Hans-Christoph Hobohm (Hrsg.)

Informationswissenschaft zwischen virtueller Infrastruktur und materiellen Lebenswelten

Information Science between Virtual Infrastructure and Material Lifeworlds

Unter Mitarbeit von Judith Pfeffing

Proceedings des 13. Internationalen Symposiums
für Informationswissenschaft (ISI 2013)

Potsdam, 19. bis 22. März 2013

vwh
Verlag Werner Hülsbusch
Fachverlag für Medientechnik und -wirtschaft

What will future publications be like?

Andreas Degkwitz

Humboldt-University Berlin
Unter den Linden 6, D-10099 Berlin
andreas.degkwitz@ub.hu-berlin.de

Abstract

The development of digital publications is making headway due to increasing data-intensive research activities and may also be noticed in subject areas like the humanities where this empirical approach is completely new and driven forth by the new technologies. At the same time, the decline of hermeneutical research can be observed. This is closely connected with the role of documents and texts in scholarly communication. Accordingly, future publications are likely to turn into objects which might be enhanced by the integration of different materials like data sets, digitized texts, pictures, audios, videos etc. The crucial point is the question which format will allow the aggregation and publication of the very complex findings of data-intensive research.

1 What is the general background?

In the State-of-the-Art-Report of DRIVER-II (Woutersen-Windhouwer/ Brandsma 2009: 7 ff.) an enhanced publication is defined "as a publication that is enhanced with three categories of information: (1) research data (evidence of the research), (2) extra materials (to illustrate or clarify), or (3) post-publication data (commentaries, ranking)". The authors of this report con-

In: H.-C. Hobohm (Hrsg.). Informationswissenschaft zwischen virtueller Infrastruktur und materiellen Lebenswelten. Tagungsband des 13. Internationalen Symposiums für Informationswissenschaft (ISI 2013), Potsdam, 19.–22. März 2013. Glückstadt: Verlag Werner Hülsbusch, 81–92.

tinue with the following assessment: “We have found that even though publishers and repositories have the building blocks and tools available, they generally do not use them to create an enhanced publication for all three categories. Instead, (1) publishers offer only one or two of these categories with their publications; (2) publishers publish extra materials (such as movies) separately from the publication, without even a link between these related objects; (3) the option to add post-publication data is rarely supported by publishers; in contrast, (4) many repositories offer the options of linking publication and research data, adding commentaries, rankings, trackbacks, etc. Only one publisher, PLoS, provides all three enhancement services for their publications. The need for enhanced publications that are embedded in the publisher or repository environment is urgent, as the number of digital scholarly objects on the internet (such as multimedia materials, data sets, and blogs) is growing enormously”.

What is the background for these statements? Knowledge and media are very closely connected and influence each other: Media indicate the understanding of knowledge. Paradigms of knowledge determine media formats. Against this background, the design of media is not only driven by technology but it is also related to the disciplinary cultures of knowledge. For centuries, the tradition of European knowledge has been aggregated and published in different formats of printed texts. The knowledge paradigm of printed texts has been closely connected with the hermeneutical and theoretical research. The potential of information technology is changing the knowledge paradigm into an empirical data-based knowledge (Hey/Tansley/Tolle 2009). This is the essence of the transition from the printed to the digital world. The new media bring forth a changed understanding of knowledge, which is data-based and transforming the hermeneutical or theoretical paradigm of research into an empirical research approach.

This process is changing the distribution of knowledge as well. The text-based formats of publications become less important for the publication of research results. For several years, data-based research and publications have been on the increase. These developments and the technical developments of digital publishing include – on a larger scale than printed formats – the possibility to enrich publications with data sets to link and process the published research results with different apps and tools: Future publications will be enhanced publications. Hence there is a demand for future publications with a significantly wider range of options printed materials have never had.

As enhanced publications, future publications might either be enriched ones, data publications or both. So far we have found enriched publications primarily in the humanities and social sciences whereas data publications have been more frequently applied in the research findings of natural sciences, technology and medicine. But currently we may also register data-based research and data publications in the humanities; namely for example in the editorial science and linguistics as well as in the TextGrid-Project[1] (Mittler 2012: 33 ff.).

But until today, as stated in the DRIVER-II-Report, the digital potential of data-based media formats has not yet been entirely used or exploited. Currently, digital publications still tend to copy the features of printed publications. However, the development of new publishing models which meet the needs and potential of digital media are urgently needed. The players within the value chain of scholarly communication (authors, editors, libraries, computers and media centres, IT-providers, publishers, researchers etc.) must accept this challenge in their collaboration as they have become a value network that creates the big chance to enhance scholarly communication by digital means.

2 What are the Drivers of Change

One of the foremost tasks of researchers is the publication of the research findings from their respective fields. The publication process has been deeply changed by computers and the internet. The emergent technologies encourage the convergence of the media and the conversion of media formats. Nowadays it is common practice to make multimedia presentations of research activities in order to illustrate the many aspects of the research projects with the help of data, digitized texts, movies, pictures, sounds, video etc. All those materials may be converged or become part of one ‘document’. Moreover the World Wide Web offers a vast potential of possibilities to communicate and to obtain any kind of information just in time, which, on the long run, will mean the termination of printed publications’ single-channel-communication.

1 http://www.textgrid.de/ <February 12, 2013>

The deep change is characterized by the migration of publications from printed texts to multimedia, data-based presentations and caused by the 'all to all communication' of the internet – instead of the linear, 'step by step process' of printed materials.

Against this background, new relations are developing between the players of the value chain in scholarly communication. The traditional players or intermediaries like libraries, publishers, book sellers etc. are taking on new business roles on the market or they will disappear. The new players have plenty of options to cope with their new roles in this value chain, which has actually become a network rather than a chain. There are also new players in that network of the publication process like media and computer centres of the universities or commercial IT-providers which develop apps and tools.

However, it is still publishers' common practice to deliver their digital content as licensed packages or subscriptions, which might lead us to the assumption that the basic value chain hasn't really changed. But the different 'roads' of the open access publishing show, that the licensing or subscription model is only one among others to deliver published materials.

In November 2008, the ITHAKA-company published the findings from the case study "Current Models of Digital Scholarly Communication" which was carried out by the Association of Research Libraries (ARL). For this study about 300 field interviews in nearly 50 universities in the United States over all disciplines were done. The case study primarily focused on the researchers' and scientists' perspective as authors and readers. The digital resources discussed in the interviews were distinguished in the following eight categories: E-only journals, reviews, preprints and working papers, encyclopedias, dictionaries, and annotated content, data, blogs, discussion forums and professional and hubs for scholarly collaboration.

The following is a brief summary of this case study based on short quotations (Maron/Smith 2008: 7 ff.):

- "… Traditions of scholarly culture relating to establishing scholarly legitimacy through credentialing, peer review, and citation metrics exert a powerful force on these innovative online projects …"
- "… Many digital publications are directed at small, niche audiences …"
- "… Some of the resources with greatest impact are those that have been around a long while …"
- "… Innovations relating to multimedia content and Web 2.0 functionality appear in some cases to blur the lines between resource types …"

- "… Projects of all sizes are still seeking paths to sustainability. For open access sites – the vast majority of the resources studied here – the challenges can be great, since subscription fees are not an option …"

The ongoing technical developments are changing this situation. The acceptance and implementation of digital publications for scholarly communication is increasing significantly. In 2008/2009 two papers of the DRIVER-II-Project were published, which concern a number of functional and technical requirements for enhanced publications. The State-of-the-Art-Report about enhanced publications provides an overview of the development of enhanced publications until 2009 and summarizes the different approaches of commercial publishers and open access repositories in dealing with this issue (Woutersen-Windhouwer/Brandsma 2009). In the paper on object models and functionalities, the technical structure (Verhaar 2008) and the functional requirements of enhanced publications are explained in depth and related to the standards of the Open-Archive-Initiative "Open Reuse and Exchange" (OAI-ORE). Both papers refer to publications in the subject areas of the natural sciences, technology and medicine, where enhanced publications already existed in 2008/2009. The recommendations and requirements these both papers contain are very useful under organizational and technical aspects in order to produce enhanced publications and to make them available as the presented resource maps may demonstrate (ibid.: 30 ff.). But the enhanced publications themselves are definitively more than purely organizational or technical challenges or solutions. Any kind of publication has its functional origin, but is not limited to it. The importance of printed books or journals is going beyond the results of a technical process. Printed materials also have a cultural impact, which enhanced publications are supposed to make as well. We are going to discuss this point later.

In 2011 and 2012 the German Research Society (Deutsche Forschungsgemeinschaft)[2] and the German Research Council (Wissenschaftsrat)[3] published their recommendations for the further development of academic infrastructures to improve research conditions towards digital research environments in the current decade. The general topics of these papers are national licenses, hosting and long term archiving of materials, open access publicshing and repositories, digitization of cultural heritage items, research data

2 http://www.dfg.de/download/pdf/foerderung/programme/lis/positionspapier_digitale_transformation.pdf <February 12,2013>

3 http://www.wissenschaftsrat.de/download/archiv/2359-12.pdf <February 12,2013>

management, virtual research environments, e-competence and e-literacy. All these measures have an immediate effect on the pre- and post-phases of publishing and on the publications themselves. They are necessary for the research processes as well as for the publication of the research results. In other words: If they are neglected in the publication of research results, all potentials of the information infrastructures will be wasted. This leads us to the conclusion that we will implicitly have to integrate the different means of information infrastructures in all phases of research work to further enhance scholarly communication and digital publishing.

3 How to Redesign Scholarly Communication?

The goal of scientific publications is the presentation and distribution of research results. The focus of these publications is consequently on aggregating findings, abstracting processes and activities, drawing conclusions and summarizing results. Publications basically reflect an interim state of knowledge or research work. Therefore, the static and/or defined status is a basic feature of publications. Printed books might illustrate this better than digital documents, which tend to produce less static but more dynamic processes and research work results. Furthermore, digital publications can be changed or revised any number of times, they may be locked and opened to different users or groups and they can be deleted or enriched in many ways. Basically, digital publications can be edited in different data formats and processed for different purposes. At each point in time, the digital means allow the authors to include and integrate processes of their work into their published results. This means that the different procedures and steps of the research work can be much better illustrated or mapped in digital publications than this could ever be accomplished in printed materials.

The procedures of the pre-publication phase (= research process) include the search for the latest state of the art, the collection and the evaluation of data, the aggregation of findings the description of conclusions and the preparation of the final results' publication. The post-publication phase covers the later enrichment of a publication, the communication about the published material and the follow-up processing. The crucial question is: which contents or processes of the pre- and post-publication phases should be pub-

lished and must be integrated in the final document? In 2013 we may take it for granted that the vast majority of researchers use their desktop devices and the internet for their work. We may also assume that the services which are necessary to produce, distribute, collect and make digital publications available – provided by libraries and publishers – are basically known.

However, the objectives for the production of digital publications in scholarly communication are diverse depending on the knowledge cultures of the different disciplines: Is the presentation of the research results only possible with the help of digital means and consequently requires digital publishing for its performance? Are there any virtual research environments which explicitly need the digital distribution of publications? Are there any changes in the understanding of knowledge, which can be clearly identified? How do researchers deal with the different data and media they use or produce during their research work? These questions concern the disciplinary cultures of knowledge and should be answered in respect of the enhanced publications' further development.

Social networks play an increasingly important role in scholarly communication – either they distribute intermediate results or they create collaborative work platforms. The influence of social networks on digital publications has not been completely explored yet and must be monitored (Purdy 2010: 50 ff.). In other words: What is the self-understanding of scholarly communication in social networks? Science communication is eventually becoming an increasingly more important medium for the distribution of research work results. Science communication may also be a part of scholarly communication, but it occurs mostly beyond the scientific context. Scholarly communication and science communication have the same objectives in some respects i.e. to raise decision makers' awareness and/or explain scientific results to the public. An additional effect may also arise from amateur scientists' contributions which may be remarkable. This suggests that science communication must also be accepted as part of future publications.

Considering the research practice in the humanities – and generally all hermeneutical and/or theoretical approaches – the understanding of documents and texts plays an important role. Considering the technical potential of enhanced publications, entities like documents or texts are not necessarily provided or needed anymore. Will future publications even abolish the traditional understanding of documents? In 2006, a group of French researchers started to discuss the notion of document and text in digital times. One of the most important consequences they drew was the re-conceptualization of the

document by the following three perspectives: sign ("vu"), text or content ("lu") and medium ("su") (Pédauque 2006, 2007). Based on this assumption, an electronic or digital document is an object, which is, beyond its functions of sign and content, some kind of unit or entity. The crucial point of the discussion is the notion of evidence and the idea of the cultural (Hobohm 2012: 4 ff., 8 ff.) and/or technical impact (Gradmann/Meister 2008: 149 f.; Dudek 2012: 198 f.) that digital documents definitively have.

4 What are the Requirements for Future Publications?

Considering the further development of future publications, a number of questions has to be answered. These questions concern the different disciplinary cultures and the relating differences in the understanding of knowledge as well as the various aims and motivations of publishing. The approach to the data is one of the central questions. What kind of disciplinary resources and data resources will be produced and used in the research projects? What are the subject-specific approaches or methods to deal with these materials? In what way do researchers handle the available or generated data or data sets: pictures, digitized materials, data collections, statistics, geo-references etc.? Which data collections or data sets have to be integrated into the publications and published as research processes or research results?

Against this background, the new roles and tasks of the different players of the 'value network' (libraries, book sellers, media and computer centres, IT-provider, publishers etc.) have to be discussed. The main topics concern the organizational and technical requirements and the business and exploitation models. These new business and exploitation models might mainly be oriented to the approaches of open-access-publishing: more author- or institution-paid publications instead of subscriptions! We have to explore the conditions and the requirements for future publications in scholarly communication: What are the appropriate service and support processes of the information infrastructure, whose importance will greatly increase facing the demands and needs of the researchers. The improvement of the publishing process and the enhancement of the publications themselves are an inevitable contribution to enhance scientific work.

In the DRIVER-Report on object models and functionalities (Verhaar 2008: 14 ff.), the requirements for storing and managing enhanced publications are identified within the DRIVER infrastructure. Accordingly, the report goes into the details of the structure of enhanced publications, which are compound objects, versioning, basic properties, long-term preservation, data model (relations/discovery), and includes a chapter about OAI-ORE. To accomplish the requirements for future publications mentioned in the report, the published research results including all the relating data should be edited and performed in the following way:

- *Performance:* Publications have a navigation to perform and to receive pictorial and textual data sets ergonomically,
- *Sharing:* Publications and all their components (data sets) have uniform resource identifier (URI) and are structured by the resource description format (RDF), so that they can be aggregated or searched as documents themselves and also in their single elements,
- *Linkage:* Publications can be integrated in various subject networks and disciplinary hubs, because they are edited in interoperable formats (e.g. XML),
- *Processing:* Publications can be processed in total as well as in their single components (data sets) by apps and tools – covering the conditions for data mining,
- *Usage:* Publications have the requirements for an access management (authentication procedures) to enable and control the enrichments and the further use
- *Mobile devices:* Publications are readable by mobile devices with their different operating systems.

Against the background of these requirements, a number of new services and support processes is needed, which libraries, media and computer centres, publishers and commercial IT provider have to implement and integrate in their service portfolios. The new services are concerning interface- and WEB-design, retro-digitization, processing and research data curation, maintenance and networking of repositories, hosting and storage of applications and data. Consequently, new business and exploitation models have to be established in order to achieve ‘open science’ and to improve scholarly communication. These service scenarios have to be integrated in the publication processes and should exploit the potential of digital media. It should be clear, that these efforts are not available for free. The accounting of services and

the licensing of generated or hosted materials, wherever it might be necessary and useful, will play an important role.

5 What will Future Publications be like?

Provided that the importance of hermeneutical and/or theoretical research is decreasing, we will be moving consequently towards data intensive, empirical approaches in research. In this case, text formats and printed books alike will probably loose their importance in scholarly communication. Indeed, somebody may assume that computer-produced documents, which aggregate or contextualise RDF-structured texts, would be some kind of enhanced documents (Gradmann/Meister 2008: 155 ff.; Dudek 2012: 197 ff.) but those documents are to a greater degree assets of research results than publications of research result themselves. In other words: The decline of texts will not be stopped by enhanced texts or patterns of RDF. The change of the research paradigm causes this decline of documents or texts. The features of future publications should exploit the potential of digital media and not be confined to 'copies' of printed books. We may find a new role for books! But what do future publications look like? This question is currently difficult to answer. On the one hand it is a matter of design and visualisation, that the access and the reception of contents and data are easy and smart. On the other hand, many support measures have to be adopted by the providers of the information infrastructure to make the future publications available and useful. But this will not work without the introduction of exploitation models and the accounting for services.

Again: What will future publications be like? If we start from the assumption, that we will have to expect future publications as objects rather than documents, we must ask ourselves what kind of presentation is able to reduce the complexity of data-based research. If documents or texts are not considered any longer, pictures and images have be taken into account. This might be surprising, but the idea of the 'image' in a technical sense might bring us to a closer understanding. A computer image covers both, the information and the structures of a file system. Future publications as digital objects will include not only data, but also coherences and structures of data, which we may call data-based content as well. In other words: If we want to present the

results of our empirical research which is based on aggregated, evaluated or processed data, we will have to illustrate or simulate our imagination of reality or that what we take for reality (Degkwitz 2012: 217 ff.). Therefore future publication will obviously be images in a very broad meaning. In every case, we need more research to create specific and suitable models for future publications. But we should keep in mind, that documents don't play the leading role anymore. We are able to create images in order to publish the data-based results of our empirical studies – and this will be our approach to publications in the future.

Acknowledgements

I would like to thank Margret Beran for her editing work.

References

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, November 2008, by order of the Association of Research Libraries (ARL). http://www.arl.org/bm~doc/current-models-report.pdf <February 12, 2013>.

Degkwitz, A. (2012). Texte, Daten, Bilder – Wissen! In: Bibliothek, Forschung und Praxis 36, 215–219.

Dudek, S. (2012). Die Zukunft der Buchstaben in der alphanumerischen Gesellschaft. Text und Dokument unter digitalen Bedingungen. In: Bibliothek, Forschung und Praxis 36,189–199.

Gradmann, S.; Meister, J. Ch. (2008). Digital Document and Interpretation. Re-Thinking "Text" and Scholarship in Electronic Settings. In: Poiesis & Praxis. International Journal of Ethics of Science and Technology Assessment 5 (2), 139–153

Hey, T.; Tansley, S.; Tolle, K. (Hrsg.). The fourth Paradigme. Data Intensive Scientific Discovery. Microsoft Research. Redmonton, Washington, Microsoft Corporation. http://research.microsoft.com/en-us/collaboration/fourthparadigm/contents.aspx <February 13, 2013>.

Hobohm, H.-Ch. (2012). Can digital libraries generate knowledge? In: Historical Social Research 37 (3), 218–229. http://www.cceh.uni-koeln.de/files/Hobohm_final.pdf <February 13, 2013>.

Mittler, E. (2012). Wissenschaftliche Forschung und Publikation im Netz. In: Füssel, S. (Hrsg.). Medienkonvergenz – Transdisziplinär. Berlin: de Gruyter, 32–80.

Pédauque, R. T. (2006). Le document à la lumière du numérique. Caen: C & F éditions. http://archivesic.ccsd.cnrs.fr/docs/00/06/22/28/PDF/sic_00000594.pdf <February 13, 2013> [English version].

Pédauque, R. T. (2007). La redocumentarisation du monde. Toulouse: Cépaduès-éditions.

Purdy, J. P. (2010). The Changing Space of Research: Web 2.0 and the Integration of Research and Writing Environments. In: Computers and Composition 27, 48–58.

Verhaar, P. (2008). Report on Object Models and Functionalities. In: DRIVER, Digital Repository Infrastructure Vision for European Research II. Leiden, 2008. http://www.driver-reposi-tory.eu/component/option,com_jdownloads/Itemid,83/task,summary/cid,54/catid,8/ <February 13, 2013>.

Woutersen-Windhouwer, S.; Brandsma, R. (2009). Report on Enhanced Publications state-of-the-art. In: DRIVER, Digital Repository Infrastructure Vision for European Research II. Amsterdam, 2009. http://www.driver-repository.eu/component/option,com_jdownloads/Itemid,83/task,summary/cid,53/catid,8/ <February 13, 2013>.