Science 2.0 and Libraries

Convergence of two sides of the same coin at ZBW – Leibniz Information Centre for Economics

Isabella Peters, Ansgar Scherp, Klaus Tochtermann

ZBW Leibniz Information Centre for Economics, Kiel/Hamburg & Kiel University, Germany
{i.peters | a.scherp | k.tochtermann}@zbw.eu

Abstract — **The "digitization" of science is currently changing the research and publication processes. This change impacts in various ways not only the day-to-day work of researchers but also library services. Libraries, however, can only manage this transition successfully if they engage in partnerships with the scientific community. That is, to jointly investigate the phenomena related to the digitization of science, draw conclusions, and develop new services. The ZBW goes beyond this idea of collaboration: Since 2014, ZBW has an in-house research group consisting of three professors, five post-docs and several PhD students. The research group covers three aspects of Science 2.0: research on the relationship between Science 2.0 and libraries, development of Science 2.0 technologies in the area of knowledge discovery, and user behavior research. This article introduces these different perspectives and shows how they synergistically contribute to fully understand all phenomena related to the transition of science.**

Keywords—Library Innovation, Science 2.0, Knowledge Discovery, Web Science

I. DIFFERENT PERCEPTIONS OF SCIENCE 2.0 AND OPEN SCIENCE

In almost all scientific disciplines the research and publication behavior is currently changing. This change has been triggered by an increased digitization of science. A good example of this change can be found with the ZBW, Leibniz Information Centre for Economics, the world's largest information center for economic literature. ZBW is member of the extra-university research organization Leibniz Association and associated with Kiel University. It is a high-tech information infrastructure conducting research in computer science and information science, and develops technologies for its own Web-based library services.

To describe the phenomenon of the ongoing digitization, different terms such as Open Science or Science 2.0 have come up. They all have in common that they call for more openness (e.g. open access), shorter publication cycles (e.g. through scientific wikis), fast feedback loops (e.g. by using social networks) or higher degree of participation and collaboration (e.g. through virtual research environments). The different terminologies reflect the focus of each research community: Open Science fosters movements towards open research as much as possible, where "open" is understood as free to use,

modify, and share by anyone for any purpose[1]. This encompasses open source software, open access publications or open research processes, i.e. making publicly available all phases of research projects [1]. Science 2.0 expands especially to how social media will impact research and publication processes. Science 2.0 research deals with the investigation of new fields for research and development, originating from the application of new participative and collaborative Internet technologies, particularly social media, in all phases of research.

Science 2.0 can enable Open Science but does not necessarily have to be open itself. For example, a research group can jointly work on a publication in a social media network, but the final publication appears in a licensed journal. Contrary, Open Science can also happen without any Science 2.0 tools.

However, with the advent of open (social) Web-based scholarly communication and Science 2.0, we face a paradox situation: on the one hand there are massive amounts of information on research – often freely – accessible, on the other hand researchers complain about information overload and lack of filters ensuring quality control. And although the research community has developed mechanisms to especially tackle latter concerns, like social recommendations and user-driven endorsements of content, it is sustainability and structures independent from the private sector which are demanded.

II. SCIENCE 2.0 AND LIBRARIES: WHAT IS THEIR RELATION?

The future roles of scientific libraries comprise two major aspects: 1) libraries will cater and enable truly new sharing processes to distribute scholarly content (e.g. publications and research data) and draw researchers' attention as well as provide new ways of access to relevant scientific work (as such still following their main mission), and 2) libraries will support mastering information overload, information integration of different content sources within and across libraries, and quality evaluation of research products by advancing search for content.

The approach followed by ZBW is currently being transferred to other libraries in Germany such as the ZBMED, Leibniz Information Center for Life Sciences, and TIB,

---

[1] http://opendefinition.org

German National Library of Science and Technology. These two libraries have already established research groups in Science 2.0 which in the near future will be further supported by newly established professor positions. The research of these professors will investigate Science 2.0 with a focus on research data visualization, data science, information provision or information delivery.

The following section summarizes the insights on the relation between Science 2.0 and libraries which have been published in [1].

Currently, libraries and digital information infrastructures provide scientists with subject-specific information at national level (on site or supra-regionally). In the future, due to Science 2.0 tools, literature provision will happen less centrally from professional-to-peer and more decentralized from peer-to-peer. This signifies an important role for the immediate provision of online available information between researchers. Scientific libraries will provide the necessary IT-infrastructure and enhance their services with existing information nodes on the World Wide Web, such as wikis, blogs, virtual research environments or profiles in social networks. All of this is necessary to provide a crucial support for this decentralized information provision. An example of such a tool is ScholarLib. The aim of ScholarLib is to make scholarly information provided by portals of scientific libraries accessible through social network sites, and vice versa [2].

Libraries will no longer exclusively act as information providers. Instead they will offer additional services (e.g. infrastructures for research data) that will support researchers in their Science 2.0-enabled publishing processes [3] and offer sustainable environments for all research products.

Libraries will master the Science 2.0 technologies that enable new paradigms for literature search. Algorithms that transport context-sensitive and individualized contents directly to researchers will be developed. One expected development is that semantic and context-related analysis of researchers' writing processes will select quality-controlled related literature and insert it into the working environment of the writer. Within this context, it is very likely that the classic library paradigm of "information pull", where researchers have to actively search for literature, will be supplemented with the paradigm of "information push", where literature is delivered proactively into the environments currently used by researchers [4]. Libraries that are tasked to provide researchers with the international research literature for their area of research, quickly and with ease, will play the important role of information providers who ensure quality and act decentralized in the background. Viral mechanisms for the dissemination of literature (such as social media and search engine optimization) and decentralized IT-services (such as social media plugins for blog platforms etc.) have great potential to define the tools of the library of the future [4].

III. SCIENCE 2.0@ZBW: HOW ZBW TACKLES SCIENCE 2.0

Right now ZBW has a unique approach to understand and meet the requirements of both Science 2.0 and Scientists 2.0. The approach is mainly driven by the changing web-based ecosystem confronting researchers with huge amounts of scientific information available, sometimes perceived as information overload, and a flood of new tools enabling advanced collaboration and communication, such as social media. By having established two strong research foci on Knowledge Discovery and Web Science the ZBW tackles Science 2.0 from two different, but complimentary angles in order to fully anticipate the Science 2.0-based changes in scholarly work. This objective is to guide researchers through this environment (covered by Web Science) and to provide tools for an enhanced and more efficient working experience (covered by Knowledge Discovery). The studies on the high-level relation between Science 2.0 and digital information infrastructures as well as its implications for policy making at different levels (e.g., funding organizations and strategic advisory bodies) completes the research group by also bridging the research foci. The overall goal is to integrate those evidence-based findings into the library's services, to support ZBW's strategic planning and development, to reflect the scientific excellence behind the library's approaches, and to advance the library community as a whole. The operationalization of this approach is based on three pillars explained in the following sections.

A. The Leibniz Research Alliance Science 2.0

In 2012, ZBW initiated the Leibniz Research Alliance Science 2.0[2]. The research program defines three grand challenges for Science 2.0 research by asking specific questions:

- New Working Habits: How does the participatory Internet with its new possibilities, particularly the social Web, change working habits of researchers? How does it impact on today's research and publication processes in different research disciplines?

- Technology Development: How can Science 2.0 support existing research processes? How can today's research processes be innovated by Science 2.0 tools?

- User Behavior Research: Which new forms of scientific communication – within the research community and between the public and the research community – does Science 2.0 facilitate? How do researchers use new Science 2.0 tools?

The objective of the Research Alliance is to investigate the research challenges related to Science 2.0 at the highest possible quality level. Within the next ten years, the alliance will follow a highly interdisciplinary research approach to find relevant answers to today's and future questions in the area of Science 2.0. The network will bundle and exploit existing synergies among its participants resulting in a notably high "group intelligence". With this group intelligence the network can deliver results and insights which are of much higher quality and which provide a much greater added value as compared to the sum of individual and isolated activities. Here, the overall goal is to provide recommendations for and guidance in similar endeavors for organizations as well as researchers facing or performing Science 2.0. Moreover, the

---

[2] http://www.leibniz-science20.de/

Research Alliance aims at developing evidence-based findings to advance the library's technologies and services.

B. Knowledge Discovery approach

Knowledge Discovery deals with the content-driven identification and localization of digital objects such as semi-structured data on the Web (i. e., Linked Open Data), documents, profiles, or communities and understanding the relationships among them. It involves the design of innovative methodologies and algorithms and their application to extensive data and document corpora of different origin and quality, also known as Big Data. Figure 1 shows a schematic depiction of the classical KDD (Knowledge Discovery in Databases) process, adopted from [5].

Figure 1: Knowledge Discovery in Databases Process

The research group applies this KDD process and thus has a technological focus with many connections to library sciences that stem from the Semantic Web. Digital libraries are strong adopters of semantic technologies, in particular so-called light-weight Linked Open Data (short: LOD). The LOD approach is a technological development to publish and interlink data of different quality and origin in the Internet. Since its emergence in 2007, the LOD approach significantly grew in popularity and leveraged the Semantic Web to success! It is used worldwide, not only in universities, research institutions, and public organizations – such as in particular libraries – but it is also adopted by very large industries such as in the sectors of media syndication and publishing as well as Internet search engines. ZBW is a LOD pioneer and co-organizer of the renowned international Semantic Web in Libraries[3] (SWIB) conference. Examples of our research in the context of LOD are an approach for increasing trust on the Semantic Web and building trust networks [6] as well as a survey on strategies for modeling LOD with particular focus on the reuse of existing vocabularies [7].

The research group on Knowledge Discovery closely collaborates with the domain experts at ZBW. For example, the group works with experts for subject indexing. Goal is the development of (semi-)automatic techniques for indexing scientific documents and its use in practice. To this end, we leverage controlled vocabularies such as ZBW's Standard Thesaurus for Economics[4] (in German: Standard-Thesaurus Wirtschaft, short: STW) to detect entities in scientific documents but also in social media [8]. The STW is a poly-hierarchical taxonomy with about 6,000 descriptors published as Linked Open Data and captures the broad spectrum of economics subjects and is connected with many other related domains such as social sciences.

Together with the domain experts, the Knowledge Discovery group investigates novel machine learning methods and metrics for multi-labeling scientific documents using the STW. We recently developed a novel multi-labeling technique based on the simple but effective kNN method in combination with using the STW for entity detection and the hypertext-induced topic selection (HITS) algorithm [9] for assessing the importance of STW concepts found in a specific scientific document [10]. The results of a 10-fold cross validation over a data corpus of about 62,000 open access documents from ZBW's EconBiz[5] literature search portal with gold standard annotations provided by experts for subject indexing of ZBW revealed very interesting results: We could predict the correct labels with an average recall of .40 (SD: .32) and an average precision of .40 (SD: .32), resulting in a F-measure of .39 (SD: .31). By this, the technique is competitive with today's approaches for multi-labeling such as Maui [11] using decision trees. Maui in its latest extension[6] yields on the same dataset only an F-measure of about .36.[7] In addition, the results also compete with commercial systems for multi-labeling using methods such as Support Vector Machines. However, in contrast to these solutions, we have the important advantage that we do not require an expensive learning phase with hand-selected documents for training the classifiers. However, as feature by design and by using the "lazy learner" kNN in our approach, we do not require an expensive training phase. In addition, any newly annotated documents can be directly leveraged for future annotation tasks. This allows for developing effective tools and applications for semi-automatically annotating scientific publications.

C. Web Science approach

The Web Science approach studies people, how they interact with the Web and what footprints they leave or networks they build when using Web- or social media-functionalities, such as sharing, liking, or creation of content or retweeting. In this approach, a special emphasis is placed on the target group of the ZBW, which is researchers from Economics and Business Studies in all career stages as well as students, and a subset of scholarly work routines, which is scholarly communication. The traces left by the researchers on the different Web- and social media-platforms are analyzed to get more information on how they use those platforms to distribute research findings, if at all, or to otherwise participate in scientific discourse. The research tool box is completed by rather qualitative methods allowing to better understand significance of quantitative data.

Since the concept of scholarly communication is strongly linked to research evaluation, publication output, and citation counts new social media-based ways of assessing the value, quality, and impact of research, often summarized as "altmetrics" [12], are also studied. Altmetrics are believed to be valuable indicators complementing the range of traditional metrics, such as h-index or impact factor, because they include new publication formats and research products (e.g., data sets, slides, videos etc.) as well as the view of a broader public which is almost inevitably confronted with research products

---

[3] http://swib.org/
[4] http://zbw.eu/stw/versions/latest/about

[5] http://econbiz.eu/
[6] https://github.com/zelandiya/maui-standalone/releases
[7] https://twitter.com/ansgarscherp/status/
590980421707571201

published on e.g. social network platforms. Hence, publications which would not have been cited, either because they were not traditional publications or because they were not recognized by other authors writing papers, can now receive countable reflections of how users interacted with them or how they were influenced.

Conducted research shows, for example, that when comparing the Impact Factors of the Top15 journals from Economics and Business Studies from the German Handelsblatt ranking with the number of readers in the social reference management service Mendeley they are ranked differently [13]. This indicates that number of citations and readers reflect different perceptions of scholarly journals. The study also revealed that the two largest groups of readers of Economics journals come from Economics and the Social Sciences; readers of Business Studies journals come from Business Administration and Economics. Moreover, Business Studies journals had by far the highest readership counts.

When economists tweet during scientific conferences, i.e. the 2014 Annual Meeting of the Verein für Socialpolitik (VfS), they mostly discuss the conference content and link their tweets to blog posts or articles [14]. The analysis of the tweeting behavior also uncovered distinct peaks in Twitter activity relating to certain conference events, e.g., panel discussion on minimum wage and immigration or talk by a delegate of the European Central Bank. Interestingly, the share of tweets containing conference content and links to articles significantly increased during those peaks (conference content from 64% to 77%; links to articles from 26% to 50%).

A survey, focus group interviews, and a panel discussion with researchers from Economics and Business studies coming from different age groups and career levels inform about actual scholarly practices and the use of social media [15]. More than a third of surveyed researchers state they use academic and professional social networks, such as Xing, academia.edu, to network with colleagues. Wikipedia is often used in research (51%), although only 6% initially answered to use the wiki for professional purpose. Here, perception and practice widely diverges among researchers. This also becomes visible when asking about use of Twitter. Only 3% of the respondents tweet for professional reasons. This is remarkable since the analysis of VfS conference tweets showed comparably high Twitter activity. Perhaps these rather contradictory results are due to the attitude towards social media engagement in scholarly communication which was discussed during the focus group interviews and the panel discussion. Social media engagement helps getting in touch with the broader public, politics, and journalism and, along with it, returns publicly funded knowledge to the society – a reason several active blogging or tweeting researchers mention when asked why they engage themselves. However, when it comes to advancing the own career they know that social media content does not substitute publications of – perceived – scientific value, traditionally published in peer reviewed journals.

IV. CONCLUSION

ZBW is the first library in Germany which has established a research group to the extent described in this paper. Either side,

the library and the researchers, benefit from each other. The researchers can conduct their research in an environment in which they have unlimited access to the library's huge content base, which allows direct access to and a continuous dialog with "real" users of the library and finally which offers plenty of opportunities to connect new technologies and algorithms with existing library services.

Still, for the library the integration is challenging. Firstly, it takes some time until research results in knowledge discovery are ready for the integration in the existing library services. Similarly, it needs some time, until user behavior and usage studies from Web Science impact on a library's strategy or on functionalities of library services. The integration challenge is also related to a well-designed knowledge transfer concept. Such a concept ensures that the researchers are released from the obligation of entering too much into non-scientific work related with the integration and maintenance of research results into library services. Additionally, the knowledge transfer concept must enable the non-scientists to understand the research results to a level which allows them do the integration.

After almost two years of scientific work in ZBW, we see convincing benefits which would have not been possible without our professors and their research groups (e.g. significant higher third party funding, radical innovations of library services, improved integration in national and European policy making bodies, increased international visibility etc.). Taking into account the challenges related to the transition in science we believe that a library can particularly well manage this transition if it does not only follow these trends but becomes an active and recognized scientific partner to also set new trends. This however will only succeed, if the library and the scientists in the library join forces in such a way that either side benefits: The library with greater innovativeness resulting in higher usage rates of their services; the scientists with unique and excellent scientific outcome which could not have been created in a mere scientific environment.

REFERENCES

[1] K. Tochtermann, "How Science 2.0 will impact on Scientific Libraries", it-information Technology, Vol. 56, No. 5, 2014, pp. 224-229, DOI: 10.1515/itit-2014-1050.

[2] P. Mutschke and M. Tamm, "Linking Social Networking Sites to Scholarly Information Portals by ScholarLib," Proceedings of ACM Web Science, 2012, URL: http://arxiv.org/abs/1205.2467.

[3] S. Vlaeminck, "Data Management in Scholarly Journals and possible Roles for Libraries – Some Insights from EDaWaX," LIBER Quarterly, Vol. 23, Issue 1, 2013, URL: http://www.edawax.de/downloads/

[4] C. Seifert, T. Borst and T. Pianos, "EEXCESS – Toolbox for managing and disseminating digital library content," Proceedings of EMTACL – Emerging Technologies in Academic Libraries, 2015, in press.

[5] U. M. Fayyad, G. Piatetsky-Shapiro and P. Smyth, "From data mining to knowledge discovery", in Advances in knowledge discovery and data mining. Menlo Park, CA: American Association for Artificial Intelligence, 1996, pp. 1-34.

[6] A. Kasten, A. Scherp and P. Schauß: A Framework for Iterative Signing of Graph Data on the Web. ESWC 2014: 146-160

[7] J. Schaible, T. Gottron, A. Scherp: Survey on Common Strategies of Vocabulary Reuse in Linked Open Data Modeling. ESWC 2014: 457-472

[8] G. Große-Bölting, C. Nishioka, A. Scherp: Generic process for extracting user profiles from social media using hierarchical knowledge bases. ICSC 2015:: 197 - 200

[9] J. M. Kleinberg, "Authoritative Sources in a Hyperlinked Environment," Journal of the ACM, Vol. 46, Issue 5, 1999, pp. 604-632.

[10] G. Große-Bölting: Comparison of different approaches for automated indexing of documents (in German: Vergleich verschiedener Verfahren zur automatischen Annotation von Dokumenten), Kiel University, April 2015. [Reported Maui results are provided by H. Schmidt]

[11] O. Medelyan, Human-competitive automatic topic indexing. Dissertation The University of Waikato, 2009.

[12] J. Priem, D. Taraborelli, P. Groth and C. Neylon, "Altmetrics: A Manifesto", 2010, URL: http://altmetrics.org/manifesto

[13] K. Nuredini and I. Peters, "Economic and Business Studies Journals and Readership Information," Proceedings of International Symposium for Information Science, Zadar, Croatia, May 2015, in press.

[14] S. Lemke, A. Mazarakis and I. Peters, "Understanding Scientific Conference Tweets," Proceedings of General Online Research, Cologne, Germany, February 2015. URL: https://conftool.gor.de/conftool15/index.php?page=browseSessions&form_session=75&presentations=show

[15] A. Mazarakis and I. Peters, "Quo Vadis German Scholarly Communication in Economics," Economics, in press, URL: http://www.economics-ejournal.org/scholeco15/presentations/quo-vadis-german-scholarly-communication-in-economics-an-analysis-on-the-use-of-social-media-isabella-peters-and-athanasios-mazarakis-zbw-2013-leibniz-information-centre-for-economics/at_download/file