung für E-Learning 2.0. In S. ZAUCHNER, P. BAUMGARTNER, E. BLASCHITZ, & A. WEISSENBÄCK (Hrsg.), *Offener Bildungsraum Hochschule. Freiheiten und Notwendigkeiten* (S. 157-168). Münster: Waxmann.

- MAYRING, P. (2008). *Qualitative Inhaltsanalyse. Grundlagen und Techniken*. Weinheim: Beltz.
- VON QUEIS, D. (1994). Karriere durch Lehre – Das Lehrportfolio zur Dokumentation der Lehrkompetenz. In B. BEHRENDT, H.-P. VOSS, & J. WILDT (Hrsg.), *Handbuch Hochschullehre* (10., aktualisierte Aufl., S. 2-34). Bonn: Raabe-Verlag.
- REINMANN, G. (2012). Der Hendl-Tipp: Finger weg von digitalen Medien in der Hochschullehre? *HDS Journal des Hochschuldidaktischen Zentrums Sachsen*, 1, 63-67. Heruntergeladen am September 27, 2012 von [https://www.hds.uni-leipzig.de/fileadmin/media/HDS\\_Journal\\_1-2012\\_Tagungsedition.pdf](https://www.hds.uni-leipzig.de/fileadmin/media/HDS_Journal_1-2012_Tagungsedition.pdf)
- SZCZYRBA, B. (2010). Die Professur als Profession – Kompetenzorientierung in Berufungsverfahren. In B. BEHRENDT, H.-P. VOSS, & J. WILDT (Hrsg.), *Neues Handbuch Hochschullehre* (J 1.7). Berlin: Raabe.
- SZCZYRBA, B., & GOTZEN, S. (2012). *Das Lehrportfolio – Darstellung, Entwicklung und Nachweis von Lehrkompetenz an Hochschulen*. Berlin: LIT.
- TRAUTWEIN, C., MERKT, M., & HEYER, J. (2012). Leitfragen zur Einschätzung der Qualität von Lehrportfolios. *ZHW-Almanach*, 2. Heruntergeladen am September 27, 2012 von <http://www.zhw.uni-hamburg.de/almanach/?p=341>
- TRAUTWEIN, C. (2012). *Struktur und Entwicklung akademischer Lehrkompetenz: Die Bedeutsamkeit individueller Lehr-Lern-Überzeugungen*. Bielefeld: WBV.
- VAN TREECK, T. (2012). Wie approaches und conceptions zu beliefs führen – Lesehilfen für Lehrportfolios. In B. SZCZYRBA, & S. GOTZEN (Hrsg.), *Das Lehrportfolio – Darstellung, Entwicklung und Nachweis von Lehrkompetenz an Hochschulen* (S. 99-112). Berlin: LIT.
- VAN TREECK, T., & HANNEMANN, K. (2012). Lehre und Praxisphasen sichtbar machen – webbasierte Lehrportfolios. In B. SZCZYRBA, & S. GOTZEN (Hrsg.), *Das Lehrportfolio – Darstellung, Entwicklung und Nachweis von Lehrkompetenz an Hochschulen* (S. 217-236). Berlin: LIT.
- VAN TREECK, T. (2010). Lehre ins Internet? Hindernisse und Erfolgsfaktoren für Social Software an der Hochschule. In I. PETERS, C. PUSCHMANN, V. TRKULJA, & K. WELLER (Hrsg.), *SOSOFT 09. Collaborative work, communication and knowledge management in theory and practice* (S. 50-58). Heruntergeladen am November 10, 2012 von <http://tinyurl.com/b2o9ov8>
- TRIGWELL, K., & PROSSER, M. (2004). Development and use of the approaches to teaching inventory. *Educational Psychology Review*, 16(4), 409-424.



## **PART THREE**

---

# **THE INTERNET AND LEGAL ISSUES**



# Law: Friend or Foe in Scientific Internet Use?

**Michael Beurskens**

Faculty of Law, Heinrich Heine University Düsseldorf

*While legal rules provide necessary protection to scientists on one hand, they impose significant restrictions on the other. Furthermore, science is governed by its own set of ethical rules. Many assumptions about the rules imposed from the outside are actually based on the more-or-less intuitive understanding of such ethics. This is a hit-and-miss-affair: Sometimes those rules meet, while they may deviate just as often.*

## Introduction

Lawyers and judges often seem to act as direct successors to priests, translating the utterings of the oracle of Delphi by explaining equally mysterious legal rules (Amsterdam, 1970). In contrast, academics serve a different religion based on other rites but with a similar goal of improving society. Often, law seems to provide unnecessary double precautions, as rules of ethical research seemingly already offer sufficient protection. Indeed, such interference in internal affairs by outsiders is intuitively frowned upon. Thus, it is unsurprising that scientists blame legal rules and their implementers for imposing unnecessary and annoying limitations on essential academic activities (Green, 1992). Because it is easy to shoot the messenger (Shakespeare, 1598), asking to kill all of the lawyers to achieve utopia is certainly not far-fetched (Shakespeare, 1623). However, further examination shows that those objections go both ways. Scientists complain about both “overregulation” and “underregulation” (Sunstein, 1990): What is there is not enough—and still too much.

This can be easily exemplified by looking at popular beliefs about intellectual property law. One tends to apply a strict perspective against others seeking to reuse the data gathered, content created, or inventions made by oneself (see the heated debate in the U.S. on granting copyright to the lecturer instead of the employing university; Triggs, 2005; Holmes & Levin, 2000; Townsend, 2003). In contrast, reuse of content made by others in teaching is considered a good practice (see aptly Lessig, 2001). To many scholars, it is counterintuitive to lose each and every right to their work when signing a publishing contract. Consequently, written agreements with publishers were and are often ignored when sending papers to colleagues or even putting them on Web sites (before self-archiving clauses became popular).

The internal system of ethical research practices at universities, in contrast, is well established and thoroughly administered by university committees and even includes written commitments and frameworks, which look and feel much like “laws.” Such rules are not limited to research on human or animal subjects; they also cover practices such as the attribution of works and data or the transparency and reproducibility of research (cf. Hudson et al., 2005 on an empirical approach to ethics in Internet Research).

When trying to determine the appropriate external (legal) rules, informal internal (ethical) codes of conduct are often confused with actual requirements imposed by law (for the lawyer’s perspective, cf. Cohen, 1934). In general, ethical means should be allowed without restrictions, whereas unethical practices should not only be frowned upon by colleagues but also sanctioned by law. However, legal rules and ethical beliefs do not necessarily go hand in hand. This may be due to questions still being debated in the scientific community or by the general public but decided by the legislators (e.g., stem cell research) or due to legal rules usually not being limited to scientists but having to cover a multitude of different cases.

Ethical rules are not the only guideline applied by laypersons trying to conform to legal requirements. Because the law itself is usually inaccessible, or at least incomprehensible, to the average researcher, they tend to fill gaps with assumptions based on information gathered from unreliable sources, including colleagues, public media, or the Internet. Again, this provides a broad fertile soil for errors—even though intuition is certainly an appropriate approach even in judicial decision-making (Wright, 2006).

If the applicable law were clear and certain and remained generally unchanged, then such confusion could be eliminated with mere education or information. However, especially with regard to the Internet, law is in a constant state of flux. Copyright cases produce unpredictable results, and fundamental differences across the national legal regimes provide an unstable basis

for the increasingly international research community (cf. Dellapenna, 2000). Indeed, unlike real property and contracts, neither copyright nor privacy law (or “data protection” law) is a required subject in the study of law, nor are they part of any bar exam. Thus, even most law professors and law students can shed little light on these topics. Similarly, it is in the best interest of a university’s administrative staff to eliminate any risk of possible liability by suggesting the most restrictive approach.

This paper will examine a few common misconceptions regarding legal rules, explain the current legal framework by giving examples from the U.S. and Germany, and provide a brief look into current developments. While the field is extremely broad, we will focus on copyright and privacy laws as the most fundamental issues involved in scientific use of the Internet.

## The Copyright Dilemma

Copyright is one of the fundamental pillars of research. Specifically, the requirement of attribution serves a dual purpose. By prohibiting plagiarism, scholars can only acquire a reputation by publishing their own ideas. Furthermore, a good reputation may form a basis for reliance, thus avoiding the need to verify information as long as the source is known to do proper research (this is an example of Akerlof’s [1970] “market for lemons”).

However, the copyright granted to scientific articles is, in most parts, identical to the copyright granted to the latest hit single, blockbuster movie, or even copy of Microsoft Word running on your computer. Whereas composers, software developers, and movie directors usually seek monetary benefits from their works, scholars generally will not be able to make a living from the articles or books they write. In contrast, if only money is at stake, there might be a good argument against perpetual property in creative products. If the public has paid enough for the cultural enrichment provided by *Jurassic Park IV*, then it should go into the public domain and be available to everyone for free. Attribution of original works, in contrast, should be perpetual—because the ability to build upon another’s research is essential for scientific progress. This simple economic difference is not embraced by the legal system, which treats popular and scholarly works largely in an identical manner. Similarly, scientific use, as such, is not generally privileged in comparison to mere private use. Furthermore, economic reasoning, or even the ethical rules of science, does not necessarily lead to a proper assumption of what is allowed or what is not allowed in copyright law.

However, there is also some good news. Unlike many areas of law (including privacy law, which we will examine below), copyright law is subject

to a rather strict framework of international treaties and conventions, which are applied uniformly across most states around the world. Still, differences persist, not only in details but also in the fundamental approach to copyright cases. This disparity favors the international character of research and is thus a useful basis for a framework for science regulation.

## Plagiarism

Plagiarism has become a heavily debated topic in Germany following the discovery that the dissertation thesis of a well-known politician was, in large part, copied verbatim from other sources without attribution (on “investigative crowdsourcing,” cf. PlagDoc & Kotynek, 2012). In the aftermath, universities invested heavily into both software and personnel to ensure early detections of fraudulent dissertations and exams.

Most cases of plagiarism are not illegal. The law and the ethical rules of plagiarism intersect imperfectly (Stearns, 1992). Indeed, our culture loves repetition—from re-telling and re-re-telling of popular stories (take a look at the Top 10 movies) to copying of designs for household goods (read up on the current legal issues between Apple and Samsung). Works may be part of the public domain, e.g., with an expired copyright, or the part copied may be so tiny as not to constitute an infringement. Indeed, it would impose an all but insurmountable challenge on most scholars (and even worse on students) to re-create everything independently on their own—reinventing the wheel once again would not further science. As scary as the word *plagiarism* may sound (coming from the Latin word *plagiarius*, referring to kidnappers of another’s slaves), it is not beyond possibility that two (great) minds reach the same idea and even use the same words to express it. Just coming to identical conclusions is no misconduct at all, while rephrasing someone else’s ideas is. Copyright law also fails when plagiarism only extends to someone else’s research data. There is no legal monopoly on mere facts. Furthermore, reusing my own work without attribution will only be considered a copyright infringement if I violate the license agreement I signed with the original publisher. “Self-plagiarism,” as such, is unknown to the law.

Although the law largely ignores plagiarism, the academic penalty could be no worse—it is “a capital offense, punishable by academic death” (Onge, 1988). This is especially problematic, as the scope of unwanted practice is highly unspecific, and there are no lawyers to resolve questions of interpretation. While there may indeed be clear-cut cases, especially involving students merely copying content from readily available online sources and handing it in as their work, things become extremely difficult in professional scientific

research. In describing a case of “scientific misconduct,” Daroff (2007) aptly states:

[...] many of the problems we encountered stem from author naivety, sloppiness, and the ambiguities involved in plagiarism and self plagiarism. One doesn't have to be a flagrant sociopath to encounter charges of misconduct or breach of ethics. (p. 532, f. 3)

Indeed, it seems that plagiarism is much like pornography—you know it when you see it (Stewart, 1964—a comparison discovered by Stearns, 1992). The issue of plagiarism is not so much the use of another's intellectual property but the subsequent fraud—basing respect by peers and students not on hard work and intelligence but on mere copying. This puts the reputation of academia as a whole in question, which is certainly unacceptable. However, because detection is slow and imprecise and has a high rate of mistakes, plagiarism often comes as a surprise. In contrast, violations of intellectual property as such are not always frowned upon by science ethics. For instance, it is a good and well-accepted practice to send digital copies of a paper to colleagues and friends, even when the publishing contract assigns such rights exclusively to the publisher.

The value of attribution in copyright law has historically been a major controversy between the Continental European *droit d'auteur* and the U.S. “copyright” systems. While the European approach is largely based on the author's natural right to his or her creation and, therefore, emphasizes attribution, U.S. copyright law is more focused on a utilitarian perspective and thus does not provide in its rules on fair use a requirement for attribution. Nevertheless, by adopting the Berne Convention for the Protection of Literary and Artistic Works (in 1989!), even the U.S. is required to ensure that every author has

the right to claim authorship of the work and to object to any distortion, mutilation or other modification of, or other derogatory action in relation to, the said work, which would be prejudicial to his honor or reputation. (Art. 6 till Sec. 1)

Nevertheless, lack of proper citations should constitute copyright infringement, even though U.S. law remains silent on the requirement of proper attribution.

Sometimes, offenders try to provide a defense of “good faith plagiarism.” One reason often expressed is that the omitted sources are not worthy of a citation, as they are not sufficiently “scientific.” Such “non-academic” sources may include anything from blog entries, tweets, comments posted in a forum, text from Wikipedia, PowerPoint presentations, or lecture materials found on the Internet to popular fiction or even texts from a different disci-

pline. Inversely, authors defend their actions by pointing out that their work was not academic, e.g., a mere slide show, some lecture notes passed out to students, or a blog entry. The reasoning again does not refer to a legal defense but shows that the ethical standard is highly imprecise. While such a close definition of quotable sources is certainly erroneous, the other variant of good faith is far more compelling: Every scholar writing a paper (such as this one) has certainly read many prior publications, most of them not in parallel to writing but beforehand. Unless one has eidetic memory, it is highly unlikely that one will be able to attribute each and every idea to a specific paper, especially when many ideas are shared among most authors (and might thus be considered “public knowledge”). Still, that defense will not be accepted because it would otherwise create a loophole that proves difficult, if not impossible, to close.

The consequences of plagiarism are extremely harsh. From a legal viewpoint, it is surprising that even a tiny bit of plagiarism will spoil a complete paper, article, or book. While the unattributed contents may only constitute a minor part of the actual product (and are often only tangential to the core theory presented), evidence of a violation of the rules of good scientific practices allegedly eliminates any benefits the work might have had for the scientific community. The law might grant damages or (in rare cases) an injunction preventing distribution of the book in its current form—but it would still respect the author’s efforts. Copying only minor parts would almost never provide cause for such an injunction. The oft-feared criminal sanctions are limited to even more extreme cases (but cf. Green, 2002, who attempts to analyze the social rules of plagiarism by comparing them to the requirements of the legal rules on theft).

Occasionally, someone tries to summarize the rules of plagiarism in simple, clear “Dos” and “Don’ts” (e.g., Gerhardt, 2006). While requiring attribution of “borrowed content” and asking for “quotation marks” seem simple enough, cases of plagiarism are not determined based on the actual behavior of the suspected author (which cannot be reproduced) but merely on the written results. The aforementioned well-publicized scandal in Germany caused many scientists to apply an extremely stringent standard of care in writing articles and books. This fear is further enforced by academic organizations, universities, and non-governmental organizations, which have designed guidelines for scholarly practices that tend to take an extremely strict stance on plagiarism. Indeed, overly careful authors on legal faculties (especially graduate students) seem to be keener to find references to add to their footnotes than to develop innovative theories or conduct their own research. Much time is wasted ensuring that everything that might have been said by

someone else before is attributed to that person—even if the author did not copy that source but merely made a “parallel invention.” It may well be possible that some authors simply choose not to publish to avoid even the appearance of unprofessionalism.

Such over-care is as worrisome as ignorance of the issue. The search for plagiarism must not turn into a witch hunt. The core issue seems to be the lack of a proper code on what constitutes plagiarism and what does not, as well as the lack of a centralized instance to interpret such rules. Furthermore, the analysis sadly always happens *ex post* and is therefore inevitably subject to hindsight bias: It seems to be rather easy to find someone who had comparable or even identical theories on a certain subject, but proving that one copied from the other is only possible with certainty if the text or large parts have been duplicated verbatim. Because only very naïve students (much less full-fledged scholars) would go that far, there is a significant margin for error in determining plagiarism. Thus, an open debate is inevitable, which should mainly be focused on the specific disciplines and not on the science community as a whole. A mere technological approach is useless except for detecting the most ignorant of copyists. Generally, preventive measures (such as an ongoing dialogue during the writing of a thesis or seminar paper) are preferable to *ex-post* sanctions.

## Reuse of Content in Teaching

In the good old days, everyone bought textbooks and read them. When Xerox invented the photocopier (in 1959), teachers (both in schools and universities) suddenly had the opportunity to create a specific selection of readings specifically catering to their lectures. Such “readers” quickly became a popular practice, and suddenly students received huge stacks of printed paper to read at home. When the Internet became widely accepted, teachers once again made use of the medium and “made available” recommended materials for download and printing (Lan & Dagley, 1999). And they lived happily ever after.

Sadly, things are never that simple in real life. Supporting self-learning by providing students with a broad selection of materials might well be a good or even commendable practice from an educational point of view. Indeed, even the Berne Convention contains a provision on the specific exceptions related to teaching:

It shall be a matter for legislation in the countries of the Union, and for special agreements existing or to be concluded between them, to permit the utilization, to the extent justified by the purpose, of [...] works by way of illustration [...] for teaching, provided such utilization is compatible with fair practice. (Article 10)

That exception has its roots in the original agreement of 1886. Still, the U.S. Copyright Act does not include a specific provision allowing for the reuse of protected works in teaching, although it covers such use under the general terms of fair use (ALA). EU-Directive 2001/29/EC expressly allows for exceptions

for the sole purpose of illustration for teaching or scientific research, as long as the source, including the author's name, is indicated, unless this turns out to be impossible and to the extent justified by the non-commercial purpose to be achieved; [and for] making available, for the purpose of research or private study, to individual members of the public by dedicated terminals [...] of works [...] contained in their collections. (Article 5)

The domestic laws implementing these exemptions vary widely (Xalabarder, 2009; Xalabarder, 2004; Ernst & Haeusermann, 2004). Furthermore, their wording is surprisingly unspecific, referring to “small parts” and “necessity.” Thus, teachers, both in schools and universities, are largely left out in the rain. Because actual cases in court are rare, and procedures usually take many years, uncertainty prevails. Nevertheless, the number of documents made available to students has been on a steady increase. For example, the University Library in Düsseldorf currently scans approximately 5,000 excerpts (mainly from books) on behalf of lecturers for roughly 600 lectures. These numbers do not take into account documents scanned directly by lecturers and their staff, as well as materials taken from online sources and converted into PDF—the University's central e-learning-platform receives more than 30,000 new documents per year (which include slides and lecture notes created by the professors themselves). Similarly, there is an extremely high demand by students for relevant materials beyond PowerPoint slides; in a recent survey, more than 80% of the students at the faculty of law giving an answer considered the direct availability of such materials as “very important” for their studies.

Still, the mist of uncertainty remains—so how do lecturers and university or the library administration cope with the risk of infringing on someone's copyright? Discussions with colleagues at Düsseldorf University show that the answers seem to be highly dependent on the actual field of research. Surprisingly, teachers at the faculty of law show the least concern and the strongest belief that their actions are covered by exceptions to copyright law. In contrast, teachers at the medical faculty are very reluctant even to talk about the use of content created by others, as there is a strong belief that any such reuse would be deemed illegal. In general, most university educators assume some kind of liability shield, even though they are unable to specify

the scope. Still, the mere belief is sufficient to cause them to provide quite a lot of reading to their students.

The policy conflict underlying content reuse in teaching is evident in states in which education is predominantly funded by the government, as in Germany (see generally Chon, 2007). Thus, states are required to pay copyright owners (or more precisely the privately held publishers) to ensure appropriate means for education. By granting exceptions to such rights, they save real cash. In contrast, they also fear that too broad exceptions might be detrimental to scientists working in their institutions, as they would be sanctioned by being unable to present their ideas in printed publications.

German law provides an excellent illustration of the issues involved. Under current law, it is illegal to provide printed copies of another's text to students (apart from exams) or even to send them texts by e-mail (whereas an exception allowing such use exists for schools, see 53 Section 3 of UrhG, i.e., the German Copyright Law). However, a well-meaning legislator created a specific exemption to the act for making short texts or short excerpts available to a specific group of students (identified by individual passwords) that actually participate (i.e., physically sit) in a course, as long as the provision of such texts is "necessary" for the purpose of teaching (see 52a Section 1 of the German Copyright Law). The only court decision available assumes that neither downloading nor printing the document is "necessary," as reading it online on screen is deemed sufficient. Furthermore, German law requires "fair compensation" for any use made; this, in turn, requires universities and libraries to take into account any use made as a flat fee agreement being declined by the relevant collection societies. Further limitations are laid down in a 2003 "Charter" of the German Library Association and the Publisher's Association (Börsenverein), which excludes the use of any content made available on the university network by the publisher itself, imposes a requirement to delete any content as soon as the course ends, limits the content to be made available to texts actually used in class (and thus excluding any articles meant to allow a student to expand his understanding), and only allows libraries to scan books that are available locally.

Still, most university professors in Germany make available any texts they consider useful. As mentioned before, legal cases are extremely rare and are most often filed against the university and not the individual teacher. The reason is twofold: Universities provide deeper pockets, and the state or the university will be required to indemnify anyone who did not act with "gross negligence." Such extreme lack of care will almost never be provable. Although the available texts go beyond the allowed threshold, there is no evidence that the introduction of the exception in 2002 has actually caused

any loss of income to the relevant publishers (indeed, their profits increased by 20%, cf. Bibliotheksverband, 2012).

As in the case of plagiarism, the law again fails to fulfill the hopes or even assumptions of its academic addressees. While the issue of plagiarism is simply ignored by law and left to science ethics, the rules allowing lecturers to make texts available to their students are unable to provide for useful learning experiences. Thus, legal rules would actually limit science in one of its core purposes—if scholars knew about those rules and obeyed them. Although some professors are evidently concerned about possible violations of copyright, many still infringe and hope for the best, or at least for secrecy. Indeed, granting students access to important source texts may well be “ethical,” even though it is still “illegal.”

### **Open Access and Creative Commons**

In recent years, the debate on “Open Access” to academic articles has gathered significant impact. This is largely due to some state and private funding requiring the publication of results under Open Access licenses. Nevertheless, traditional journals not only prevail but even thrive. Open Access is remarkable in the context of this paper for two reasons. First, its supporters try to adapt copyright to fit the needs imposed by academia. Second, they do so by creating another framework that sits somewhere between “ethics” and “law.” This becomes clear when one examines the licenses discussed for “Open Data.” As mentioned before, data as such is not protected by law; nevertheless, the proponents of such licenses try to create an enforceable framework for their needs.

Just as many authors never read their publishing agreements, few people actually spend the time reading the “legal code” of a Creative Commons license. Fewer still know that beyond the “Unported” version, there are indeed “ported” versions adapted to the needs of a specific jurisdiction (e.g., there is a “CC3.0-BY-DE” containing a version of the Creative Commons license adapted to the specific needs of Germany). These go beyond mere translation but try to adapt the rules to the requirements of the respective copyright laws—thus leading not to a common German Language CC-Version but requiring German, Austrian, and Swiss versions of each license. These “ported” versions take into account the specific requirements of each domestic law (e.g., regarding limitations of liability, protection of “non-creative” databases under the European Union Database Directive, and so on).

Many authors still believe that publishing their work on the Internet (whether as part of a blog, a “working paper,” or whatever) without any li-

cense is sufficient. Others refer to one of the Creative Commons licenses available. Still, even with a system as simplified as Creative Commons, misinterpretations abound—especially regarding the “NonCommercial”-variants.

Indeed, publication practice seems to be governed mainly by outside influences (such as journals requiring new articles not published before) and unwritten ethical rules (mainly referring to attribution). While Creative Commons Licenses give their users peace of mind and generally perform their task extremely well, they suffer from the same issues as copyright law itself: Because it has to cover all kinds of works, it goes beyond what is needed, while leaving questions specific to academia unanswered.

## The Privacy Debate

The “right to be let alone” (Warren and Brandeis, 1890), or at least ensure that personal data is not made available to everyone else in the world, seems quite intuitive at first glance. However, issues arise when one tries to apply that rule to real-life situations (see, e.g., Bruckman, 2002, who rejects analogies to practices in traditional media). This specifically relates to research on social networks, such as Twitter or Facebook. This is also where intuition begins to fail—because users make their information freely available to just about anyone, a right to privacy seems illusory. Similarly, ethics provide little certainty or even guidance (Burk, 2008).

No modern scholar would believe that any data made available on the Internet (and specifically on social networks) are public and thus not subject to legal protection. Nevertheless, the Dos and Don'ts regarding the use of such data are not only subject to laws but also to an increasing ethical debate. The relevant European Union Directive 95/46/EC refers to anonymization of data in its recital 26:

whereas the principles of protection shall not apply to data rendered anonymous in such a way that the data subject is no longer identifiable; whereas codes of conduct within the meaning of Article 27 may be a useful instrument for providing guidance as to the ways in which data may be rendered anonymous and retained in a form in which identification of the data subject is no longer possible.

While this sounds fine in theory, we know that “deanonymization” is possible based on a very small amount of separate identifiers (cf. Ohm, 2010, also covering the relevant exceptions in the U.S.). The law remains ignorant of this possibility, which may in turn make any publication of data gathered on social media (even if much attention is paid to anonymization) possibly illegal. For the future, Ohm suggested several possible paths—strictly punish-

ing those who cause the actual harm, waiting for some (unlikely) technology to allow for perfect automatic anonymization or banning any re-identification technologies (much like technologies used to circumvent copy protection measures). The privately held social network providers have so far been unable to perform their role as gatekeeper and clearinghouse for academic use (Livingston, 2011). Still, it would be neither ethical to keep all research data secret nor would a potential breach of confidentiality regarding personal data be accepted under an ethical framework. Far more practical, yet also largely based on assumptions is the analysis by Yakowitz (2011). She correctly emphasized the value of research data and the danger of shifting to an “opt-in” model in research. Nevertheless, automatic anonymization may work in some contexts, but it is certainly impossible, e.g., in Twitter research. Thus, any authority or court eventually reviewing data publication would be required to apply a balancing test—with uncertain results.

Trying to resolve the issue in a 12-page paper is certainly impossible. Once again, a clear line rule is needed. Balancing tests are impractical in everyday social science research. Such rules should be developed and proposed by actual scientists—and not imposed upon them by a regulatory authority. There is a general danger of generalizing the issue to cover related interest groups, such as journalists (Scassa, 2010). However, much like the current rule, which is based on pre-existing academic practices, a new rule should be based on the actual needs of scientists and not exclusively on external policies, as observed in copyright law.

## Summary and Outlook

Law and ethics are largely independent regimes. Rules of legal interpretation fail in determining “ethical” conduct, whereas legal rules will never be able to decide fully whether a practice is acceptable in academia. Conflicts between legal and ethical regimes are difficult to resolve. The issue is further emphasized when the legal framework lacks specificity and relevant practical experiences in its implementation (cf. Kaplow, 1992).

Intuition often provides the only, yet uncertain, guideline for determining proper scientific conduct. The situation not only causes significant risk for researchers and teachers alike, but it also leads to wasteful and often unnecessary precautions. A clear and definite framework, which can be interpreted by established methods, seems highly desirable. Developing such frameworks suffers from collective action problems, as well as divergent legal frameworks in different countries and diverging ethical standards in different disciplines. Nevertheless, such a task is not only in the interest of individual

scholars but also in the best interest of legislators and universities. Imposing outside rules on the academic process is unlikely to have a positive impact on actual practice.

## Acknowledgements

This research was supported by the Düsseldorf University Strategic Research Fund.

## References

- AKERLOF, G. A. (1970). The market for ‘lemons’: Quality uncertainty and the market mechanism. *Quarterly Journal of Economics*, 84(3), 488-500.
- AMERICAN LIBRARY ASSOCIATION (n. d.). Fair use and electronic reserves. Retrieved November 10, 2012 from <http://tinyurl.com/7pc4v2l>
- AMSTERDAM, A. G. (1970). The Supreme Court and the rights of suspects in criminal cases. *New York University Law Review*, 45, 785-793.
- Attribution 3.0 Germany (CC BY 3.0). Retrieved September 29, 2012 from <http://creativecommons.org/licenses/by/3.0/de/deed.en>
- BIBLIOTHEKSVERBÄNDE, ARBEITSGEMEINSCHAFT WISSENSCHAFTLICHER VERLEGER UND BÖRSENVEREIN (2003, October 3). Gemeinsame Charta zum Verständnis von §52a UrhG (cited as Charter). Retrieved November 10, 2012 from <http://bit.ly/SDhZ9A>
- BRUCKMAN, A. (2002). Studying the amateur artist: A perspective on disguising data collected in human subjects research on the Internet. *Ethics and Information Technology*, 4, 217-231.
- BURK, D. L. (2008). Information ethics and the law of data representations. *Ethics and Information Technology*, 10, 135-147.
- CHON, M. (2007). Intellectual property “from below”: Copyright and capability for education. *University of California, Davis Law Review*, 40, 803-854.
- COHEN, F. S. (1934). Modern ethics and the law. *Brooklyn Law Review*, 4, 33-50.
- DEUTSCHER BIBLIOTHEKSVERBAND E.V. (2012, June 6). Alles wieder zurück auf Anfang: Das drohende Auslaufen von § 52a Urheberrechtsgesetz (UrhG) wirft Forschung, Lehre und Studium um Jahre zurück. Retrieved September 29, 2012 from [http://www.bibliothekverband.de/fileadmin/user\\_upload/DBV/positionen/2012\\_06\\_Stellungnahme\\_\\_52a\\_final.pdf](http://www.bibliothekverband.de/fileadmin/user_upload/DBV/positionen/2012_06_Stellungnahme__52a_final.pdf)
- DAROFF, R. D. (2007). Scientific misconduct and breach of publication ethics: One editor’s experience. *Medicine and Law*, 26, 527-533.

- DELLAPENNA, J. W. (2000). Law in a shrinking world: The interaction of science and technology with international law. *Kentucky Law Journal*, 88(4). Retrieved November 10, 2012 from <http://tinyurl.com/ampheot>
- ERNST, S., & HAEUSERMANN, D. M. (2006). Teaching exceptions in European Copyright law—Important policy questions remain. Retrieved September 29, 2012 from [http://papers.ssrn.com/sol3/papers.cfm?abstract\\_id=925950](http://papers.ssrn.com/sol3/papers.cfm?abstract_id=925950)
- GERHARDT, D. R. (2006). Plagiarism in cyberspace: Learning the rules of recycling content with a view towards nurturing academic trust in an electronic world. *Richmond Journal of Law and Technology*, 12(3), 10-61. Retrieved September 29, 2012 from <http://jolt.richmond.edu/v12i3/article10.pdf>
- GREEN, H. P. (1992). Constitutional implications of federal restrictions on scientific research and communication. *University of Missouri-Kansas City Law Review*, 60, 619-643.
- GREEN, S. P. (2002). Plagiarism, norms, and the limits of theft law: Some observations on the use of criminal sanctions in enforcing intellectual property rights. *Hastings Law Journal*, 54, 167-242.
- HUDSON, J. M., BRUCKMAN, A., GELLERSEN, H., SCHMIDT, K., BEAUDOUIN-LAFON, M., & MACKAY, W. (2005). Using empirical data to reason about internet research ethics. In H. GELLERSEN, K. SCHMIDT, M. BEAUDOUIN-LAFON, & W. E. MACKAY (Eds.), *Proceedings of the Ninth European Conference on Computer Supported Cooperative Work , 18-22 September 2005, Paris, France* (pp. 287-306). Dordrecht: Springer.
- HOLMES, G., & LEVIN, D. A. (2000). Who owns course materials prepared by a teacher or professor? The application of copyright law to teaching materials in the Internet age. *Brigham Young University Education and Law Journal*, 2000(1), 166-189.
- KAPLOW, L. (1992). Rules versus standards: An economic analysis. *Duke Law Journal*, 42(3), 557-629.
- LAN, J., & DAGLEY, D. (1999). Teaching via the Internet: A brief review of copyright law and legal issues. *International Forum on Information Technology in Education*, 1(11), 25-30.
- LESSIG, L. (2001). *The future of ideas*. New York: Random House.
- LIVINGSTON, J. S. (2011). Invasion contracts: The privacy implications of terms of use agreements in the online social media setting. *Albany Law Journal of Science and Technology*, 21, 591-636.
- OHM, P. (2010). Broken promises of privacy: Responding to the surprising failure of anonymization. *University of California-Los Angeles Law Review*, 57, 1701-1777.
- ONGE, K. R. S. (1988). *The melancholy anatomy of plagiarism*. Lanham: The University Press of America.
- PLAGDOC, & KOTYNEK, M. (2012, June 8). Reflections on a swarm. Retrieved November 10, 2012 from <http://tinyurl.com/axbbm2l>

- SCASSA, T. (2010). Journalistic purposes and private sector data protection legislation: Blogs, tweets and information maps, *Queen's Law Journal*, 35(2), 733.
- SHAKESPEARE, W. (1598). *Henry IV, part 2*.
- SHAKESPEARE, W. (1623). *Henry VI, part 2*.
- SUNSTEIN, C. R. (1990). Paradoxes of the regulatory state. *University of Chicago Law Review*, 47, 407-441.
- STEARNS, L. (1992). Copy wrong: Plagiarism, process, property, and the law. *California Law Review*, 80(2), 513-553.
- STEWART, P. (1964). *Jacobellis v. Ohio*, 378 U.S. 184, 197 (U.S. Supreme Court), concurring opinion.
- TOWNSEND, E. (2003). Legal and policy responses to the disappearing "teacher exception," or copyright ownership in the 21st century university. *Minnesota Intellectual Property Review*, 4, 209-283.
- TRIGGS, C. (2005). Academic freedom, copyright and the academic exception. *Workplace*, 7, 60-79.
- WARREN, S. D., & BRANDEIS, L. D. (1890). The right to privacy. *Harvard Law Review*, 4, 193-218.
- WRIGHT, R. G. (2006). The role of intuition in judicial decision-making. *Houston Law Review*, 42, 1381-1424.
- XALABARDER, R. (2004). Copyright exceptions for teaching purposes in Europe (Working Paper). Retrieved November 10, 2012 from <http://tinyurl.com/bkxelsf>
- XALABARDER, R. (2009). Study on copyright limitations and exceptions for educational activities in North America, Europe, Caucasus, Central Asia and Israel. Retrieved November 10, 2012 from <http://tinyurl.com/a7tyxwu>
- YAKOWITZ, J. (2011). Tragedy of the data commons. *Harvard Journal of Law & Technology*, 25, 1-67.



# **Do Easily Copied Internet Media in the Library Lead to Plagiarism?**

**Gabriel Gomez**

Chicago State University

*Copying is such an integral part of today's digital media, much of it widely distributed on the Internet, and this may have something to do with the seeming increase in plagiarism. While the library may be the ideal source for students to gather information, most students are well acquainted with the Internet from their personal use and, sadly, the practices they develop there may conflict with the important and at times complex practices that define proper and improper copying within academic settings like the library. This paper takes a qualitative look at ownership, copyright, fair use, and plagiarism from the standpoint of the library to see if mistakes and misunderstandings arising from casual Internet use may well lead to increased incidents of plagiarism.*

## **Introduction**

Copying, and digital copies in particular, are central to today's information rich environment, a reality that through its many possibilities, can lead to or support plagiarism (Ma et al., 2008, p. 199). One cause said to explain the current seeming increase in plagiarism summarized by Selwyn (2008) was the intriguing idea that changes in our information environment such as "the structure and nature of the internet itself" (p. 466) make plagiarism inherent to that environment. Interestingly, Selwyn's interpretations of his own data

do not suggest that the internet is necessarily causing new or different patterns of plagiarism amongst the student body or, indeed, that online sources constitute a necessarily 'new' form of plagiarism. (p. 476)

Much has been written about plagiarism, its changing definition, its proliferation, technological means to combat it, and even its moral or psychic underpinnings, some of it leading to conflicting views. It is not possible to offer a comprehensive examination of all these items here, as even just a comprehensive view of plagiarism could fill a book. Related issues like copyright infringement and Creative Commons licensing are similarly large topics.

Instead this paper will explore, through an ethical enquiry, in a qualitative manner, issues that have made copying a normal, value-free experience in our information environment, an issue that has led to a seeming increase in academic plagiarism in the U.S. In fact, one aspect of this value-free aspect of copying and digital media is addressed by Guindon (2006), who noted that "by accessing a Web page you are actually making multiple copies of it on your computer" (p. 164). While such copying is indeed neutral, problems might arise when this ease of copying found online or in digital media comes to support plagiarism.

The library should be the nexus between students and information, but students have developed their own ideas and practices regarding online information before they ever enter a college or university library, often with little concern for ownership or citation. In academia, plagiarism usually refers to the conscious use of someone else's ideas, writing, or research as one's own; to avoid this problem, proper citation is vital to academic use in the library and within academia in general. Merriam-Webster's online dictionary defines the verb *plagiarize* as "to steal and pass off (the ideas or words of another) as one's own: use (another's production) without crediting the source." Sadly, in my experience as an educator, many students do not share the concerns of the educators they must work with on the matter of proper citation and plagiarism, something Selywn (2008) and others have noted. For example, in Selwyn's survey, fifty-nine percent of respondents admitted to plagiarism.

As the work of a U.S.-based professor who educates school librarians, this paper will address matters like ownership, copyright, and the doctrine of fair use from the standpoint of a U.S. academic library. It will do so only as they might impact the act of plagiarism. In this area, though, I am indebted to a number of authors, in particular Pressman (2008).

I teach online and rely on online resources heavily, and I wonder if that reliance on the Internet is why my students seem to need ever-more-explicit

instruction about the very basics of citation and plagiarism. My students often treat media or information that is freely available and free of cost to them online as if it is free of restrictions on how they can use it, a trend previously noted by Deborah R. Gerhardt, who wrote:

A student may understandably assume that anything available for free is not owned by anyone, and therefore, may be freely used. (2006, p. 6)

Information or media is usually sponsored or paid for in a variety of ways (see, e.g., Kaser, 2000). Information is often owned or governed by copyright law; rules and practices must govern its use. Even in the case of sources like Wikipedia, where information is funded by the Wikimedia Foundation, ethical rules apply to information use in U.S. academic settings. All information created by an author or producer, regardless of where it comes from, even if it uses a Creative Commons license, must be cited.

Although this seems simple, common misunderstandings about the proper way to use information occur. These misunderstandings can be distinct from a different idea, namely that the Internet and media or content on it should, in fact, be freer in terms of restrictions, if not utterly free of cost. This last understanding will be addressed very briefly in the last section of this paper, as it may have some interesting bearing on plagiarism as well.

## **Between Academic Rules and Free Information**

Two widely divergent views about information appear to bracket the discussion on plagiarism. In education, plagiarism is often seen as a relatively clearly defined problem that should also be easily and widely understood (Fish, 2010). Of course even respected researchers or academics can be confused or even disagree on this matter, but one thing is certain. Proper citation and plagiarism are widely discussed issues that every academic must strive to understand. On the other side of the discussion on plagiarism is an evolving attitude related to the proliferation of information in digital form that seems to liberate information. This view finds a more extreme expression in the idea that the technological developments represented by the Internet should make information as free and unfettered as possible, an idea with as many variable meanings as might be found among the different exponents of the slogan "Information wants to be free" (Wikipedia, 2012). Neither general library practice nor the basic aims of copyright would allow one to dismiss either side of this debate completely. Instead, a balance between a basic respect for authors and ownership and the need of the public to have access to infor-

mation should inform both library and copyright practices. But this is speaking in very general terms.

To address plagiarism in the specific realm of education, I found Stanley Fish's (2010) article very helpful. Fish wrote:

And if you're a student, plagiarism will seem to be an annoying guild imposition without a persuasive rationale (who cares?).

For students, plagiarism may well affect only a small portion of their information use and then only in a moral sense, because as Fish noted, plagiarism is not a crime. While this is true, information misuse can be a legal matter.

Distinctions between plagiarism and copyright misuse should be clear to education professionals. At least one would hope so, but it is not necessarily the case. There are indeed some reasons why many in academia might be confused.

For example, in the U.S., proper citation involves naming authors or creators and publishers, and this can be seen as describing ownership. However, citation is not primarily about designating owners. Instead, citing authors or creators and publishers directs readers to the original source of the material borrowed, and that source of the material may not be the owner of the material, an issue I will return to later.

In addition, for proper citation, the date of publication is also required, and students are often taught by educators to find that date next to a copyright symbol. Again, in citation, the date is meant to refer to the date the expression was created. Using the copyright date does not necessarily refer to the very large legal matter of copyright itself.

Citation practices like these, where ownership and copyright seem to be central, can be confusing for educators and even more confusing for students because, to reiterate, citation has a larger and slightly different aim than describing ownership and referring to copyright law. To summarize again, citation in the U.S. is primarily concerned with directing readers or viewers toward the source of borrowed expressions or ideas. I wonder if students always understand this basic aim as they try to keep track of authors, publishers, and dates.

If students might be confused at this juncture, it is important to remember that they already possess ideas about how to use online information. And as Pfannenstiel (2010) noted:

Problems arise when students use their habitual or everyday online research and writing strategies to manage information in academic contexts. (pp. 41-42)

In that everyday online use, as Pfannestiel described, copying without emphasis on citation, plagiarism, or copyright infringement concerns seemed to be quite normal. Of course, this state of affairs is reversed in a U.S. academic setting.

If students live in an information-rich world, one they access outside of academic confines, then plagiarism can be seen to affect only a small portion of that world. Furthermore, definitions of plagiarism can be difficult to separate from ideas about ownership or copyright because of how citation is done, and this can lead to confusion. Finally, common information and media uses and practices where citation, plagiarism, and copyright infringement concerns do not play a central role have to be unlearned and then replaced by new norms for an academic setting.

The next section explores information use in the library from a U.S. perspective to demonstrate an important point, namely that information use is a complex matter for scholars everywhere and national law and practice are only part of a complex equation.

## **Library Information Use in the United States**

Even if students understand how the aims of citation differ from the broader and related issues of copyright and copyright infringement, that does not mean students understand all they need to know concerning these matters. There is a great deal more to know and it can be a challenge, even for information professionals. Proper material use is vital to library professionals and this use involves understanding ownership, copyright, fair use, and more. Library professionals rely on a large body of information defining information and media use from organizations like the American Library Association (ALA) and its many sub-organizations and working groups. With the support of such entities, libraries and librarians then define proper use and even shifting or competing understandings on the meaning and aims of copyright protection in general.

However, as already noted, for students in the U.S., the matter of copyright and ownership are related to plagiarism in that proper citation requires an author, a publisher, and a date. If we accept that a more casual relationship to information has become formative of today's students, it follows that such students might approach information in the library in a more simple instrumental sense that might be summarized by the questions "Can I get what I need, can I take it with me, can I share it or use it, and if so, how?" Such questions are relevant to the library because at the heart of copyright are the seemingly conflicting needs of creators who have a right to benefit from their

work, and the general societal good of having knowledge widely available (Pressman, 2008). Ideally, all who use information in an educational setting and not just librarians, will understand the importance of a balance between these two apparently conflicting aims as the first step in using information legally and ethically.

### **Ownership and Accessibility**

In U.S. libraries, the relationship of a user to library materials is also critical. Two issues stand out. First, libraries collect material that users may use, and within limits, copy for further personal, educational, or professional use, primarily through note taking, the use of copy machines, or via downloading and digital copying. Second, libraries must protect the copyrights of their materials not just to benefit the copyright holders and to protect themselves from being accused of legal wrongdoing, but also to help the general public for whom copyright is meant to ensure access to knowledge. This second issue may explain why

[t]hough librarians know that information is anything but free, those who use libraries have come to believe that it is free. (Kaser, 2000)

Returning briefly to the issue of copying, individual libraries find their own way in this matter with the help of larger organizational and supra-institutional entities. In addition, libraries and librarians interpret the law and the agreements they sign covering the materials they purchase because the ability to make copies can also vary according to the copyright owner of the material. Of course, fair use understanding and practices also play a role here, something addressed more fully in the next section of this paper. From these various sources and understandings, some general guidelines emerge. Material use and copying in the library are meant strictly for those users who can access the library. It does not mean that the user owns the material and is then allowed to use the material in every way imaginable.

This is the beginning of the problem with copyrights for many young users, particularly students. They associate using material with owning material, but this can be misleading when confronting intellectual property.

[T]here is no such thing as ownership of ideas. Holders of copyrights own only specific expressions of their creations. Second, copyright is only a limited monopoly on reproduction. Works are protected for a certain duration, after which they become part of the public domain. Finally, when you purchase a book or a music CD, you become the actual owner of the product, the author (or the corporation that holds the rights) has very little to say about the way you use the work

as long as you don't try to reproduce it in flagrant infringement of the copyright laws. (Guindon, 2006, p. 160)

Therefore, assuming one can go into a library, either virtually or in person, then one can use the ideas found there. However, ownership of the expression of the ideas is protected and one cannot do certain things with the material from which one has learned unless the copyright has expired. To do so would be to act as if one had the rights referred to above, which includes the ability to take credit for, or to make money from the expression (the actual word, images, and so on) made by the author or creator. In addition,

Owning a copy of a work is different than owning the copyright to that work. A student who purchases a book has the right to lend it to a neighbor, display it on the coffee table, re-sell it on Amazon.com™, or use it as a doorstop. But he does not have the right to scan a chapter from that book and load it onto a course Web site. While the book becomes his physical property, it does not become his intellectual property. Likewise, an e-mail or presentation that Student A forwards to Student B does not become the intellectual property of Student B, but remains that of Student A. An article retrieved from a database has its own copyright nuances. Technically speaking, libraries do not "own" the articles within subscription databases. They are simply paying for access to them, which is similar to "renting" them for a period of time. (AlSaffar, 2006, p. 13)

Separating ownership from copyright can be a difficult task. If students do not care about plagiarism, the nuances of copyright protection may seem even more obscure. Of course, there is a relationship between plagiarism and copyright. For U.S. academics, acknowledging the source of an idea is the basis of ethical use. Ethical use is built on a strict regimen of citation, one that generally effects or puts into action copyright law and practice.

The distinction between use and ownership is further complicated because libraries have often replaced the outright purchase of magazines, journals, and similar article-length works with electronic subscriptions, something AlSaffar likened to renting. This can lead to various levels of access with confusing or disturbing results. Most researchers have probably experienced this when they discovered they could gather a list of articles but then could not access the full text of all articles on that list. This occurs because the library has only paid for full access to the database listing the articles and some of the articles listed, but not all of them. Individual libraries have various services or strategies to help the user in such instances, ranging from using interlibrary loans to going online or asking a librarian for help. A user might even have to go to another library or purchase a copy outright. Serious researchers are likely to perform any actions or follow any advice needed to

recover the missing article, but students with less formal information use practices may not.

Ultimately, scholars and educators may work hard to understand the relationships and distinctions between ownership, copyrights, accessibility, and plagiarism reviewed here. The non-academic user of information may see little reason to understand these distinctions and then put into practice means that ensure authors, creators, or owners are correctly cited and adequately compensated for their efforts. These means will be investigated more closely in the next section.

### **Fair Use in the Library**

Students must understand that they can access information or even own specific copies of information, but in either case, fair use guidelines must actually guide and define acceptable copying and citation within an academic setting to avoid both misuse and plagiarism. Here again, libraries often develop their own understanding and guidelines regarding fair use with the help of professional organizations like the ALA and their own interpretations of the law. Understandings arrived at in this way are often made available to users, frequently on Web sites or in handouts. As with copyrights, professionals in the same field might arrive at slightly different understandings about fair use, something that might confuse students if they go from one library to another.

To clarify fair use for library users, graphic representations are also often created by libraries to recommend amounts of material that might be used when quoting or copying text, images, sound, music, film, or video. Here is a small text sample from a table in guidelines developed by the Stanford University Library regarding how much poetry that library believes might be copied in line with fair use.

Up to 10% or 1,000 words, whichever is less, of a copyrighted text work. For example, you may use an entire poem of less than 250 words but no more than three poems by one poet or five poems by different poets from the same anthology. (Stanford University Libraries and Academic Information Resources, 2010)

In this Web site table, similar numeric amounts exist for music and lyrics, animation, video, film, photographs, databases, and data tables. Ultimately, although libraries work hard to make fair use easy to understand and deploy, students must use such materials to develop their own complex sense of responsible use regarding intellectual property, a sense that should arise from an understanding of the basic concept of fair use. If they can manage this, they should be capable of successfully deploying any material for educational purposes. But how likely is this considering the wider information environ-

ment of most students? Richard Kaser (2000), the former executive director of the National Federation of Abstracting & Information Services, wrote the following:

When the man on the street says “fair use,” he is speaking another language than the one we as publishers—and our lawyers—know. He’s speaking the same language as the one you speak when you photocopy for a friend the interesting newspaper article you read this morning. He’s speaking the same language that you speak when you loan the video tape you made of last Sunday’s game to a pal or spin off that new CD onto tape so you can listen to it in your car. He’s speaking the same language that you speak when you help your child write a term paper, in which you quote and cite various sources verbatim, without requesting permission.

Though Kaser’s tone may be sarcastic and the media he lists are somewhat dated, his basic point is one many students may agree with, namely that fair use understandings rarely guide non-education information and media use. However, in an academic setting, university libraries generally develop clear recommendations on fair use.

This is where students often run into problems when they copy material for assignments. Without condoning any such misuse it is possible to recognize that students who copy informally in the way Kaser suggested may not always mean to steal information, break the law, or plagiarize. Everyday use may be quite legal, but in an academic setting, to use material without citation remains completely unethical. Even so, students may do it because the information is available to them and they can make the copies of it they like without incurring much cost. If they understand that it is not ethical, they often cannot explain exactly how or why it is wrong

due to not having an adequate enough understanding of the topic (plagiarism) to detect it in concrete examples. Unintentional plagiarizing could very well be the result. (Hochstein et al., 2008, p. 63)

To know exactly why such actions are wrong requires a fair amount of effort and knowledge. Even then, using information or media is no simple matter. Once books, sound recordings, films, television programs, and newspapers were distinct as well as easily sold and consumed. Today, those media come together online, allowing a library or Internet user to copy and distribute a wide variety of information in a way that challenges traditional information and media use. Students accustomed to a rich information environment, one they seemingly access for “free,” might well come to understand copying online digital information as just one way that they can use digital media.

## Free Media

Free information is an illusion. But it would take a far greater act of magic than I can conjure to take away this popular myth. (Kaser, 2000)

Kaser went on to explain that the idea of free information persists even though people often pay to access the source of that information, that is, they will pay to access the Internet. Kaser's point about free information is important. The idea of it has persisted for quite some time in the U.S. even as Kaser summarized how a large portion of information seen as free was actually paid for over the last century.

Just as importantly, he indicated there was a willingness on the part of information users to pay for information, but generally speaking, only if it was not too expensive. Significantly, the company he mentions that successfully sold its documents charged a dollar, a price that has since become extremely important in the era of commercially available downloaded music, as seen in examples such as iTunes' 99-cent price. What is so helpful in Kaser's work is the idea that information free of cost, or information at a low cost seems to have incredible staying power in the process of consuming information. Furthermore, the idea that consumers would resist prices they think are too high is also important, something the popular video rental company Netflix found when it raised prices in 2011, and in so doing, "lost 800,000 US subscribers" ("Netflix stock still hurt by price hike," 2012, July 13).

Raising prices beyond what consumers deem fair is seen as a serious problem for digital media companies. Indeed, the title of the aforementioned article highlights that the stock price of Netflix has not recovered from consumer resistance to its attempt at a price increase. Just as importantly, where information is seen as free, as on the Internet, breaking away from that model can also be very difficult, a point explored again at the end of this article.

From my students who assumed they could use Internet material as they saw fit, I have heard the following three basic rationales: 1) it is free, 2) they have paid for their access to it, or 3) it is technically possible. Among these reasons it is hard to find any concern about the confusing state of affairs between ownership, copyright, fair use, plagiarism, or even the basic idea that one can possess or own copies of information or media and still not own the rights to such material.

This confusing state of affairs affects not just my students. It is widely recognized by many for whom copyright ownership is extremely important. It is a recognition that has led to changes in the music, film, and television industries, as well as in journalism, and now publishing—changes that might involve combating what is seen as copyright infringement and piracy with

punishment, education, pricing, or some combination of these three elements. According to Davis (2011), the preferred solution is education in line with successful pricing strategies.

That emphasis on education over enforcement is not coincidental. Publishers have watched how the recording and movie industries have struggled with content piracy and seem determined to follow a different track. While there are tools built on a litigation business model, many vendors are focused on the commercial opportunities that come from seeing copyright infringement as a sign of demand to be monetized. (p. 36)

This is the trend in digital media and information today. Those who produce information in a variety of once-separate media are searching for ways to make pricing and selling of information and media appealing so that piracy or misuse no longer occurs. And for good reason, because as Comas-Forgas and Sureda-Negre (2010) stated:

The Internet as a 'source of sources' or a 'library of libraries' is no longer a metaphor; the idea has become a reality that affects all fields based on information exchange, locating content and accessing and producing knowledge. (p. 228)

In that library of libraries, the average library user or student is often confused about the relationship between copying, plagiarism, and copyright ownership infringement. Part of the confusion arises from students' familiarity with that library of libraries in their private lives, and not through educational resources like the school or academic library. Another part of the problem is the ability to see copying, plagiarism, and copyright ownership infringement as distinct, something complicated by the fact that all three can easily collide in a single action. For example, a student used to the casual copying practices spawned by frequent Internet use may copy some information into an assignment just as they might do with a text message, an e-mail, a Facebook post, or some other action. Copying in this manner in an educational endeavor can easily lead to unintentional plagiarism, and worse, the frequency of such actions offers cover to those who deliberately plagiarize, because such plagiarizers often say, "Everyone is doing it," something found by Selwyn in his 2008 study of British university students (p. 475).

It is conceivable that students do not realize that practices from their private use of information and media lead to unethical information use in an academic setting. Furthermore, acclimating those students to the complex, if essential, practices of education regarding plagiarism may not be a simple process, and this complicates the very definition of plagiarism. Plagiarism must be seen as a set of practices that exist in a narrow band of information use for many students. This is one of the reasons so much has been written on

the topic of plagiarism in the digital age. If plagiarism is part of the educational realm or “our house,” as Fish (2010) called it, then it becomes easier to see that teachers must teach students that the rules in “our house” are very different from the rules where they live.

There is one final issue which must be addressed. If information and media consumption practices have developed wherein copying, copyright misuse, and plagiarism seem less important or even irrelevant, there are those who would go further. Some have actively resisted either paying for information or media. In the words of a resister named Zac Shaw,

Asking today’s music consumers to kindly start paying for recorded music again because it’s the ethical thing to do isn’t only unviable—it’s not the ethical thing to do anymore. Free Culture is an ethic, and I think I can speak for my generation when I say we believe it to be the high ground over the way the music industry used to be run. (“In defense of free music: A generational, ethical high road over the industry’s corruption and exploitation,” 2012, June 19)

Among the reasons Shaw resists paying are “perpetual copyright and destruction of fair use and the public domain.” Shaw refers to “Free Culture,” which is also the name of a book by Lawrence Lessig (2004), and some would say, a movement. It is not necessary at this point to get into the merits of either Shaw’s or Lessig’s position, although Lessig in particular makes some persuasive arguments, whether or not one agrees with him.

What is important to note is that the very existence of resisters or even a movement of resisters to the current state of affairs regarding copying, copyright, fair use, and the public domain makes even more complex the troubling problem of plagiarism in our information-rich environment. It is easy to imagine a student who does not care about the narrow ethical issue of plagiarism; now it is possible to imagine students who confront educators with the pronouncement that plagiarism is simply a means to control that which should be free in both a commercial and in a larger philosophical sense. Because of the free culture movement, students may believe that all information “wants to be free.”

## References

- ALSAFFAR, J. (2006). Copyright concerns in online education: What students need to know. *Journal of Library Administration*, 45(1/2), 1-16.
- COMAS-FORGA, R., & SUREDA-NEGRE, J. (2010). Academic plagiarism: Explanatory factors from students’ perspective. *Journal of Academic Ethics*, 8(3), 217-232.
- DAVIS, K. N. (2011). The copyright evolution. *Information Today*, 28(10), 1-36.

- FISH, S. (2010, August 9). Plagiarism is not a big moral deal. *The New York Times*. Retrieved November 11, 2012 from <http://tinyurl.com/35lgqjx>
- GERHARDT, D. R. (2006). Plagiarism in cyberspace: Learning the rules of recycling content with a view towards nurturing academic trust in an electronic world. *Richmond Journal of Law and Technology*, 12(3), 10-61. Retrieved September 29, 2012 from <http://jolt.richmond.edu/v12i3/article10.pdf>
- GUINDON, A. (2006). A very short history of copyright: Adopting the user's perspective. *Canadian Journal of Information & Library Sciences*, 30(3/4), 153-174.
- HOCHSTEIN, D. D., BREWER, J., STEINKE, M. D., & TAYLOR, J. (2008). Examining the issue of academic plagiarism: What do students at Wright State University Lake Campus know about plagiarism? *Association for Regional Campuses of Ohio Journal*, 1459-1481.
- Information wants to be free. (2012, September 2). In *Wikipedia*. Retrieved September 29, 2012 from [http://en.wikipedia.org/wiki/Information\\_wants\\_to\\_be\\_free](http://en.wikipedia.org/wiki/Information_wants_to_be_free)
- KASER, R. T. (2000). If information wants to be free ... Then who's going to pay for it? *D-Lib Magazine*, 6(5). Retrieved September 29, 2012 from <http://www.dlib.org/dlib/may00/kaser/05kaser.html>
- LESSIG, L. (2004) *Free culture: How big media uses technology and the law to lock down culture and control creativity*. New York: The Penguin Press.
- MA, H., WAN, G., & LU, E. (2008). Digital cheating and plagiarism in schools. *Theory into Practice*, 47(3), 197-203.
- Netflix stock still hurt by price hike. (2012, July 13). *The Australian Financial Review*. Retrieved November 10, 2012 from <http://tinyurl.com/ao6ev2m>
- PFANNENSTIEL, A. (2010). Digital literacies and academic integrity. *International Journal for Educational Integrity*, 6(2), 41-49.
- Plagiarize. (2012). *Merriam-Webster Online*. Retrieved September 29, 2012 from <http://www.merriam-webster.com/dictionary/plagiarize>
- PRESSMAN, R. R. (2008). Fair use: Law, ethics and librarians. *Journal of Library Administration*, 47(3/4), 89-110.
- SELWYN, N. (2008). 'Not necessarily a bad thing ...': A study of online plagiarism amongst undergraduate students. *Assessment & Evaluation in Higher Education*, 33(5), 465-479.
- SHAW, Z. (2012, June 19). In defense of free music: A generational, ethical high road over the industry's corruption and exploitation. Retrieved November 2, 2012 from <http://tinyurl.com/cl2nsov>
- STANFORD UNIVERSITY LIBRARIES AND ACADEMIC INFORMATION RESOURCES (2010). Proposed educational guidelines on fair use. Retrieved September 29, 2012 from [http://fairuse.stanford.edu/Copyright\\_and\\_Fair\\_Use\\_Overview/chapter7/7-c.html](http://fairuse.stanford.edu/Copyright_and_Fair_Use_Overview/chapter7/7-c.html)



# Scientists and Librarians Create an Environmental Toxicology Data Repository

Deborah Keil<sup>I</sup> and Kenning Arlitsch<sup>II</sup>

<sup>I</sup> Department of Pathology, University of Utah, <sup>II</sup> University of Utah J. Willard Marriott Library

*This paper proposes the development of an environmental toxicology data repository for the Wasatch Front, the major metropolitan area in the state of Utah. This repository, developed by scientists and librarians, will address gaps in national-level repositories that cannot take into account unique local issues caused by geography, industry, military installations, and population. The repository will enable local clinicians and academic researchers to collect data, describe it using a standard metadata schema, preserve it for the future, and make it accessible through common search engines.*

## Introduction

The Wasatch Front in Utah encompasses several cities in north central Utah and is home to most of the state's population. The area is characterized by its rugged beauty and topographical contrasts, with high, snow-capped mountains to the east transitioning to the Great Basin Desert in the west. The Wasatch Mountain range rises more than 1,500 meters from the valley floor in Salt Lake City to heights of 3,600 meters, making it one of the most precipitous elevation gains in proximity to a major urban area. The Great Salt Lake is another dominant feature of the landscape, providing a migratory stopover for birds as well as supporting a multi-million dollar brine shrimp industry, the only life the lake supports. The Great Basin is a vast area of alternating

desert basins and high mountain ranges that stretches across Nevada to the Sierra Mountains in California. One of its defining characteristics is that there are no significant rivers and none of its water escapes to either ocean. Eighty percent of Utah's human population—just over two million people—lives along the Wasatch Front, and this population depends on the 500 inches of annual average snowfall on the mountains that produces life-giving water. (“Wasatch Front,” 2012).

The writer Wallace Stegner (1960) referred to the American West as the “geography of hope” but also warned of its fragility under man's heavy hand. The geography of the Wasatch Front, which is so stunning, also creates environmental conditions that exacerbate the effects of man-made pollution and, consequently, affect human health. The topography creates a natural phenomenon known as temperature inversions, where warm air at the elevation of the mountains presses cold air to the valley floors during winter high-pressure systems, trapping the emissions of industry, automobiles, and homes and sometimes creating some of the worst air quality in the nation (Garber, 2011). The presence of major extractive industries, military installations, and a rapidly growing population all contribute to environmental pollutants. Air, water, and soil are compromised, and the long-term health effects of these conditions remain unclear. Some health conditions may be correlated: The region has been identified as exceeding national population averages in the incidence of asthma (May, 2011; KSL, 2007) and autism (Baio, 2012), while statistically significant increases in lung and bronchial cancer were identified near Hill Air Force Base (Ball et al., 2008).

The Wasatch Front is home to nationally renowned medical institutions, the Utah State Health Department and world-class universities with researchers who investigate environmental pollutants and related health issues. However, there is no local infrastructure to facilitate the exchange of information among these regional experts and institutions. The identification of potential cause-and-effect relationships of illnesses due to environmental chemical contaminants depends on the availability of reliable human health and exposure data. Developing a cooperative infrastructure for sharing and combining data could identify local themes and prioritize action according to regional environmental health data trends.

## **Research Problem**

Most environmental toxicology data repositories have a national scope and fail to consider the unique regional effects of local topography, industry, and population. This paper proposes an environmental toxicology data repository

focused on the Wasatch Front that could function as a model for other regionally based repositories. The goal is to merge toxicologically relevant and human health data in a regional centralized repository. Such a data repository on a flexible, interoperable, and accessible platform would provide for the development of computational risk assessment models to address regional environmental insults and improve health outcomes.

## **Potential Sources of Environmental Contaminants on the Wasatch Front**

### **Extractive Industries**

One of the largest open-pit copper mines in the world has been in operation in the Salt Lake valley since the early part of the twentieth century and has fuelled a significant part of the area's economy (Arrington & Hansen, 1963). Open-pit hard rock mining is environmentally destructive and produces large amounts of waste (known as tailings) because only a small portion of the total mined material contains the sought-after ore. The tailings contain heavy metals such as lead, zinc, and cadmium and can contaminate the soil, creating potential health hazards for humans and wildlife. Vegetables grown in mine tailings have been shown to accumulate heavy metals in the edible parts of the plants (Cobb et al., 2000). Additionally, metal smelters release gases such as carbon dioxide (CO<sub>2</sub>), sulphur dioxide (SO<sub>2</sub>), and nitrogen oxide (NO<sub>x</sub>) (Dudka & Adriano, 1997).

On the west side of the Great Salt Lake is U.S. Magnesium Corporation, formerly known as MagCorp. The company was sued by the U.S. Department of Justice (DOJ) on behalf of the Environmental Protection Agency (EPA), and the DOJ (2001) noted that for

many years MagCorp has ranked Number One on the EPA's toxic release inventory, based on its chlorine emissions.

The federal government sued again in 2005, alleging that the plant was "the nation's worst polluter" and that it "illegally manufactured and dumped carcinogenic polychlorinated biphenyls, or PCBs" (Henetz, 2005). MagCorp filed for bankruptcy, and its subsequent restructuring as U.S. Magnesium helped it to rebuff the lawsuits. U.S. Magnesium has made substantial improvements in its emissions, though it is still considered one of the top five polluters in Utah (Fahys, 2008).

## **Military Installations**

The Wasatch Front is home to several military installations, such as the Dugway Proving Ground and the Tooele Army Depot in the desert to the west of Salt Lake City and Hill Air Force Base to the north.

The U.S. Army manages the Dugway Proving Ground, and its mission is to test biological and chemical weapon systems. Dugway also serves as a training ground for fighter pilots from Hill Air Force Base. In 1968, six thousand sheep died in Skull Valley, thirty miles from Dugway, due to apparent exposure from an accidental release of the VX nerve agent (Boffey, 1968). More recently, scientists at Dugway have worked with anthrax spores (Matsumoto, 2003).

The Tooele Army Depot once stored approximately 42% of the United States' chemical weapons stockpile, including sarin, mustard, and VX blister and nerve agents, some of which "could persist in the environment long after an accidental release" (Carnes & Watson, 1989). While these weapons are mainly known for their acute lethality, the "possibility of long-term brain dysfunction after exposure to a nerve agent has also been raised" (Carnes & Watson, 1989). In 1991, President George H. W. Bush reversed decades-old U.S. policy by "foreswearing the use of chemical weapons for any reason," effectively eliminating the need to maintain a stockpile of these weapons (Foote, 1994). As a result of this policy change and because of the dangers of the long-term storage of these aging weapons, a plan to incinerate them was developed, and by 2012, the chemical weapons stockpile was reduced by 99%.

Hill Air Force Base is located in Ogden and is one of the largest employers in Utah. It has been listed as a Superfund site by the Environmental Protection Agency since 1987 (U.S. Environmental Protection Agency, 2012) due to soil contamination from polychlorinated biphenyls (PCBs) and a degreasing solvent known as trichloroethylene (TCE). While the PCBs have largely been removed from the soil, the TCE contamination has spread into the aquifer adjacent to the base and continues to migrate as a subterranean plume (Jackson & Dwarakanath, 1999). TCE is a known carcinogen, and the cleanup is expected to last another 65 years (Dougherty, 2010).

## **Home and Automobile Emissions**

The population density along the Wasatch Front can be expected to contribute to particulate emissions. Home furnaces and fireplaces, as well as gasoline and diesel motor vehicles produce pollution that adds to industrial pollution and affects air quality, adding to the industrial production of particulate

matter (PM) (Utah Department of Environmental Quality: Division of Air Quality, 2012). Wood smoke from natural or human-started wildfires also add particulate pollution (Long et al., 2002), which can vary in size from fine (2.5 microns and smaller diameter) to coarse (10 microns). Fine PM can pass deep into the lungs and has been linked to cardiovascular disease and specifically to heart disease events triggered by short-term exposure along the Wasatch Front (Pope et al., 2006). Particulate air pollution increases mortality in the region (Pope et al., 1999).

## **Human Biomonitoring**

Population-based biomonitoring programs in the United States began with the efforts of the National Human Monitoring Program (NHMP), administered by the U.S. Environmental Protection Agency, and National Health and Nutrition Examination Survey (NHANES), administered by the United States Centers for Disease Control and Prevention. In the last decade, the National Human Exposure Assessment Survey (NHEXAS), administered by the U.S. Environmental Protection Agency, and NHANES (1999-2000) have expanded the list of chemicals measured to upward of 140 (National Research Council (U.S.) Committee on Human Biomonitoring for Environmental Toxicants, 2006). This process provides valuable data about background exposure levels and trends in the population as a whole. However, it is less effective for addressing local issues, and it does not consider environmental monitoring data.

The U.S. is geographically expansive with great population diversity, and each region may be defined by unique health concerns. It would seem reasonable to implement a local biomonitoring project to identify unique local issues and to address health problems. The coordination of a local surveillance program may provide for more agility in human biomonitoring, as the approach would be scaled to greater responsiveness for local environmental concerns. Environmental issues may be uniquely assessed for the community by combining local human biomonitoring data, hospitalization data, and environmental contaminant sources. Based on current toxicological data combined with an expanding population located near a number of extractive industries, military installations, and unique geologically restrictive formations, the potential risks to human health call for a Wasatch Front regional human biomonitoring program.

Within the scope of environmental toxicology, it is increasingly important to collect exposure information and articulate this with corresponding human health. The National Research Council (NRC) of the National Academy of

Science (NAS) emphasizes the importance of collecting human exposure, population susceptibility, environmentally relevant hazard, and biomonitoring data (National Research Council Committee on Toxicity Testing and Assessment of Environmental Agents, 2007). Although the literature characterizing environmental chemicals in controlled laboratory studies has grown substantially, the primary limitations in human health risk assessment lie in linking this laboratory-based data to real-world exposures.

The Wasatch Front serves as an excellent model for such an undertaking. Health disparities include increased rates of autism, asthma, and lung and bronchial cancers, and this disease pattern differs from those identified nationally. By identifying these local trends, unique health priorities for this region can be established. Sharing data and conclusions with policy and health professionals facilitates the ability and the urgency to address local issues.

## **Data Management**

The rapidly accumulating literature on data management points to significant challenges and opportunities for scientists and librarians working collaboratively. Health- and science-related disciplines are generating an unprecedented deluge of data, and the scientific community is struggling to harness larger data sets and increasingly complex data in a transparent and integrated fashion. Currently, some exposure assessment data that identifies sources and chemicals is housed in government organizational “siloes” such as the National Center for Environmental Assessment (NCEA), administered by the U.S. EPA, National Exposure Research Laboratory (NERL), administered by the U.S. EPA, and Regional EPAs. The challenge is to create interoperability between these repositories to more effectively prioritize environmental health threats. Achieving multi-scale integrations is necessary to advance translational science.

Recently, National Institute of Environmental Health Sciences (NIEHS), administered by the U.S. National Institutes of Health, has made available Chemical Effects in Biological Systems (CEBS), administered by the U.S. NIH, the first public repository for toxicology data. CEBS was originally developed in 2002 to house genomic data and was recently expanded to include experimental design with animal, human, and cell culture data in toxicology. Adapting or linking this model to a local repository may maximize local information for a community.

## Challenges

Most scientists struggle to manage the data sets that support their research, particularly as the ability to produce data increases. Numerous barriers to effective management have been described; some are technical in nature, while others are non-technical. Technical barriers include storage, metadata, and software, while non-technical barriers can include fear of competition, lack of trust, lack of incentives, lack of control (Feijen, 2011), and concerns about data quality (Research Information Network & National Endowment for Science Technology and the Arts, 2010).

Librarians must also confront the challenges of providing high-quality data management services. A unique set of skills is required to address the finer points of repository management, discipline-specific metadata, search engine optimization, and digital preservation; many librarians are unprepared.

## Opportunities

There are obvious advantages that come from managing data in a formalized structure, including minimizing loss and maximizing retrieval and reuse. Recent research also demonstrates that authors who make available the data sets that support their publications enjoy a higher rate of citations (Piwowar et al., 2007). Making data available has been characterized as necessary to advance scientific research and solve global problems by some authors (Faniel & Zimmerman, 2011), as well as promoting efficiency in research and scholarly rigor by others (Research Information Network & National Endowment for Science Technology and the Arts, 2010).

There are many advantages to combining toxicology and health data in a central or coordinated repository. The interpretation of a study may be broadened when considered in the perspective of related studies, and new relationships or queries may be generated. As the data are combined in a flexible format within an institutional repository, a future foundation for computational modeling may be established. Currently, limited computational models are available to query between data sets. As technology advances, these data are poised for utilization in future discoveries and applications of risk assessment.

## Scientists' Data Management Needs

Data management needs for most scientists can be divided into two broad areas: working data and archival data. Working data are shared and used by co-researchers as they conduct their research. Access to the data is generally limited to members of the team, and data at this stage is closely guarded. The

technology needs include restricted workspaces that allow team members to share files and version control features. It may also include the need for electronic laboratory notebook software.

Data sets become archival after researchers have published their findings and agreed to release all or a subset of their data. Many universities have created institutional repositories to permanently archive the scholarly publications of their researchers, and institutional repositories may be appropriate for data sets as well.

### **Wasatch Front Repository Proposal**

The environmental toxicology repository proposal for the Wasatch Front consists of two parts: 1) a repository run by the University of Utah built on an existing infrastructure, and 2) inclusion of that repository (as well as others) in the Mountain West Digital Library (MWDL).

The University of Utah manages an institutional repository known as USpace (<http://uspace.utah.edu>), whose mission is to capture the intellectual output of the University of Utah faculty, staff, and students. USpace consists of several collections, such as electronic theses and dissertations, scholarly papers, and university administrative records. Adding another collection for data sets is feasible.

Other researchers, particularly those unaffiliated with the University of Utah may not wish to store their data in a university-owned repository. The MWDL was founded in 2002 under the auspices of the Utah Academic Library Consortium and is a distributed digital library portal that harvests only metadata (not objects) via Open Archives Initiative—Protocol for Metadata Harvesting (OAI-PMH) from approximately eighteen remote hosting sites, mostly academic libraries. Those sites, in turn, support the digital collections of dozens of partners—many of which are not libraries—that lack their own infrastructure. Search results in the MWDL portal link users to the hosting site where the particular object or collection exists; this practice allows owning institutions and researchers to maintain the control and identity of their collections. The MWDL has a formalized partnership agreement and a metadata application profile.

Scientists wishing to make available their toxicology data and publications related to Wasatch Front environmental conditions could have the option of submitting to USpace or to another repository that is OAI-PMH-compliant and adheres to the MWDL Dublin Core Application Profile (Walters et al., 2011). Metadata describing the data sets would be harvested into the aggregated MWDL index hosted at the University of Utah, but the data sets and publications themselves would remain with the hosting site.

## Conclusion

The geography of the Wasatch Front compounds the effects on human health of population growth, as well as industrial and military environmental pollution. National environmental toxicology data repositories do not consider the unique local characteristics that appear to have a detrimental effect on air and soil quality and may correlate to higher incidences of asthma, autism, and other environmental health-related problems.

The creation of a local environmental toxicology data repository may help to address some of the serious health issues and may serve as a model for other regionally focused data repositories. The established political and technical infrastructures of the USpace institutional repository at the University of Utah and the regional MWDL can be leveraged to quickly establish this repository. The approach is intended to be economical and adaptable to other regional areas that are supported by an academic library. Librarians themselves may recognize this proposal as an opportunity for greater involvement in data management and to work more closely with scientists and other researchers.

## References

- ARRINGTON, L. J., & HANSEN, G. B. (1963). *The richest hole on earth: A history of the Bingham Copper Mine*. Logan, UT: Utah State University Press.
- BAIO, J. (2012). Prevalence of autism spectrum disorders—Autism and developmental disabilities monitoring network. *Morbidity and Mortality Weekly Report*, 61(3), 1-19.
- BALL, W., LEFEVRE, S., JARUP, L., & BEALE, L. (2008). Comparison of different methods for spatial analysis of cancer data in Utah. *Environmental Health Perspectives*, 116(8), 1120-1124.
- BOFFEY, P. M. (1968). Nerve gas: Dugway accident linked to Utah sheep kill. *Science*, 162(3861), 1460-1464.
- CARNES, S. A., & WATSON, A. P. (1989). Disposing of the US chemical weapons stockpile: An approaching reality. *The Journal of the American Medical Association*, 262(5), 653-659.
- COBB, G. P., SANDS, K., WATERS, M., WIXSON, B. G., & DORWARD-KING, E. (2000). Accumulation of heavy metals by vegetables grown in mine wastes. *Environmental Toxicology and Chemistry*, 19(3), 600-607.
- DOUGHERTY, J. M. (2010, March 28). No further action needed in Hill Air Force Base PCB cleanup. *Deseret News*. Retrieved September 29, 2012 from

- <http://www.deseretnews.com/article/700018954/No-further-action-needed-in-Hill-Air-Force-Base-PCB-cleanup.html?pg=all>
- DUDKA, S., & ADRIANO, D. C. (1997). Environmental impacts of metal ore mining and processing: A review. *Journal of Environment Quality*, 26(3), 590.
- Experts puzzled over some high numbers in Utah asthma report. (2007, June 11). *The Salt Lake Tribune*. Retrieved November 10, 2012 from <http://bit.ly/RP8QvN>
- FAHYS, J. (2008). EPA: U.S. Magnesium wastes endangers workers, families, birds. *Salt Lake Tribune*. Retrieved November 10, 2012 from <http://tinyurl.com/bearqwl>
- FANIEL, I. M., & ZIMMERMAN, A. (2011). Beyond the data deluge: A research agenda for large-scale data sharing and reuse. *International Journal of Digital Curation*, 6(1), 58-69.
- FEIJEN, M. (2011). What researchers want. *SURF*, 57(2), 1-32.
- FOOTE, W. (1994). The chemical demilitarization program—Will it destroy the nation's stockpile of chemical weapons by December 31, 2004? *Military Law Review*, 146, 1-93.
- GARBER, H. (2011). Air quality on the Wasatch Front is a public health crisis. *Utah Physicians for Healthy Environment*. Retrieved November 10, 2012 from <http://tinyurl.com/b5cs885>
- HENETZ, P. (2005, May 14). Feds sue plant for alleged PCB spill. *Salt Lake Tribune*. Retrieved September 29, 2012 from [http://www.sltrib.com/ci\\_2734847](http://www.sltrib.com/ci_2734847)
- JACKSON, R. E., & DWARAKANATH, V. (1999). Chlorinated decreasing solvents: Physical-chemical properties affecting aquifer contamination and remediation. *Ground Water Monitoring & Remediation*, 19(4), 102-110.
- LONG, R. W., SMITH, R., SMITH, S., EATOUGH, N. L., MANGELSON, N. F., EATOUGH, D. J., POPE, C. A., & WILSON, W. E. (2002). Sources of fine particulate material along the Wasatch Front. *Energy & Fuels*, 16(2), 282-293.
- MATSUMOTO, G. (2003). Anthrax powder: State of the art? *Science*, 302(5650), 1492-1497.
- MAY, H. (2011, March 11). Asthma: Why do some Utah kids struggle harder to breathe? *Salt Lake Tribune*. Retrieved September 30, 2012 from <http://www.sltrib.com/sltrib/news/51289555-78/healthyforwhom-asthma-utah-percent.html.csp>
- NATIONAL RESEARCH COUNCIL (U.S.) COMMITTEE ON HUMAN BIOMONITORING FOR ENVIRONMENTAL TOXICANTS (2006). *Human biomonitoring for environmental chemicals*. Washington, D.C.: National Academies Press.
- NATIONAL RESEARCH COUNCIL COMMITTEE ON TOXICITY TESTING AND ASSESSMENT OF ENVIRONMENTAL AGENTS (2007). *Toxicity testing in the 21st century: A vision and a strategy*. Washington, D.C.: National Academies Press.
- PIWOWAR, H. A., DAY, R. S., & FRIDSMA, D. B. (2007). Sharing detailed research data is associated with increased citation rate. *PLoS one*, 2(3), e308.

- POPE, C. A., HILL, R. W., & VILLEGAS, G. M. (1999). Particulate air pollution and daily mortality on Utah's Wasatch Front. *Environmental Health Perspectives*, 107(7), 567-573.
- POPE, C. A., MUHLESTEIN, J. B., MAY, H. T., RENLUND, D. G., ANDERSON, J. L., & HORNE, B. D. (2006). Ischemic heart disease events triggered by short-term exposure to fine particulate air pollution. *Circulation*, 114(23), 2443-2448.
- RESEARCH INFORMATION NETWORK & NATIONAL ENDOWMENT FOR SCIENCE TECHNOLOGY AND THE ARTS (2010). *Open to all? Case studies of openness in research*. Retrieved September 29, 2012 from <http://apo.org.au/research/open-all-case-studies-openness-research>
- STEGNER, W. (1960). Wilderness letter. Retrieved September 29, 2012 from <http://www.colorado.edu/AmStudies/lewis/west/wilderletter.pdf>
- U.S. DEPARTMENT OF JUSTICE (2001). U.S. files suit against Magnesium Corporation of America for toxic pollution near Great Salt Lake. Retrieved September 29, 2012 from <http://www.justice.gov/opa/pr/2001/January/021enrd.htm>
- U.S. ENVIRONMENTAL PROTECTION AGENCY (2012). Superfund program—Hill Air Force Base. Retrieved November 10, 2012 from <http://tinyurl.com/avk7tvd>
- UTAH DEPARTMENT OF ENVIRONMENTAL QUALITY: DIVISION OF AIR QUALITY (2012). Air pollutants: Particulate matter. Retrieved September 29, 2012 from [http://www.airquality.utah.gov/Public-Interest/about\\_pollutants/about\\_pm.htm](http://www.airquality.utah.gov/Public-Interest/about_pollutants/about_pm.htm)
- WALTERS, C., WILLEY, K., KELLER, T. W., LAMPERT, C., MATTHEWS, R., MCINTYRE, S., MOWER, A., RIDING, E., ROBERTSON, J., SOUTHWICK, S., STRACK, G., & WOOLCOTT, L. (2011). *Mountain West Digital Library Dublin core application profile*. Retrieved November 10, 2012 from <http://tinyurl.com/a488nyn>
- Wasatch Front. (2012, September 25). In *Wikipedia*. Retrieved September 29, 2012 from [http://en.wikipedia.org/wiki/Wasatch\\_Front](http://en.wikipedia.org/wiki/Wasatch_Front)



## **PART FOUR**

---

# **THE INTERNET AND NEW MODES OF SCHOLARLY COMMUNICATION**



# Notes towards the Scientific Study of Public Communication on Twitter

**Axel Bruns and Jean Burgess**

Australian Research Council Centre of Excellence for Creative Industries and Innovation, Queensland University of Technology, Brisbane

*Twitter is now well-established as an important platform for real-time public communication. Twitter research continues to lag behind these developments, with many studies remaining focused on individual case studies and utilizing home-grown, idiosyncratic, non-repeatable, and non-verifiable research methodologies. While the development of a full-blown “science of Twitter” may remain illusory, it is nonetheless necessary to move beyond such individual scholarship and toward the development of more comprehensive, transferable, and rigorous tools and methods for the study of Twitter on a large scale and in close to real time.*

## Introduction

Social media platforms such as Twitter are playing a significant role in public communication—first among private individuals and now increasingly also among media organizations, journalists, governments, and politicians in conversation or debate with their citizens, consumers, and users. Researchers working with Twitter data at various levels of scale and complexity have already generated rich insights into the use of this social media platform: for personal communication, politics, journalism, crisis communication, and so on, almost *ad infinitum* (see, for example, Crawford, 2009; boyd et al., 2010; Hermida, 2010; Marwick & boyd, 2011). Within humanities and social sci-

ence approaches to media and communication, as has been the case with the study of earlier new media technologies and forms (especially television and its audiences), we see an eclectic mix of methods and radically different scales of analysis. In this paper, we argue for the importance of transferable methods, enabling meaningful comparative work across research teams and national traditions, if not disciplines, and, it is hoped, for a more systematic coordination of multi-method approaches.

Our ability to compare the findings of Twitter research across individual case studies, in fact, is hindered by the lack of a standard set of communicative measures and metrics that may be applied in the analysis of Twitter datasets—if we are to pursue more “scientific” approaches to Twitter research grounded in humanities and social science approaches to questions of media and communication, the development of such metrics will be an important contribution. In the following sections of the paper, we provide examples from our own research of how relatively simple metrics, particularly when used comparatively, at scale and over time, can yield analytically productive insights into longstanding questions of media and communication studies: Who are the main actors engaged around a topic or event? How might we think about the communicative and / or power relations among those actors? What are the main themes or frames associated with the social media communication around a topic or event?

This work is set against the much broader backdrop of what David Berry and others have called the “computational turn”—a “third wave” of digital humanities that sees the shift from computational tools to a new computational paradigm, changing the ontologies and epistemologies of humanities research (Berry, 2012). Such a shift is represented, for example, by the work of Franco Moretti on large-scale, corpus-based literary analysis in the mid-2000s, as well as Richard Rogers’s (2009) call to employ “natively” digital methods to diagnose patterns of social change via the digital traces that can be gleaned via the Internet rather than using the Internet to carry out traditional social science or humanities enquiries—for Rogers, this is the distinction between “virtual” and (natively) “digital” methods. In what follows, we present the techniques resulting from our development of “natively” digital approaches to communication via the use of the Twitter application programming interface (API) and discuss their applications.

## Twitter Communication Metrics

The development of metrics for understanding public communication on Twitter naturally begins with a review of the data points the Twitter API

already offers, directly or indirectly. In addition to the tweet text itself, the data and metadata that the API offers to describe a single tweet include a number of other key points of interest:

- **text:** Contents of the tweet itself, in 140 characters or less
- **to\_user\_id:** Numerical ID of the tweet recipient (for @replies)
- **from\_user:** Screen name of the tweet sender
- **id:** Numerical ID of the tweet itself
- **from\_user\_id:** Numerical ID of the tweet sender
- **iso\_language\_code:** Code (e.g., en, de, fr) of the sender's default language
- **source:** Client software used to tweet (e.g., Web, Tweetdeck)
- **profile\_image\_url:** URL of the tweet sender's profile picture
- **geo\_type:** Format of the sender's geographical coordinates
- **geo\_coordinates\_0:** First element of the geographical coordinates
- **geo\_coordinates\_1:** Second element of the geographical coordinates
- **created\_at:** Tweet timestamp in human-readable format
- **time:** Tweet timestamp as a numerical Unix timestamp

Further information can be extracted from the tweets themselves. An examination of the syntax of each tweet, for example, can reveal whether it should be classified as belonging to one of the following categories of communicative activity:

- **Original tweets:** Tweets that are neither @reply nor retweet
- **Retweets:** Tweets that contain RT @user... (or similar)
  - **Unedited retweets:** Retweets that start with RT @user...
  - **Edited retweets:** Retweets that do not start with RT @user...
- **Genuine @replies:** Tweets that contain @user, but are not retweets
- **URL sharing:** Tweets that contain URLs

(Any one tweet will be either an original tweet, retweet, or @reply, but tweets from each of these categories may also contain URLs.)

Although basic, such simple approaches to categorizing tweets are already able to generate significant insights into the interaction patterns that may be observed for public communication on Twitter: Our minute-by-minute examination of the #royalwedding hashtag that accompanied the 29 April 2011 British royal wedding, for example, clearly points to key moments in the day, such as the newlyweds' first public kiss on the balcony of

Buckingham Palace, which resulted in a sharp spike in original tweets—and expressing viewers’ personal reactions to the moment—and simultaneous drops in retweeting, @replying, and link-sharing activities. Further, overall tweeting volumes also indicate the times at which major international television networks began and ended their coverage (see Bruns, 2011).

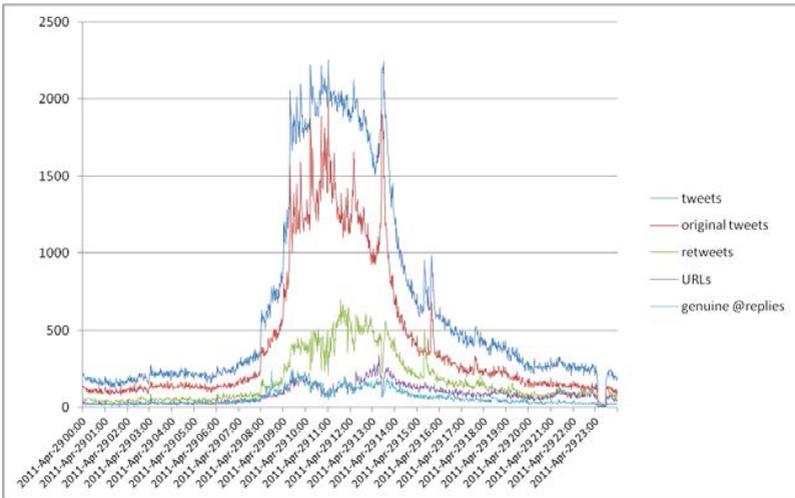


Figure 1. Minute-by-minute activity in the #royalwedding hashtag, 29 April 2011 (times in GMT)

Once they are based on such standard metrics, such analyses of individual hashtagged events may then also be usefully compared across a range of different events, to identify shared or divergent patterns between activities of the same time. Bruns and Stieglitz (forthcoming) do so for a large number of hashtag datasets, and detect clear correlations between the wider communicative context within which specific hashtags operate and the communicative patterns that may be observed within these hashtags themselves (see Figure 2).

Their analysis of some 40 different hashtag events points to two clear, divergent patterns of hashtag activity: on the one hand, hashtags that are associated with breaking, unforeseen news events and crises (#egypt, #londonriots, #qldfloods) are characterized by a substantial level both of tweets containing URLs, and of retweets; user practices here can be described as a form of gatewatching (Bruns, 2005), with users actively seeking out and



emotional responses (#headdesk)—might lead to the identification of additional types of hashtag use. Such work is only possible, of course, if standard metrics are applied to the study of such further communicative events on Twitter.

## Twitter User Metrics

In addition to the development of such metrics for the description of communicative patterns in hashtagged conversations, additional standardized measures may also be established to examine the make-up and activities of the user communities—the ad hoc publics (Bruns & Burgess, 2011)—that form around such hashtags. In the first place, it is possible to use the distinction between tweet types that we have outlined above to describe the tweeting profile of each participating user: to examine, for example, the balance between original tweets, @replies, and retweets they have sent, and to correlate this with the number of @replies and retweets they have received in turn. Such an analysis may be used, for instance, to distinguish accounts that merely retweet other users' messages, or post their own, from those that genuinely engage with others by @replying.

At hashtag level, however, such metrics may also be examined in connection with other communicative patterns. Central to such analysis is a further division of the hashtag community into its more or less active components: based, for example, on a simple division of the total contributor base for a hashtag into its leading, most active twitterers and other, less active groups, it becomes possible to determine the extent to which a small number of highly active participants dominate exchanges, and to examine differences in tweeting patterns across these groups of more or less active users. Our analysis of the well-established #auspol hashtag for the discussion of Australian politics, for example, shows that, of the more than 26,000 users who participated from February to December 2011, the most active one percent of users accounted for nearly two thirds of all tweets (the top ten percent posted more than 90% of all #auspol tweets)—and that this leading group was considerably more likely to engage in @replying than the less active user groups. For other hashtags (such as those for widely televised, world events like #royalwedding), activity patterns are vastly different—here, the lead users account for a much smaller proportion of posts, and it is the “long tail” that contributes the bulk of all tweets (see Figure 3).

Again, the comparative work that is able to extend such analyses of individual hashtags to generate a more comprehensive view of how centralized or distributed individual Twitter events are, and how this correlates with the

type of hashtag event in each case, depends crucially on the establishment of a standard set of metrics to describe these activity patterns. Such standardization does not preclude hashtag-specific analysis, or aim to privilege the development of purely quantitative aggregate figures on hashtag usage over in-depth, qualitative study; rather, it serves as a crucial enabler for further qualitative research by pinpointing those leading users, key tweets, and other exceptional patterns that are most worthy of deeper analysis.

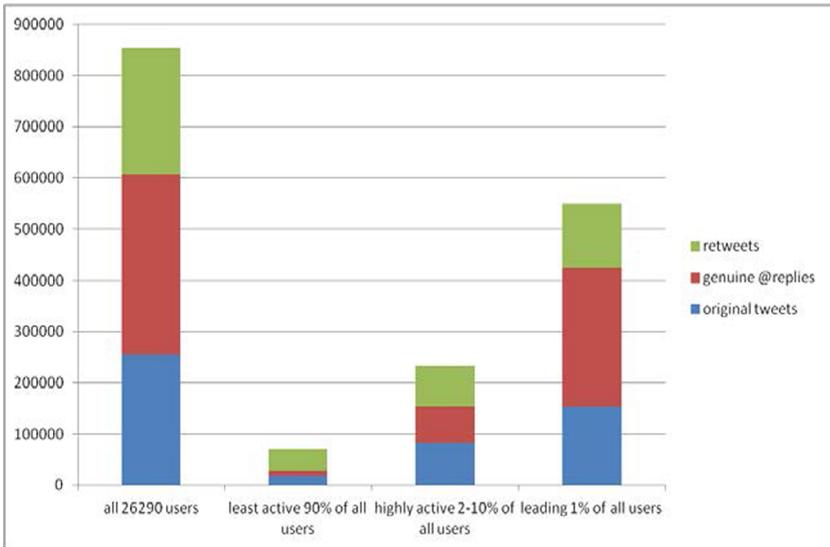


Figure 3. Contributions to #auspol made by the different groups of more or less active users (February to December 2011)

## Beyond the Hashtag

The establishment of such standardized metrics for the study of Twitter interactions through hashtags enables new forms of comparative research that detects shared patterns and practices that transcend individual hashtags themselves. However, such work does not manage to overcome the fundamental limitations associated with hashtag-based approaches themselves: these necessarily cover only the tip of a communicative iceberg, and miss out, in particular, on a substantial amount of follow-on communication as users respond to hashtagged tweets but do not themselves include the hashtag in their

@replies. More broadly, too, hashtag-based studies are appropriate only in communicative contexts where clearly established hashtags do exist—they are able to examine the particular form of political discourse that takes place in tweets carrying the #auspol hashtag, for example, but not the everyday political exchanges that take place, unhashtagged, right across the Australian Twittersphere.

Hashtag studies have been a popular tool for Twitter researchers in recent years not least because it is comparatively easy to capture a hashtag dataset, while the establishment of a representative or even comprehensive sample of general Twitter activity is considerably more difficult, especially for large populations of Twitter users (see, e.g., Bruns & Liang, 2012): the former requires researchers to track just a single keyword, using readily available tools, while the latter must build on dedicated technology to identify and follow the public tweets of a potentially very large number of Twitter users on an ongoing basis. To date, few studies of Twitter populations at this level of comprehensiveness exist; future attempts to undertake them will have to wrestle especially with the prohibitive pricing regime for high-volume data access which Twitter has now established.

To the extent that they may be successfully carried out, such studies may again utilize the standard metrics outlined above, however. User metrics may be used, for example, to examine the distribution of diverse communicative approaches across a larger population of users, and could lead to the development of a systematic typology of Twitter users as described by their activity patterns (from users who specialize in posting original tweets only through to those who engage exclusively in retweeting the messages of others); as an aside, this could potentially also be used to automatically identify spambots and similar accounts with highly unusual tweeting patterns.

Tweet metrics, on the other hand, may be used on a population-wide basis to examine common diurnal patterns of Twitter activity (for example, to examine whether @replying or link-sharing take place more frequently at specific times of the day), or to highlight particular moments of heightened activity within the dataset. Where such analysis is possible in close to real time, it may enable the automatic detection of breaking news or crisis events, for example—similar to, but substantially extending beyond, the insights that Twitter’s “trending topics” already provide. Additionally, of course, tweet metrics may also be applied to the tweets sent by specific identified subsets of the overall Twitter population whose activities are being tracked; here, they generate insights that are comparable to those arising in hashtag studies, but may be able to transcend the inherent limitations introduced by focusing only on explicitly hashtagged tweets.

Finally, a more comprehensive study of Twitter activities among an identified population of users must also take into account more strongly the established follower / followee networks of these users. While studies proceeding from an analysis of a shared hashtag may assume that participating users are connected in the first place by their shared interest in the hashtag (which enables them to see one another's tweets even if they are not following each other), a population-wide study of Twitter patterns must build on the assumption that only the followers of a given user will be likely to see the tweets posted by that user.

This further complicates the analysis of such population-wide activity patterns; at the same time, however, the baseline patterns that a longitudinal study of Twitter use may be able to establish will also serve as an important point of comparison for the analysis of shorter-term hashtag events as we have outlined it above. Hashtag-based work alone may show the total volume of tweets responding to a certain event or issue, or it may pinpoint certain users as leading contributors to the discussion; only in comparison to these baseline patterns, however, does it become possible for researchers to determine just how exceptional the hashtagged volume of tweets was, or how far from their standard patterns of interaction a user might have diverged in tweeting about a specific topic.

## **Conclusion and Reflections**

In this paper, we have catalogued some recently developed and potentially transferable methods and metrics for the study of public communication on Twitter, as a particularly prominent example of how social media platforms are remediating and transforming communication within the changing media ecology. In doing so, we have demonstrated how a range of metrics and analytical techniques that address routine research questions in media and communication studies can help to make sense of the social media “data deluge.”

However, there remain many new challenges for humanities and social science-inflected disciplines seeking to build on and extend data-driven approaches to Internet communication. Two of the most significant of these concern methodology and disciplinary practices. First, media and communication researchers need to develop (and not just outsource) the appropriate technical skill and broader “code literacy” sufficient to engage knowledgeably and critically with these methods—with broad consequences for the content and pedagogy of research training and PhD programs in particular. Second, there is much room for further development of multi-method approaches, integrating and innovating upon traditional qualitative methods

(including close textual analysis and ethnographic approaches) in a “big data” context, bearing in mind the critical “provocations” for big data recently proposed by danah boyd and Kate Crawford (2011).

Beyond the practical methodological issues raised by the burgeoning field of data-driven media and communications research lie the political and pragmatic issues arising from competing regimes of data access, usage, and control. For example, Twitter.com is effectively asserting monopoly rights on Twitter data through various technical and legal means, including the ban on Web-based export of Twitter archives (making the widely used archiving service Twapperkeeper ineffective for research purposes) and the choking off of access to its “firehose” except via prohibitively expensive commercial providers such as Gnip, or by prior arrangement (as in the “gift” of historical Twitter data to the Library of Congress). At the same time, the “open science” and “open data” movements propose a set of norms for scientific research that would ask us to make our original or processed datasets freely available for the use of our peers or the public (Rees, 2010)—creating a very complex set of problems for social science researchers who rely on third-party proprietary data such as Twitter archives.

With appropriate critical reflection, humanities and social science approaches to the “scientific” study of public communication, such as those discussed in this paper, may in fact offer a “special case” of the politics of knowledge associated with the current turn to “big data” and computational methods, because of their entanglement—even at the level of data collection—with the shifting business models of social media platforms, shifting and variable regulatory structures in relation to data access and use, and public anxieties surrounding the control and use of our social data at a moment when “personal” information and public communication are converging.

## References

- BERRY, D. (Ed.). (2012). *Understanding digital humanities*. London: Palgrave Macmillan.
- BOYD, d., GOLDBER, S., & LOTAN, G. (2010) Tweet, tweet, retweet: Conversational aspects of retweeting on Twitter. Paper presented at the 43rd Hawaii International Conference on System Sciences, Koloa, Kauai. Retrieved September 30, 2012 from <http://www.computer.org/portal/web/csdl/doi/10.1109/HICSS.2010.412>
- BOYD, d., & CRAWFORD, K. (2011, September 21). Six provocations for big data. Paper presented at A Decade in Internet Time: Symposium on the Dynamics of the Internet and Society, Oxford. Retrieved September 20, 2012 from [http://papers.ssrn.com/sol3/papers.cfm?abstract\\_id=1926431](http://papers.ssrn.com/sol3/papers.cfm?abstract_id=1926431)

- BRUNS, A. (2011, August 12). Twitter and the royal wedding, pt. 1: Something processed. Retrieved November 10, 2012 from <http://tinyurl.com/a3fbd14>
- BRUNS, A. (2005). *Gatewatching: Collaborative online news production*. New York: Peter Lang.
- BRUNS, A., & BURGESS, J. (2011, August 25-27). The use of Twitter hashtags in the formation of *ad hoc* publics. Paper presented at the European Consortium for Political Research conference, Reykjavik. Retrieved September 30, 2012 from [http://snurb.info/files/2011/The%20Use%20of%20Twitter%20Hashtags%20in%20the%20Formation%20of%20Ad%20Hoc%20Publics%20\(final\).pdf](http://snurb.info/files/2011/The%20Use%20of%20Twitter%20Hashtags%20in%20the%20Formation%20of%20Ad%20Hoc%20Publics%20(final).pdf)
- BRUNS, A., & LIANG, Y. E. (2012). Tools and methods for capturing Twitter data during natural disasters. *First Monday*, 17(4). Retrieved September 30, 2012 from <http://firstmonday.org/htbin/cgiwrap/bin/ojs/index.php/fm/article/view/3937/3193>
- BRUNS, A., & STIEGLITZ, S. (forthcoming). Quantitative approaches to comparing communication patterns on Twitter. *Journal of Technology in Human Services*.
- CRAWFORD, K. (2009). Following you: Disciplines of listening in social media. *Continuum: Journal of Media & Cultural Studies*, 23(40), 525-535.
- FULLER, M. (Ed.). (2008). *Software studies: A lexicon*. Cambridge, MA: MIT Press.
- HERMIDA, A. (2010). Twittering the news: The emergence of ambient journalism. *Journalism Practice*, 4(3), 297-308.
- MARWICK, A., & BOYD, d. (2011). I tweet honestly, I tweet passionately: Twitter users, context collapse, and the imagined audience. *New Media and Society*, 13(1), 114-133.
- REES, J. (2010, March) Recommendations for independent scholarly publication of data sets (Creative Commons Working Paper). Retrieved September 30, 2012 from <http://neurocommons.org/report/data-publication.pdf>
- ROGERS, R. (2009, May 8). The end of the virtual—digital methods. Retrieved September 30, 2012 from [http://www.govcom.org/rogers\\_oratie.pdf](http://www.govcom.org/rogers_oratie.pdf)



# Scholarly Blogging: A New Form of Publishing or Science Journalism 2.0?

Cornelius Puschmann<sup>I</sup> and Merja Mahrt<sup>II</sup>

<sup>I</sup> School of Library and Information Science, Humboldt University of Berlin,

<sup>II</sup> Department of Communication, Heinrich Heine University Düsseldorf

*This paper examines scholarly blogging as an emergent phenomenon among academics of different disciplinary backgrounds, as well as science enthusiasts and practitioners wishing to communicate about topics related to a specific academic field with a broader public. We give a brief historical account of scholarly blogging, paired with a review of academic literature about the phenomenon. Results from a survey conducted among bloggers active on scilogs.de, a German-language science blogging platform, show that considerable differences exist between conceptualizations of scholarly blogging as “publishing 2.0,” i.e., a replacement for traditional venues of scholarly communication, and blogging as a new form of science journalism. Building on this differentiation, we ask what relevance scholarly blogs have today and in the future, both from the internal perspective of science and from the external vantage points of funders, lawmakers, and civil society.*

## Scholarly Blogs: Issues of Definition

Scholarly blogs are most commonly defined as blogs written by academic experts that are dedicated in large part to scientific content. This working definition is less straightforward than it may initially seem, since neither what

an expert is (e.g., are graduate students or high school science teachers experts?) nor what constitutes scholarly content (do peer-reviewed articles, scholarly monographs, or a well-written Wikipedia entry all constitute scholarly content in one way or another?) cannot be defined unambiguously. Such issues notwithstanding, most researchers who have investigated science blogs seem to broadly follow one of two routes, defining scholarly blogs as blogs containing scholarly content or (more often) written by scholars with some kind of institutional academic affiliation (cf. Shema et al., 2012).

A brief look at the emergence of scholarly blogs in the early 2000s helps us to better understand them as an extremely heterogeneous phenomenon with a wide range of functions. An early contribution to research into scholarly blogging came from Mortensen and Walker (2002), who described blogs as tools for writing and knowledge management primarily used by PhD students. In a later publication, Walker (2006) discovered that her own usage had changed considerably over a longer period of time. While her blog had initially been an ideal place for trying out new ideas and discussing them with peers outside of strict academic hierarchies, it had turned to a public stage on which to present material that was ready for public (and peer) scrutiny. Similar observations came from Gregg (2009) in her critical assessment of the use of blogs by junior researchers. Building on Walker's typology, she characterized blogs as a subcultural form of expression favored by young academics as part of constructing a professional identity.

In contrast to the ethnographic approaches of Mortensen and Walker (2002; see also Walker, 2006), Davies and Merchant (2007), and Gregg (2009), a number of strongly content-driven studies exist. These approach science blogging less from the perspective of actors and more from the vantage point of information, text, and genre. Bar-Ilan (2005), who analyzed content-based statistics related to 15 academic blogs, came to the conclusion that their authors were interested primarily in distributing information and sparking discussion (rather than experimenting with new ideas). An equally data-centric approach was used by Luzón (2009) in her study of hyperlinks in academic blogs. She found that links are overproportionally used in academic blogs compared to private online journals.

Science blogs in the strict sense (i.e., scholarly blogs pertaining to the natural sciences, particularly physics, chemistry, and life sciences) were discussed by Bonetta (2007). Her short piece presented two popular blogs that have since been incorporated into the publisher-sponsored platforms PLoS Blogs and Nature Network. In Bonetta's characterization, the function of blogs is not to serve as a space for personal reflection and debate with peers, but as a tool to present science and scientific findings to a lay public in a

comprehensible way. Similar arguments came from Wilkins (2008), who assigned blogs (both written by scholars and by science journalists) the role of a mediator between academia and the general public:

Blogging is also a way to demythologize science. Unlike laws and sausages, the public should see science during its manufacture, but the lay public is generally ill-equipped to interpret what they see, and science bloggers play a crucial role here. (p. 411)

Wilkins thus saw blogs as a modern means of conducting science communication, rather than for articulating thoughts in progress or communicating with peers. Much of the difference in these characterizations is owed to diverging disciplinary traditions. A content analysis of 11 academic blogs by Kouper (2010) showed that frequently the claim of addressing a lay audience in the mode of science journalism is not redeemed by scholarly bloggers. Both the choice of topics and the linguistic presentation of the material are rarely suitable to complete laypersons; at least intermediary knowledge of the issues presented is necessary. Instead, scholarly blogs appear often to be read by scholars or by people with a decided interest in academic information, whether they are affiliated with an institution of higher learning or not. Scholarly blogs appear to also serve a function comparable to a “virtual water cooler” (Kouper, 2010) around which experts share and debate context-specific information in a more or less informal manner. This seems hardly compatible with the assumption that blogs should follow the lead of science journalism in catering to the general public with the mission of educating it about science or providing a means of critical evaluation or public control of scientific work and practices. Yet for many bloggers, presenting and discussing the results of scholarly research with the public is of at least some significance (Colson, 2011).

The most significant contribution on the motivations of blogging academics to date comes from Kjellberg (2010). In her qualitative assessment of Swedish, Danish, and Dutch researchers of different scholarly disciplines, she highlighted the complementary function of blogs for the distribution of content and personal knowledge management. According to Kjellberg’s subjects, an important feature of blogs is that they allow publishing spontaneously and without rigorous stylistic and formal constraints or the requirements of editors and publishers (see also Davies & Merchant, 2007). Bloggers (academics and non-academics alike) carefully consider their audience and make stylistic and thematic choices according to the assumed makeup of their readership. But because it is never truly possible to know who is in the audience, a degree of uncertainty remains about the appropriateness of these choices.

The level of acceptance of scholarly blogs varies significantly from one disciplinary and cultural context to another. The French platform *hypotheses.org* is an example of a successful attempt to establish blogs as an institutionally recognized element of scholarly communication. It hosts over 200 *carnets de recherches* (research notebooks), which undergo formal peer review before being admitted to the platform. In contrast to the alternative conceptualization of scholarly writing implicit in many blogs—freer, less constrained by tradition and convention, less elitist than traditional publishing in journals and monographs—blogs are integrated into the entrenched ecosystem of scientific communication in this approach. They are, in other words, adapted to the needs of scholars, rather than being seen as instruments of change to overcome the status quo in academic communication.

## Motives of Scholarly Blog Readers

Research on the readers of scholarly blogs is still emerging at this point. Identifying who regularly reads academic blogs is largely speculative. Representative data on the use of Web 2.0 material suggest that in the U.S., for instance, about a third of the population reads blogs (Pew Research Center, 2010), while in other countries, the proportion may be much smaller: for example, less than 10% in Germany (Busemann & Gscheidle, 2011). Results for the use of different genres of blogs are missing, and it is difficult to categorize blogs along the genre paradigm.

Internationally, there is exemplary research on the reach and impact of scholarly blogs. For China, existing studies of the academic blogosphere point to small networks with strong reciprocal relationships between bloggers (Wang et al., 2010). As in other countries, blogs in the context of scholarly information do not appear to be widely read, although some academic blogs figure among the so-called A-list blogs that garner a million visits or more per month (Batts et al., 2008). A non-representative survey of blog readers by Yu (2007) provided at least tentative indicators for the use of scholarly blogs. Science and education attract a medium level of interest, markedly below the level of interest for entertainment. A central motive of reading blogs, however, is seeking information. Following the uses and gratifications approach, Kaye (2005, 2010) analyzed the motives of blog readers systematically. Her research pointed in a similar direction: Easy access to a wide range of information is the central motivation of readers. Although her surveys likely included regular readers of scholarly blogs, their exact motives in relationship to other types of content published in blogs remain unclear.

Recently, Littek (2012) conducted a survey among readers of two German-language academic blogging platforms. She distinguished between readers with an academic background, science journalists, and laypersons. All three groups appreciate academic blogs as an informative, but also informal and sometimes entertaining, format. Laypersons also ascribe high quality to the information provided by bloggers. Science journalists are a little more critical of this aspect and do not see blogs as a replacement for journalistic coverage of research. Academic readers, on the other hand, think that blogs can provide more accuracy and higher quality than science journalism.

Access to specific information from a trustworthy source is thus an important motive for different kinds of readers of blogs. The diverging viewpoints about the strengths and weaknesses of blogs among different groups of readers have implications for the approach of bloggers, namely whether they knowingly or unknowingly follow certain demands or prioritize a given objective over others.

## **Motives of Scholarly Bloggers**

While far from complete, more research has been carried out on the motives of scholarly bloggers than on those of readers. Following up on Kjellberg's qualitative approach, we conducted a Web-based survey of scholarly bloggers active on the German platform scilogs.de from May 7 to June 3, 2012. SciLogs is run by commercial popular science publisher Spektrum der Wissenschaft and hosts over 60 blogs in total. Users were recruited via a call for participation from the platform management published in the platform's internal newsgroup and via e-mail. Reminders were sent two weeks after the start of the survey and three days before its end. We received responses from 44 authors, providing us with a fairly large sample of the platform's active bloggers. Bloggers answered mostly standardized questions on their blogging habits and histories, their academic backgrounds, and their opinions about academic blogging. Opinions were recorded through statements that respondents rated on a gradable scale (e.g., from "strongly agree" to "strongly disagree"). A few demographic variables were also obtained, but the questionnaire was carefully designed to assure anonymity to the greatest possible extent, even among such a small group of people.

The majority of respondents were either between 30 and 39, or 40 and 49, and a large portion of participants were male (73%), while only few female bloggers were represented in our sample (23%, with 4% declining to specify gender). SciLogs has a marked bias toward the natural sciences, with 59% of respondents reporting to be from that area. Of the respondents, 20% came

from the humanities, while 7% hailed from the social sciences, and 5% associated themselves with life sciences, engineering, or a combination of fields. Sixty percent reported to have blogged for over two years, and 50% reported writing for another blog in addition to their SciLog, most often with a focus on similar scientific themes.

SciLogs, in contrast for instance to the French platform hypotheses.org, is not exclusively a site for full-time academics with permanent work contracts at publicly funded universities or research institutes. Forty-three percent of the participants reported being employed in an academic position, while smaller percentages associated themselves with journalism, PR, or described themselves as self-employed.

An equally diverse picture emerged with regard to the SciLoggers' academic career status. Forty-five percent of the respondents reported having a PhD, but only 2% hold the *Habilitation*, a postgraduate degree that was in the past formally required to be eligible for a tenured professorship in Germany and that is still widely regarded as the hallmark of an academic career. Forty-three percent say the highest position that they have achieved in their academic career is that of a *Mitarbeiter*, a usually non-tenured research or teaching position that is generally held prior to achieving the status of *Professor*. Fourteen percent of respondents have achieved a permanent position as *Mitarbeiter*, while 9% have achieved some level of professorship (assistant, associate, or full). Many SciLoggers either have not yet decided to pursue an exclusively academic career, have recently taken up this aim, or do not / no longer seek such a career.

The diversity of the SciLogs community is also reflected in the bloggers' aims and in the views they have of their readers. Over 60% of respondents see presenting their field of research to a general audience as an important goal of their blog, while about half see establishing a thematic presence online as important. Considerably fewer bloggers (35%) want to bring grievances or controversies to the public's attention or express themselves creatively (30%; multiple goals could be selected).

The respondents do not see their blogs as appropriate outlets for original research. Only a fourth of those surveyed want to present results of their work in their blog, while over 50% regard this aim as not relevant at all to their blogging. While over 60% see discussion and the exchange of ideas as pivotal to their blog, publishing texts or essays written in other contexts is a potential use of the blog to only 15% of users. The alternative communicative aims, differences in envisaged readership, and divergent genre associations of the bloggers in relationship to traditional scholarly communication play out in full force in relationship to the strategic goals of the respondents. Over 80%

state that advancing their own career inside the institutional academic system was not a relevant factor in their decision to take up blogging. Answering questions about science and research is a relevant motive for 30% of those polled, while repaying a debt to society plays a role for 35%. Interestingly, the responses concentrated on the respective ends of the scale—either bloggers see the societal function of blogs as fairly important or as not important at all, but very few respondents were undecided on this question. Thirty-five percent of the respondents blog because they enjoy controversies, highlighting the function of blogs as places of debate and opinion rather than neutrality and impartiality. Only a small portion is interested in documenting a specific phase of their research or project through their blog, a marked contrast to the approach to blogging taken by the scholars that Walker (2006), Kouper (2010), and Kjellberg (2010) described. Whereas these early scholarly blogs realized functions aimed at the author (learning to write, finding one's voice, reflecting problems, documenting research), the surveyed SciLoggers clearly see themselves as communicating with a wider audience. They aim to educate a general readership about broad scientific issues, not to use their blogs as a sort of virtual notebook, or to debate a specific aspect of their research with a small network of colleagues. For 80%, the public at large is the main audience, followed by people with an interest in the blogger's area of expertise. Colleagues are somewhat less important (44%), as are students (42%). Funding bodies and decision-makers at institutions and companies are considerably less relevant target audiences (9% and 13%, respectively; respondents could indicate multiple target groups), emphasizing the conceptualization of blogging as a public activity. For a majority (80%), the motives they have for blogging have not changed over time.

The strong emphasis on public communication in a privileged, yet alternative, communicative arena (outside the lecture hall, yet with a clear claim to authority and expert status), paired with the tendency to enjoy controversy, aligns itself with the socio-demographics of the bloggers, who are predominantly male, middle-aged, well-educated, and at least in part still in the process of establishing themselves academically. Our findings raise the question of whether a platform with more diversity in relationship to age, gender, and academic seniority would produce different styles of scholarly blogging.

The surveyed SciLoggers see blogging as a strongly interactive phenomenon that transcends the much-lamented ivory tower of scholarship, and they value debate accordingly. Over 80% find that commentators seek informed debate in their comments. Seventy-five percent find that commentators ask questions, while very few of them point out mistakes or criticize the blogger in ways that he / she finds overwrought. Only a small percentage find com-

ments to frequently be very negative or aggressive, although many respondents indicate they have encountered some negative comments. Twenty percent indicate that they have never deleted a comment, while 36% do it very rarely. Only 5% say they frequently delete comments.

Another contrast to the primary conceptualization of the scholarly blog as a digital lab notebook becomes evident when polling bloggers about their main themes. Twelve percent say they blog primarily about their own work, while 36% blog about research from their own field but conducted by others. Thirty-four percent do both. Surprisingly, 18% indicate that they write about neither, allowing the assumption that they write about scientific issues, but not based on current research results. An equal number of respondents say that a topic for a post comes to their attention based on a scholarly publication (19%) rather than a story in the mainstream media (19%), indicating a split between these points of departure inside SciLogs. Fewer bloggers want to broadly comment on a topic of interest (15%), correct something they have read (14%), or discuss their current research (9%). The last point is the least important motive for writing a post, even less frequent than blogging because someone asked the blogger to discuss a particular issue.

Overall, the SciLog authors have a fairly critical view of mainstream (science) journalism. The number of respondents who have at some point criticized journalists in their blog compared to the number of those who have never done so is roughly equal. Few respondents feel that their own research is presented inaccurately by journalists, but over 60% feel that this is the case sometimes or frequently with the research of others from their field, and they feel that journalists report scientific issues in a sensationalist fashion (57%).

Unsurprisingly, SciLoggers are also avid blog readers, with 98% reporting that they also read other blogs, although their enthusiasm for other forms of informal science communication is low. Over 50% report reading either [scilogs.de](http://scilogs.de) or [scienceblogs.de](http://scienceblogs.de), pointing to a considerable language bias toward German-language blog sources. The language split is noteworthy especially because scholarly publishing in the natural sciences is predominately conducted in English, creating a language barrier between the results of scientific research and the general population, in addition to the considerable background knowledge necessary to contextualize complex scientific problems.

The majority of respondents would like to see a wider uptake of open scholarly communication, along the lines of their own efforts. Over 50% strongly agree that scientists should communicate more with audiences outside institutional academia, and over 80% agree somewhat or fully with this statement. Seventy-five percent believe that tenure processes should take public science communication more strongly into account. Interestingly, a

majority does not see scholarly blogs as a replacement for science journalism. Twenty-three percent are undecided about this statement, while only 3% agree to it, highlighting the niche role of blogging—not a replacement for science journalism, but also not comparable to traditional academic publishing.

## **Conclusions: Waiting for the Big Picture to Emerge**

Blogs are actively used in a variety of scholarly contexts by academic communities around the globe. They are used by individual academics to document their research, discuss ideas with peers, educate and communicate with a wider audience beyond their immediate work context, and promote their research and often themselves before a wider public. Blogs are also used by science organizations, journalists, and enthusiasts, who often have in-depth academic training, to communicate about scholarly issues.

The use of blogs by journalists, science organizations, and enthusiasts undoubtedly has great potential for furthering the public's understanding of science and for fostering excitement and support for scientific issues. What is unresolved is the question of which aspects of traditional scholarly publishing blogs will be able to replace. A number of problems have so far prevented blogs from achieving success, and it is unclear what role, if any, they should play in this area in the future.

Our survey of the SciLogs authors highlights the diversity of the scholarly blogging community and how the actors, norms, and conventions on one platform may differ from those on another. Whereas *hypotheses.org* aims to transplant traditional institutionalized scholarship into blogs, SciLogs strives to open up a new space in which scientific issues are presented and debated by an interested public. This is done without the elitism that underpins institutionalized academia, but such a “revolutionary” approach carries the consequence that this kind of scholarly blogging has little impact on the entrenched system of scientific communication. Unless they seek debate (and sometimes controversy) and enjoy educating (or, negatively put, lecturing) a lay audience, career scientists have little incentive to take up blogging.

From the vantage point of policy-makers, this may well be an issue worth addressing in the future, since science is under constant pressure from the emancipated public to become more transparent and accountable. Scientists themselves have very little reason to support change: Either they are too junior to experiment with new and untested formats of scholarly publishing, or they are senior and have a stake in the existing publishing ecology. Civil society also has little leverage. While there is an increasing demand to change

things, scholars have little reason to care, because the general public has very scarce immediate influence on how science publishing is done.

Platforms such as SciLogs highlight the need for timely information about scientific issues, presented by experts, in a language, format, and discursive space that enable a lay audience to participate. Beyond merely presenting scientific issues, they are opened up by the SciLoggers for debate, making them (at least potentially) the subject of a broader social consensus that is politically significant in relationship to controversial issues such as climate change or the use of nuclear energy. Yet it is questionable whether institutionalized scholarship is willing or able to engage with the broader public in this fashion, given that its established genres of communication have a strong inward orientation; in other words, they contribute to discourse inside the academy, but not beyond it. Blogging as a paradigmatically new form of scholarly communication may well fail to penetrate the walls of the ivory tower, notwithstanding its partial success outside its confines.

## Acknowledgements

This research was supported by the Düsseldorf University Strategic Research Fund. We would like to express our sincere gratitude to the SciLogs community and Spektrum der Wissenschaft's editor-in-chief, Dr. Carsten Könneker, for his extensive support and feedback.

## References

- BAR-ILAN, J. (2005). An outsider's view on "topic-oriented" blogging. In D. SATISH, & K. R. PRABHAKAR (Eds.), *Blogs—Emerging communication media* (pp. 113-126). Hyderabad, India: ICFAI University Press.
- BATTS, S. A., ANTHIS, N. J., & SMITH, T. C. (2008). Advancing science through conversations: Bridging the gap between blogs and the academy. *PLoS Biology*, 6(9), e240.
- BONETTA, L. (2007). Scientists enter the blogosphere. *Cell*, 129(3), 443-445.
- BUSEMANN, K., & GSCHIEDLE, C. (2011). Web 2.0: Aktive Mitwirkung verbleibt auf niedrigem Niveau. *Media Perspektiven*, (7-8), 360-369.
- COLSON, V. (2011). Science blogs as competing channels for the dissemination of science news. *Journalism*, 12(7), 889-902.
- DAVIES, J., & MERCHANT, G. (2007). Looking from the inside out: Academic blogging as new literacy. In M. KNOBEL, & C. LANKSHEAR (Eds.), *A new literacies sampler* (pp. 167-197). New York: Peter Lang.

- GREGG, M. (2009). Banal Bohemia. Blogging from the ivory tower hot-desk. *Convergence: The International Journal of Research into New Media Technologies*, 15(4), 470-483.
- Hypotheses—Platform for academic blogs in the humanities and social sciences. (n. d.). Cléo, France: Centre for Open Electronic Publishing. Retrieved October 1, 2012 from <http://hypotheses.org/>
- KAYE, B. K. (2005). It's a blog, blog, blog world: Users and uses of weblogs. *Atlantic Journal of Communication*, 13(2), 73-95.
- KAYE, B. K. (2010). Going to the blogs: Toward the development of a uses and gratifications measurement scale for blogs. *Atlantic Journal of Communication*, 18(4), 194-210.
- KJELLBERG, S. (2010). I am a blogging researcher: Motivations for blogging in a scholarly context. *First Monday*, 15(8). Retrieved October 1, 2012, from <http://firstmonday.org/htbin/cgiwrap/bin/ojs/index.php/fm/article/view/2962/2580>
- KOUPER, I. (2010). Science blogs and public engagement with science: Practices, challenges, and opportunities. *Journal of Science Communication*, 9(1), 1-10.
- LITTEK, M. (2012). *Wissenschaftskommunikation im Web 2.0*. Frankfurt am Main: Peter Lang.
- LUZÓN, M. J. (2009). Scholarly hyperwriting: The function of links in academic weblogs. *Journal of the American Society for Information Science and Technology*, 60(1), 75-89.
- MORTENSEN, T., & WALKER, J. (2002). Blogging thoughts: Personal publication as an online research tool. In A. MORRISON (Ed.), *Researching ICTs in context* (pp. 249-279). Oslo: InterMedia.
- PEW RESEARCH CENTER (2010). Generations 2010 (Pew Internet and American Life Project). Retrieved November 2, 2012 from <http://tinyurl.com/cxdebk7>
- SHEMA, H., BAR-ILAN, J., & THELWALL, M. (2012). Self-citation of bloggers in the science blogosphere. In A. TOKAR, M. BEURSKENS, S. KEUNEKE, M. MAHRT, I. PETERS, C. PUSCHMANN, T. VAN TREECK, & K. WELLER (Eds.), *Science and the Internet* (pp. 183-192). Düsseldorf: Düsseldorf University Press.
- WALKER, J. (2006). Blogging from inside the ivory tower. In A. BRUNS, & J. JACOBS (Eds.), *Uses of blogs* (pp. 127-138). New York: Peter Lang.
- WANG, X., JIANG, T., & MA, F. (2010). Blog-supported scientific communication: An exploratory analysis based on social hyperlinks in a Chinese blog community. *Journal of Information Science*, 36(6), 690-704.
- WILKINS, J. S. (2008). The roles, reasons and restrictions of science blogs. *Trends in Ecology & Evolution*, 23(8), 411-413.
- YU, H. (2007). *Exploring the Chinese blogosphere: The motivations of blog authors and readers* (master's thesis). National University of Singapore. Retrieved November 2, 2012 from <http://tinyurl.com/cytd32h>



# Self-Citation of Bloggers in the Science Blogosphere

Hadas Shema<sup>I</sup>, Judit Bar-Ilan<sup>I</sup>, and Mike Thelwall<sup>II</sup>

<sup>I</sup> Department of Information Science, Bar-Ilan University, Israel, <sup>II</sup> Statistical Cybermetrics Research Group, University of Wolverhampton

*In the age of social media, scientists are no longer limited to traditional forms of scholarly communication and dissemination of research. Social networks, blogs, and micro-blogs allow scientists to promote their work and gain recognition among an audience wider than their peers alone. We researched four categories of blog posts from the blog aggregator ResearchBlogging.org (RB), which aggregates peer-reviewed research, to determine whether RB bloggers cite their own research, and to what extent. We found that the population of self-citers in our sample is very homogenous: The average self-citer is male, has earned a PhD, and is affiliated with a university or a research institute.*

## Introduction

Scholarly discourse has existed for hundreds of years prior to the World Wide Web. However, the Web in general and social media in particular have given rise to faster, more transparent forms of communication between researchers. Wikis, blogs, micro-blogs, and social networking sites are all being used for scholarly discourse. Unfortunately, these new methods of communication have not yet been thoroughly researched. In this article we focus on the study of research blogs, because these allow extended informal discus-

sions about academic research and hence may shed light on how science is communicated and evaluated.

Despite the informal nature of blogs, Kjellberg (2010) found that science bloggers emphasized the importance of citing blog sources in ways similar to those in formal academic discourse. Researchblogging.org, an aggregator of science blogs, aggregates posts with citations to peer-reviewed research using HTML code that creates a structured academic citation. Research bloggers who cite peer-reviewed research can register with the aggregator, and when they mark relevant posts in their blog, these posts appear on the aggregator's site. The site's editors ensure that posts follow the guidelines and are of appropriate quality.

In our previous investigation into RB bloggers (Shema et al., 2012), we studied 135 bloggers of 126 blogs who had at least twenty entries posted on the RB aggregator between January 1, 2010 and January 15, 2011. We found them to be highly educated, with 32% having obtained a PhD and 27% being graduate students. Fifty-nine percent of the bloggers were affiliated with a university or a research institute. Bloggers cited research from high-impact multidisciplinary and niche journals, such as *Nature* and *The New England Journal of Medicine*. Groth and Gurney (2010) found a similar preference for high-impact journals in chemistry posts aggregated by RB.

The current investigation used RB to study another aspect of research blogging: self-citation. Our goal was to determine to what extent bloggers cited their own peer-reviewed research, and what kinds of bloggers cite their own peer-reviewed research.

## Related Research

Self-citation is a common, well-known phenomenon in scientific literature. Aksnes (2003) analyzed over 45,000 articles from the National Citation Report (NCR) for Norway between the years 1981-1996, as well as citations to these articles until the year 2000. Over 70% of the articles cited at least once received one or more self-citations. Articles cited less than five times had a larger share of self-citations than those cited 46-50 times (29.9% and 19.4%, respectively).

Aksnes noted that the decrease in the share of self-citations was inevitable, since an article can only be cited a limited number of times by its author(s). The number of self-citations increased with the number of authors. Articles with ten authors had 6.7 self-citations on average, while articles with one author had only 1.5. Discipline-wise, Aksnes found clinical medicine to

have the lowest level of self-citation (17%) and the highest level to be in chemistry and astrophysics (31% each).

A study by Fowler and Aksnes (2007) analyzed further the Norwegian NCR (years 1981-2000) but took a different approach than Aksnes' 2003 study. Fowler and Aksnes used an author-oriented, rather than publication-oriented, methodology for over 19,000 Norwegian researchers. This change in methodology lowered the self-citation percentage to an average of 11% in comparison with 21% in Aksnes' previous study. According to Fowler and Aksnes, a self-citation yields, on average, 3.65 citations from other authors in ten years. In addition, they showed that very productive authors tend to cite themselves more often than less productive ones. Both Fowler and Aksnes and Costas et al. (2010), who had similar findings, suggested that this might be explained by productive authors having a larger pool of potential articles to cite, as well as more present opportunities to refer to past articles.

Ohm (2007) wrote about his experience as a guest-blogger in a popular law blog, *The Volokh Conspiracy*. He blogged for a week about two of his articles, which were freely available at the Social Science Research Network (SSRN) repository. SSRN presents for each article the number of abstract views, article downloads, and download rank. In 2.25 days since the Volokh publication, the number of views for both of Ohm's articles doubled, and the number of downloads increased by 74% for the first article and 63% for the second. After 2.25 days, Slashdot, one of the most popular technology sites, linked to Ohm's post. In 21 hours, the number of views for both articles doubled, and the number of downloads increased 137% for the first article and 142% for the second.

When we visited Ohm's list of articles in SSRN, we found that his article "Broken Promises of Privacy: Responding to the Surprising Failure of Anonymization" has been downloaded almost nine thousand times (for comparison, the next article in Ohm's list has been downloaded about 2,500 times; data was collected in July 2012). A Web search found the article was covered in a New York Times blog called Bits (2010, July 27). However, we could not determine whether it was the article coverage that led to the relatively large number of downloads, or if the article was found worthy for coverage and downloads because of its quality and / or content.

Similarly, Terras (2012) blogged and tweeted about each of her refereed articles that was available from her university repository (she had more than two thousand Twitter followers at that time). After noticing that the number of downloads went up sharply after each post or tweet, she decided to promote three articles about the same research project, but not mention a fourth

project article. The promoted articles were downloaded at least eleven times more than the non-promoted one.

While these case studies cannot be generalized, they nevertheless show that self-citing of articles using social media has the potential to increase the impact of scientific articles. In this study, our objective was to learn more about self-citation of refereed articles in blogs.

## Methods

We studied four categories of RB: 1) computer science / engineering, 2) ecology / conservation, 3) philosophy, and 4) mathematics to determine how many bloggers cited their own peer-reviewed research in posts. The first three categories were samples from January 1, 2010 to April 30, 2012. The mathematics category, due to its small size, was studied in full (i.e., from October 11, 2007 to April 30, 2012).

We manually extracted self-citations from the four categories. Then we removed blogs and posts that were unreachable, either because of a broken link or because they became invitation-only / private; duplicate posts were removed as well. We searched each post for the name of the author. Posts with no known authors were also removed from the sample. We classified as “anonymous” commercial or academic society blogs whose posts had no individual by-lines (e.g., Aurametrix, Sage Insight).

Bloggers who wrote in more than one blog were counted only once, and only bloggers who wrote posts included in the sample were considered as authors (if a blog had other authors, we did not take them into account). Blogs with anonymous authors were not included in the overall number of blogs. There were 304 blogs in the four categories (each blog was counted once, regardless of how many categories it appeared in), of which fifty (16%) were anonymous. The RB system allows an unlimited number of tags, so one post can appear in several categories. In such cases, posts were considered for each category separately (e.g., a post with both mathematics and philosophy tags was counted once in each category). Bloggers who self-cited themselves in more than one category were also counted for each category separately.

We collected personal information about the bloggers (name and gender) by searching their blogs’ RB profiles. Since some of the authors are not individually registered in those profiles, we checked individual by-lines of every post, as well as the “About” and “Profile” parts of every blog. In the case of self-citers, we used both their blogs and their university homepages (if they had them) and sometimes LinkedIn profiles to determine their levels of education.

## Results

We studied the shares of self-citing bloggers and posts in each discipline, as well as the bloggers' gender balance and their levels of education. Table 1 shows the number of valid posts per category, as well as the number of self-citing posts and their percentage out of the total number of posts in each category. The largest number of self-citing posts, in absolute numbers, was in computer science (35), while the largest percentage of self-citing posts was in mathematics (10%). The ecology category had, in absolute numbers, the second-largest group of self-citing posts (31), but they constituted just 5% of the category. Using a test for differences in proportions, the proportion for ecology was significantly different from the other three proportions ( $p < 0.05$ ), but the other three were not significantly different from each other.

Table 1. Posts and self-citing posts in the four categories

Category	Posts	Self-citing posts
ecology	612	31 (5%)
computer science	407	35 (9%)
philosophy	182	17 (9%)
mathematics	173	18 (10%)

Table 2 shows the number of non-anonymous bloggers, as well as the number of self-citers and their percentage out of the overall number in each category. The number of posts referring exclusively to self-cited articles could not easily be determined, because bloggers sometimes cited their articles in a scholarly structure (so it could be automatically identified by RB) but gave only a URL or a simple mention to other peer-reviewed material discussed. Computer science had the largest number of self-citers as well as the highest percentage. The mathematics category had the lowest number and percentage of self-citers, though its percentage of self-citing posts (see Table 1) was the highest. Only the difference between computer science and mathematics was statistically significant ( $p < 0.05$ ).

Table 2. Bloggers and self-citers in the four categories

Category	Bloggers	Self-citers
ecology	132	17 (13%)
computer science	93	19 (20%)

philosophy	64	11 (17%)
mathematics	89	8 (9%)

## Education

The majority of the self-citers in the sample had a PhD (36; 80%); five (11%) had a master's degree, three (7%) were graduate students, and one (2%) was an M.D. (see Figure 1). Thirty-seven out of the forty-five self-citers (82%) were affiliated with a university or a research institute. Previous research (Shema et al., 2012) showed a wider variety in levels of education among RB bloggers: 27% were graduate students and only 32% had a PhD. They also had a lower level of affiliation (59%) with academic institutes.

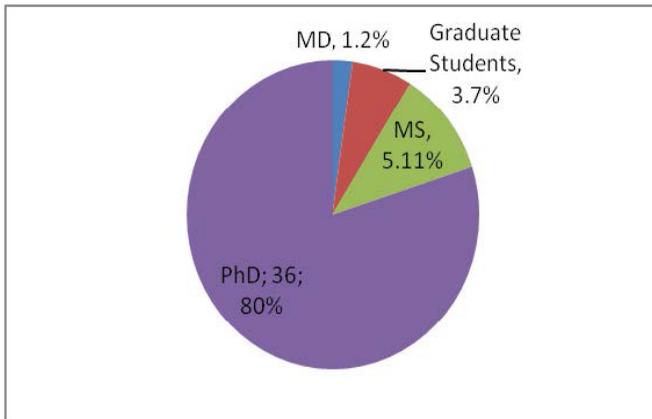


Figure 1. Distribution of self-citing bloggers by level of education

## Gender

In our previous study of science blogs, we found that over 70% of the blogs were written by one or two male authors (see Figure 2). Only 22% of the blogs were written or co-written by female authors (Shema et al., 2012).

Table 3 shows the gender disparities in self-citing. The percentages of male and female bloggers were calculated out of the overall number of bloggers, while the percentages in the female citers' column were calculated out of the number of female bloggers. Considerably fewer women than men cited themselves, even when their low percentage in the general blogger population was taken into account. The difference was statistically significant in the

ecology and computer categories ( $p < 0.05$ ). When we removed duplicates (bloggers who self-cited in more than one category), we found forty-five self-citers in all categories, of them thirty-nine (87%) male and six female (13%).

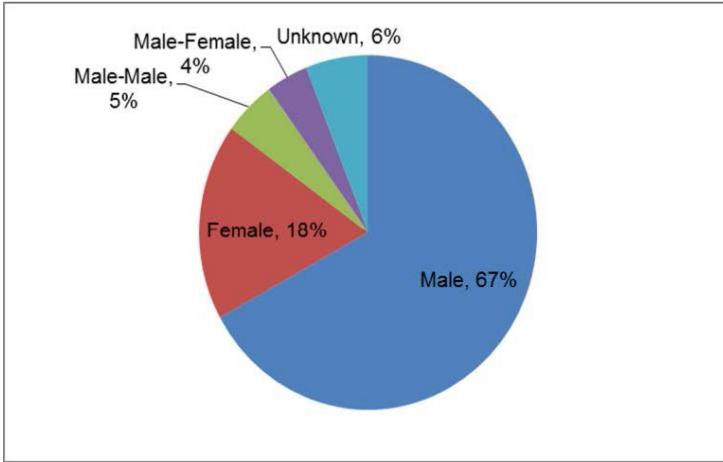


Figure 2. Distribution of gender among bloggers (Shema et al., 2012)

Table 3. Gender disparities among known bloggers and self-citers

Category	Known bloggers (male)	Known bloggers (female)	Male citers	Female citers
ecology	85 (64%)	47 (36%)	14	3 (6%)
computer science	70 (75%)	23 (25%)	18	1 (4%)
philosophy	48 (76%)	15 (24%)	9	2 (13%)
mathematics	74 (83%)	15 (17%)	8	0 (0%)

These gender disparities of the bloggers were in line with those found in studies of Wikipedia contributors, and Wikipedia editing is a little like science blogging in the sense that both are discussing knowledge. Glott et al. (2010) found that around 13% of the contributors to Wikipedia were women. Lam et al. (2011) found that the initial percentage of women contributors in their sample was about 16%, but dropped to around 6% for contributors who had made more than five hundred edits.

## Limitations

Our study had several limitations. Unlike the well-documented references in journals, blog citations are transient, and links obsolesce with time. For example, blogs move to a blog network or leave it, become invitation-only, or disappear from the Web altogether. RB blog posts keep accumulating over time, and new self-citations along with them. Hence our blogs, bloggers, and the number of self-citation may have changed since they were assessed. Moreover, we assumed the bloggers' RB profiles and other information sources to be authentic and up-to-date, but this might not be always the case.

The sample was limited to only four categories and to non-anonymous RB bloggers and posts. Our characterization might therefore only be true for the sample categories and RB blogs, rather than for the general science blogs population. It could be that the large differences between the number of posts in the sample showed biases towards disciplines in which RB is well known and towards bloggers that promote their blogs by submitting them to RB.

## Discussion

In this study we focused on the question of self-citation in blogs. The self-citing bloggers were a highly homogenous group: male (87%), having a PhD (80%), and affiliated with a university or a research institute (82%). Men cited themselves significantly more in ecology and computer science. The rate of self-citing posts was low overall but varied according to discipline, with mathematics having the highest percentage of self-citing posts (10%), computer science and philosophy having a slightly lower percentage (9%), and ecology having the lowest (5%). Only the ecology category had a significantly different proportion of self-citations. The percentages of self-citers were higher: 20% in computer science, 17% in philosophy, 13% in ecology, and 9% in mathematics. Only the difference between computer science and mathematics was statistically significant.

It is important to note that self-citations in blogs are very different than those in scientific discourse. In the academic world, formal publishing is a necessity ("publish or perish"), while blogs are more of an extra-curricular activity. Refereed articles with multiple authors are more likely to be cited, since most, if not all of those authors will continue to publish in the same area. However, these authors are not likely to all have blogs. Even authors who are science bloggers can blog without referring to their own research, while academic publications often build on the authors' previous work. The

bloggers have the freedom to post anonymously, while academic authors do not.

A science blogger does not have to be a published author in a peer-reviewed journal. Bloggers who are undergraduates, professional science writers, and so forth might not have peer-reviewed publications to cite.

We suggest that bloggers who have earned a PhD and are affiliated with a research institute are likely to have authored more refereed publications than those who did not and therefore have more of them to cite. This is in line with Fowler and Aksnes' (2007) and Costas et al.'s (2010) findings about the positive correlation between productivity and self-citing.

With the increased emphasis of societal impact of research, it is quite plausible that in the future, more and more scientists will be blogging, tweeting, and depositing in open access repositories, to gain the attention of larger audiences.

## Acknowledgement

This study was supported by the FP-7 EU funded project ACUMEN on assessing Web indicators on research evaluation.

## References

- AKSNES, D. W. (2003). A macro study of self-citation. *Scientometrics*, 56(2), 235-246.
- COSTAS, R., LEEUWEN, T. N., & BORDONS, M. (2010). Self-citations at the meso and individual levels: effects of different calculation methods. *Scientometrics*, 82(3), 517-537.
- FOWLER, J., & ASKNES, D. W. (2007). Does self-citation pay? *Scientometrics*, 72(3), 427-437.
- GLOTT, R., SCHMIDT, P., & GHOSH R. (2010). Wikipedia survey—overview of results (Report). Retrieved November 10, 2012 from <http://tinyurl.com/66zazh5>
- GROTH, P., & GURNEY, T. (2010). Studying scientific discourse on the Web using bibliometrics: A chemistry blogging case study. In *Proceedings of the WebSci10: Extending the frontiers of society on-line, April 26-27th, 2010, Raleigh, NC*. Retrieved November 10, 2012 from <http://tinyurl.com/ao9h368>
- KJELLBERG, S. (2010). I am a blogging researcher: Motivations for blogging in a scholarly context. *First Monday*, 15(8), Retrieved October 2, 2012, from <http://firstmonday.org/htbin/cgiwrap/bin/ojs/index.php/fm/article/view/2962/2580>

- LAM, S. K., ANURADHA, U., ZHENHUA, D., SHILAD, S., MUSICANT, D. R., LOREN, T., & JOHN, R. (2011, October 3-5). WP: clubhouse? An exploration of Wikipedia's gender imbalance. Paper presented at WikiSym'11, Mountain View, California. Retrieved October 2, 2012 from <http://grouplens.org/system/files/wp-gender-wikisym2011.pdf>
- OHM, P. (2007, April 14). Do blogs influence SSRN downloads? Empirically testing the Volokh and Slashdot effects (Research paper Nr. 07-15). University of Colorado. Retrieved November 10, 2012 from <http://tinyurl.com/b4w2xry>
- Part II: Answers to questions about Internet privacy. (2010, July 27). *The New York Times*. Retrieved November 10, 2012 from <http://tinyurl.com/34uowps>
- SHEMA, H., BAR-ILAN, J., & THELWALL, M. (2012) Research blogs and the discussion of scholarly information. *PLoS ONE*, 7(5), e35869.
- Slashdot: News for nerds, stuff that matters. (n. d.). Retrieved October 2, 2012 from <http://slashdot.org/>
- TERRAS, M. (2012, April). The verdict: Is blogging or tweeting about research papers worth it? Retrieved November 10, 2012 from <http://tinyurl.com/cdnkxz4>
- Volokh conspiracy. (n. d.). Retrieved October 2, 2012 from <http://www.volokh.com/>

# Semantic Change of the Publication-Concept?

**Alexander Tokar**

Department of English Linguistics, Heinrich Heine University Düsseldorf

*The semantic feature [quality control through other experts in the appropriate research field] constitutes one of the distinctive features of the concept of academic publication. That is, only those texts that have been positively evaluated by other scholars can be regarded as academic publications. However, on the Internet any scholar can easily publish the results of his or her research without subjecting these results to the scrutiny of other scholars (e.g., by uploading a manuscript on a homepage or posting it on a blog site). Do those publications also qualify as academic publications? Has the concept of academic publication recently undergone a semantic change?*

## Introduction

The Internet (especially, the so-called Web 2.0) has given rise to new modes of scholarly communication. These include, for example, scholarly blogging, scholarly tweeting, presenting the results of one's research in a wiki, and the like. One of the most important differences between these new Internet-based forms of scholarly communication and traditional academic communication is that the former do not rely on quality control through other experts in the appropriate field. This means that any scholar can post the results of his or her research to, e.g., his or her blog page without subjecting these results to the scrutiny of reviewers, which is characteristic of traditional scholarly communication (e.g., submitting a manuscript to a refereed journal or an edited volume). The central question raised by this article is whether this

democratization of scholarly communication, which has been brought about by the Internet, has resulted in a change in the concept of academic publication. That is, do present-day scholars regard as legitimate publications only those texts that have been positively evaluated by other scholars? Or: Does any text that a scholar uploads to the Internet automatically qualify as a publication?

This article utilizes the following structure: The first section discusses the most important theoretical issues pertaining to the notion of semantic change. The focus is on mechanisms, outcomes, and causes of semantic change. The second section expands on what has already been said above: In which respects are Internet publications different from traditional academic publications? The third section presents the results of an empirical investigation aimed at answering the question of whether the concept of academic publication has recently undergone a semantic change. Finally, the article discusses the results of this investigation and proposes a new research question and a methodology for further study of this topic.

## **Semantic Change: Mechanisms, Outcomes, and Causes**

Meaning change is traditionally classified into a number of subcategories. First of all, with regard to its mechanisms, semantic change can be classified into metonymy and metaphor. The difference between these mechanisms is that while an output meaning that has come into existence via metonymization exhibits a more or less real, “objective” link to its input meaning, an output meaning that has been a product of metaphorization is only perceptually similar to its input meaning (see, e.g., Hock, 1986, p. 285). For example, as I have argued elsewhere (Tokar, 2009, pp. 52-65), the use of the word *friend* in the context of online social networking Web sites such as Facebook is an example of metonymization: Facebook friends can include Facebook users’ offline friends, but very often they also include people whom Facebook users do not regard as friends in the literal, real-life meaning of this word, i.e., as people whom they “know well and regard with affection and trust” (WordNet). By contrast, the semantic development of the word *firewall*, which originally meant “a wall designed to prevent the spread of fire in a building,” but over the course of time has also come to signify “a piece of software designed to protect computers from viruses and especially hackers’ attacks,” is an example of metaphorization: Firewalls in buildings are considerably different from computer firewalls, but we perceive a functional similarity between them (i.e., that both are protection devices).

Secondly, with regard to its outcomes, a semantic change can be classified into an instance of full-idiomatization and that of quasi-idiomatization (Mel'čuk, 2001, pp. 447-460). The former gives rise to an output meaning that does not contain its input meaning. For example, the output meaning "a computer firewall" does not contain the input meanings "fire" and "wall," which are inherent in the components *fire* and *wall*: A computer firewall is never literally a wall that prevents the spread of fire in a building. In contrast to full-idiomatization, quasi-idiomatization produces output meanings that contain their input meanings plus some additional (often unpredictable) meanings. A case in point is the use of the phrase *list of publications* or *publications list* in the context of scholars' institutional and private homepages. What is interesting here is that an online-based publications list often does not only list the publications of a particular scholar, but also provides information as to how those publications can be accessed (e.g., downloadable PDF files, links to journals' Web sites, and sources alike). An online-based publications list is, thus, a quasi-idiom in relation to its pre-Internet counterpart. The meaning of the former contains the input meaning "a list of some scholar's publications" plus the additional meaning "information as to how those publications can be accessed."

Finally, with regard to its causes, semantic change can be classified into semantic change caused by lexical gaps and semantic change due to linguistic conservatism. The former can again be exemplified by the metaphorization of *firewall* and many other Internet terms (e.g., *to surf*, *to visit a Web site*, *a browser*, *a bookmark*, etc.). As Meyer et al. (1997, p. 3) pointed out, software developers usually give preference to metaphoric expressions when dealing with lexical gaps because metaphors allow "computer users to see a potentially complex concept in terms of a well-known and simple one" and, precisely because of this, "aid users in understanding and remembering new concepts." In addition to this, semantic change represents the default strategy of dealing with lexical gaps because it is cognitively easier to modify the meaning of an existing expression than to coin an entirely new expression (Tokar, 2012, p. 124, 129; see also Jansen, 2005). As far as semantic change due to linguistic conservatism is concerned, consider the following quote from Ullmann (1970):

It often happens that language is more conservative than civilization, material as well as moral. Objects, institutions, ideas, scientific concepts change in the course of time; yet in many cases the name is retained and thus helps to ensure a sense of tradition and continuity. (p. 198)

A good illustration of this is the phrase *publications list*, mentioned above. As stated earlier, an online-based publications list is more than a list

of some scholar's publications. It is a Web page that (often) enables its visitors to access (at least some of) the publications listed there. Nevertheless, the original expression *publications list* has not been abandoned in favor of some other word or phrase that better describes what a list of publications on the Internet usually is.

The term "semantic change due to linguistic conservatism," proposed by Ullmann, is not a very successful terminological solution since it suggests that, for example, scholars who post their publications lists online and allow other Internet users to access them use the term *publications list* because of their conservatism. That is, they are consciously aware of the fact that their online-based publications lists are more than lists of their publications, but in order "to ensure a sense of tradition and continuity" (in the sense of Ullmann), they have (deliberately!) decided to retain the original expression *publications list*. On the contrary, the majority of Internet users are actually unaware of the semantic development undergone by *publications list*. This is because an online-based publications list is, as stated above, a quasi-idiom in relation to its pre-Internet counterpart. That is, the former is, like the latter, a list of some scholar's publications (the input meaning is thus retained), but in contrast to the latter, the former is also a location on the Internet where some of the publications can be accessed (this is the additional idiomatic meaning). Given the preservation of the input meaning "a list of some scholar's publications" in the output meaning, it is extremely unlikely that Internet users have ever considered the possibility of replacing *publications lists* with some other expression.

## Traditional Publications versus Internet Publications

As Stefik (1996) pointed out, in the offline world,

writers write and editors determine which books are worth publishing. They control access to the printing presses and the distribution channels of publishers. Publishers publish the books, have them printed, and ship copies to wholesalers, libraries, and bookstores. (p. 6)

In stark contrast to this, on the Internet, "writers can be their own publishers" (Stefik, 1996, p. 9). That is, nobody has to subject their work to the scrutiny of editors and publishers in order to be able to "publish" it on a personal Web site or a blog page. One consequence of this is that the Internet abounds in "a vast array of digital works of indeterminate quality and value" (Stefik, 1996, p. 9) or, as the Russian journalist Mikhail Leontyev once put it, in a "rubbish-heap" of content of very poor quality (an interview with Sergey

Korzun on July 23, 2008). The second consequence, which is more important for this article, is that this “elimination” of editors and publishers on the Internet has resulted in a semantic change of the terms *publish* and *publication*. Consider, for example, the use of the verb *publish* in the following two sentences:

A photograph of my dog, taken by me and published on my website, was used on another website without my permission. (<http://tinyurl.com/ced4p2b>)

RSS feeds allow you to receive on your computer updates on the last posts published on my blog. (<http://www.beppegrillo.it/en/aiuto.php>)

What these sentences serve to illustrate is that an Internet publication is anything that can be uploaded to the Internet: e.g., a photograph of one’s dog, a blog post, a video, etc. At first glance, it may seem that the verb *publish* is used here in its literal meaning “to make public” (Oxford English Dictionary), “to make generally known,” “to disseminate to the public” (Merriam-Webster Online). Indeed, when, for example, bloggers post information on the Internet, they make it public (i.e., generally known). However, as was recognized by Stefik, in the offline world, due to the fact that editors and publishers determine which books are worth publishing, not every piece of information can be made public. For instance, we can hardly imagine a traditional publisher such as Cambridge University Press publishing blog posts or photographs of users’ dogs (unless editors find a special reason why this might be interesting to readers).

As with online-based publications lists, Internet publications whose quality has not been determined by editors and publishers belong to Ullmann’s category of semantic change due to linguistic conservatism. That is, even though publishing practices on the Internet have become considerably different from those in the offline world, both are regarded and referred to as instances of one and the same process: *publishing*. In other words, for an ordinary language user, it does not make a big (semantic) difference whether someone *publishes* an article in an academic journal or whether someone *publishes* a photo of a dog on his or her blog. This is due to the fact that, as in the case of publications lists, an important aspect of the input meaning “a traditional publication” has remained part of the output meaning “an Internet publication”: Both publishing an article in a journal and publishing a photo of a dog in a blog post involve disseminating the objects of publishing to other people. Hence, the above-cited dictionary definitions of *to publish* as “to make public,” “to make generally known,” “to disseminate to the public.”

However, notice that the case of Internet publications that have not been subjected to any quality control before publication is different from that of

Internet-based publications lists in that the former does not involve the addition of new semantic content. The case of Internet publications involves the removal or, as we said earlier, the “elimination” of some semantic content in the input meaning. That is, the information “editors and publishers warrant the quality of publications they disseminate to the public” has been removed from the input meaning “a traditional publication,” thereby giving rise to the output meaning “anything posted to the Internet without any quality control through other Internet users.”

This conceptual “elimination” has been committed by people such as Tim Berners-Lee, who invented and created the World Wide Web. That is, the fact that the Internet has, from its earliest days, remained a publication platform for virtually anyone is directly connected to the fact that Internet creators and later, providers of various Web 2.0 services, which mainly contain user-generated content (which usually does not undergo any quality control on the part of service providers!), have removed the semantic content “quality control” from the input meaning “a traditional publication.” Had it not been for this removal, active access to the Internet (i.e., the one that involves not only the consumption but also the production of Internet content) would now be in the hands of a relatively small number of people. A somewhat similar example, which I discussed elsewhere (Tokar, 2009, p. 8; see also Stefik, 1996, pp. 115-120), is electronic mail. If e-mail creators had not removed, for example, “envelopes” and “stamps” from the input meaning “traditional mail,” sending an e-mail would now involve putting an e-mail message into an electronic “envelope” and placing a digital “stamp” on it, for which we would probably have to pay the providers of e-mail services.

Having said this, let us now focus on the use of the term *publication* in academic contexts. As was indicated above, an academic publication (e.g., a monograph, an article in a refereed journal or in an edited volume, etc.) is one whose quality has been controlled by at least one other expert working in the same field as the author of the publication. (Hence the term “peer review,” defined by Merriam-Webster Online as “a process by which something proposed (as for research or publication) is evaluated by a group of experts in the appropriate field.”) The most prestigious academic publications are those that have been positively peer-reviewed (usually anonymously) by more than one expert in the appropriate field. The semantic feature [quality control through other experts in the appropriate field] can thus be regarded as one of the distinctive features of the concept of academic publication. (Distinctive features are necessary conditions that must be fulfilled by an entity in order to qualify as an instance of some concept. For example, being female and being someone’s parent are the necessary conditions that are fulfilled by all mothers:

Any entity that is both female and someone's parent qualifies as a mother; see, e.g., Cruse, 2004, p. 250.) However, despite the fact that the feature [quality control through other experts] constitutes one of the necessary conditions of an academic publication, a number of online publications lists mention (and very often also provide electronic access to) "publications" that either have been rejected by other experts or have not been evaluated by them at all (i.e., texts which have never been submitted to a journal or an edited volume). These are traditionally referred to as *unpublished manuscripts*. What is interesting about this term is that, on the one hand, the presence of the adjective *unpublished* corroborates the analysis of the feature [quality control through other experts] as one of the distinctive features of the concept of academic publication: If scholars who mention unpublished manuscripts in their publications lists were not aware of this fact, they would definitely not label these manuscripts *unpublished manuscripts*. (These manuscripts would, for them, be, in no essential respect, different from published manuscripts, which have been positively evaluated by other scholars). On the other hand, however, the fact that publications lists include unpublished manuscripts suggests that the concept of academic publication is now, indeed, undergoing a semantic change. If a scholar mentions an unpublished manuscript in his or her publications list, that person wants to emphasize the belief that the manuscript is worth publishing even if it has not been positively evaluated by other experts in the same field. The author thus claims that he or she has sufficient authority to decide that a particular manuscript can be published in a refereed journal or an edited volume. And he or she also invites other Internet users to convince themselves that this is, indeed, the case (by reading the full text of an unpublished manuscript).

## The Sociology of Semantic Change: A Case Study

Traditional linguistic studies on semantic change usually do not go beyond the mere documentation of the fact that the meaning of some expression is changing / has changed. Sociological aspects such as, for example, the question of how many members of a particular linguistic community are taking part in a change of meaning of a particular expression (i.e., the question of how many speakers use that expression in a semantically novel way) are typically not considered. This is because an answer to this question requires a manual analysis of a very large collection of both spoken and written texts produced by a very large number of members of a particular linguistic community. In other words, an analyst would have to consider all instances of the use of the expression under analysis in that corpus of spoken and written

texts. A methodological alternative to this is represented by so-called elicitation tests, which aim at eliciting native speakers' grammaticality or acceptability judgments. That is, for example, in the case of a semantic modification, a linguist can invent sentences in which the expression under analysis is used in a semantically novel way and then ask subjects whether those sentences are grammatical / acceptable for them. Both corpus-based investigations and elicitation tests are typically large-scale studies requiring a considerable amount of time.

Fortunately, answering the question of how many scholars are participating in a change of the concept of academic publication does not require a large-scale corpus investigation or / and an elicitation test. To answer this question, a researcher only has to count the overall number of publications lists that include unpublished manuscripts. Additionally, it makes sense to count the overall number of traditional academic publications citing unpublished manuscripts. These numbers can be seen as indicators of the acceptance (or the non-acceptance) of unpublished manuscripts in academia.

With respect to a particular university, these results can be obtained in the following way: Enter the URL of the university under investigation (or the URL of one of the faculties of that university) to the search mask "Enter a site URL" and the phrase *unpublished manuscript* to the search mask "Query" at [http://www.google.com/enterprise/search/products\\_gss.html](http://www.google.com/enterprise/search/products_gss.html). Google Site Search, a Web tool available at this address, will then yield all occurrences of the phrase *unpublished manuscript* on the Web pages of the university or faculty under investigation. Then classify the occurrences of the phrase *unpublished manuscript* into those that occur in scholars' publications lists and those that involve citations in traditional publications. (Many university Web sites provide their own search masks, enabling visitors to search for specific content located on the Web pages of a particular university. This is, thus, an alternative to Google Site Search.)

Before presenting the results of a Google Site Search for the occurrences of *unpublished manuscript* on the Web pages of my own university, it must be noted that the Heinrich Heine University Düsseldorf comprises five faculties: the Faculty of Medicine, the Faculty of Mathematics and Natural Sciences, the Faculty of Arts and Humanities, the Faculty of Business and Economics, and the Faculty of Law. According to the Web page "University Facts and Figures" (2010), "today, around 20,000 students, more than 1,700 lecturers and 900 further employees study, teach, and work on our campus."

The results of a Google Site Search for the occurrences of the phrase *unpublished manuscript* on the Web pages of the five faculties of my university are as follows: There are no occurrences of *unpublished manuscript* on the

Web pages of the Faculty of Medicine. Likewise, there are no occurrences of the German equivalent *unveröffentlichtes Manuskript*. I also conducted a search for the adjectives *unpublished* and *unveröffentlicht* only, expecting to find them in collocations such as *unpublished PhD thesis*, *unpublished work*, etc. But, again, Google Site Search yielded no results for the Faculty of Medicine.

There is only one occurrence of the phrase *unpublished manuscript* on the Web pages of the Faculty of Mathematics and Natural Sciences. It occurs in the References lists of a PDF document available on the Web pages of the faculty. The PDF document contains a description of a research project involving Heinrich Heine University Düsseldorf and another university in North Rhine-Westphalia. There is one occurrence of the German adjective *unveröffentlicht* in the context of a publications list of a lecturer (with a doctorate) employed by the faculty. *Unveröffentlicht* occurs in the context of *Projekt-Abschlussbericht*, i.e., a project completion report written by the lecturer in question. No PDF or Word file containing the full text of the project completion report is provided.

There are nine occurrences of the phrase *unpublished manuscript* on the Web pages of the Faculty of Arts and Humanities. Of these nine occurrences, one is a “publication” included in a publications list of a professor of the faculty. The full text of the manuscript can be downloaded as a PDF file. Another occurrence is a “publication” included in a publications list of members of a research group working at the faculty. No PDF or Word file containing the full text of the manuscript is provided. The remaining seven occurrences of *unpublished manuscript* can be found in the References lists of various documents (articles, abstracts, conference programs, etc.) available on the Web pages of the faculty. In addition to *manuscript*, the adjective *unpublished* can relatively often (in comparison with what is listed for other faculties) be found before other nouns. Thus, there are *unpublished papers* (five occurrences), *unpublished bachelor thesis / unpublished BA-Thesis* (two occurrences), *unpublished M.A. Project* (one occurrence), *unpublished PhD dissertation / unpublished doctoral dissertation* (six occurrences), *previously-unpublished material* (one occurrence), *unpublished articles* (one occurrence), *unpublished research reviews* (one occurrence), *unpublished script of lecture* (one occurrence), *unpublished report* (one occurrence), *unpublished work* (one occurrence), *unpublished texts* (one occurrence), and simply *unpublished* (one occurrence). Of these occurrences of the adjective *unpublished*, four are “publications” included in publications lists of members of the faculty. (Two of them are university professors; the other two are non-tenured lecturers with doctorates.) All other occurrences can be found in

various documents (e.g., articles, course descriptions, PowerPoint slides, etc.) available on the Web pages of the faculty. Finally, it must be mentioned that there are two occurrences of the German adjective *unveröffentlicht* on publications lists of members of the faculty and one occurrence of *unveröffentlichtes Manuskript* in the References list of a doctoral dissertation whose full text can be downloaded as a PDF file.

There are no occurrences of either the phrase *unpublished manuscript* or simply the adjective *unpublished* on the Web pages of the Faculty of Business and Economics. Likewise, there are no occurrences of the German phrase *unveröffentlichtes Manuskript* or the adjective *unveröffentlicht*.

There are no occurrences of either the phrase *unpublished manuscript* or the adjective *unpublished* on the Web pages of the Faculty of Law. Likewise, there are no occurrences of the German phrase *unveröffentlichtes Manuskript* or the adjective *unveröffentlicht*.

## Discussion and Outlook

If we use the number of occurrences of the term *unpublished manuscript* on institutional homepages as an indicator of a semantic change of the concept under investigation, then the results obtained clearly indicate that the semantic change of the concept of academic publication as outlined in the previous sections of this article is a fairly marginal phenomenon in academia (at least, as far as my university is concerned). Thus, only a very small number of academic employees of the Heinrich Heine University Düsseldorf regard themselves as qualified authorities to decide that their work is worth publishing. The most striking results are those for the Faculties of Medicine, Business and Economics, and Law, whose staff members seem to be completely unaware of the possibility of “publishing” unpublished work online. With regard to the Faculty of Mathematics and Natural Sciences and the Faculty of Law, the low number of unpublished manuscripts could perhaps be attributed to the more dominant preprint / working paper culture of those faculties, i.e., the practice of posting unpublished work at online-based document servers and repositories such as, for example, arXiv and The Social Science Research Network (SSRN).

Both preprints and working papers are similar to unpublished manuscripts in that neither the former nor the latter have ever been accepted for publication in traditional academic journals. However, a preprint is usually a document that has been submitted to some journal for peer review, but whose author(s) has / have not yet been notified about the reviewers’ decision. A working paper is a document which, according to its author, contains some

interesting ideas but which, in its current form, does not have the potential to be positively evaluated by reviewers of a journal. The author of a working paper thus uploads his or her currently unpublishable work to an academic repository, hoping that at some point in the future, thanks to other users' comments and his / her own further research on the same topic, this work will become publishable. In semantic terms, we can perhaps argue that the concept of a preprint is defined by the following semantic feature: [the expectation that the article will, in the near future, pass the quality control of a traditional journal]. And a working paper is defined by the semantic feature [the expectation that the article will, in the more distant future, be able to pass the quality control of a traditional journal]. In other words, the authors of both preprints and working papers share the traditional view that the feature [quality control through other experts in the appropriate research field] constitutes one of the distinctive features of the concept of academic publication. Accordingly, the occurrences of the expressions *preprint* and *working paper* on Web pages of the Faculty of Mathematics and Natural Sciences and the Faculty of Law cannot serve as indicators of semantic change of the concept of academic publication. To the contrary, they can serve as indicators of the non-change of the concept under study. However, as was conjectured above, the popularity of preprints and working papers among academics in these faculties can be one of the explanations for the non-popularity of unpublished manuscripts there. If other (especially senior) colleagues label their unpublished work *preprints* and *working papers*, why should I label my own unpublished work in a different way?!

The results for Heinrich Heine University Düsseldorf presented in the preceding section cannot but raise a number of (important) questions. First of all, are these (or similar) results true for other universities in Germany and (ideally) worldwide? The preliminary answer to this question, which can be given at the moment of writing, is "Yes." Thus, I conducted a similar Google Site Search for the occurrences of the phrases *unpublished manuscript* and *unveröffentlichtes Manuskript* (as well as related terms in English and German) on Web pages of several randomly chosen universities in Germany and English-speaking countries. The central conclusion that can be drawn from these searches is very similar to what was said before: Unpublished manuscripts represent a marginal phenomenon in academia worldwide. (Due to space limitations in this volume, these results will be presented elsewhere.)

Another important question is: Why are unpublished manuscripts more popular among academics at the Faculty of Arts and Humanities than among academic employees of other faculties? One possible explanation is that research in humanities is perceived (especially, if compared to research in so-

cial and natural sciences) as fairly subjective. Just think of the many linguistic theories such as, for example, structuralism, functionalism, generative grammar, cognitive grammar, etc., all of which can often fairly plausibly account for one and the same linguistic phenomenon (often in a very different way). It is very possible that on such occasions a structuralist analysis will not satisfy a generativist and a generative approach will not be enthusiastically accepted by a functionalist. In other words, an article submitted to a linguistic journal by a structuralist may not be positively evaluated by a reviewer of the generative persuasion. Similarly, an article submitted to a linguistic journal by a generativist may not be positively evaluated by a reviewer of the functionalist persuasion. This fairly probable scenario in humanities (at least in linguistics) considerably enhances the possibility of an author of a rejected article simply uploading his or her rejected work to the Internet (labeling it an *unpublished manuscript*), hoping that this work will be positively evaluated by colleagues.

Other reasons explaining the (relative) popularity of unpublished manuscripts among employees of the Faculty of Arts and Humanities can surely be identified as well, but this requires a large-scale qualitative study (i.e., conducting qualitative interviews with the members of staff of this faculty), which goes beyond the scope of the present investigation.

## Acknowledgements

This research was supported by the Düsseldorf University Strategic Research Fund. The author also wishes to express his gratitude to Wiebke Ostermann for her help with data collection during the first phase of this project. My special thanks go to Cornelius Puschmann, Jean Burgess, Axel Bruns, Ralph Schroeder, and Michael Beurskens for commenting on an earlier version of this article. I alone am responsible for any remaining errors and shortcomings.

## References

- arXiv.org e-Print archive. (n. d.). Retrieved October 2, 2012 from <http://arxiv.org/>
- CRUSE, D. (2004). *Meaning in language: An introduction to semantics and pragmatics* (2nd ed.). Oxford: Oxford University Press.
- Friend. (2012). *WordNet: A lexical database for English*. Princeton: Cognitive Science Laboratory of Princeton University. Retrieved October 2, 2012 from

- <http://wordnetweb.princeton.edu/perl/webwn?s=friend&sub=Search+WordNet&o2=&o0=1&o8=1&o1=1&o7=&o5=&o9=&o6=&o3=&o4=&h=>
- Google site search. (2012). Retrieved October 2, 2012 from [http://www.google.com/enterprise/search/products\\_gss.html](http://www.google.com/enterprise/search/products_gss.html)
- HOCK, H. (1986). *Principles of historical linguistics*. Berlin: Mouton de Gruyter.
- JANSEN, S. (2005). *Sprachliches Lehngut im „world wide web“: Neologismen in der französischen und spanischen Internetterminologie*. Tübingen: Gunter Narr.
- KORZUN, S. (2008, July 27). [Interview with Mikhail Leontyev, a Russian journalist]. *Bez durakov [Without fools]*. Retrieved October 2, 2012 from Echo of Moscow Web site <http://echo.msk.ru/programs/korzun/528795-echo/>
- MEL'ČUK, I. (2001). *Kurs obščej morfologii: Morfoložičeskie znaki [Course in general morphology: Morphological signs]*. Moskva: Jazyki Slavjanskoj Kul'tury.
- MEYER, I., ZALUSKI, V., & MACKINTOSH, K. (1997). Metaphorical Internet terms: A conceptual and structural analysis. *Terminology*, 4(1), 1-33.
- Peer Review. (2012). *Merriam-Webster Online*. Retrieved October 2, 2012 from <http://www.merriam-webster.com/dictionary/peer%20review>
- Publish. (Third edition, September 2007; online version June 2012). *Oxford English Dictionary*. Oxford: Oxford University Press. Retrieved August 31, 2012 from <http://www.oed.com/view/Entry/154072>
- Publish. (2012). *Merriam-Webster Online*. Retrieved October 2, 2012 from <http://www.merriam-webster.com/dictionary/publish>
- Social Science Research Network (SSRN). (2012). Retrieved October 2, 2012 from <http://www.ssrn.com/>
- STEFIK, M. (1996). *Internet dreams: Archetypes, myths, and metaphors*. Cambridge, MA: MIT Press.
- TOKAR, A. (2009). *Metaphors of the Web 2.0: With special emphasis on social networks and folksonomies*. Frankfurt am Main: Peter Lang.
- TOKAR, A. (2012). *Introduction to English morphology*. Frankfurt am Main: Peter Lang.
- ULLMANN, S. (1970). *Semantics: An introduction to the science of meaning*. Oxford: Blackwell.
- University facts and figures. (2010). Retrieved October 2, 2012 from <http://www.uni-duesseldorf.de/home/en/about-hhu/further-links/facts-and-figures.html>



## **PART FIVE**

---

# **THE INTERNET AND SCHOLARLY IMPACT**



# Citations in Web 2.0

**Katrin Weller and Isabella Peters**

Department of Information Science, Heinrich Heine University Düsseldorf

*Citations are a classic dimension of scientific communication. This paper looks at two different scenarios in which citation analysis can be applied to novel Web 2.0 environments: One case study deals with citations on Twitter and the other with analyzing blog posts and social bookmarking systems.*

## Introduction

Scientific communication is a process that, among other things, involves citing other scholars' publications. Therefore, it is not surprising that citation analysis has become one key method for investigating relevance and importance in academia (see, e.g., Cronin, 1984). Citation analysis can thus have practical implications for scientists' work and life, because it is used to evaluate the impact of individual scientists, working groups, institutions, or scientific journals, and may be the basis for decisions about funding grants and job appointments (Stock, 1994; Stock, 2001). Furthermore, citations help scientists to filter the enormous amount of scientific literature and allow browsing and searching in publication databases such as Web of Science and Scopus, thus becoming part of information retrieval strategies. Accordingly, the quality, comparability, and adequacy of applied methods in citation analysis are of high importance. The scientific disciplines of informetrics (Tague-Sutcliffe, 1992) and, more specifically, scientometrics deal with these key challenges and establish procedures for measuring and comparing scientific output based on publications and scientific reputation based on citations (Haustein, 2012; Leydesdorff, 1995).

With the growing importance of the Internet, the principles of informetrics have also been applied in Web environments, thus shaping the new discipline of webometrics (Thelwall, 2008). The fundamental principle of the Internet is a connection via hyperlinks; hyperlinks interlink Web sites with each other and thus build the World Wide Web. Smith (2004) showed that Web links resemble patterns of classic citations in printed publications. References or footnotes in printed publications and links on Web sites are the keys for finding relevant information in both search engines and bibliographies. In addition to these hyperlink structures, recent Web 2.0 tools come with a number of other important functionalities that enable novel forms of social interaction. They have brought about new aspects that can be measured in webometrics (e.g., those relating to access and usage, Web publication behavior, and user interrelations).

Scientific discussions are also increasingly being held in various Web 2.0 environments such as blogs, forums, and Twitter. Gray et al. (2008) pointed out that scholars were acting as authors in different Web 2.0 environments, including wikis, podcasts, and blogs—a development that challenges the classic understanding of the authorship concept in scientific communication. Gray et al. (2008) also discussed the difficulties of quoting and citing Web 2.0 sources in scientific publications. Currently, activities outside classic publication channels such as scientific journals are rarely considered in official evaluations of scientists' impact and scope. Yet, with the growing importance of using the Internet in scientific communication, there is a need for discussing combinations of scientometric and webometric indicators. So far, the most notable effort to promote and discuss alternative scientometric indicators for Web environments has been the altmetrics initiative (Priem et al., 2010). The authors of the altmetrics manifesto argued for the development of new metrics that would enable filtering and browsing of the growing amount of information on the Web. Priem and Hemminger (2010), furthermore, provided an overview on Web 2.0 services, which might be of interest for new scientometric indicators (e.g., measuring publication impact on the basis of social mentions). This paper represents our own contribution to this ongoing discussion.

In citation analysis, one typically distinguishes citations from references, which actually are two sides of the same coin (Stock, 2001). Slightly inconsistently, citation is also used as a broader term that subsumes both the dimension of citations as well as the dimension of references; this fact often leads to confusion and inconsistent use. If an author cites an exact passage from a text, this is called a quotation. If a publication includes a formal mention of another work, there is a linkage between these two publications that

can be looked at from two perspectives. From the cited work's perspective, this linkage is a citation, received by the cited author. From the citing work's perspective, the linkage is a reference: The citing author has referred to another work (usually in the References section or as a footnote).

The Web 2.0 has created lots of new types of references. Let us have a look at some examples: The microblogging service Twitter allows users to easily cite other users' tweets by retweeting them (boyd et al., 2010) and including additional hyperlinks. Blogs may also include hyperlinks as references. Furthermore, trackbacks or pingbacks automatically inform bloggers when other blogs cite them (Kim & SangKi, 2008). With social bookmarking, users indicate interests in scientific publications via sharing URLs or Web resources. These are only some of the Web 2.0-related examples, at which we will have a closer look in the subsequent sections. There are various others, but they are beyond the scope of this article. In summary: On the one hand, various new forms of social content may receive citations, because people may cite YouTube videos, SlideShare slides, or podcasts. On the other hand, various types of Web 2.0 contents include references to either classic publications (e.g., a blog post linking to a journal article) or to other types of social content (e.g., a tweet referencing a blog post).

We will now present the results from two different case studies. First, we will look at types of citations that can be found on Twitter. Second, we will analyze the linking behavior of scientific bloggers and the visibility of bloggers' publications in different social bookmarking systems (for example, Mendeley) and bibliographic databases (e.g., Scopus). Both offer preliminary results in the area of citations in Web 2.0 and should encourage future research in this area.

## **Citation Analysis in Twitter**

Priem and Costello (2011) defined citations in Twitter as "direct or indirect links from a tweet to a peer-reviewed scholarly article online" and distinguished first and second-order citations, based on whether there is an "intermediate webpage between the tweet and target resource." They collected tweets from 28 academics and found that, of all URLs in these tweets, 6% were links to peer-reviewed articles (either directly or via an intermediate page), which could be counted as citations. We have argued that linking to a peer-reviewed publication is only one possible dimension of citing with Twitter and used different, alternative definitions (Weller et al., 2011; Weller & Puschmann, 2011). The basis of our definition is the distinction between

external citations and internal citations. Tweets may either include references to external resources or to information available on Twitter.

All URLs in tweets can be considered as a citation act: The tweet includes a reference in the form of a URL, and a certain Web site obtains a citation through this tweet. For some scientometric analyses, references to scientific publications are of the greatest interest, and the approach of Priem and Costello (2011) for counting those URLs might suffice. Yet, references to scientific blog posts, news articles, or presentation slides may also be valuable information. For example, Thelwall et al. (in press) look at links to science-related YouTube videos. Moreover, for general informetric analyses, all sorts of references to URLs are of relevance and should thus be considered as types of citations.

As Twitter itself is a channel for communicating and publishing pieces of information, we can also find a different type of citation behavior: Quite frequently, Twitter users directly quote other peoples' tweets. Tweets are either copied completely, or users copy parts of an existing tweet and add their own comment. In many cases, the users also mention the original author—this clearly resembles citation practices in scientific communication. Because these copied tweets have often been labeled as “Retweets” or “RT” by Twitter users, Twitter has established retweeting as a genuine Twitter functionality (Kooti et al., 2012). Retweets can thus be interpreted as a form of inter-Twitter citations (internal citations). A user who retweets another user's tweet publishes a reference: The retweeted user receives a citation. In general, users retweet for different reasons, such as information diffusion, or use retweets as a “means of participating in a diffuse conversation” (boyd et al., 2010). Retweet analyses can help to identify influential Twitter users, interesting topics on Twitter, and information diffusion—much as citation analysis can do in classic publication databases. Because Twitter has now largely standardized the format of retweets (when the specific retweet button is used on Twitter), retweet analyses can be performed more easily and become more reproducible. However, for altmetric analyses, some technical challenges remain when users manually modify retweeted statements.

## **Selected Results**

Having defined these two different types of Twitter citations, we will now take a closer look at actual Twitter data to see how they are applied in scientific communication. We looked at different sets of “scientific tweets,” i.e., tweets that can be interpreted as scientific communication. In our cases, these tweets were either collected based on specific hashtags for scientific conferences or based on the tweets' authors (Weller et al., 2011). We chose single

conference hashtags and used a list of almost 600 Twitter users who identify themselves as scientists or people closely related to academia (Weller & Puschmann, 2011). Table 1 includes basic information for the three datasets and summarizes the proportions of internal and external citations in these tweets. These datasets reveal high citation activities in science-related tweets. Whereas only three percent of general tweets are retweets (boyd et al., 2010), the conference tweets and the scientists' tweets all have more than 20% RTs.

Table 1. The three test datasets and the proportion of internal and external citations

<b>Dataset</b>	<b>#www2010</b>	<b>#mla09</b>	<b>scientists</b>
Description	World Wide Web Conference (WWW 2010), Raleigh, NC, USA. April 26-30, 2010.	Modern Language Association Conference (MLA 2009), Philadelphia, PA, USA. Dec. 27-30, 2009.	Tweets collected from 589 selected science-related Twitter users.
Data collection period	4/13/10 to 5/14/10	12/15/09 to 1/14/10	1/7/10 to 8/31/10
No. of tweets	3,358	1,929	410,609
No. and % of external citations (URLs)	1,338 40%	525 27%	227,550 55%
No. and % of internal citations (RTs)	1,121 33%	413 21%	92,225 22 %
No. and % of retweets that include URLs	530 47%	270 65%	58,525 63%

Notice also a very high number of external citations in scientists' tweets: Fifty-five percent of the tweets contained at least one URL. Some tweets also included more than one URL, so the number of total URLs in the datasets is even higher. For example, in the #www2010 dataset, 1,338 tweets include at least one URL. There is a total number of 1,460 URLs in the dataset. These URLs may reference the same Web sites. In the #www2010 dataset, there are 574 unique Web sites linked by 1,460 URLs. For the conference datasets, we have considered the cited URLs and manually classified them into 10 categories (see Figure 1). Users in the #mla09 dataset almost never cited actual scientific publications in their tweets. More frequently cited were blog posts and press articles. For #www2010, the distribution is more balanced. Finally, our analysis showed that internal and external citations on Twitter are also highly interwoven. More than half of the retweets (63%) in the scientist dataset included URLs (65% for #mla09 and 47% for #www2010; see Table 1). This finding suggests that Twitter is heavily used for re-sharing information resources.

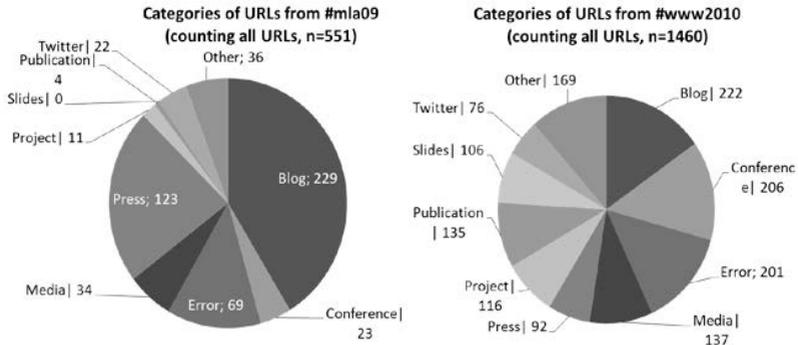


Figure 1. URLs from #mla09 and #www2010 by categories

## Citations in Blogs and Social Bookmarking Services

Blogs are typically personal Web sites where published posts are displayed in reverse chronological order (see, e.g., Puschmann, 2010). They serve as easy-to-publish media and are therefore increasingly used by scholars (Luzón, 2009) to discuss the latest research with their peers and other audiences (Mahrt & Puschmann, in press) and as a means of self-reflection (Reinmann, 2008) or education. Linking is a fundamental part of blogging practice, with pingbacks and trackbacks informing bloggers when their blog was cited by another blog. Additionally, blog posts often contain URLs to various Web resources (creating external citations) or to sites within the same blog(-platform), which might be regarded as an instance of self-citation (see, e.g., Shema et al., 2012). Luzón (2009) analyzed linking behavior and link types of 15 academic blogs and found that over 50% of links point to pages within the same blog.

In social bookmarking services (for example, Delicious), users, browser-independently, save and tag Web resources, such as blogs or Web sites, for later retrieval. Scholarly social bookmarking services also allow the saving of bibliographic information for scholarly products (Reher & Haustein, 2010). Analogous to citation counts, bookmarks to publications can be seen as indicators of how interested a community is in a given publication (Haustein, 2012). Groth and Gurney (2010) analyzed which and how chemical journal articles are discussed on ResearchBlogging.org (e.g., in terms of the impact factor of the journal), whereas Shema et al. (2012) investigated the de-

mographics and topical foci of bloggers from the same platform. Bar-Ilan et al. (2012) studied publication lists and publications found in Scopus and in social bookmarking systems of 57 presenters from the 2010 Leiden Social Technology Indicators (STI) Conference. They found that in Mendeley, more than 80% of the 1,136 sampled articles were saved by users. Moreover, Mendeley bookmarks are significantly correlated ( $r = 0.45$ ) to Scopus citations. Correlations among Mendeley, CiteULike, and Web of Science for 1,613 Nature and Science articles were also processed by Li et al. (2012). Web of Science and Mendeley showed a moderate correlation of  $r = 0.55$  and CiteULike a correlation of  $r = 0.34$ . Ninety-two percent of the sampled articles were also bookmarked by at least one user in Mendeley and 60% by one or more CiteULike users. In the following sections, we will explain our research questions and present the results of our own study on blogs and social bookmarking systems to compare them with the results found in related work.

## Data Collection

Scientific blogs were our key information source in this study because they determined the selection of the analyzed authors. We used two blog portals, Scienceblogs.com and Scienceblogs.de, which host blogs of scientific writers. We only considered authors who are affiliated with universities or other research institutions. This limitation resulted in 33 English-language authors and 11 German-language bloggers. Because some blogs are maintained by more than one author, we combined the authors of each blog and analyzed data from 30 English and 10 German blogs indicated by their respective authors' names. For all of the chosen blogs, we manually collected the blog's name, the name(s) of author(s), the blog's starting date, and the number of blog posts, comments, and unique commentators. Moreover, we automatically extracted the URLs of the blog posts to analyze linking behavior of bloggers. The analysis is based on 19,721 blog posts. For author-based citation statistics, we employed the same approach as Bar-Ilan et al. (2012) and used Mendeley, BibSonomy, and CiteULike to extract social bookmarking data for each article that a blog author had written. To gain article-based metrics as well as bookmarking statistics, we first searched for the official publications lists of chosen bloggers on institutional or private Web sites. Here, we worked with individuals and not blogs. We considered publications lists found on institutional or private Web sites as a gold standard, because we assumed that scientific authors are strongly interested in regularly maintaining their publications lists to be visible in the scientific community. However, some authors did not have any publications lists, so we had to create such

lists from publications found in the analyzed social bookmarking systems. We also cross-checked social bookmarking systems to find articles missing on the publications lists and to determine the share of “official” papers (recorded in self-maintained publications lists) in social bookmarking systems. Authors without publications lists or articles saved in social bookmarking systems were excluded from analyses. Authors were also excluded when author disambiguation was too difficult because self-maintained publications lists could not be found on the Web (e.g., Jessica Palmer). In sum, we analyzed 936 publications found on personal publications lists and social bookmarking systems by 41 authors. To compare social bookmarking data with traditional author metrics provided by bibliographic databases, the number of publications and citations found in Scopus was also collected. We chose Scopus as the source for citation data because it allows users to search for authors by first and last name. Because Scopus only indexes a selection of available journals and other publication formats, we only gained data from 678 publications, meaning that about 28% of the publications of the analyzed bloggers could not be found in Scopus. This value is slightly higher than those reported by Bar-Ilan et al. (2012) and Li et al. (2012), probably because of the smaller dataset used in our study.

## Results

The use of URLs is common practice in blogs, as shown in Figure 2. Especially heavy bloggers distribute URLs via blog posts (e.g., Lambert).

However, the shares of outgoing URLs linking to Web sites outside the blogs and to other blog posts (i.e., self-citation) differ fundamentally among blogs. Table 2 shows the 10 most linked top-level domains from scienceblogs.com and scienceblogs.de. Other social media platforms, such as Wikipedia, YouTube, or Twitter, and news platforms (e.g., The New York Times or Spiegel) are mostly referenced in blog posts, besides self-reference to scienceblogs.de or scienceblogs.com, which are the top-link destinations in our dataset. The results for self-citations correspond to those found by Luzón (2009) for scienceblogs.de but are lower for scienceblogs.com, which might be explained by our automatic analysis focussing on top-level domains.

Surprisingly, it turned out that self-maintained publications lists are not complete or updated frequently by authors. Twenty-two percent of the publications from authors of scienceblogs.com and 25% of publications from authors of scienceblogs.de are only findable via author-name searches in other sources (i.e., Scopus, CiteULike, Mendeley, and BibSonomy). The detailed analyses of the three social bookmarking systems showed that, for

both author groups, Mendeley is the service where most of the publications can be found (53% in scienceblogs.com and 42% in scienceblogs.de).

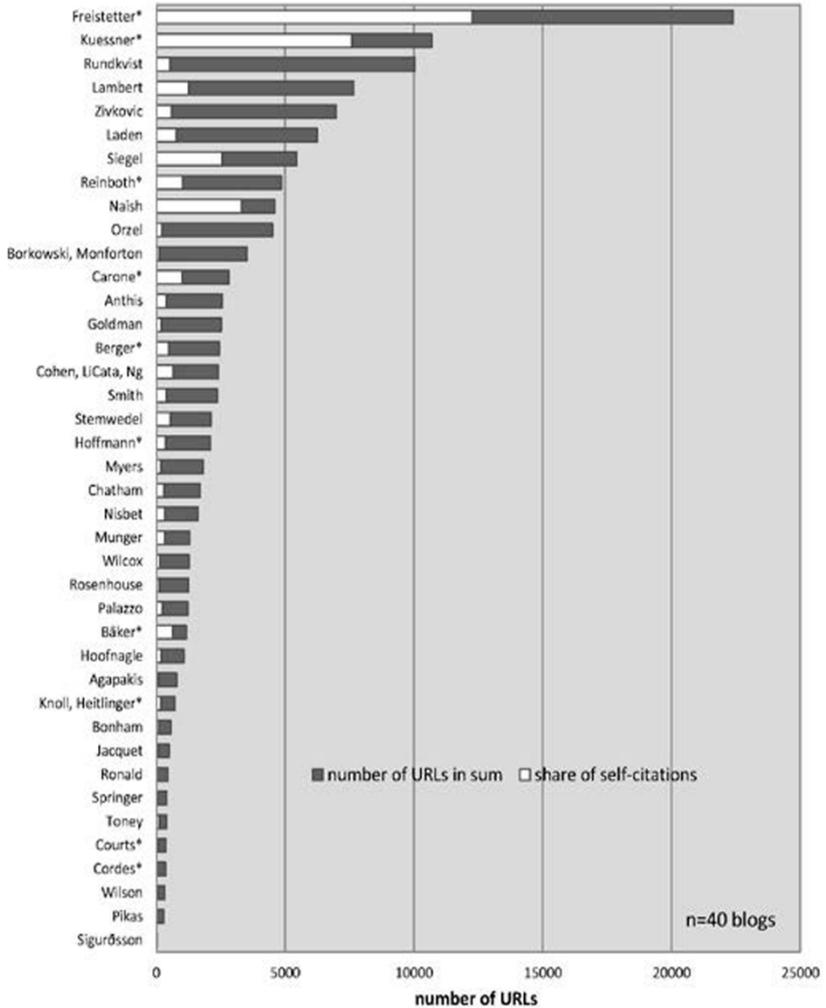


Figure 2. Number of URLs in blog posts and self-citations. \* = Scienceblogs.de authors

Table 2. Link destinations from scienceblogs.de and scienceblogs.com

outgoing links from blog posts (scienceblogs.de)			outgoing links from blog posts (scienceblogs.com)		
destination	absolute	%	destination	absolute	%
scienceblogs.de	2509	52,45	scienceblogs.com	18041	23,40
de.wikipedia.org	3709	7,78	technorati.com	3008	3,90
en.wikipedia.org	882	1,85	blogger.se	2873	3,73
amazon.de	517	1,08	en.wikipedia.org	2430	3,15
flattr.com	393	0,82	delicious.com	2044	2,65
esowatch.com	342	0,72	amazon.com	1088	1,41
arxiv.org	304	0,64	nytimes.com	746	0,97
spiegel.de	238	0,50	researchblogging.org	632	0,82
youtube.com	225	0,47	del.icio.us	625	0,81
twitter.com	202	0,42	ncbi.nlm.nih.gov	519	0,67

Because of our small dataset, we used Kendall's  $\tau$  for calculating correlation values between Scopus citation counts and bookmark numbers from Mendeley, CiteULike, and BibSonomy. Table 3 shows the correlation values for scienceblogs.de authors; Table 4 displays values for scienceblogs.com authors. For scienceblogs.com, we worked with only 29 authors, because one author had no publications indexed in Scopus. Our findings for all 936 publications from both author sets conform to those of Bar-Ilan et al. (2012); the highest significant correlation is between Mendeley and Scopus at  $\tau = 0.483$  (see Table 5). The results indicate that users bookmark and cite in a similar way and that often cited papers are also more likely to be bookmarked. Conversely, social bookmarking systems cover 28% more articles than Scopus, meaning that users of bookmarking systems create via bookmarks a more holistic view of scientific authors and reward more products of scholarly practice (e.g., blog posts).

Table 3. Correlations between the number of citations and bookmarks for 11 scienceblogs.de authors and 198 Publications \*\*Correlation is significant at the 0.01 level (two-tailed)

Kendall's $\tau$	bookmarks Mendeley	bookmarks CiteULike	bookmarks BibSonomy
citations (Scopus)	0.636**	0.397	-0.189

Table 4. Correlations between the number of citations and bookmarks for 29 scienceblogs.com authors and 738 publications. \*\*Correlation is significant at the 0.01 level (two-tailed).

Kendall's $\tau$	bookmarks Mendeley	bookmarks CiteULike	bookmarks BibSonomy
citations (Scopus)	0.463**	0.355**	0.219

Table 5. Correlations between the number of citations and bookmarks for 40 scienceblogs.com/.de authors and 936 publications. \*\*Correlation is significant at the 0.01 level (two-tailed).

Kendall's $\tau$	bookmarks Mendeley	bookmarks CiteULike	bookmarks BibSonomy
citations (Scopus)	0.483**	0.367**	0.107

## Conclusion and Outlook

In Web 2.0, citations and references can appear in various formats, and the analysis of citation structures can be applied to different forms of scientific communication on the Web. We have seen that scientists communicate via Twitter and blogs and make use of references in both services. We distinguished between internal and external citations on Twitter, which are interwoven. Slightly differently, we had to distinguish external links from self-citations in blog posts. Furthermore, the visibility of scientific publications in social bookmarking systems was discussed, and different bookmarking systems were compared in terms of coverage. Mendeley is the most popular social bookmarking service and should therefore be fed with publications to make them more visible to the community. Further research should comprise detailed analyses of blog posts', tweets', and scientific articles' content in order to reveal whether bloggers blog and tweet about the same topics that they study professionally. The next step will be to measure the impact of authors on the blogosphere or Twittersphere and determine how indicators should be transferred into the field of scientometrics.

## Acknowledgements

Thanks to Cornelius Puschmann, for gathering and analyzing Twitter data, and to Evelyn Dröge, for the manual categorization of URLs. Thanks to Stefanie Hausteil, Lisa Beutelspacher, Pascal Chave, Parinaz Maghferat, and Jens Terliesner for support in data collection on blogs and social bookmark-

ing. This research was supported by the Düsseldorf University Strategic Research Fund.

## References

- BAR-ILAN, J., HAUSTEIN, S., PETERS, I., PRIEM, J., SHEMA, H., & TERLIESNER, J. (2012, September 5-8). Beyond citations: Scholars' visibility on the social web. Paper presented at the 17th International Conference on Science and Technology Indicators, Montreal, Canada. Retrieved October 3, 2012 from <http://arxiv.org/pdf/1205.5611v1.pdf>
- BOYD, d., GOLDR, S., & LOTAN, G. (2010). Tweet, tweet, retweet: Conversational aspects of retweeting on Twitter. In *Proceedings of the 2010 43rd Hawaii International Conference on System Sciences* (pp. 1-10). New York: Curran. Retrieved October 3, 2012 from <http://www.danah.org/papers/TweetTweetRetweet.pdf>
- CRONIN, B. (1984). *The citation process: The role and significance of citations in scientific communication*. London: Taylor Graham.
- GRAY, K., THOMPSON, C., CLEREHAN, R., SHEARD, J., & HAMILTON, M. (2008). Web 2.0 authorship: Issues of referencing and citation for academic integrity. *Internet and Higher Education*, 11(2), 112-118.
- GROTH, P., & GURNEY, T. (2010, April 26-27). Studying scientific discourse on the web using bibliometrics: A chemistry blogging case study. In *Proceedings of the Web Science 2010: Extending the frontiers of society on-line*. Retrieved October 3, 2012 from [http://journal.webscience.org/308/2/websci10\\_submission\\_48.pdf](http://journal.webscience.org/308/2/websci10_submission_48.pdf)
- HAUSTEIN, S. (2012). *Multidimensional journal evaluation. Analyzing scientific periodicals beyond the impact factor*. Berlin: De Gruyter Saur.
- KIM, G., & SANGKI, H. S. (2008). A study of online (digital) reputation in blogosphere based on relationship and activity. In *Proceedings of the International Conference on Cyberworlds September 22-24, 2008, Hangzhou, China* (pp. 173-179). Los Alamitos, California: IEEE Computer Society.
- KOOTI, F., YANG, H., CHA, M., GUMMADI, K. P., & MASON, W. A. (2012). The emergence of conventions in online social networks. In *Proceedings of the Sixth International Association for the Advancement of Artificial Intelligence Conference on Weblogs and Social Media*. (pp. 194-201). Palo Alto, California: The AAAI Press. Retrieved November 11, 2012 from <http://tinyurl.com/asdetlq>
- LEYDESDORFF, L. (1995). *The challenge of scientometrics: The development, measurement, and self-organization of scientific communication*. Leiden: DSWO.
- LI, X., THELWALL, M., & GIUSTINI, D. (2012). Validating online reference managers for scholarly impact measurement. *Scientometrics*, 91(2), 461-471.

- LUZÓN, M. (2009). Scholarly hyperwriting: The function of links in academic weblogs. *Journal of the American Society for Information Science and Technology*, 30(1), 75-89.
- MAHRT, M., & PUSCHMANN, C. (in press). Schnittstelle zur Öffentlichkeit oder virtueller Elfenbeinturm? Inhaltsanalytische Befunde zur Leserschaft von Wissenschaftsblogs. In *Proceedings of the Conference on Public Science and New Media*. Karlsruhe: Zentrum für angewandte Kulturwissenschaft.
- PRIEM, J., & COSTELLO, K. L. (2011). How and why scholars cite on Twitter. *Proceedings of the American Society for Information Science and Technology*, 47(1), 1-4.
- PRIEM, J., & HEMMINGER, B. M. (2010). Scientometrics 2.0: Toward new metrics of scholarly impact on the social Web. *First Monday*, 15(7). Retrieved November 11, 2012 from <http://tinyurl.com/2akb3kw>
- PRIEM, J., TARABORELLI, D., GROTH, P., & NEYLON, C. (2010, October 26). Altmetrics: A manifesto. Message posted to <http://altmetrics.org/manifesto/>
- PUSCHMANN, C. (2010). *The corporate blog as an emerging genre of computer-mediated communication: Features, constraints, discourse situation*. Göttingen: Universitätsverlag Göttingen.
- REHER, S., & HAUSTEIN, S. (2010). Social bookmarking in STM: Putting services to the acid test. *Online*, 34(6), 34-42.
- REINMANN, G. (2008). Lehren als Wissensarbeit? Persönliches Wissensmanagement mit Weblogs. *Information Wissenschaft & Praxis*, 59(1), 49-57. Retrieved November 11, 2012 from <http://tinyurl.com/bf3z4vr>
- SHEMA, H., BAR-ILAN, J., & THELWALL, M. (2012) Research blogs and the discussion of scholarly information, *PLoS One*, 7(5), e35869.
- SMITH, A. G. (2004). Web links as analogues of citations. *Information Research*, 9(4). Retrieved October 3, 2012 from <http://informationr.net/ir/9-4/paper188>
- STOCK, W. G. (1994). *Wissenschaftsevaluation: Die Bewertung wissenschaftlicher Forschung und Lehre*. Munich: ifo Institut für Wirtschaftsforschung. Retrieved October 3, 2012 from <http://tinyurl.com/am68ocr>
- STOCK, W. G. (2001). *Publikation und Zitat: Die problematische Basis empirischer Wissenschaftsforschung*. Köln: Fachhochschule Köln. Retrieved October 3, 2012 from <http://www.fbi.fh-koeln.de/institut/papers/kabi/volltexte/band029.pdf>
- TAGUE-SUTCLIFFE, S. (1992). An introduction to informetrics. *Information Processing & Management*, 28(1), 1-3.
- THELWALL, M. (2008). Bibliometrics to webometrics. *Journal of Information Science*, 34(4), 605-621.
- THELWALL, M., KOUSHA, K., WELLER, K., & PUSCHMANN, C. (in press). Assessing the impact of online academic videos. In G. WIDEN, & K. HOLMBERG (Eds.), *Social information research*. Bingley: Emerald Publishing.

- WELLER, K., DRÖGE, E., & PUSCHMANN, C. (2011). Citation analysis in Twitter: Approaches for defining and measuring information flows within tweets during scientific conferences. In M. ROWE, M. STANKOVIC, A.-S. DADZIE, & M. HARDEY (Eds.), *1st workshop on Making sense of Microposts (#MSM2011)* (pp. 1-12). Retrieved October 3, 2012 from [http://ceur-ws.org/Vol-718/paper\\_04.pdf](http://ceur-ws.org/Vol-718/paper_04.pdf)
- WELLER, K., & PUSCHMANN, C. (2011, June 14-17). Twitter for scientific communication: How can citations / references be identified and measured? Paper presented at the Web Science Conference, Koblenz. Retrieved October 3, 2012 from [http://journal.webscience.org/500/1/153\\_paper.pdf](http://journal.webscience.org/500/1/153_paper.pdf)

# **Google Scholar versus Google Scholar: Among Publish or Perish, Scholarometer, and My Citations, Which Citation Count Tool Is Telling Which Truth?**

**Ulrich “Tibaut” Houzanme**

School of Library and Information Science, Indiana University Bloomington

*Three tools that use Google Scholar (GS) as a data source were used to identify citations from 30 of the most influential information scientists (15 from the U.S. and 15 from the UK). Scholarometer is the best tool to recommend overall. Though My Citations ranked second overall, it is the best tool when the data are available. Publish or Perish was clearly the least effective, with numerous author name disambiguation and discipline categorization problems.*

## **Introduction**

Citation or reputation computing relies not only on formulas and sources of the data, but also on the tools used to harness such data. A recent addition to commercial citation sources such as Web of Science (WoS) and Scopus, Google Scholar (GS) has continued to improve its coverage and effectiveness. As its reputation has increased, it is more frequently used, not just to complement WoS and Scopus, but also as a viable alternative. Its identified advantages over WoS and Scopus (e.g., languages, breadth of coverage, and far more topics and sources of publications) and its free availability on the Internet make it an attractive source for citation harvesting (De Sutter & Van Den Oord, 2012). Applications are built to harvest GS data and present them

with some sort of computing, sophisticated metrics, and, of course, different features and interfaces. With regard to this, the three most prominent applications that use GS as a data source are Publish or Perish (PoP), Scholarometer, and Google's own My Citations service. For all three, GS is the only source of their citations data.

PoP, as one of the very first tools to support citation count using GS data, has been around since 2007 (Harzing, 2010). It is a downloadable application that works on the Microsoft Windows platform. Once installed, the user can query GS and obtain total citations, lists of publications, h-indices, and other measurements. Specifically, it provides author, journal, general citation, and impact measures along with a multi-query center. Seven main subdivisions of scientific fields are available for selection to reflect the field of inquiry. The software manufacturer notes, however, that subject areas selection is not functioning, due to an upgrade of GS, as Tarma Software (2012) reported:

In May 2012 GS redesigned its interface and integrated the advanced search page in its general search page. In doing so it removed the option to select specific subject areas. As a result subject area filtering is now no longer possible, neither in GS, nor in Publish or Perish. ("Subject area selection no longer functional," 2012, June 29)

The major consequence of this is the fact that no discipline-specific search is available, and all fields' search is included, which will amplify and skew the result. The hurdles then, using PoP, are time consumption and disambiguation issues inherited from GS data, issues that have not been fixed by the software.

Scholarometer ([www.scholarometer.indiana.edu](http://www.scholarometer.indiana.edu)) "provides a service to scholars by computing citation-based impact measures" (Hoang et al., 2010) that use a crowdsourcing approach, whereby researchers can contribute to building an emergent semantic network that allows the study of interdisciplinary annotations and trends. Scholarometer is a browser add-on for Google Chrome and Firefox that offers a wide range of citation analysis computing in categorized disciplines.

GS's My Citations ([www.scholar.google.com/citations](http://www.scholar.google.com/citations)), also called GS Author Citation Tracker, or GSACT, has been evaluated by Jacsó (2012), who found that it provided too little improvement too late over GS. Authors sign up with a Google account and, once their e-mail account is verified, they can manage their publications lists by accepting or rejecting works the system suggests that they have published. The author has the option to make the profile public. This leads to availability of citations and other measures (h-index, i-index, etc.); this approach depends directly on the researcher's willingness to share. When publications and citations are made public and linked

to researchers' profiles, other members of the scientific community can examine and dispute any misplaced publication. This could lead to more transparency and sustained honesty. The author is also responsible for the many tasks required at the data cleansing stage of citation collections, solving many verification and disambiguation issues debated at length by Smalheiser and Torvik (2009).

Perhaps worthy of mention is the Research Impact Evaluation Tool, or ResEval (Imran et. al., 2009), which was not accessible when this analysis was conducted, and thus was excluded (<http://project.liquidpub.org/reseval/>).

Overall, Meho and Yang (2007) have identified a strong link between the data source (WoS, GS, Scopus) and the citation count. In this study, the data source is the same (GS), but the tools are different. One would expect that a same author would have the same citation count, regardless of the application tool. Do these three tools actually produce equivalent results? The main question investigated here, however, is which GS citation analysis application is the best.

## Literature Review

The literature review focuses on two aspects: The sources of citation indexing and the tools' basic requirements in terms of author name disambiguation.

### Sources of Citation Index

The prominent research citation sources (WoS, Scopus, GS) have been used either in confirmation of, comparison with, or as complements to one another (Jacsó, 2005a; Meho & Yang, 2006, 2007; Bar-Ilan, 2008, 2010). Meho and Yang (2007) uncovered a significant overlap among the three main data sources (58.2% overlap between Scopus and WoS and only 30.8% overlap between the union of the previous two and GS), which shows that 48.3% of the literature indexed in GS is available neither in WoS nor in Scopus.

GS has been available since 2004, WoS, as an online searchable database, since 2002, and Scopus since 2004. Over the past decade, many researchers have compared them. Not only has the content of GS been found to be deeper and broader, it has also been found to be "scholarly" (of research quality), in comparison to commercial databases available through libraries (Howland et. al., 2009; Kousha & Thelwall, 2008). However, as it came about only in 2009, Microsoft Academic Search (<http://academic.research.microsoft.com>) has probably not had enough time for similar review, and little is known of its comparative depth and breadth of coverage. Initial reports, however, sug-

gest that its interface is more appealing and its results are cleaner. Nevertheless, Microsoft Academic Search was not included in this analysis because it does not use GS as a data source.

This research focuses solely on GS, based on evidence in the literature of its broad coverage (William, 2008; Bornmann et al., 2009; De Sutter & Van Den Oord, 2012). Some researchers report that GS lacks accuracy, even as they praise its comprehensiveness (Garcia-Perez, 2010; Beel & Gipp, 2010). In any case, GS is subscription-free (unlike WoS and Scopus) and is widely available to researchers on the Web. In addition, it is viewed as an indispensable tool in terms of compiling comprehensive, complete, and fair citation counts, as a complement to WoS and Scopus (Meho & Yang, 2007; Bar-Ilan, 2008; Garcia-Perez, 2010; De Sutter & Van Den Oord, 2012). There remains the question of whether GS should be used alone or alongside other data sources.

### **Author Name Disambiguation and the Tools**

For the tools to be considered of high quality, they must address the major issues related to literature attribution to the right authors. Smalheiser and Torvik (2009) have identified in their landmark study four challenges with author name disambiguation: 1) a single author publishing under different names, including orthographic variants and misspellings, name change over time for social, religious, gender, or other reasons, 2) different authors carrying the same name, 3) lack of metadata such as nationality, birth date, and so on, to help disambiguate authors, 4) multi-authors, multi-disciplinary, and multi-institutional publications / collaborations make it hard to identify all authors or determine the right discipline of the publication, or the most deserving institution of the collaboration. The focus of this article will be on discovering the tools that directly or indirectly incorporate these challenges in their design or approach for accurate publications and citations counts, which, in turn, have an impact on the ranking of the tools.

### **Research Question**

In light of the tools available and the context of this inquiry, what makes the difference when all tools share the same GS data source? In this particular case, it would not be unreasonable to expect the same result for citation count when the tools share the same data. However, there is a difference between the results, which leads us to the following questions:

- Is the difference attributable to limitations in the data available to the applications?
- Do the tools disambiguate authors effectively?
- Do differences in tool features / approaches influence the results?

These questions will be addressed using researchers in library and information science (LIS) as the test case. The results are therefore not generalizable to all disciplines and cannot help answer questions such as what causes the difference for every discipline or even which factors explain the difference within the same discipline. Previous research has also focused on comparisons within a single discipline because of the complexity of citation data and the available data sources.

Though the researcher will not undertake extensive disambiguation and data cleansing, he or she will use a relatively significant statistical sample that provides a meaningful conclusion. The underlying intention is to assess how well the tools perform with very little human intervention. Similar studies can be replicated in other disciplines to help compare the findings across fields.

## **Methodology**

Previous studies that compared results of citation analysis have focused on both the data sources and the tools used, their features, interfaces, and error rates. All those variables had an impact on the conclusions (Jacsó, 2005b, 2005c; Meho & Yang, 2006, 2007; Bar-Ilan, 2008; Howland et al., 2009; Bornmann et al., 2009; Bar-Ilan, 2010). Of special interest are studies and methods that focused on a single source or the citation index as a secondary data source (Davis & Shaw, 2011).

As a preliminary attempt to compare GS applications, this study focused on LIS researchers. The objective is to explore the comparability of the applications, not to uncover general trends in all fields of study. The three applications analyzed will be ranked from most to least accurate, as they report the citation counts for the 30 most influential scholars in LIS.

## **Study Sample and Its Significance**

A purposive sampling method has been used, and one that spans two continents and presents a longitudinal case study (Choemprayon & Wildemuth, 2009): In this case, the study examines citations of the work of five of the

most influential LIS researchers in the United States and five from the United Kingdom. In using these scholars, who were identified in previous studies as among the most prolific, the expectation is to provide enough data that could potentially be indexed by GS and thus to provide enough data to the applications. The somewhat limited sample in this pilot study does not carry enough weight to draw conclusions about the impact of these scholars' research, but it will help elucidate the general characteristics of the tools that could inform more ambitious studies. Using the three GS applications, lifetime citation counts were extracted on September 25, 2012 for Nicholas J. Belkin, Tefko Saracevic, Marcia J. Bates, Christine Borgman, Blaise Cronin, Gary Marchionini, Raya Fidel, Katherine W. McCain, Amanda Spink, Howard D. White, Michael K. Buckland, John M. Budd, Andrew Dillon, Peter Hernon, and Carol C. Kuhlthau (Cronin & Meho, 2006), and for Peter Willett, Stephen Robertson, Mike Thelwall, David Ellis, Nigel Ford, Maurice Line, Tom Wilson, Keith van Rijsbergen, Cyril Cleverdon, Stevan Harnad, Michael Lynch, Brian Vickery, A. E. Cawkell, David Bawden, and Jack Meadows (Oppenheim, 2006).

## **Results and Discussions**

The analysis focused on the evaluation of the tools' conformance with literature recommendations, on the practical side of the tools' evaluation that includes the availability, accuracy, and intrinsic differences or similarities within the tools and effort level demanded by the tool, as well as on an overall consideration of all these factors.

### **Tools' Conformance Levels with Literature Recommendations**

My Citations includes most of the disambiguation challenges identified by Smalheiser & Torvik (2009), such as name variations suggestion, different names for the same author, identifying metadata, collaboration, and discipline weight. This could be explained by the fact that authors' self-identification, e-mail verification through institutional domain name, and assumed correction of data on her / his profile take care of most of the issues related to accuracy, except in the case of different identities related to the same author, which has not been solved yet. So, from a purely conceptual or theoretical standpoint, My Citations appears to meet more requirements than Scholarometer or PoP, and in second place is Scholarometer, because it takes care of some of these disambiguation issues, such as name variation suggestion and discipline categorization; it also provides some level of collaboration data. The third posi-

tion obtained by PoP, in theory, reflects the lack of efficiency with data accuracy. Whether these observations are confirmed in practice will depend on the assessment of data availability, accuracy, and the level of effort invested in the process. So, with practical experimentation in terms of citation harnessing, the performance of each of the three tools has been evaluated on the three grounds enunciated, and the following results, observations, and analysis ensued.

### Availability of Citation Data

Top 15 U.S. Information Scientists (Cronin & Meho, 2006)	PoP Ver. 3.7 # of Papers	PoP Ver. 3.8 # of Citations	GS My Citation # of Papers	GS My Citation # of Citations	Scholarometer # of Papers	Scholarometer # of Citations
Nicholas J. Belkin	186	8546			155	7705
Tefko Saracevic	236	8471	184	8260	147	7813
Marcia J. Bates	158	5692			236	11454
Christine Borgman	361	6114			260	5762
Blaise Cronin	355	4788			229	4499
Gary Marchionini	320	8079			252	7920
Raya Fidel	112	2627			76	2514
Katherine W. McCain	87	3109			71	3060
Amanda Spink	340	10002			264	9802
Howard D. White	991	30319	99	3718	108	4564
Michael K. Buckland	207	3660			81	1831
John M. Budd	485	7015			96	1401
Andrew Dillon	837	16380			201	5172
Peter Hernon	512	3654			171	2989
Carol C. Kuhlthau	108	5490			34	2212
Top 15 U.K. Information Scientists (Oppenheim, 2006)	PoP Ver. 3.7 # of Papers	PoP Ver. 3.8 # of Citations	GS My Citation # of Papers	GS My Citation # of Citations	Scholarometer # of Papers	Scholarometer # of Citations
Peter Willett	1000	31312	506	18031	224	17125
Stephen Robertson	1000	49338	39	130	112	3234
Mike Thelwall	287	6391	284	6487	283	6425
David Ellis	1000	108943			91	6967
Nigel Ford	298	8570	136	3272	116	3146
Maurice Line	500	2774			13	123
Tom Wilson	1000	37082			171	1762
Keith van Rijbergen	52	239			22	74
Cyril Cleverdon	76	2074			11	637
Stevan Harnad	569	12555	482	12297	204	10259
Michael Lynch	1000	104575			185	33958
Brian Vickery	282	5938			46	605
A.E. Cawkell	156	501			108	363
David Bawden	436	3651			237	2887
Jack Meadows	259	2670			130	1441

Figure 1. Results of citation count extraction from PoP, GS My Citations, and Scholarometer

As seen in Figure 1, Scholarometer and PoP have citation data available for each of the 30 information scientists. My Citations lacks citation reports on about 75% of the researchers. One possible explanation is that these established scholars have not signed up with My Citations yet. They may not perceive great personal benefits, although their participation would help the following generations benefit from access to their publications and profile on My Citations. PoP's limitation of 1,000 publications does not prove usable for prolific researchers. And as the number of publications is limited, so are the total citations obtained. Considering the fact that some data, even if further investigation is required, is better than no data at all, then with regard to citation data availability, Scholarometer is most recommendable, with PoP as a distant second choice. My Citations has some caching up to do.

### **Accuracy of Available Citation Data**

As was indicated above, the accuracy of the citation data is closely linked to the disambiguation challenges addressed with each tool. Selecting the appropriate discipline with PoP produced indifferent results, because it did not increase or decrease the total citations counts obtained when all disciplines were selected.

Also, when the most prolific information scientists are shown to have published between 300 and 600 titles, PoP's over 600 and potentially more than 1,000 publications show a crucial lack of disambiguation. For that reason, and for those addressed in the data availability analysis, PoP's citation count is not accurate, particularly for the most productive scholars, who may have authored over 1,000 publications.

In addition, some total citation counts (Figure 1), such as David Ellis's and Michael Lynch's, are way beyond the norm for most cited LIS scholars, which calls for caution. Also, the data flagged in the first shade of gray (Figure 1) shows a limitation of PoP data to 1,000 publications. As a consequence, both PoP's publication and citation count data appear inflated.

Scholarometer's citation counts (Figure 1) are trailing below almost every PoP count. Based on the observations above, Scholarometer has a more trustworthy status because of its output compared to that of PoP.

The somewhat limited output of publications and citations counts with My Citations, in comparison with Scholarometer's and PoP's data, could be explained by the disambiguation built into the application and the discipline-specific search it supports.

Only in one case (Peter Willett) did My Citation retrieve more data than Scholarometer, but the difference of almost the double of the total number of publications, coupled with about the same number of total citations, is pretty

apocryphal and deserves further investigations outside the scope of this research.

### **Intrinsic Similarities or Differences between the Tools Results**

This comparison of intrinsic differences would be useful for tool selection, whereby a backup or comparison tool would provide meaningful and different information not covered by the second tool or other tools retained or domesticated (Schroeder & Dimitrina, 2009). The ultimate benefit for such an approach would be to compare and contrast so as to be able to investigate the stark differences and why they exist.

Table 1. Spearman's correlation coefficient between citation counts from My Citations, PoP, and Scholarometer

	Spearman's rank correlation coefficient		
	# of Citations		
	GS My Citations	PoP Ver. 3.7	Scholarometer
GS My Citations	1	-0.085	0.942
PoP	-0.085	1	-0.257
Scholarometer	0.942	-0.257	1

The citation count results of all the tools, where there is an entry or data available for each of the three tools, show that Google Scholar and My Citations have a strong correlation (0.94). This means that using these two tools together would not yield much difference when the data are available and a choice of either one without the other would be acceptable.

However, PoP and Scholarometer's highly negative correlation (-0.257) suggest that the tools produce sufficiently different results, which can be used to contrast with one another and might yield enough material for comparison.

The negative contrast of PoP and My Citations (-0.085) also suggests a difference that can be contrasted and compared for accuracy verification. It is not, however, as strong a difference as between PoP's and Scholarometer's results, in a normal circumstance. This observation is impeded by the aforementioned fact that PoP has major flaws that can lead to a considerable waste of time trying to investigate. Because of this, this paper recommends using, for the sake of more meaningful results comparisons, My Citations and Scholarometer at the same time.

Because the tools share GS as a data source, using the two least correlated tools, in a truly acceptable norm, offers an opportunity for challenges and investigations that will likely reveal meaningful results.

## **Effort Level in the Use of the Tools for Citation Data Harvesting**

While the question of how "little effort" should be spent collecting citation data was asked, it is not a full usability question. And though no usability test and survey have been conducted, it is the author's report of personal experience that is the case here. Individual experience could vary with operating systems, terminal access, and software or application. This comment is of an individual nature and needs more investigation with a study along the lines of the specific reasons for the comments. This being said, the author finds the least difficulty with My Citations, as the data are readily available or not available; PoP also seems simple to use once the software is installed. Scholarometer appears simple, though the wait time and the inability to navigate away from the browser do not make for time-efficiency, particularly when one has many scholars' citation data to search. This reporting relates to the author's experiment with the tool, and it is not believed that the observation has a heavy weight. Therefore, further studies focusing on usability are suggested for a more informed conclusion and recommendations.

## **Overall Theoretical and Practical Evaluation**

Combining both theoretical and practical aspects of the tools evaluation, including missing data in My Citations, Scholarometer scores better and is preferable, but its data must still be disambiguated.

When all data are available on all the scholars, My Citations would be preferable because of the disambiguation built into it, though more investigation of the accuracy would be recommended to draw safer conclusions.

Though this paper can recommend the use of Scholarometer, I caution against a blind trust and exhort verification of data accuracy. In cases where the data are available in all three tools, I would give a slight edge to My Citations because of the prior conclusion, though I still would recommend investigating data accuracy, as well as verification of publication and citation integrity.

Overall, because of the complexity of the data, the missing data, and data inflation, it is not easy to declare a de facto winner, though Scholarometer is preferable considering all aspects at the same time. For that matter, circum-

stances are diverse and the performances can only be appreciated based on the specific circumstance.

## Conclusions

To date, no study, before this one, has endeavored to compare the accuracy and completeness of the three GS-based applications investigated. Comparing them to one another is equal to comparing GS to GS, with the differences existing in the application design and behavior.

And to summarize this experiment, by using three different applications for retrieving and counting GS citations, My Citations scores good theoretical points because of the two takes on validation (system and the author her/himself). However, it lacks entries for about 75% of the scholars, which renders its recommendation difficult.

PoP did a good job until Google changed the setting of GS, making PoP's results hard to use without a significant amount of time to disambiguate. Data limitations and, particularly, inflation render it not recommendable at this time.

Scholarometer, with built-in disambiguation features, despite wait time for results that lowers productivity in the case of a massive-scale project, with discipline selection choice, altogether provides more acceptable results with few citation counts over the chart to investigate.

In short, the overall evaluations of the disambiguation features, the accuracy of the data, the availability of citations, and to a lesser degree, the ease of use appear more favorable to Scholarometer.

A combination of Scholarometer's and My Citations' use, when the data are available, will provide enough meaningful contrast for sources of investigation.

These conclusions are not applicable in a context of researchers cheating the citation system (Labbé, 2010). Nevertheless, this experiment can be replicated in other disciplines, and the outcomes used in meta-studies can help determine overall trends and the most reliable tools. As Albion (2012, p. 1) concluded from a study on citation count rates in the field of education, "Valid comparisons depend upon the availability of discipline-specific benchmarks."

Longitudinal studies across disciplines and studies of other open access citation sources (Meho & Sugimoto, 2009), such as Microsoft Academic Search tool, will help further elucidate the value of the available tools. In this era of Web-delivered, language-independent, and no-fee citation data access and study, researchers have access to remarkable sets of tools to continue

applying the paradigm Sir Isaac Newton described in the 17th century: build on the shoulders of giants.

## Acknowledgments

The author expresses his gratitude to Professors Debora Shaw, for a useful gift of knowledge and precious time reviewing this paper; Cassidy Sugimoto, for discussions that have sparked my interest in the topic; Mike Thelwall, for advice and time, on short notice, in a busy summer 2012; Blaise Cronin, for strategic advice, Katy Börner and Howard Rosenbaum, for encouragement to pursue this topic; Staša Milojević, for guidance and patience. My thanks are also due to the anonymous reviewer(s), who helped with pertinent remarks, the German delegation to ASIS&T 2011 New Orleans (Danke schön!), and all my friends who facilitated my trip to Dusseldorf die Altstadt! Last, but not least, I thank Fausta for her support, proofreading, and patience, alongside Jaden and Jerick, for allowing me time away to deliver.

## References

- ALBION, P. R. (2012). Benchmarking citation measures among the Australian education professoriate. *The Australian Educational Researcher*, 39(2), 221-235.
- BAR-ILAN, J. (2008). Which h-index? A comparison of WoS, Scopus and Google Scholar. *Scientometrics*, 74(2), 257-271.
- BAR-ILAN, J. (2010). Citations to the "introduction to informetrics" indexed by WOS, Scopus and Google Scholar. *Scientometrics*, 82(3), 495-506.
- BEEL, J., & GIPP, B. (2010). Academic search engine spam and Google Scholar's resilience against it. *Journal of Electronic Publishing*, 13(3). Retrieved November 11, 2012 from <http://tinyurl.com/bkebjhh>
- BORNMANN, L., DANIEL, H.-D., MARX, W., SCHIER, H., RAHM, E., & THOR, A. (2009). Convergent validity of bibliometric Google Scholar data in the field of chemistry: Citation counts for papers that were accepted by *Angewandte Chemie International Edition* or rejected but published elsewhere, using Google Scholar, Science Citation Index, Scopus, and Chemical Abstracts. *Journal of Informetrics*, 3(1), 27-35.
- CHOEMPRAYON, S., & WILDEMUTH, B. M. (2009). Case studies. In B. M. WILDEMUTH, *Applications of social research methods to questions in information and library science* (pp. 54-59). Westport, CT: Libraries Unlimited.

- CRONIN, B., & MEHO, L. (2006). Using the h-Index to rank influential information scientists. *Journal of the American Society for Information Science and Technology*, 57(9), 1275-1278.
- DAVIS, C. H., & SHAW, D. (Eds.) (2011). *Introduction to information science and technology*. Medford, NJ: Information Today.
- DE SUTTER, B., & VAN DEN OORD, A. (2012). To be or not to be cited in computer science. *Communications of the ACM*, 55(8), 69-75.
- GARCIA-PEREZ, M. A. (2010). Accuracy and completeness of publication and citation records in the Web of Science, PsycINFO, and Google scholar: A case study for the computation of h-indices in psychology. *Journal of the American Society for Information Science and Technology*, 61(10), 2070-2085.
- HARZING, A.-W. (2010). *The publish or perish book: Your guide to effective and responsible citation analysis*. Melbourne: Tarma Software Research Pty Ltd.
- HARZING, A.-W., & VAN, W. R. (2008). Google Scholar as a new source for citation analysis. *Ethics in Science and Environmental Politics*, 8(1), 61-73.
- HOWLAND, J. L., WRIGHT, T. C., BOUGHAN, R. A., & ROBERTS, B. C. (2009). How scholarly is Google Scholar? A comparison to library databases. *College and Research Libraries*, 70(3), 227-234.
- JACSÓ, P. (2005a). Google Scholar: The pros and the cons. *Online Information Review*, 29(2), 208-214.
- JACSÓ, P. (2005b). Comparison and analysis of the citedness scores in Web of Science and Google Scholar. *Lecture Notes in Computer Science*, 3815, 360-369.
- JACSÓ, P. (2005c). As we may search: Comparison of major features of the Web of Science, Scopus, and Google Scholar citation-based and citation-enhanced databases. *Current Science*, 89(9), 1537-1547.
- JACSÓ, P. (2008). The pros and cons of computing the h-index using Google Scholar. *Online Information Review*, 32(3), 437-452.
- JACSÓ, P. (2011). Google Scholar duped and deduped: The aura of “robometrics.” *Online Information Review*, 35(1), 154-160.
- JACSÓ, P. (2012). Google scholar author citation tracker: Is it too little, too late? *Online Information Review*, 36(1), 126-141.
- KOUSHA, K., & THELWALL, M. (2008). Sources of Google Scholar citations outside the science citation index: A comparison between four science disciplines. *Scientometrics*, 74(2), 273-294.
- LABBÉ, C. (2010, April 14). Ike Antkare one of the great stars in the scientific firmament. Retrieved November 11, 2012 from <http://tinyurl.com/a7tj69>
- MEHO, L. I., & SUGIMOTO, C. R. (2009). Assessing the scholarly impact of information studies: A tale of two citation databases—Scopus and Web of Science. *Journal of the American Society for Information Science and Technology*, 60(12), 2499-2508.

- MEHO, L. I., & YANG, K. (2006). A new era in citation and bibliometric analyses: Web of Science, Scopus, and Google Scholar. Retrieved October 5, 2012 from <http://arxiv.org/pdf/cs/0612132v1.pdf>
- MEHO, L. I., & YANG, K. (2007). Impact of data sources on citation counts and rankings of LIS faculty: Web of Science versus Scopus and Google Scholar. *Journal of the American Society for Information Science and Technology*, 58(13), 2105-2125.
- OPPENHEIM, C. (2007). Using the h-index to rank influential British researchers in information science and librarianship. *Journal of the American Society for Information Science and Technology*, 58(2), 297.
- SCHROEDER, R., & DIMITRINA, S. (2009). Social scientists and the domestication of e-research tools. Retrieved November 11, 2012 from <http://tinyurl.com/beqx9kt>
- SMALHEISER, N. R., & TORVIK, V. I. (2009). Author name disambiguation. *Annual Review of Information Science and Technology*, 43(1), 1.
- TARMA SOFTWARE (2012, June 29). PoP user's manual: Subject area selection no longer functional. Retrieved November 11, 2012 from <http://tinyurl.com/aap4q4u>
- WILLIAM, H. W. T. (2008). Google Scholar search performance: Comparative recall and precision. *Portal: Libraries and the Academy*, 9(1), 5-24.

# The H-Index: What Is It, How Do We Determine It, and How Can We Keep Up With It?

**Timothy Ireland, Kathy MacDonald, and Peter Stirling**

University of Waterloo, Canada

*Much research has been undertaken about the h-index. What started out as a simple compound metric based on an individual's publications and citation counts has the potential to become increasingly complex and difficult to measure. We outline a simple but effective step-by-step process for creating comprehensive citation counts of an author's publications, and subsequently a more robust and accurate h-index based on results combined from multiple sources.*

## Introduction: What Is It?

In 2005, a physicist named Jorge E. Hirsch developed a simple premise in an effort to quantify the scientific output of an individual researcher.

I propose the index  $h$ , defined as the number of papers with citation number  $\leq h$ , as a useful index to calculate the scientific output of a researcher. (Hirsch, 2005, p. 16569)

To easily determine the h-index of a researcher, examine the number of times each paper has been cited and put them into descending order. Thus, if an individual has eight papers that have been cited 33, 30, 20, 15, seven, six, five, and four times, the individual's h-index would be six. The first paper, 33, gives a one—one paper has been cited at least once. The second paper gives a two—two papers have been cited at least twice. The third paper gives a three, and we continue all the way up to six with the sixth highest paper.

Tokar, A., Beurskens, M., Keuneke, S., Mahrt, M., Peters, I., Puschmann, C., van Treeck, T., & Weller, K. (Eds.). (2012). *Science and the Internet* (pp. 237-247). Düsseldorf: Düsseldorf University Press

The final two papers have no effect in this case because they have been cited less than six times.

Hirsch based his argument on the premise that the h-index is useful for comparing different researchers in similar fields. If the h-index is similar for the two people, their overall influence in the scientific field is similar, independent of the number of papers written or the number of overall citations. Likewise, an individual's h-index should increase linearly over time. It should be noted that the h-index is not the sole indicator of an individual's research impact and that its value varies between disciplines. Those disciplines not heavily invested in journal article publication and citation metrics as a measure of impact may find the h-index less useful.

The simplicity and ease of use / understanding led to the h-index metric being included in Thompson Reuters' ISI Web of Science (WoS) and Elsevier's Scopus "less than two years after its formation" (Zhang et al., 2011).

In the advancement and promotion process for faculty members at many Canadian universities, scholarly output is often an influential factor in determining tenure and promotion. Gathering citation counts for every article, conference proceedings, book chapters, and patents can be daunting. The added challenge is that the same article may be indexed in multiple databases in which some citing articles are the same between databases and some are unique.

WoS and Scopus collect and organize citation counts and can calculate an individual's h-index. Google Scholar does it via Google Scholar Citations. However, each source may determine a different value of the h-index for each individual. Sometimes the variation in the h-index between sources can be large. A person could take the highest citation or h-index counts from one of these databases and use them in tenure and promotion documentation or grant applications, but it may not be a full accounting of a person's h-index and may not be as accurate as it could be. Combining citation counts from various research databases gives a larger citation count and therefore a higher h-index. One can do this in such a way that it can be self-sustaining in terms of maintaining up-to-date citation counts and therefore h-index and provide documentation / proof of citation counts and h-index calculation. This paper outlines a step-by-step process on how to do this.

## **Brief Literature Review**

There has been much research into the h-index and its variations. A recent review of the literature related to the publication, testing, and popularity of the h-index was summarized in an article by Zhang et al. (2011). The litera-

ture also points to the creation of variants of the h-index (such as the presence or absence of self-citations) to improve the metric. Likewise, the h-index has been applied to researchers from various fields and countries such as optometrists in Australia (Efron & Brennan, 2011), earth sciences (Mikki, 2010), psychology (Bador & Lafouge, 2011), chemical engineering (Prathap, 2011), medicine (Sanni & Zainab, 2011), and information science and library science (Levitt & Thelwall, 2009; Li et al., 2010; Meho & Yang, 2007; Oppenheim, 2007). There is extensive research into the calculation of an individual's h-index on different databases, as well as research into combining the results of databases (García-Pérez, 2010).

The research has suggested three main databases that should be used to determine the h-index of an individual scholar: WoS, Scopus, and Google Scholar (GS) Citations. Each provides an h-index based solely on the information it contains (content indexed), and as such is influenced by the strength and weakness of each database (Bar-Ilan, 2008). Substantial research has been done comparing the results of these three databases. Likewise, there has been considerable research into determining the accuracy of the h-index within each of these three databases. It has been established that an accurate h-index usually requires a compilation of multiple indexes (García-Pérez, 2010; Jacsó, 2008; Meho & Yang, 2007).

## **Question: How Do We Determine It?**

### **Determining the H-Index Using Extensive and Exhaustive Searching**

There is established methodology for creating de-duplicated and federated searches from databases to determine a more accurate h-index for an individual.

One study investigated 25 library and information science faculty members. The reported time for the project was over 3,300 hours (Meho & Yang, 2007), averaging about 132 hours per person. Overall, it would have taken a single person well over a year to complete the project. By the end of the project, the information would likely be out of date because some citations would have been added during the completion time. The study is comprehensive; however, most researchers would be unwilling or unable to dedicate that amount of time to determining their own h-index.

Could there be an easier, faster process that would not become outdated as soon as it was complete?

## Challenges

A quick search using the established h-index databases (WoS, Scopus, and GS) reveals obvious differences in the results of the h-index for three researchers associated with the University of Waterloo (see Table 1). (The CVs of the individuals were provided to ensure accuracy.) If one of these researchers were to ask what their h-index is, what is the correct reply? Why are the numbers so different?

Table 1. A comparison of articles, total citations, and h-indices from WoS, Scopus, and GS for three researchers on June 15, 2012

	<b>scholar 1</b>	<b>scholar 2</b>	<b>scholar 3</b>
field of study	physics	psychology	psychology
first article published	2002	2003	1986
items in WoS	15	16	158
total citations	109	86	4032
WoS h-index	6	6	33
items in Scopus	19	23	146
total citations	106	247	3960
Scopus h-index	5	8	31
items in GS	43	28	195
total citations	177	544	7572
h-index in GS	8	11	40

In Table 2, comparing information in Gold Rush (i.e., a service that can be used to compare the holdings of various databases), we note that there is a substantial difference in the journals included within WoS and Scopus (GS information was unavailable). The Gold Rush search confirms that searching additional research database(s) adds the potential of finding overlooked unique journals not included in WoS or Scopus.

Within the current academic environment, research is no longer encompassed within a single subject area; it is more interdisciplinary. For researchers considered in Table 1, the addition of other databases in physics and psychology may yield additional citations. A search in multiple databases, indexing both the subject and specific journals, will increase the probability of finding additional citations. Ulrich's Web (i.e., a database that lists information about journals) provides information on which databases a specific journal may be indexed. Importing citations from multiple databases will also create duplicates, which will need to be managed carefully. The more suc-

cessful (as a measure of both publications and citations) an individual is, the more challenging this task becomes.

In some situations, author order and publication format are important, and there has been research into this (Levitt & Thelwall, 2009). More prestige may be given to authors who fall toward the front of the author listing. Likewise, what should be included as a citation may vary.

Table 2: Gold Rush comparison of WoS and Scopus on June 19, 2012

	<b>Unique Journals</b>	<b>Similar Journals</b>	<b>Total Journals</b>
WoS	2828	13053	15881
Scopus	14453	13053	27506

## How Do We Keep Up With It?

### Outline of Our Citation Tracking System

Developing research skills to determine individuals' scholarly impact may be essential to advancing an academic career (Hirsch, 2007). We now describe our methodology for citation tracking and determining a suitable h-index to be used for tenure and promotion applications, as well as tracking an individual's personal research influence. The process we outline can be set up easily for graduate students and faculty in the early years of their careers. Faculty members with many publications and a high h-index will find the process more time-consuming. The value of this citation tracking system is that it:

- creates a current list of all the academic output for an author, which may increase the accuracy of attributable citations;
- tracks author order and type of material published (for example, first author in a peer-reviewed journal);
- lists the author's citations in a transparent manner. This information can then be made publicly accessible;
- provides proof of combined citation counts and h-index calculation as each publication and all citing publications are collected and organized. This information can then be made publicly available;
- collects papers that cite their work so authors can monitor their impact and identify potential collaborators or competitors; and
- uses article citation alerts, automated e-mail, and RSS notification, making it easy to keep the database current.

## **Overview of the Process**

The following steps outline our process for collecting and de-duplicating citations from multiple databases to calculate a more accurate h-index.

### **Step 1: Set Up the Author's Personal H-Index Tracking Database(s)**

Using citation management software (CMS), set up two databases (the following process uses RefWorks). The first database (DB1) will include all of the academic output to be tracked. The second database (DB2) will contain all the citations to each of the works contained in DB1. An alternative option is to keep both academic output and citations of those works in the same RefWorks account using the folder level to collect author publications by publication type and the subfolder level to contain the citing works. For clarity, this paper will describe only the two database option.

### **Step 2: Populate DB1**

Identify which research databases have citation index functionality and contain the author's publications. Some databases with this functionality include WoS, Scopus, PsycINFO, Social Science Research Network (SSRN), GS, etc. If the researcher is compiling a small number of publications, individual titles can be searched. However, a more useful strategy is to use the author finder search option. For example, in WoS, there is a tab with an author search / author finder option that refines by field, institution, and date range to narrow the author sets to a reviewable number. Scopus has a similar tab.

If searching a very common name results in too many author sets to sort efficiently, or if the name is not listed in the author sets, search for last name combined with each article title, or name and topic(s). If necessary, repeat the search using last name and the research topic relevant to the author's work.

A research profile can be set up with some research databases. This helps database indexers identify authors and thereby increases the accuracy of their citations. For example, WoS has a profile system called Researcher ID under the additional resources tab. Scholar Universe is another independent researcher profile system. These methods help databases credit authors with appropriate citations.

Once all the relevant articles are identified, the citations need to be imported into DB1.

### **Step 3: Populate DB2**

DB2 contains all the relevant citations to all of the author's works in DB1. To track the citations, folders and subfolders will be used. The folders are labelled according to the publication types that are relevant for the creator's intended use. For example, most tenure and promotion documentation requires a tally of the citation counts for all refereed papers and separate citation counts for first-author papers compared with co-authored papers. The structure includes one folder each for all refereed first-author papers, for all refereed co-authored papers, for all conference proceedings, for book chapters, one for books, for patents, and for other publications such as government reports, white papers, thesis, dissertation, audio-visual modules, submissions to royal commissions, etc.

Subfolders can then be created within the publication type folders for each article, book, book chapter, conference proceeding, or other publication. For example, in the first-author folder, one should have a subfolder designated for each first-authored article. Each subfolder is labeled by the title of the publication. It may be beneficial to add the year if there are articles with similar titles published over a number of years. If useful, include a folder for posters, invited speaker presentations, patents—whatever is relevant for the discipline and stage of academic advancement. An efficient method of creating subfolders is to copypaste the title or the first part of the title of the article from DB1.

### **Step 4: Import the Citations into the Appropriate Subfolder in DB2**

Return to an appropriate research database and do two things. First, identify the number of times the article, book, or conference proceeding has been cited in that database and export those citing articles into the designated subfolder for that particular article, book chapter, etc. Second, set up citation alerts for all publications in DB1 in each of the research databases used. If citation alerts are unavailable, saved searches may be available. Look for a “set alert,” “e-alert,” or “feed (RSS)” button. RSS feeds can be incorporated directly into a RefWorks account. The use of citation alerts allows for a relatively self-sustaining method for tracking.

If using GS, set up a direct export to RefWorks within Scholar Preferences or use RefGrab-It (i.e., a feature of RefWorks that allows citations on a Web page to be imported into RefWorks) to import multiple GS records. When reviewing records found in GS, consider what should be included in RefWorks. Consideration should be given to including or excluding certain

items, such as papers in other languages, advertisements, fact files, peer-reviewed trade publications, Web pages, posters, etc. GS does pick up citations to books, government publications, and white papers, etc. that may not appear in traditional databases, but nevertheless demonstrate the impact of a scholar's work.

### **Step 5: Move to the Next Appropriate Database and Repeat Steps 2 through 4**

### **Step 6: Add Outstanding Items to DB1 and Add Originals if Desired**

It is possible that not all of an author's academic output appears in a database, e.g., a non-governmental organization publication. These publications can be manually entered into RefWorks. RefWorks also offers the option to add attachments to citations. If desired, a text file, video clip, PDF, or other attachment may be added to the citation.

### **Step 7: Remove Unwanted Duplicates**

One of the functionalities of RefWorks is the ability to identify duplicates automatically. DB1 should be free of all duplicates.

DB2 is a little more difficult in that duplicates are allowed in different subfolders. It is permissible to receive a citation for more than one item at the same time. For example, an article may cite several different articles from the same author with a single paper. Thus, duplicates are allowed, but not within the same subfolder.

### **Step 8: Harvest the Results**

Determine final citation counts for each type of publication using the organize folders view, that is, all A1 (first-author), all CP (conference proceedings), etc. These numbers can then be transferred to tenure and promotion documentation, grant proposals, etc., or ranked to determine one's h-index.

### **Step 9: Keep the Database Up-to-date**

Monitor e-mail or RSS alerts and add new articles citing the author's work to the appropriate citation subfolder to maintain a current record of citations for each publication, thus building the list of citations in preparation for an advancement process or other future use. Every time a new citation is added, the appropriate de-duplication process should be undertaken.

## **Step 10: Make Both DB1 and DB2 Public**

DB1 and DB2 can be published or shared online, using RefShare, and made available for downloading into other CMS packages. This may increase both the frequency of an author's work being cited, and help ensure it is cited correctly. It also offers a transparent look into the compilation of an individual's h-index.

## **Conclusions and Future Research**

To determine an individual's comprehensive h-index, searching multiple databases is required. A CMS streamlines the process immensely. Using a CMS creates a current list of all the academic output for an author. It builds a repository of publications with citations taken from one or multiple research databases. In doing so, authors have the opportunity to catch and correct any errors that database has made with respect to their publications, thus increasing the accuracy of citations to the author's work. The CMS contents can be shared by posting to a personal Web site or a departmental / university repository. The process allows for tracking various types of publications (first-authored papers, co-authored papers, patents, etc.) and their respective citation counts, which are required in most tenure and promotion documentation and some grant proposals. The author's citations can be presented in a transparent manner by making the citation database publicly accessible. The process also allows authors to collect papers that cite their work so authors can monitor their impact and identify potential collaborators or competitors. Through the use of article citation alerts, automated e-mail, and RSS notification, authors can keep their databases up-to-date and therefore have a current record of their citations counts and h-index.

There may be concerns about the creation of a system that offers different results than WoS or Scopus (or even GS). If an h-index from WoS and Scopus are seen as authoritative, it is not a stretch to look at combining the results of these two databases (and removing duplicate records). Our research has shown that authors with an h-index of 31 in WoS and 33 in Scopus ended up with a calculated index of 34 using our method.

Likewise, the inclusion or exclusion of a cited reference type will quite obviously affect the h-index of a researcher. The outstanding challenge is determining exactly what should be included and what should be excluded. Further research and discussion about the types of publications to be included in an h-index calculation is required.

## Acknowledgements

Our thanks are due to Sandra Keys and Annie Bélanger for assistance editing this paper.

## References

- BADOR, P., & LAFOUGE, T. (2011). Comparative analysis between impact factor and h-index for psychiatry journals. *Canadian Journal of Information and Library Science*, 35(2), pp. 109-121.
- BAR-ILAN, J. (2008). Which h-index? A comparison of WoS, Scopus and Google Scholar. *Scientometrics*, 74(2), pp. 257-271.
- EFRON, N., & BRENNAN, N. A. (2011). Citation analysis of Australia-trained optometrists. *Clinical and Experimental Optometry*, 94(6), pp. 600-605.
- GARCÍA-PÉREZ, M. A. (2010). Accuracy and completeness of publication and citation records in the Web of Science, PsycINFO, and Google Scholar: A case study for the computation of h-indices in psychology. *Journal of the American Society for Information Science & Technology*, 61(10), pp. 2070-2085.
- HIRSCH, J. E. (2005). An index to quantify an individual's scientific research output. *Proceedings of the National Academy of Sciences of the United States of America*, 102(46), pp. 16569-16572.
- HIRSCH, J. E. (2007). Does the h-index have predictive power? *Proceedings of the National Academy of Sciences of the United States of America*, 104(49), pp. 19193-19198.
- JACSÓ, P. (2008). Testing the calculation of a realistic h-index in Google Scholar, Scopus, and Web of Science for F. W. Lancaster. *Library Trends*, 56(4), pp. 784-815.
- LEVITT, J. M., & THELWALL, M. (2009). The most highly cited library and information science articles: Interdisciplinarity, first authors and citation patterns. *Scientometrics*, 78(1), pp. 45-67.
- LI, J., SANDERSON, M., WILLET, P., NORRIS, M., & OPPENHEIM, C. (2010). Ranking of library and information science researchers: Comparison of data sources for correlating citation data, and expert judgments. *Journal of Informetrics*, 4(4), pp. 554-563.
- MEHO, L. I., & YANG, K. (2007). Impact of data sources on citation counts and rankings of LIS faculty: Web of Science versus Scopus and Google Scholar. *Journal of the American Society for Information Science and Technology*, 58(13), pp. 2105-2125.
- MIKKI, S. (2010). Comparing Google Scholar and ISI Web of Science for earth sciences. *Scientometrics*, 82(2), pp. 321-331.

- OPPENHEIM, C. (2007). Using the h-index to rank influential British researchers in information science and librarianship. *Journal of the American Society for Information Science & Technology*, 58(2), pp. 297-301.
- PRATHAP, G. (2011). Correlation between h-index, Eigenfactor™ and article influence™ of Chemical Engineering journals. *Current Science*, 100(9), p. 1276.
- SANNI, S. A., & ZAINAB, A. N. (2011). Evaluating the influence of a medical journal using Google Scholar. *Learned Publishing*, 24(2), pp. 145-154.
- ZHANG, L., THUIS, B., & GLÄNZEL, W. (2011). The diffusion of h-related literature. *Journal of Informetrics*, 5(4), pp. 583-593.



**PART SIX**

---

**THE INTERNET AND THE FUTURE  
OF SCIENCE**



# **Publishing against the Machine: A New Format of Academic Expression for the New Scientist**

**Adam Sofronijevic**

University of Belgrade, University library “Svetozar Markovic”

*This paper presents possibilities for profound transformation of academic communication. The changing role of humans in scientific communication is analyzed on the basis of ongoing technological developments. Machine analysis and production of scientific texts are discussed, and increasing efficiency in scientific communication is advocated.*

## **Introduction**

Communication is crucial for scientific work in two aspects: First, it allows for dissemination of new ideas and findings, thus making progress in science possible. Second, the resulting products of this communication in written form (e.g., journal articles) are the main evaluation tool for judging the entire scientific process behind them. This makes communication important to the scientific work, because only excellence in both fields leads to continuous scientific progress. It seems that in science, too, it is not only about who we are, but also about the clothes we are wearing.

One must remember that good clothing alone may not get one far in the scientific world, but also that one will not be allowed to walk naked. The tremendous importance of scientific communication today seems not to be matched with the same amount of interest, funding, and dynamics as the research area of scientific work is. While the corpus of world knowledge doubles with amazing dynamics, we see no such dramatic changes in scien-

tific communication. One can only imagine how it would be if every five or eight years we would have new forms of scientific communication that would allow for twice as efficient dissemination of knowledge and twice as good evaluation tools and methods. All kinds of amazing discoveries and breakthroughs, made on almost a daily basis, are communicated nearly in the same ways as 40, 50, or 100 years ago. While we have computers and the Internet, we are still stuck with humans writing essays word by word, explaining what they have discovered and thought. Machine-enabled help in text production is negligible when compared with machine-enabled help in research.

This may come as no surprise when one contemplates the complex nature of human communication. The strength of computing machines is in computing and thereby solving structured problems. Now, we are on the verge of a revolution that will enable machines to take over the burden of solving more sophisticated problems, also in the area of communication. The possibilities for pattern matching and pattern recognition of contemporary machines will continue to increase. This will make jobs deemed creative in the past available for machines. In order to benefit, scientists need to look into processes in other fields to make the transition as smooth as possible. Competing with machines in performing work in areas where machines excel humans, such as some fields of production and analysis of structured communications, will be inefficient.

## **Machines Are Taking Over: Other Industries**

Changes in technologies play a somewhat surprising role in our lives and in our business environment. To explain it, we need new perspectives and ideas. To master it for everyday purposes, we need new approaches, skills, and competences. Changes that may be dramatic in some industries are unnoticed in other fields due to the dynamics of the environment. The rush to cope with change occupies many available resources. Brynjolfsson and McAfee (2011) discussed the role of human labor in light of ongoing technological developments. More and more human jobs will be replaced by emerging technologies. They are capable of performing work tasks previously considered accomplishable only by human creativity. How will we manage new machines? Will entrepreneurs use machines to provide new work opportunities, replacing those destroyed? So far they have been successful in this task, and every new technological breakthrough brought not only the destruction of old ways, but also gave rise to new workplaces. A new dimension of technology improvement was discussed as early as 1963:

We are being afflicted with a new disease [...] technological unemployment. This means unemployment due to our discovery of means of economizing the use of labor outrunning the pace at which we can find new uses for labor. (Keynes, 1963)

Brynjolfsson and McAfee (2011) claim that these questions are more than ever central for the societies and economies of today. The pace at which technology change destroys workplaces is increasing. One must wonder whether new work opportunities will arise at an appropriate pace or whether high unemployment is inevitable. Certainly, people had to acquire new skills and competences after machines took over their old jobs. Some examples follow to illustrate the previous discussion.

Until a few years ago, vehicle driving was considered impossible to even be considered as a job suitable for machines. The complexity of traffic situations and the need for what was considered intuitive decision-making were deemed beyond the processing power of any machine (Levy & Murane, 2004). Nowadays, computers can drive vehicles autonomously. This puts millions of jobs, including five million truck drivers in the U.S. alone, in jeopardy (see Brynjolfsson & McAfee, 2011). Academic courses on programming autonomous driving vehicles are offered online (e.g., Artificial Intelligence (cs373) Programming a Robotic Car).

One has to wonder how many jobs in higher education will be replaced by computers. As of May 2012, grading robots can do a fantastic job providing timely and accurate grades for more than 90,000 students enrolled at Udacity—Introduction to Computer Science Course (CS101): Building a Search Engine. We see attempts at building on economies of scale, when, for example, MIT and Harvard join forces to form edX. With such a dynamic development environment, machines seem certain on winning at least some jobs in higher education with humans remaining important at those places where interaction and intuition are important in the educational process.

With jobs being taken over by machines in a diversity of industries, one may wonder if there is a general type of job in which machines excel. Licklider (1960) predicted these issues more than fifty years ago:

Man-computer symbiosis is an expected development in cooperative interaction between men and electronic computers. It will involve very close coupling between the human and the electronic members of the partnership. The main aims are 1) to let computers facilitate formulative thinking as they now facilitate the solution of formulated problems, and 2) to enable men and computers to cooperate in making decisions and controlling complex situations without inflexible dependence on predetermined programs. In the anticipated symbiotic partnership, men will set the goals, formulate the hypotheses, determine the criteria, and per-

form the evaluations. Computing machines will do the routinizable work that must be done to prepare the way for insights and decisions in technical and scientific thinking. Preliminary analyses indicate that the symbiotic partnership will perform intellectual operations much more effectively than man alone can perform them. Prerequisites for the achievement of the effective, cooperative association include developments in computer time sharing, in memory components, in memory organization, in programming languages, and in input and output equipment.

Thus, skills are discussed in the framework of a symbiotic human-machine collaboration. This clearly represents a plea for humans not to compete with machines. They should develop skills in which they excel and use machines for jobs in which they perform better to increase human productivity. A general conclusion can be derived from Licklider's early insight. The division of work among machines and humans is based on the complexity of the tasks involved. As machine resources grow stronger, more jobs will be given to them to perform. As the need for human activity shrinks, the pressure to develop skills in areas that are extremely complex grows. Thus, human intuition becomes more and more important.

The career of Steve Jobs, late CEO of Apple, Inc., provides an example for the importance of human intuition and its successful implementation for business results. Jobs always emphasized the importance of intuition in business decision-making and even refused to do market research for new products. He relied solely on his own intuitive insight into his customers' nature (Isaacson, 2011).

Levy and Murane (2004) provided a detailed list of skills in which humans or machines excel. Their prediction is that only the knowledge sector will provide new jobs for humans, including entrepreneurs, researchers, computer programmers, educators, and consultants. One thing these jobs have in common is that they require problem-solving skills for which there are no rule-based solutions. In spite of all improvements achieved or predicted, machines have not yet reached the necessary level of storage, speed, and processing power needed to cope with such tasks.

Complex communication is one area in need of significant improvements. Levy and Murane (2004) defined it as "interacting with humans to acquire information, to explain it, or to persuade others of its implications for action." They provide examples: a manager motivating the people whose work he / she supervises; a sales person gauging a customer's reaction to a piece of clothing; a biology teacher explaining how cells divide; an engineer describing why a new design for a DVD player is an advance over previous designs. They predict that machines will eventually supersede humans at jobs requiring application of deductive rules, such as arithmetic operations or boarding

pass recognizing, and application of inductive rules (i.e., pattern recognition), such as predicting a mortgage default or recognizing a spoken name. Basically, Levy and Murane (2004) claim that tasks requiring well defined structures that can be accomplished by following a set of rules will be taken over by the machines.

Another interesting insight shared by both Levy and Murane (2004) and Brynjolfsson and McAfee (2011) is that non-routine manual tasks will be an area where humans will be better than machines in the foreseeable future. Therefore, a labor market curve described by demand and level of education will not be linear anymore, but U-shaped, because besides demand for high-level experts, a huge demand for a human workforce will appear in the area of unskilled labor demanding execution of non-routine manual tasks.

The game of chess provides a striking example. Bordering science, arts, and sports, chess was long considered a human-only, creative activity. Everything changed at the end of the 20th century, when chess-playing machines managed to gain such dominance that duels between them and humans became boring. Not even the best human players could ever win against a strong, purposely built chess-playing machine. This changed the notion of creativity forever. The definition is changing, as machines take over fields that are considered creative and reserved for humans only. The most important concept for management that originated in chess is “free-style chess.” This is a tournament where players act as teams and use computers. Since the 2005 competition, tournaments have not been won by teams made up of grandmaster players or the most powerful machines. Instead, in 2005 a team of two human players of average ability using three average machines won the tournament (Rasskin-Gutman, 2009). Their competitive advantage over other teams was a highly optimized process, based on their knowledge of humans to organize themselves using machines to work as a seamless team. One could also say that they made best use of their intellectual capital in producing the best results in a competition based on strict rules and having a straightforward goal. Their experience goes a long way in explaining communication between humans and machines and their interfaces.

The reasons for rapid changes in labor profile demand and technology can be explained with the help of Moore’s law (Brynjolfsson & McAfee, 2011). The observation that the number of components on a chip doubles each year has held true for almost 60 years (Moore, 1965). The increase in complexity leads to an exponential rise in efficiency. The real speed and scope of improvements can be noticed only in later stages of their development. Kurzweil (2000) illustrated such improvements by a story about a prize offered by an emperor to the inventor of the chess game. The inventor asked for

a grain of rice on the first square of the chess board and twice that amount on the next square. For the first part of the board he received an amount equivalent to that raised at an average rice field, but in the second part of the chess board, the exponential rise showed its strength: The inventor ended up with a pile of rice equal in size to the Himalayas. Brynjolfsson and McAfee (2011) argue that technology development has entered “the second part of its chess board” and that dynamics will increase in the near future. As evidence, everyone can witness the online education revolution at [edxonline.org](http://edxonline.org) or [udacity.com](http://udacity.com). However, Cowen (2011) proposed a “technology plateau theory.” He believes that there are indeed counter-effects to the potentials for growth provided by current technologies.

## **Scientific Communication: Machine-Assisted Reading and Writing**

Scientific communication comprises two basic aspects or processes: reading and writing—or in terms that are more machine-friendly—analysis and production of the text. I will recapitulate the current state of affairs in several areas of machine-related text production and analysis.

Text analysis has recently become a very dynamic and productive field of research. More importantly, this field corresponds to the growing needs of researchers. One of the main problems has been defined in Takeshima and Watanabe (2010) as the difficulty for researchers to read and understand scientific papers effectually and effectively. To deal with this problem, many different ways of employing machine help have been devised. Takeshima and Watanabe (2010) focused on supporting the understanding process. They based their work on the fact that figures and tables reflect important contents of papers. Subsequently, they developed a method to extract sentences specific to figures or tables. Schafer et al. (2008) described methods for extracting interesting factual relations from scientific texts. The extracted relations are simplified, and the resulting “quirples” are stored in a database from where they can be retrieved by a relation-based search. More recently, Schafer and Kiefer (2011) described breakthroughs that have been made in deep parsing of long sentences. Such deep parsers provide the possibility to answer questions and explore definitions in the near future. Integration of annotation tools and natural language analysis tools can provide useful functions in text analysis and in preparing machine text production. Advances made in this area are diverse and significant (see Rupp et al., 2007). Considerable progress has also been made in accurate statistical parsing of realistic texts (Briscoe & Carroll, 2002) and even in finding predominant word senses

in untagged text (McCarthy et al., 2004). Recently, sentence fluency has been analyzed, and means to evaluate this important feature, especially when dealing with machine-produced texts, have been presented (Chae & Nenkova, 2009).

Certain fields have witnessed faster growth of machine involvement in text analysis and creation. One such field is medicine, where vast amounts of data make it impossible for human-only activities to be efficient. For instance, Cao et al. (2009) presented the subject of question answering that differs from information retrieval in providing summaries rather than lists of documents, thus saving users additional work. Grau et al. (2009) presented a solution to automatically extract knowledge from papers in a specific corpus of kidney-related scientific papers. Such extraction may be of great help in scientific areas where data are abundant.

Machine translation is a fast-growing field due to the possibilities for profit. Advances in machine analysis and production of text in conjunction to this field are considerable. I will mention just some of the advances. Zhang and Clark (2011) provided a model for the problem of word-ordering, which is one of the biggest obstacles to smoother machine-translated text. The problem of evaluating machine-produced translation is complex because of the need for automated evaluation. Automatic metrics such as BLEU fail to achieve satisfactory levels of correlation with human judgments at the sentence level. Kulesza and Shieber (2004) proposed a new class of metrics based on machine learning.

Advances in machine transliteration have also been made (e.g., Li et al., 2009). This particular area is of great importance for providing accurate synthesis of different affiliations in a large citation database, which is of enormous importance for evaluating scientific results.

## **Look into the Future: New Scientist and Communication**

It is hard to imagine the scientist of tomorrow who will not use all technological advances at hand to be more efficient and effective in scientific communication. Therefore, we must think that in the future, all kinds of machines will be at one's disposal, providing help in different aspects of scientific communication. How can they help? Based on the advances in text production and analysis described above, we may define several most likely possibilities.

In text analysis, machine readability of texts is of the highest importance. Therefore, such forms of text will be highly used. I will describe one of these forms, which is often called a "nano-publication." Such nano-publications

may provide for machine readability of the shortest possible scientific statements with possibilities for referencing. A nano-publication is a very short declaration connecting two concepts by means of a third and providing metadata about this relation—i.e., conditions under which the relation is viable, author, timestamp, etc. (Groth et al., 2010). Originating in the life sciences, nano-publications seem to be envisioned and increasingly shaped as a tool for efficient publication of datasets. Nano-publications are depicted in more detail in Mons and Velterop (2009). Beyond the advantages of machine readability and possibilities for referencing, nano-publications may also be important in providing an important field for human employment, especially in transitional and developing countries. Therefore, the use of this scientific communication tool has a twofold importance: On the one hand, nano-publications foster efficiency by using machine readability to their advantage, while, on the other hand, the need for referencing leaves sufficient incentives for the employment of human scientists in their production. In discussing alternative forms of publications, one must also consider incentives for publication (i.e., the potential rewards for publishing). One interesting idea in this area is to shape future scientific communication to make it more suitable for applying a micro-credit system (Casati et al., 2011). That system may involve more finely graded rewards for publishing and communicating advances to a general body of knowledge than the system that is in place today allows.

Machine translation is another area in which scientific communication could be improved if more machines were involved. Some areas are more prone to structured text forms and therefore will benefit more from advances in this area. Huge scientific communities in China, India, and Russia are on the rise. As English is not a native language to them, the enormous size of these communities makes their members more prone to intra-community communication than to dialog with the international community. In contemplating possibilities for machine translation usage in scientific communication, we have to consider the differences between languages used by people and machines. These languages will have to converge if machine translation is to be used on a wider scale. Recent findings (Branigan et al., 2010) suggest that there is already some convergence and that strong evidence for the alignment of human and machine languages are available in interactions recorded between humans and machines. Evidence from different areas of human and machine interaction suggest that there is a strong difference on the part of the humans in evaluating relations with other humans and with machines (Weibel, 2008). My opinion is that humans seek pragmatic results, especially in a down-to-earth area such as science. This may cause humans to adapt to the language style of machines and make use of the advantages this

may provide. The structured nature of language in most scientific fields (Ahmad, 2012) will favor this process.

Possibilities for using machines in scientific processes are numerous. Eureqa is a software tool that provides equations based on data fed to it. It identifies the simplest mathematical formulas describing the underlying mechanisms that produced the data. It is free to download and use.

Autonomous scientific discovery has been considered impossible without at least some human intervention. This seems to change too. King et al. (2009) reported on a laboratory robot that was created by the computational biology research group at Aberystwyth University. This machine is the first one in history to discover new scientific knowledge independently of its human creators. It achieves this by using techniques from artificial intelligence to automate all aspects of the scientific discovery process: generating hypotheses, designing experiments to test these hypotheses, running the physical experiments using robotic systems, analyzing and interpreting the resulting data, and repeating the cycle.

Another unexpected area in which machines may take over jobs that so far have been reserved for humans is original text production. Certain more structured types of text, such as sport results news, may already be produced by machines. In the future, more and more genres will be produced by machines or involve machine participation. Scientific communication is one such area. Structured pieces like abstracts, literary reviews, etc. are possible candidates for machine involvement. A series of texts produced by machines are being published by Forbes online. These texts are highly structured financial reports based on data fed to the text-writing machine. They are almost undistinguishable from human-produced text. Narrative Science is the company based in Chicago that developed the text-producing machine. The early reactions to this development can be illustrated with the help of the following titles of some recent newspaper articles and blog posts: “The robot journalist: an apocalypse for the news industry?” (Bell, 2012, May 13), “Stock advice: Hiring software as analyst” (Fernandez, 2012, July 6), and “Can an algorithm write a better news story than a human reporter?” (Levy, 2012, April 24). As one can see, the fear of skills becoming obsolete drives the first reactions. Such attitude may lead to very inefficient results in all areas of future human-machine interactions and especially in the field of communication regarding science.

## Conclusion

A scientist destined to work in an environment providing possibilities of machine assistance in text analysis and production will have to be a manager to an extent far beyond the needs of today. Beyond managing his / her own time and perhaps a team of humans, a new scientist will have to manage a team of humans and machines performing work tasks best suited for each of them and avoid doing work better / faster done by others. To make time for these additional tasks, some activities performed today will be left to machines. Structured tasks are natural candidates for this. As many activities related to scientific communication are highly structured, a growing amount of activities related to both text analysis and text production will be left to machines.

## References

- AHMAD, J. (2012). Stylistic features of scientific English: A study of scientific research articles. *English Language and Literature Studies*, 2(1), 47-55.
- Artificial intelligence (cs373): Programming a robotic car. (2012, April). Udacity. Retrieved November 11, 2012 from <http://tinyurl.com/cgblk82>
- BELL, E. (2012, May 13). The robot journalist: an apocalypse for the news industry? *The Guardian*. Retrieved November 11, 2012 from <http://tinyurl.com/cr3f8mv>
- BRANIGAN, H. P., PICKERING, M. J., PEARSON, J., & MCLEAN, J. F. (2010). Linguistic alignment between people and computers. *Journal of Pragmatics*, 42(9), 2355-2368.
- BRISCOE, T., & CARROLL, J. (2002, May 29-31). Robust accurate statistical annotation of general text. In *Proceedings of the Third International Conference on Language Resources and Evaluation, Las Palmas* (pp. 1499-1504). Retrieved October 3, 2012 from <http://www.lrec-conf.org/proceedings/lrec2002/pdf/250.pdf>
- BRYNJOLFSSON, E., & MCAFEE, A. (2011). *Race against the machine: How the digital revolution is accelerating innovation, driving productivity, and irreversibly transforming employment and the economy* (Kindle ed.). Lexington, MA: Digital Frontier Press.
- CAO, Y., ELY, J., ANTIEAUM L., & YU, H. (2009). Evaluation of the clinical question answering presentation. In *Proceedings of the Workshop on Current Trends in Biomedical Natural Language Processing* (pp. 171-178). Stroudsburg: The Association for Computational Linguistics. Retrieved October 3, 2012 from <http://aclweb.org/anthology-new/W/W09/W09-1322.pdf>

- CASATI, R., ORIGGI, G., & SIMON, J. (2011). Micro-credits in scientific publishing. *Journal of Documentation*, 67(6), 958-974.
- CHAE, J., & NENKOVA, A. (2009). Predicting the fluency of text with shallow structural features: Case studies of machine translation and human-written text. In *Proceedings of the 12th Conference of the European Chapter of the Association for Computational Linguistics* (pp. 139-147). Stroudsburg: The Association for Computational Linguistics. Retrieved November 11, 2012 from <http://bit.ly/S9Lboa>
- COWEN, T. (2011). *The great stagnation* (Kindle ed.). Boston: Dutton.
- edX. (2012). Retrieved October, 3 2012 from <https://www.edx.org/>
- GRAU, B., LIGOZAT, A., & MINARD, A. (2009). Corpus study of kidney-related experimental data in scientific papers. In *Proceedings of the Workshop on Biomedical Information Extraction* (pp. 21-26). Stroudsburg: The Association for Computational Linguistics. Retrieved November 11, 2012 from <http://tinyurl.com/b7ryaf6>
- GROTH, P., GIBSON, A., & VELTEROP, J. (2010). The anatomy of a nanopublication. *Information Services and Use*, 30(1), 51-56.
- FERNANDEZ, T. (2012, July 6). Stock advice: Hiring software as analyst. *Securities Technology Monitor*. Retrieved November 11, 2012 from <http://bit.ly/KYFbuL>
- Introduction to computer science (cs101): Building a search engine. (2012, April). Udacity. Retrieved November 11, 2012 from <http://tinyurl.com/c7h7g47>
- ISAACSON, W. (2011). *Steve Jobs* (Kindle ed.). New York: Simon and Schuster.
- KEYNES, J. M. (1963). *Essays in persuasion*. New York: W. W. Norton & Co.
- KING, D., ROWLAND, J., OLIVER, S. G., YOUNG, M., AUBREY, W., & CLARE, A. (2009). The automation of science. *Science*, 324(5923), 85-89.
- KULESZA, A., & SHIEBER, S. M. (2004, October 4-6). A learning approach to improving sentence-level MT evaluation. In *Proceedings of the 10th International Conference on Theoretical and Methodological Issues in Machine Translation, Baltimore*. Retrieved November 11, 2012 from <http://tinyurl.com/aub8rad>
- KURZWEIL, R. (2000). *The age of spiritual machines: When computers exceed human intelligence*. New York: Viking Adult.
- LEVY, F., & MURANE, R. J. (2004). *The new division of labour: How computers are creating the next job market*. Princeton: Princeton University Press.
- LEVY, S. (2012, April 24). Can an algorithm write a better news story than a human reporter? Retrieved November 11, 2012 from <http://tinyurl.com/7txfscv>
- LI, H., KUMARAN, A., ZHANG, A., & PERVOUCHINEM, V. (2009). Whitepaper of NEWS 2009 machine transliteration shared task. In *Proceedings of the 2009 Named Entities Workshop: Shared Task on Transliteration* (pp. 19-26). Retrieved October 3, 2012 from <http://www.mt-archive.info/NEWS-2009-Li-2.pdf>
- LICKLIDER, J. C. R. (1960). Man-computer symbiosis. *IRE Transactions on Human Factors in Electronics*, HFE-1, 4-11. Retrieved October 3, 2012 from <http://groups.csail.mit.edu/medg/people/psz/Licklider.html>

- MCCARTHY, D., KOELING, R., WEEDS, J., & CARROLL, J. (2004). Finding predominant word senses in untagged text. In *Proceedings of the 42nd Annual Meeting of the Association for Computational Linguistics, Barcelona* (pp. 280-287). Stroudsburg: The Association for Computational Linguistics. Retrieved October 3, 2012 from [http://acl.ldc.upenn.edu/acl2004/main/pdf/95\\_pdf\\_2-col.pdf](http://acl.ldc.upenn.edu/acl2004/main/pdf/95_pdf_2-col.pdf)
- MONS, B., & VELTEROP, J. (2009). Nano-publication in the e-Science era. Retrieved November 11, 2012 from <http://tinyurl.com/9wm6p4b>
- MOORE, G. E. (1965). Cramming more components onto integrated circuits. *Electronics*, 38(8), 114-117.
- RASSKIN-GUTMAN, D. (2009). *Chess metaphors: Artificial intelligence and the human mind*. Cambridge, MA: MIT Press.
- RUPP, C. J., COPESTAKE, A., CORBETT, P., & WALDRON, B. (2007). Integrating general-purpose and domain-specific components in the analysis of scientific text. In *Proceedings of the UK e-Science Programme All Hands Meeting, Nottingham*. Retrieved November 2, 2012 from <http://tinyurl.com/cnx6jsl>
- SCHAFFER, U., & KIEFER, B. (2011). Advances in deep parsing of scholarly paper content. In R. BERNARDI, S. CHAMBERS, B. GOTTFRIED, F. SEGOND, & I. ZAIHRAYEU (Eds.), *Advanced language technologies for digital libraries* (pp. 135-153). Berlin: Springer.
- SCHAFFER, U., USZKOREIT, H., FEDERMANN, C., MAREK, T., & ZHANG, Y. (2008). Extracting and querying relations in scientific papers on language technology. In *Proceedings of the Sixth International Conference on Language Resources and Evaluation* (pp. 3040-3046). Retrieved October 3, 2012 from [http://www.lrec-conf.org/proceedings/lrec2008/pdf/773\\_paper.pdf](http://www.lrec-conf.org/proceedings/lrec2008/pdf/773_paper.pdf)
- TAKESHIMA, R., & WATANABE, T. (2010). Extraction of co-existent sentences for explaining figures toward effective support for scientific papers reading. In R. SETCHI, I. JORDANOV, R. J. HOWLETT, & L. C. JAIN (Eds.), *Knowledge-based and intelligent information and engineering systems* (pp. 230-239). Berlin: Springer.
- WEIBEL, D., WISSMATH, B., HABEGGER, S., STEINER, Y., & GRONER, R. (2008). Playing online games against computer vs. human-controlled opponents: Effects on presence, flow, and enjoyment. *Computers in Human Behavior*, 24, 2274-2291.
- ZHANG, Y., & CLARK, S. (2011). Syntax based grammaticality improvement using CCG and guided search. In *Proceedings of the 2011 Conference on Empirical Methods in Natural Language Processing, Edinburgh* (pp. 1147-1157). Stroudsburg: The Association for Computational Linguistics. Retrieved November 11, 2012 from <http://tinyurl.com/agcyp67>

# Developing Scientific Software: The Role of the Internet

**Aleksandra Pawlik, Judith Segal, Helen Sharp, and Marian Petre**

The Computing Department, The Open University

*In this article, we describe how scientists use the Internet when they develop scientific software. We present the findings from 27 interviews with scientists-developers and professional software developers who develop scientific software. Based on the analysis of the empirical data, we discuss three aspects of the role of the Internet in scientific software development practice: 1) the use of the Internet in addressing the gaps in scientists' software development knowledge, 2) the use of network-based tools and methods to manage the software development process, and 3) communication between scientists-developers via the Internet.*

## Introduction

Software development is an inseparable part of research in many scientific domains. As research progresses, it raises new questions and challenges that the existing software may not be able to address. At the same time, advanced domain knowledge is necessary to understand what the software is supposed to do. For these two reasons, in many cases scientists develop scientific software themselves. The software that they develop is not their primary goal but rather the means to an end (Basili et al., 2008; Kelly, 2007; Segal, 2009).

That ultimate end is to progress scientists' own research (Segal, 2007). Typically, these scientists do not have the same background in software development that professional software developers would have (Sanders & Kelly, 2008; Segal, 2008). Understanding the software development practices of scientists may reveal ways in which software engineers and scientists themselves can better support development of scientific software. As the Internet is more and more present in the research world, the focus of this paper is on the role of the Internet in scientists' software development practice. Since the study had an exploratory character, we did not design it on the basis of any pre-existing theoretical assumptions. The aim was to provide a broad understanding of how the Internet fits into scientific software development.

A brief scan of related publications reveals that there is very little literature that discusses the role of the Internet in the software development practices of scientists. Research about the Internet and science mainly discusses the ways the Internet facilitates collaboration between scientists and scientific institutions (e.g., Olson et al., 2008). When it comes to software development practices, it has been noted that scientists advance their knowledge about software development in various ways, including self-study resources available on the Internet (Sanders & Kelly, 2008). In 2006, Wilson noted that scientists shared and discussed their source code primarily via e-mails. However, more recent studies (Hannay et al., 2009; Nguyen-Hoan et al., 2010) indicate that other tools dedicated to supporting software development, such as version control systems, are in use by scientists.

Our study provides insight into and information about the role of the Internet in scientific software development practice. We believe that our work will address the dearth of literature on scientific development practices and the Internet.

## **Methodology**

Since the aim of our study was to explore and understand real-life practices of scientists developing software, we used a qualitative approach (Robson, 2002). The methodology of our study follows the guidance for using interviews in software engineering research as described by Seaman (1999). We used semi-structured interviews to obtain the data that would answer our research questions and to explore the research area, allowing the information to emerge from the data. The interviewer always had an interview guide to make sure that no topic that we wanted to cover was omitted. All interviews were recorded and transcribed. We interviewed 27 participants, out of whom:

- 20 were scientists who were developing software that was also used by other scientists,
- two were scientists who developed software for their own use (however they were happy to share their code if they were asked),
- two were scientists with a degree in computer science developing software used by other scientists,
- one was a professional software developer engaged in a scientific software development project,
- two were scientists who mainly used scientific software developed by other scientists.

The last two participants' views and opinions were informative in our study as these scientists-users often had to refer to the same resources as the scientists who developed the software. The opinions of these two participants helped us to form a better overview of the role of the Internet in scientific software development.

The data collected during the interviews were then analyzed using thematic analysis (Boyatzis, 1998); that is, the data were coded using codes formed during a bottom-up analysis, and then these codes were grouped to form themes. The themes then helped to articulate the findings of this study.

## **Findings**

In this section, we will present the findings from our study. First, we will discuss the role of the Internet in progressing software development skills among the scientists. Then we present the findings on how the Internet facilitates a scientific software development process and how it supports communication between scientists-developers.

### **Addressing Gaps in Software Development Knowledge**

Our study showed that the Internet was one of the main resources used for addressing gaps in the scientists' software development knowledge. Out of 27 interviewed scientists, 24 were either almost exclusively self-taught when it comes to software development skills or combined self-teaching with a one- or two-semester course in programming. These courses were usually a part of their undergraduate degree. The self-teaching was based on multiple materials, and among these materials, the Internet appeared to be one of the main sources of information. The way these online resources were used varied. It could be regular checking up on whether there are any new solutions in

the existing technologies that the scientists used. As one of the participants reported:

Every now and then I go through it [Python doc] and I go, “Hey, there’s a new trick [that] has [been] developed in Python that I want to be using”—mainly the online resources.

Another strategy to progress one’s knowledge related to software development is to go straight to a search engine when, for example, an issue with source code appears. Google seems to serve well as the first reference point that can be used to find the answer. In fact, as one of the participants reported, Google was probably the only place he could seek support:

What I would do, I would write some code, run it, [and] when I get an error I would then take the error and put that error into Google and then see what the error was. Then if I got anything that was reasonable as an answer back, I would try to figure that out. Try the solutions that were on the Internet. Other than that, there wasn’t really much more I could do or many more resources I could use.

This strategy of pasting in the code of the error or the error message itself (whether it was an output from a compiler or a runtime error) was common for scientists-developers with varying levels of experience in software development. Two participants said that they would rather check the details of programming language syntax or a particular compiler output on Google than use a textbook on their desk.

The participants did not explicitly discuss the ways they evaluate the reliability of the sources they use. Only two interviewees explained to us their methods for assessing the correctness of the implementation solution or particular syntax details that they found online. As one of them put it:

If it comes down to coding, resources for C++ or something like that, either it works or it doesn’t. Whether it’s trustworthy or not doesn’t matter to me because I will run a test that will test the program and check a bit and get some output additional information to check if the command works as they claim probably double check against other sites [the interviewee means other online resources], if it agrees and gives the same information.

Our findings show that the Internet is used on a regular basis by scientists in order to address the gaps in their software development knowledge. The scientists appear to simply choose solutions that are “good enough” to address an issue with which they are dealing at a given moment. This finding is consistent with the model of scientific software development proposed by Segal (2008).

## Use of Network-Based Tools and Methods Facilitating the Software Development Process

The Internet provides a number of tools and methods to help manage scientific software development projects as well as to support the work coordination within the development team. Sixteen of our participants explicitly discussed how they use the Internet in coordinating their work. The ability of managing tasks via the Internet seems to be an ideal fit for scientific software development projects, which often represent collaborations between groups or individuals at different physical locations. Twenty-four of our participants were involved in projects in which the development team was distributed. Even if the core development team was co-located and the developers could discuss everything in person, there were usually external source code contributors or people writing documentation who did not even work in the same country.

The approaches to Internet-supported management ranged from using advanced tools dedicated to managing projects to less formal management via e-mails. One of the participants described in detail how Trac (i.e., a project management system) was adapted for one of the software development projects in which he was engaged:

We described what the project was trying to achieve. We divided that into tasks for each task. We put the name of the person who was by principle the leader of the task. They could nominate other people to actually share the work and then associate tickets with tasks. And we had milestones, so for a demonstration or deliverable we had a milestone. The milestone then was based on completing the tickets.

Another participant described how his team tried out different tools to address the needs of their project. These systems were used both for managing the project and also for managing the project Web site content.

We started off using media-wiki. [...] for the project website we used the system [the user could not remember the exact name], which is fine, but it's more for managing small projects and labs. It's great for secretaries to use. But to interact with it problematically is a little bit more tricky, and its authentication realm is different. This was OK, but we wanted something better. We tried media-wiki and then went to Google-wiki. I think now we're going to change to the system called Drupal [[www.drupal.org](http://www.drupal.org)].

Summarizing, the scientists use a variety of network-based tools and methods to facilitate the software development process. These tools and methods seem to be suitable for scientific software development projects in which scientists-developers are often not co-located. Additionally, these tools

and methods encourage scientists-developers to prioritize their tasks, assign roles and responsibilities, and set up goals and milestones.

## **Communication via the Internet between Scientists-Developers**

Out of 25 of our participants who were developing scientific software, only one was exclusively a solo developer at the time of the interview. All others were engaged in projects that included other developers. In all cases, communication among the developers of scientific software was carried out via the Internet.

E-mails were the most commonly used means of communication. All participants who were developers either explicitly mentioned using e-mails while working on the software, or we found implicit evidence for that (for example, during the interview, the scientists showed us an example of a design document circulated via e-mail).

E-mails were used for a number of purposes: for sending bits of source code, for circulating minutes of meetings, for discussing details of implementation, and so on. If a given project had other tools supporting development management and collaboration, these tools and e-mails tended to be used for different purposes. As one of the participants described it:

When something needs to be acted upon, we all have an e-mail dialog if it's not supposed to be permanent, if it's just the decisions required. But if it's more permanent, then it will all go on a Web site and we like these content management systems.

Apart from exchanging e-mails, the scientists also intensively used other tools such as Skype, Internet messengers, mailing lists, or even a forum to communicate with others engaged in the same project. For projects in which scientists-developers were not co-located, communication via the Internet was very frequent. In cases when the core development team was co-located, communicating via the Internet was not that often; however, it did take place, such as when one of the core developers was temporarily at a different location or an "occasional" developer needed to consult the main team.

## **Discussion**

In this section, we will discuss the benefits and challenges of the use of the Internet in scientists' software development practices. We will also provide some suggestions on the ways some of these challenges could be addressed.

## Addressing Gaps in Software Development Knowledge

The immediate availability of information on the Internet addressing the gaps in scientists' software development knowledge may be very useful for them. There may be two potential reasons for this usefulness. First, for scientists-developers, software is usually only a means to an end rather than the main goal itself (Segal, 2007). This means that as soon as they find a solution for their software development problem, they can get back to their main task, which is advancing their research. Hence, scientists prefer a solution that is simply "good enough" over a more sophisticated and flexible one. Second, scientists typically lack formal education and experience in software development that professional software developers have. The Internet may possibly help them to find a solution, even if they are not sure what type of solution they want and where to look for it. When they get stuck with a compiler or runtime error, they can search for a solution via an Internet search engine (most often using Google); instead of spending hours trying different fixes and consulting textbooks or colleagues, a scientist may go online to find a solution to a bug.

However, there may be certain challenges related to using the Internet for addressing the gaps in the scientists' knowledge about software development. First, there is the question of the appropriateness of the solutions that the scientists-developers find online. When it comes to fixing problems with implementation, one of the reported strategies was to "see if it works and (sometimes) compare the solution with other resources." Assuming that this strategy actually gives some confidence in the particular solution, there is still uncertainty on how a quick fix may affect other parts of the software or how it may affect further software development and maintenance. A quick fix that deals with a given bug may generate problems in the future. Second, there is the question of trustworthiness and reliability of the sources from which the scientists learn about various aspects of software development. In our research, the scientists did not discuss to any extent which sources were, according to them, trustworthy and reliable. The participants generally did not describe how they assess if the Internet resources that they use suit their needs best. We cannot be sure whether this is due to the fact that the verification process was so obvious for them that they did not even think about mentioning it or whether they simply accepted a solution that was "good enough" (Segal, 2007). Summarizing, the Internet can certainly be a great help for progressing one's skills in scientific software development, but more research about how to mitigate the risks discussed above is needed. The first step could be finding ways to raise scientists' awareness of these risks. Another

approach could be finding an effective way of creating and running “knowledge centers” that could gather information about reliable resources useful for scientists developing software. Such knowledge centers could also be platforms for exchanging ideas and experiences among scientists-developers. Some actions to gather the information useful for scientists-developers and to assist them in communication between each other have been already taken by the Software Sustainability Institute (SSI) in the United Kingdom. This institute organizes workshops and trainings for scientists developing software as well as for professional software developers who are involved in scientific software projects. SSI also collects information about sources used and recommended by the scientific software community. These sources are entered into a freely available “knowledge base.”

### **Use of Network-Based Tools and Methods Facilitating the Software Development Process**

Many tools and methods that support the software development process may help the scientists-developers to organize and coordinate their work in scientific software projects. The participants in our research mentioned that they were engaged in some projects in which the software development process itself was not clearly laid out and organized. In our study, we observed that introducing tools such as a version control or Trac imposes the assigning of roles and responsibilities and requires setting up milestones and goals. These changes, if they took place, were perceived as advantageous by our participants.

The challenge related to the use of network-based tools and methods facilitating software development process is that their variety may become overwhelming. Setting up, for example, a version control, an issue tracker, and a wiki will be meaningless if the developers involved in the project do not use them. In fact, as one of our participants commented, if a scientist-developer is involved in five projects at a time and all five of them have a wiki or an issue tracker, it is highly likely that he will not contribute to any of them. And scientists-developers may sometimes be involved in multiple projects at the same time. We think that more research is needed to establish a balance between using various tools that support the software development process and the actual needs and character of a given scientific software development project. Maybe it could be possible to create a kind of step-by-step evaluation of project’s needs that could help scientists-developers to identify which tools would be useful for them.

## **Communication via the Internet among Scientists-Developers**

The main benefit of communication via the Internet is the fact that it may accommodate the distributed nature of most scientific software development projects. Scientists who are physically based at different institutions in different locations can communicate every day, not only verbally discussing various matters (as it could be done over the phone) but also sharing source code, demonstrating program behavior using a shared desktop, or having a team teleconference.

The main deficit of communication via the Internet is that it cannot replace face-to-face communication. It may be tempting to assume that in the Internet era, scientists-developers' meetings in person are unnecessary and since they are often time- and finance-consuming, such meetings should be cancelled. But this assumption may be wrong. One of our participants who worked in a project engaging scientists from not only different countries but also different continents said that he found Internet communication only effective with people whom he knew in person. Hence, as he reported, he spent about half of his time traveling between different locations to discuss the developed software and, what was equally important for him, to establish working relationships with other developers.

From what the participant said, it was clear that these relationships could be built and strengthened thanks to the informal interaction and the possibility to talk in person with other developers. This finding is consistent with Olson and Olson (2000), who found that some elements that are very important in collaborative software projects such as "informal hall time before and after [the formal meeting]," "implicit cues," and "spatiality of reference" are not and cannot be well supported by technology. The authors pointed out that in projects in which the elements of this collaboration are present (that is, in face-to-face interaction), there are fewer misunderstandings and productivity tends to be better. Improved data flow, whether it is sending a massive piece of source code, raw data, or a team videoconference stream, cannot entirely replace face-to-face communication. It is important to raise awareness among scientists-developers that despite having more and more powerful collaboration and communication tools, they still need to plan for meetings in person.

## **Conclusion**

Our study provides evidence that the Internet can bring many benefits to scientific software development practices. It may help scientists-developers

to keep focus on their main aim, which is advancing their research by easing the process of software development. The Internet is a vast source of knowledge that is easily accessible. This may help scientists to speed up the progress with software development, and the saved time may be allocated to advancing their research. The Internet supports collaboration and communication in scientific software development projects that tend to be of distributed nature. At the same time, our findings indicate that there are some risks involved when it comes to using the Internet in scientific software development. These risks may not seem very apparent and obvious at first glance, but they may in fact have a negative long-term impact on scientific software.

## References

- BASIL, V. R., CARVER, J. C., CRUZES, D., HOCHSTEIN, L. M., HOLLINGSWORTH, J. K., SHULL, F., & ZELKOWITZ, M. V. (2008). Understanding the high-performance-computing community: A software engineer's perspective. *IEEE Software*, 25(4), 29-36.
- BOYATZIS, R. E. (1998). *Thematic analysis: Coding as a process for transforming qualitative information*. Thousand Oaks, California: Sage Publications.
- HANNAY, J. E., MACLEOD, C., SINGER, J., LANGTANGEN, H. P., PFAHL, D., & WILSON, G. (2009). How do scientists develop and use scientific software? In *Proceedings of the 2009 ICSE Workshop on Software Engineering for Computational Science, Vancouver* (pp. 1-8). Washington, DC: IEEE Computer Society. Retrieved November 11, 2012 from <http://tinyurl.com/atqnr51>
- KELLY, D. (2007). A software chasm: Software engineering and scientific computing. *IEEE Software*, 24(6), 120-119.
- NGUYEN-HOAN, L., FLINT, S., & SANKARANARAYANA, R. (2010). A survey of scientific software development. In *Proceedings of the 2010 ACM-IEEE International Symposium on Empirical Software Engineering and Measurement* (Art. 10). New York: ACM Press.
- OLSON, G. M., & OLSON, J. S. (2000). Distance matters. *Human-Computer Interactions*, 15(2), 139-178.
- OLSON, G. M., ZIMMERMAN, A., & BOS, N. (Eds.). (2008). *Scientific collaboration on the Internet*. Cambridge, MA: MIT Press.
- ROBSON, C. (2002). *Real world research: A resource for social scientists and practitioner-researchers*. Malden: Blackwell.
- SANDERS, R., & KELLY, D. (2008). Dealing with risk in scientific software development. *IEEE Software*, 25(4), 21-28.

- SEAMAN, C. (1999). Qualitative methods in empirical studies of software engineering. *IEEE Transactions on Software Engineering*, 25(4), 557-572.
- SEGAL, J. (2007, September 23-27). Some problems of professional end user developers. Paper presented at IEEE Symposium on Visual Languages and Human-Centric Computing, Idaho. Retrieved October 3, 2012 from <http://oro.open.ac.uk/17674/1/segal-ProfessionalEndUserDevelopers.pdf>
- SEGAL, J. (2008, May 13). Models of scientific software development. Paper presented at First International Workshop on Software Engineering in Computational Science and Engineering, Leipzig. Retrieved October 3, 2012 from <http://oro.open.ac.uk/17673/1/SegalICSE08R.pdf>
- SEGAL, J. (2009). Software development cultures and cooperation problems: A field study of the early stages of development of software for a scientific community. *Computer Supported Cooperative Work (CSCW)*, 18(5-6), 581-606.
- WILSON, G. (2006). Where's the real bottleneck in scientific computing? Scientists would do well to pick up some tools widely used in the software industry. *American Scientist*, 94(1), 5.



# Öffentlichkeit und Neue Medien: das Projekt „InsideScience“

**Thorsten Greiner<sup>I</sup>, Jesús Muñoz Morcillo<sup>II</sup>, Caroline Y. Robertson-von Trotha<sup>II</sup>, und Klaus Rümmele<sup>I</sup>**

<sup>I</sup> Karlsruher Institut für Technologie, Presse, Kommunikation und Marketing,

<sup>II</sup> Zentrum für Angewandte Kulturwissenschaft und Studium Generale am Karlsruher Institut für Technologie

*Wissenschaftskommunikation ist keine Einbahnstraße – sie braucht den Dialog mit der Öffentlichkeit. Neue Video- und Onlineformate leiten ihn ein, indem sie Komplexes veranschaulichen und die Forscherinnen und Forscher sehr authentisch zeigen. Die Verbreitung über Web 2.0-Kommunikationskanäle und einen semantischen Wissensraum ermöglichen der Öffentlichkeit einen leichten und interaktiven Zugang zu den Filmen.*

## **Einführung**

Am Karlsruher Institut für Technologie (KIT) experimentieren die Akteure im Projekt „InsideScience“ (<http://www.kit.edu/insidescience>) mit neuen Video- und Onlineformaten, um komplexe Aspekte der Forschung zu vermitteln und mit der Öffentlichkeit zu interagieren. Themen sind die Computergestützte Theoretische Teilchenphysik und Humanoide Roboter. Mit ihnen beschäftigen sich zwei Sonderforschungsbereiche am KIT. (Die Deutsche Forschungsgemeinschaft fördert die Sonderforschungsbereiche „Computergestützte Theoretische Teilchenphysik“ und „Humanoide Roboter – Lernende und kooperierende multimodale Roboter“. Sie finanziert auch das Projekt „InsideScience“.)

## Das Konzept der Öffentlichen Wissenschaft

### Neue Prämissen im Internetzeitalter

Die theoretische Grundlage des Projekts bildet die Überarbeitung des Konzepts der Öffentlichen Wissenschaft im Zeitalter der Neuen Medien und des Internet. Im deutschen Sprachraum taucht die Formel „Öffentliche Wissenschaft“ als Begriff für dialogbasierte Wissenschaftskommunikation erstmals 1999 in einer Festschrift des Instituts für Angewandte Kulturwissenschaft in Karlsruhe auf, dem Vorgänger des ZAK | Zentrum für Angewandte Kulturwissenschaft und Studium Generale. Darin ist die Rede von der Entwicklung einer profilierten Öffentlichen Wissenschaft nach dem angelsächsischen Vorbild der public science in Form von Tagungen und Vortragsreihen als Voraussetzung für die Entstehung einer Diskussionsplattform, auf der sich Wissenschaftler, Politiker und die interessierte Öffentlichkeit austauschen (Robertson-Wensauer, 1999, S. 20-23). Der Begriff „public science“ stand für die Impulse des Bodmer-Reports der Royal Society (The Royal Society, 1985), berücksichtigte aber zugleich die Mängel des Reports, allen voran die Annahme, dass eine quantitativ verbesserte Kommunikation automatisch für mehr gesellschaftliche Akzeptanz sorgt. Sie hat sich mittlerweile als falsch erwiesen.

Der bekannte Wissenschaftspublizist Heinz Haber verwendete bereits Ende der 1960er Jahre den Begriff „Öffentliche Wissenschaft“ im Sinne einer Wissenschaftsvermittlung für Fachfremde im Gegensatz zu den damaligen populärwissenschaftlichen Formaten (Haber, 1968; Möhn, 2000). Der Begriff führte zwei immer wiederkehrende und oft umstrittene Argumente in die damalige Hochschul- und Technokratie-Debatte ein: dass der Kenntnisstand der Bevölkerung für demokratische wissenschaftspolitische Entscheidungen wichtig ist und dass die Wissenschaft öffentlich werden soll, um ihre Finanzierung zu legitimieren (Kohring, 2004, S. 162).

Spätestens mit Faulstich (2006) kommt der Terminus in der Fachliteratur an und wird dort als ein in der Tradition der Aufklärung stehendes Projekt aufgefasst. Faulstich reiht noch die zahlreichen Formen der traditionellen, frontalen und unidirektionalen Wissenschaftsvermittlung auf, wie etwa Wissenschaftssendungen im Fernsehen oder Science Center.

Das Augenmerk der Karlsruher Öffentlichen Wissenschaft liegt auf der dialogbasierten Kommunikation zwischen Wissenschaft und Öffentlichkeit. Sie pflegt eine Mehrkanalkommunikation, die auf einem Modell der Interdependenz gründet. Dieses Alleinstellungsmerkmal erleichtert erheblich die Anpassung an die Neuen Medien.

## Im Dialog: Wissenschaft und Öffentlichkeit

Ein weiterer wichtiger Trend der Wissenschaftskommunikation im Zeitalter der Neuen Medien, den das Projekt aufnimmt, zeigt sich an Wissenschaftlerinnen und Wissenschaftlern, die den direkten Austausch mit der Öffentlichkeit über Kanäle des Web 2.0 suchen. Einer von ihnen ist Dr. Alexander Knoll. Der Biologe am KIT bloggt seit 2006 auf *Alles was lebt* (<http://scienceblogs.de/alles-was-lebt/>):

Wir Blogger posten aus Leidenschaft und aus der tiefen Überzeugung, dass Wissenschaft der Gesellschaft, die uns finanziert, verpflichtet ist und damit auch dem Dialog mit der Öffentlichkeit. Ein anderer Aspekt ist, dass wir Wissenschaftler immer ein Stück weit um Rechtfertigung kämpfen. Also: Wieso forschst ihr über ein Unkraut und tut als Botaniker nicht lieber etwas gegen den Welthunger? Ich kann dann nicht sagen: Lies meine Papers! Dafür braucht es andere Ansätze. Das Bloggen hat einen persönlichen Faktor, der Interessierte direkt erreicht – womit ich auch das Bild vom verwirrten, weltfremden Forscher korrigiere. Und wir beugen dem wissenschaftlichen Analphabetismus vor. (Rauch, 2011, S. 73)

Auch für Stefan Rahmstorf (2010), Professor am Potsdam-Institut für Klimafolgenforschung, sind „Wissenschaftskommunikation in einer unsicheren Welt“ und „Kommunikation für die kritische Öffentlichkeit“ am besten möglich in seinem Blog, in dem er seine kritische Sicht auf die oberflächliche bis einseitige Berichterstattung in Medien, zum Beispiel zum Thema Klimawandel, ausbreitet. Rahmstorf hat für sich die Social Media-Kanäle als Möglichkeit erkannt, gegenzusteuern und seine Argumentation ungefiltert darzulegen.

Abstrahiert man von diesem exponierten Beispiel, so ist allgemein der Trend festzustellen, dass Kommunikation mit dem Aufschwung von Blogs, Netzwerken und Portalen sozialer und schneller wird. Er führt außerdem dazu, dass Dialog und Interaktion zwischen Wissenschaft und Öffentlichkeit wichtiger werden. Zudem bekommen die wissenschaftlichen Akteure in der Kommunikation mehr als früher ein Gesicht: Wer spricht, wird als Person sichtbar und bleibt nicht hinter der Institution verborgen. So wird die Kommunikation authentischer und erreicht wichtige Zielgruppen auf anderen, neuen Wegen besser als mit klassischen Instrumenten, wie der Presseinformation, der Magazinreportage oder dem Radio-Interview, die schon alleine die für soziale Netzwerke typische Art der Ansprache (personalisiertes Auftreten, Umgangssprache, usw.) in der Regel nicht erlauben. Diese neue Kommunikationspraxis prägt auch das Projekt „InsideScience“ mit.

## **Öffentliche Wissenschaft im Projekt „InsideScience“**

Um die Kommunikation zwischen Wissenschaft und Gesellschaft zu ermöglichen, hat das Projektteam unterschiedliche Videoformate realisiert, die über die Darstellung der Forschung am KIT hinaus tiefe Einblicke in die Grundlagen der Teilchenphysik und der Robotik ermöglichen. Neben einführenden Beiträgen entstanden vertiefende und erklärende Filme, die einen Dialog nähren können. In sozialkritischen Beiträgen regte das Team zudem Diskurse an, die über eine partizipative Dissemination im Internet ein Stimmungsbild ermitteln oder eine direkte Interaktion erreichen wollen. Der Gefahr des Digital Divide begegnete das Team mit Filmvorführungen und der Ausstellung von informellen Lernsettings (Wissensräumen).

In diesem Kontext versteht sich auch die Evaluation des Projektes als Teil des Kommunikationsprozesses: Zu jedem Video und auf fast jeder der verwendeten Plattformen bietet das Team Feedback-Kanäle zum Kommentieren und zur Bewertung der Inhalte an. Ein Moderator betreut sie, der oft als Vermittler zwischen Wissenschaftlern und Interessenten fungiert. Die Evaluation ergänzt das Team um traditionelle Methoden wie standardisierte Fragebögen und Leitfadeninterviews zu den Filmen.

## **Von der Skizze bis zum Film: Die Videoproduktion**

### **Hand in Hand: Wissenschaftler und Medienschaffende**

Junge Wissenschaftlerinnen und Wissenschaftler aus der Computergestützten Theoretischen Teilchenphysik und aus der Forschung rund um Humanoide Roboter produzierten zusammen mit Journalisten, Regisseuren und Kameraleuten von der Dienstleistungseinheit Presse, Kommunikation und Marketing (PKM) und vom ZAK am KIT rund 30 Videos, die Forschung verständlich machen. Gemeinsam entwickelten sie Ideen für die Filme und schrieben die Drehbücher, diskutierten die Aktionen der Wissenschaftlerinnen und Wissenschaftler vor der Kamera und den Einsatz von Animationen. In Redaktionsrunden waren Produzenten und Wissenschaftler gleichrangige Partner bei der Feinarbeit an den Filmen.

Um diese Zusammenarbeit vorzubereiten, führte das Projektteam lange und intensive Gespräche mit den Wissenschaftlerinnen und Wissenschaftlern. Es entwickelte sich ein Austausch, der im Alltag der institutionellen Wissenschaftskommunikation oft nicht möglich ist. So gelang es dem Team, mehrere der Wissenschaftlerinnen und Wissenschaftler in den beiden Sonderforschungsbereichen zu öffnen für das Medium Film, für den Auftritt vor der Kamera und für den Dialog mit der Öffentlichkeit.

## **Authentizität und Anschaulichkeit – Auftritt und Animation**

Die Videos arbeiten mit einem Film-Mix: In realen Sequenzen beschreiben die Forscher ihre wissenschaftlichen Ansätze und Ergebnisse. Sie treten authentisch auf und wagen zugleich selbst den Spagat, ihre komplexe Forschung mit einfachen Mitteln – zum Beispiel kleinen Zeichnungen auf einer Tafel – zu erklären. Nicht immer gelang dies bei der ersten Aufnahme – doch am Ende war jeder der beteiligten Wissenschaftler in einem der Videos zu sehen.

Animierte Sequenzen wiederum zeigen das Innenleben von Maschinen oder gebrauchen visuelle Metaphern, um sehr abstrakte Methoden und Fortschritte zu illustrieren. Sie gestatten Einblicke in theoretische Modelle – zum Beispiel über den Vergleich eines Higgs-Felds mit Honig („Wie bitte entsteht überhaupt Masse? Die Suche nach dem Higgs-Teilchen“, 2011, Dezember 13) – oder hinter die Fassade einer Maschine, die äußerlich unverändert erscheint, in deren Inneren aber Forscherinnen und Forscher entscheidende Fortschritte erzielt haben – wie etwa im Humanoiden Roboter Armar („Programmieren durch Vormachen: Wie Roboter von Menschen lernen“, 2012, Juli 2).

Es gab durchaus kritische Reaktionen auf die Filme: Wissenschaftlerinnen und Wissenschaftler aus dem Bereich der Humanoiden Roboter fanden sie zu einfach, um korrekt zu sein. Auf Besucher der Vorführungen, Internetnutzer und Journalisten wiederum wirkten manche Filme zur Theoretischen Teilchenphysik immer noch zu kompliziert. Der Reichweite der Filme schadete das nicht (siehe Kapitel zur Dissemination).

## **Medientraining für Wissenschaftlerinnen und Wissenschaftler**

ZAK und PKM boten den Wissenschaftlerinnen und Wissenschaftlern ein Medientraining an. Es half ihnen, mit den Medien umzugehen, sicher und authentisch mit der Öffentlichkeit zu kommunizieren und vor der Kamera aufzutreten. Die Teilnehmerinnen und Teilnehmer erlebten sich in verschiedenen Interview- und Gesprächssituationen und bekamen individuell Feedback. Hinzu kamen praktische Übungen, die den Wissenschaftlerinnen und Wissenschaftlern Verhaltensoptionen für den Fall aufzeigten, dass sie auf Skepsis und Kritik stoßen, die sich auf ihren Forschungsgegenstand bezieht.

## Im Dialog mit der Öffentlichkeit: Disseminationsstrategien

### Verbreitung über Portale, soziale Netzwerke und Blogs

Um die Öffentlichkeit zu erreichen, verfolgte das Projekt eine online-basierte Strategie der Dissemination: Es verbreitete die Videos über die KIT-Homepage, über Wissenschaftsportale mit großer Reichweite, wie Spektrum der Wissenschaft ([www.spektrum.de](http://www.spektrum.de)), über Social Media-Kanäle und über Videoplattformen. Gerade bei den Zugriffen über YouTube schnitt „Inside Science“ im Vergleich zu Projekten anderer Einrichtungen im deutschsprachigen Raum sehr gut ab:

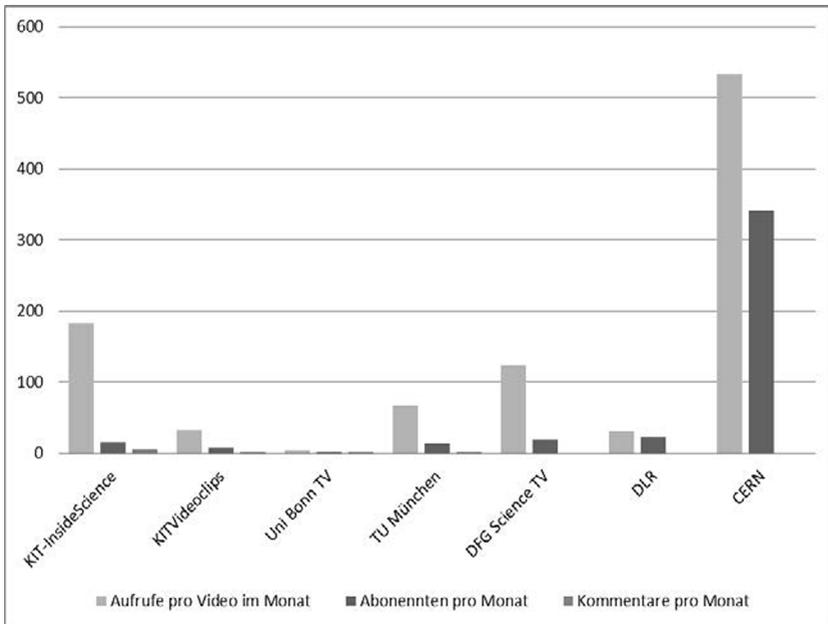


Abbildung 1. Resonanz von Forschungseinrichtungen auf YouTube, Stand 09.05.2012. (Das CERN steht als Referenz für Best Practice in anderen Ländern.)

Die Interaktion über Social Media-Kanäle entwickelt sich langsam. Seit dem Start des Facebook-Accounts am 29. März 2011 (mit eigenem Video-Content seit Dezember 2011) wurden 77 likes registriert (Stand 29.8.2012). Dies entspricht einem durchschnittlichen Zuwachs von circa fünf likes pro

Monat. Die Viralität der Beiträge ist wesentlich besser bei eigenem Video-Content und Ankündigungen von Veranstaltungen als bei der Verbreitung von Interviews oder empfohlenem Content aus dem Netz.

Die Projektpartner teilen zudem in einem Blog (InsideScience – Weblog, <http://insidescience.kit.edu/weblog>) Eindrücke vom Entstehungsprozess der Videos mit der Öffentlichkeit – auch dies mit dem Ziel, mit ihr in einen Dialog zu treten. Rund 40 Artikel sind bislang auf dem Blog erschienen, verfasst haben sie hauptsächlich Medienschaffende und Kulturwissenschaftler. Die Wissenschaftler der Sonderforschungsbereiche treten vor allem in Interviews in Erscheinung. Die beliebtesten Blogbeiträge sind Ankündigungen von Veranstaltungen oder Texte mit wissenschaftlichem Anspruch. Der Blog funktioniert auch als Experimentierfeld für die Verbreitung von Webvideos in Verbindung mit vertiefenden Beiträgen, die zur Diskussion anregen sollen.

### **Schülerprojekte und Filmvorführungen**

Daneben setzte das Projektteam auch auf einen Face-to-Face-Austausch mit der Öffentlichkeit. Es lud Schüler und Schülerinnen ein, mit Wissenschaftlerinnen und Wissenschaftlern zu reden und ihr eigenes Video über Teilchenphysik zu produzieren. Außerdem diskutierten die Projektpartner ihre Resultate bei einer Konferenz in Karlsruhe mit Künstlern, Wissenschaftlern, Studierenden und Besuchern, sowie bei einer Filmpremierre in einem Kino der Stadt.

### **Der semantische Wissensraum**

Bei Tagungen und öffentlichen Veranstaltungen präsentierten die Projektpartner ein Instrument, das der Öffentlichkeit ermöglicht, die Videos interaktiv zu sehen: den semantischen Wissensraum (Wissensraum Elementarteilchenphysik, <http://inside-science.forschung.kit.edu/180.php>). Das Team verwirklichte ihn in einer ersten Stufe mit dem freien Mind Map-Werkzeug Spicynodes, später mit der Plattform Flash. Der Wissensraum stellt eine Online-Umgebung dar, die auch als informelles Lernsetting in öffentlichen Räumen funktionieren kann. Im Vordergrund stand die Kontextualisierung des produzierten Materials, um ein funktionales Video-Archiv zu bilden, das für eine heterogene Öffentlichkeit attraktiv ist. Die Kontextualisierung von Wissen in Archiven ist ein Thema, das die Forschung im Bereich des digitalen Kulturerbes seit einigen Jahren beschäftigt (Grau, 2004; Warnke, 2003). Das InsideScience-Projektteam griff auf pädagogische Konzepte der Wissenskonstruktion und auf Überlegungen der Medientheorie und der Medienkunst zurück (Fleischmann & Reinhard, 2004).

## **Wissensraum zur Theoretischen Teilchenphysik**

Für das Thema Theoretische Teilchenphysik hat das Team insgesamt vier Varianten des Wissensraums implementiert: eine textbasierte Version für das Museum und öffentliche Räume, eine ebenso textbasierte YouTube-Version für Internetnutzer und zwei weitere Piktogramm-basierte Varianten für öffentliche Räume und für den Einsatz im Internet. Für die Ausstellung des Wissensraums bei der InsideScience-Tagung am ZKM im Dezember 2011 wurde zudem eine interaktive Bodenprojektion installiert. Sie bestand aus einem Partikelfeld, das auf die Bewegungen der Besucher wie ein Higgs-Feld reagierte. Sobald eine Person im Raum stand, sammelten sich langsam die Partikel um den Besucher herum und bildeten somit virtuell die Masse der Person nach. Über dem Partikelfeld schwebte eine halbtransparente Kugel mit dem Schriftzug *InsideScience*. Links und rechts erschienen die wichtigsten Begriffe der Teilchenphysik. Die Bodenprojektion war so konfiguriert, dass die Kugel den Besucher zu einem von drei aufgestellten Terminals verfolgte, während sich das Higgs-Feld wellenartig verformte.

## **Wissensraum zur Forschung über Humanoide Roboter**

Auch die Filme über Humanoide Roboter kann die Öffentlichkeit in einem Wissensraum für das Web 2.0 und das Museum erleben. Das Team hat die Inhalte auf einer frei beweglichen Grafik mit anklickbaren Elementen verteilt, die eine Platine mit Mikrochips zeigt. Der Betrachter kann sie entlang der Leiterbahnen erkunden. Die Grafik ist so gestaltet, dass sie aus der Ferne betrachtet einem Fingerabdruck ähnelt. Sobald der digitale Fingerabdruck angeklickt oder (im Falle der Museumsvariante) berührt wird, vergrößert sich das Bild zur Übersichtsgröße, der Besucher kann nach Inhalten suchen. Hinter jedem Mikrochip befinden sich Informationen zum entsprechenden Thema, das dazugehörige Video ist in einem kontextuellen Videoplayer eingebettet. Die Idee des Fingerabdrucks im Platinen-Stil, die sich an La Mettries Begriff der Mensch-Maschine anlehnt, kommt auch in den Haupt- und Vertiefungsbeiträgen vor. Das Fingerabdruck-Motiv ist zugleich die favorisierte Darstellung für Buttons, die auf Feedback-Formulare verlinken.

## **Evaluation**

Dialogorientierte Konzepte brauchen andere Evaluationsmethoden als eine Top-down-Kommunikation zur Vermittlung von Wissen, zumal wenn sie sich vornehmlich auf Online- und Bewegtbildformate stützen. Für eine Eva-

luation der Produktion und Verbreitung der InsideScience-Filme eignen sich hybride Vorgehensweisen, die sowohl standardisierte Evaluationsmethoden, wie den klassischen Fragebogen, als auch Online-Feedbackfunktionen und automatisierte Statistiken beinhalten. Das Projektteam hat vor diesem Hintergrund folgende Evaluationsmethoden konzipiert:

- Untersuchung von Nutzerverhalten anhand automatisierter Datenerhebungen auf Web-Videoplattformen;
- Meinungserfassung durch Interpretation von Kommentaren in verschiedenen Web-Umgebungen (Videoportal, Blog, Social Media und informelle Lernsettings);
- Online-Feedback in den oben genannten Web-Kontexten und klassische Methoden, wie standardisierte Fragebögen und Leitfadeninterviews bei Filmpremieren (Stichproben mit 150 bis 170 Personen, siehe Abbildungen 2 und 3), Ausstellungen, Workshops und Filmprojekten mit direkter Beteiligung der Zielgruppe.

Dieser Methodenmix sichert die Qualität des Projektes durch Bürgerbeteiligung. Zudem trägt er zur Untersuchung bislang wenig beachteter Kommunikations- und Austauschprozesse bei und eröffnet somit eine Diskussion über die Evaluationspraxis partizipativer Bewegtbildkommunikation.

Der an Erfahrungen in der E-Partizipation (vgl. Gräf, 2010; Kukartz, 2009; Rosa-Luxemburg-Stiftung, 2004; Initiative ePartizipation, 2005; Initiative D 21, 2010) und an die Evaluationspraxis von Public Communication of Science and Technology-Aktivitäten (Neresini & Pellegrini, 2008) angelehnte Methodenmix versucht sowohl die formalen Aspekte der Filmproduktion von „InsideScience“, als auch das Potenzial der Filmbeiträge zu verbessern, Diskussionen anzustoßen und zu analysieren. So wurde zum Beispiel festgestellt, dass die Einbindung von Wissenschaftlern in das Projekt über die Dreharbeiten hinausgehen soll, weil Internetnutzer oft komplexe Fragen stellen; andererseits hat die bisherige Verbreitung auf YouTube gezeigt, dass qualitative Diskussionen auf Foren oder moderierte Videoblogbeiträge ausgelagert werden müssen, um sich entfalten zu können.

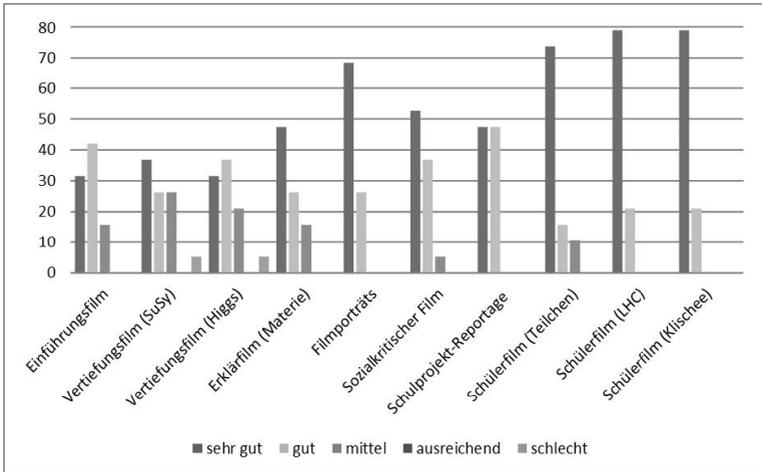


Abbildung 2. Bewertung der Verständlichkeit von Videos zur Teilchenphysik (in Prozent), Befragung von 19 Personen während einer Film Premiere mit 150 Besuchern im Dezember 2011

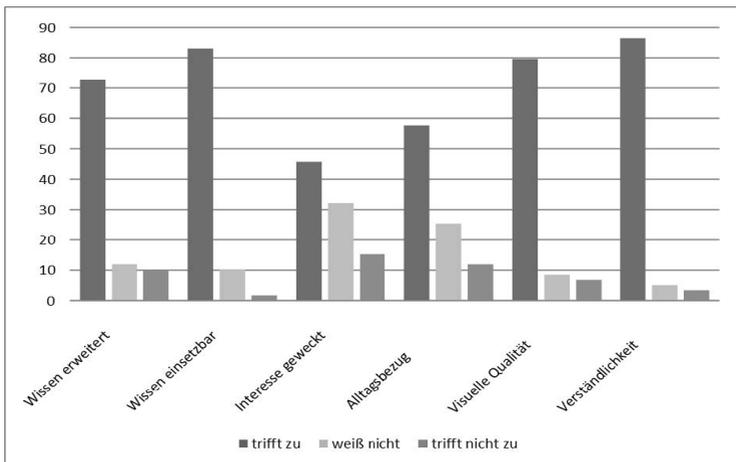


Abbildung 3. Wahrgenommene Wirkungen der Videos zu Humanoiden Robotern (in Prozent), Befragung von 59 Personen während einer Film Premiere mit 170 Besuchern im Juni 2012

## Referenzen

- FAULSTICH, P. (Hrsg.). (2006). *Öffentliche Wissenschaft. Neue Perspektiven der Vermittlung in der wissenschaftlichen Weiterbildung*. Bielefeld: transcript.
- FLEISCHMANN, M., & REINHARD, U. (Hrsg.). (2004). *Digitale Transformationen. Medienkunst als Schnittstelle von Kunst, Wissenschaft, Wirtschaft und Gesellschaft*. Heidelberg: Whois.
- GRÄF, L. (2010). *Online Befragung. Eine praktische Einführung für Anfänger*. Berlin: Lit.
- GRAU, O. (2004). Für den erweiterten Dokumentationsbegriff. Datenbank für virtuelle Kunst. Heruntergeladen am November 2, 2012 von <http://tinyurl.com/c6kkb8x>
- HABER, H. (1968). Öffentliche Wissenschaft. *bild der wissenschaft online*, 5, 744-753.
- INITIATIVE D 21 (2010). Digitale Gesellschaft, die digitale Gesellschaft in Deutschland – sechs Nutzergruppen im Vergleich. Heruntergeladen am November 11, 2012 von <http://tinyurl.com/ybbredu>
- INITIATIVE ePARTICIPATION (2005). Elektronische Bürgerbeteiligung in deutschen Großstädten 2005. Zweites Website-Ranking der Initiative eParticipation. Berlin. Heruntergeladen am November 11, 2012 von <http://tinyurl.com/abrwy17>
- KOHRING, M. (2004). Die Wissenschaft des Wissenschaftsjournalismus. Eine Forschungskritik und ein Alternativvorschlag. In C. MÜLLER (Hrsg.), *SciencePop. Wissenschaftsjournalismus zwischen PR und Forschungskritik* (S. 161-183). Graz: Nausner & Nausner.
- KUCKARTZ, U., EBERT, T., RÄDIKER, S., & STEFER, C. (2009). *Evaluation Online. Internetgestützte Befragung in der Praxis*. Wiesbaden: VS Verlag.
- MÖHN, D. (2000). Textsorten und Wissenstransfer / Text Types and the Transfer of Knowledge. In K. BRINKER, G. ANTOS, W. HEINEMANN, & S. F. SAGER (Hrsg.), *Text und Gesprächslinguistik / Linguistics of text and conversation. Internationales Handbuch zeitgenössischer Forschung / International handbook of contemporary research* (S. 561-573). Berlin: Walter de Gruyter.
- NERESINI, F., & PELLEGRINI, G. (2008). Evaluating public communication of science and technology. In M. BUCHI, & B. TRENCH (Hrsg.), *Public communication of science and technology* (S. 237-248). New York: Routledge.
- Programmieren durch Vormachen: Wie Roboter von Menschen lernen. (2012, Juli 2). Heruntergeladen am November 11, 2012 von <http://bit.ly/WRzCK2>
- RAHMSTORF, S. (2010, November 29-Dezember 1). Wissenschaftskommunikation in einer unsicheren Welt: Kommunikation für die kritische Öffentlichkeit. Vortrag auf dem Dritten Forum Wissenschaftskommunikation, Mannheim.
- RAUCH, T. (2011). Nischensport mit Zukunftspotenzial: Interview mit Alexander Knoll über Lust und Verpflichtung an einer neuen Form des Dialogs mit der Öffentlichkeit. *lookKIT. Magazin für Forschung, Lehre, Innovation*, 2, 72-74.

- ROBERTSON-VON TROTHA, C. Y. (2007). Öffentliche Wissenschaft – ein notwendiger Dialog. In J. KLAUS, & H. VOGT (Hrsg.), *Wissensmanagement und wissenschaftliche Weiterbildung. Dokumentation der Jahrestagung der deutschen Gesellschaft für wissenschaftliche Weiterbildung und Fernstudium an der Universität Karlsruhe* (TH) (S. 7-20). Hamburg: DGWF.
- ROBERTSON-VON TROTHA, C. Y. (2009). Fachübergreifende Lehre und Schlüsselkompetenz als Programm. 60 Jahre Studium Generale und 20 Jahre Angewandte Kulturwissenschaft an der Universität Karlsruhe (TH). In C. Y. ROBERTSON-VON TROTHA (Hrsg.), *Schlüsselqualifikationen für Studium, Beruf und Gesellschaft* (S. 97-134). Karlsruhe: Universitätsverlag.
- ROBERTSON-WENSAUER, C. Y. (1999). Einleitung. Wozu ‚Angewandte Kulturwissenschaft‘ an einer technischen Hochschule? In C. Y. ROBERTSON-WENSAUER (Hrsg.), *Interfakultatives Institut für Angewandte Kulturwissenschaft. Universität Karlsruhe (TH). 1989-1999 Zehn Jahre interdisziplinäre Institutsarbeit* (S. 19-23). Karlsruhe: Interfakultatives Institut für Angewandte Kulturwissenschaft.
- ROSA-LUXEMBURG-STIFTUNG (Hrsg.). (2004). Neue Tendenzen bei Bürgerbeteiligungsprozessen in Deutschland. Veränderte Rahmenbedingungen, Praktiken und deren Auswirkungen. Berlin. Heruntergeladen am Oktober 5, 2012 von [http://www.rosalux.de/fileadmin/rls\\_uploads/pdfs/allg\\_Texte/Kommunalpolitik/StudieBuergerbetEnd.pdf](http://www.rosalux.de/fileadmin/rls_uploads/pdfs/allg_Texte/Kommunalpolitik/StudieBuergerbetEnd.pdf)
- THE ROYAL SOCIETY (Hrsg.). (1985). Public understanding of science (Bericht). Luton: Inprint of Luton Limited. Heruntergeladen am Oktober 5, 2012 von [http://royalsociety.org/uploadedFiles/Royal\\_Society\\_Content/policy/publications/1985/10700.pdf](http://royalsociety.org/uploadedFiles/Royal_Society_Content/policy/publications/1985/10700.pdf)
- WARNKE, M. (2003). Daten und Metadaten. Online-Ressourcen für die Bildwissenschaft. *zeitenblicke*, 2(1). Heruntergeladen am Oktober 5, 2012 von <http://www.zeitenblicke.de/2003/01/warnke/warnke.pdf>
- Wie bitte entsteht überhaupt Masse? Die Suche nach dem Higgs-Teilchen. (2011, Dezember 13). InsideScience. Heruntergeladen am Oktober 5, 2012 von <http://www.youtube.com/watch?v=URB0nHNuUZE>



