



Profiles

Profile of Sören Auer, Professor for Data Science and Digital Libraries at Leibniz University of Hannover and Director of TIB German National Library of Science and Technology, and Vision Speaker at the 33rd Annual NASIG Conference

Christian Burris, Profiles Editor



Photo courtesy of Dr. Sören Auer

Dr. Sören Auer will be one of three Vision Speakers at the upcoming 33rd Annual NASIG Conference in Atlanta, Georgia. He is the Director of the Technische Informationsbibliothek (TIB) - known as the German National Library of Science and Technology in English - since July 1, 2017. In addition to these duties, he leads the "Research and Development" program area and the "Data Science and Digital Libraries" research group at TIB. His research areas include the topics of data science, digital libraries and open knowledge. I concluded my interview with Dr. Auer by e-mail on Friday, February 16, 2018.

How did you get involved in the field of digital libraries?

As a researcher I became interested in improving how we share our findings. As a young PhD student in computer science I was already playing around with open-source software like Open Journal Systems (OJS) and Fedora. Later on I started developing my own data wiki called OntoWiki, because I thought (and still think) that we need more structure and semantics. Last year I was appointed director of TIB, the German National Library and Information Center for Science and Technology, and digital libraries (in a wider sense) are now my key focus area.

Could you describe the nature of semantic data and linked data?

It is widely accepted that data plays a key role in our society nowadays, maybe even to a similar extent as natural resources, industrial products, or services. However, it is important that many stakeholders can participate in the generation, processing, and use of data. To achieve this, we need to establish a common understanding of the meaning of data. Linked data and semantic technologies help in this regard, since they provide the base structure, identification system, and semantics to develop domain-specific vocabularies in a distributed, but still interlinked, way.

Where do the fields of data science and digital libraries intersect?

Data science can help digital libraries to leverage the opportunities of intuitive digital information flows. The

vision of representing scholarly knowledge in a way that facilitates intuitive interaction dates back to the era of emerging electronic information processing. Vannevar Bush remarked in his influential 1945 essay "As We May Think" that "publication has been extended far beyond our present ability to make real use of the record." Later J. C. R. Licklider proposed in Libraries of the Future the concept of "procognitive systems," which capture the semantic relations and content within documents and data across disciplines so that they can then be queried and interpreted by users. However, so far, we have made relatively little progress in realizing their vision. Now, the document-oriented workflows in science have reached (or already exceeded) the limits of adequacy, as highlighted, for example, by recent discussions on the increasing proliferation of scientific literature, the deficiency of peer-review, and the reproducibility crisis. Despite improved and digital access to scientific publications in the last decades, the fundamental principles of scholarly communication remain unchanged and continue to be largely document-based: Researchers produce essays and articles that are made available in online and offline publication media as roughly granular text documents. Data science and semantic technologies in particular can help to complement the document-oriented information flows with knowledge- and data-driven ones.

What is the role of Open Access in digital libraries?

From my perspective Open Access is absolutely crucial but should be realized according to fair and reasonable conditions. In Germany, we currently have a large debate about this topic because negotiations with publishers about switching to an OA-based model turn out to be very difficult, and some publishers exploit their market position to the maximum extent. I think we as researchers need to put much more emphasis on publishing our research results in such a way that they are accessible to the whole society and not only to those able to pay subscriptions. Also, we should make sure that OA publishing is reasonably priced in order not to further reduce the resources for research. A shift

from venue-based to more article-based scientometric indicators hopefully can help in this regard.

What is covered in the field of open knowledge?

From my perspective, all types of content being licensed according to the Open Definition (<http://opendefinition.org>) is open knowledge. This includes software, data, OA publications, and much more. A key aspect of creating and curating open knowledge is a collaborative approach, where people in a community interact and jointly create something valuable. Examples in this regard are Wikipedia or OpenStreetMap, and we are working on fostering a similar collaboration for open educational resources with the SlideWiki platform.

What are some of the challenges that you have encountered in your field?

From a technological perspective, designing innovative systems in a scalable and sustainable way with a focus on usability is definitely a challenge. Even after my term at Fraunhofer, or working closely with industry, I feel that there is no silver bullet, but a portfolio of many small methods and persistent attention to the problem is necessary. On a social level, it is always interesting to observe that there are many different viewpoints on certain issues, depending on the background and context of a person and community. Once we are able to understand this and the reasons for certain arguments or concerns, we are able to bridge between these communities and achieve much more than in isolation. A particular challenge for digital libraries, for example, is that unlike traditional libraries, who had almost a monopoly on knowledge exchange for centuries, in the digital world libraries compete with global players. Only when we are able to collaborate as a digital library community on open knowledge, open-source, and infrastructure cost sharing (e.g., as arXiv, DataCite, or ORCID demonstrate), we can be successful in the digital world.

Where do you see the field of digital libraries in five years?

I hope that digital libraries will be able to anticipate and drive the transition from document- to more knowledge-based information flows. I envision that at some point we will be able to represent the world's research knowledge in a vast, distributed knowledge graph, which can be queried to answer research questions, such as overviews and comparisons of approach addressing a certain research problem. We recently published a position paper on this issue: <https://zenodo.org/record/1157185>.

Do you have any additional comments?

I'm looking forward to meeting the NASIG community and discussing these ideas in more detail.