Beyond Infrastructure – Modelling Scholarly Research and Collaboration

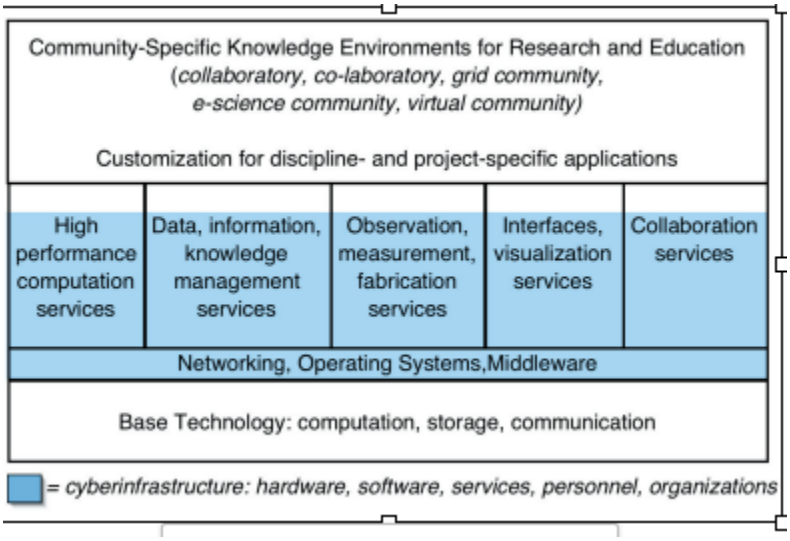
Susan Schreibman, Stefan Gradmann, Steffen Hennicke, Tobias Blanke, Sally Chambers, Alastair Dunning, Jonathan Gray, Gerhard Lauer, Alois Pichler, Jürgen Renn, et al.

Beyond Infrastructure

Modelling Scholarly Research and Collaboration

Over the past decade considerable research has been carried out into creating infrastructure to support digital scholarship -- from the "Atkins report" (Atkins et. al. 2003) commissioned by the National Science Foundation, to the more specific humanities/social science focused "Our Cultural Commonwealth" (Unsworth et. al. 2006), to the funding of large community-based infrastructure projects such as the Mellon-funded Bamboo and EU-supported DARIAH (Blanke et. al., 2011b).

The Atkins report introduced the following layered vision of the way technical research infrastructures are related to each other:



This "mother of all eScience layer cakes" introduced the hitherto canonical division between the blue area of supporting Cyberinfrastructureand the white area of discipline-specific applications. Most initiatives following the Atkins report were to focus more or less exclusively on the cyberinfrastructure layer.

The model of thought introduced by the two American-commissioned reports (Atkins and Unsworth) has been adopted in Europe, starting with the e-Science initiative in the UK that focused on the use of Grid technology (and

which evolved in parallel with the NSF activity), and the German D-Grid initiative.

But it has become clear, however, that the focus on building infrastructure, while essential to support digital humanities scholarship, needs to be accompanied by a concomitant methodological emphasis. Rockwell (2010) pointed this out in the section of his contribution to "Dangers of Infrastructure" ; i.e. that in building infrastructure we need to be aware of two major pitfalls:

- **Research infrastructure is not research** just as roads are not economic activity. We tend to forget when confronted by large infrastructure projects that they are not an end in themselves. [...].
- **Infrastructure projects can become ends in themselves** by developing into an industry that promotes continued investment. To sustain infrastructure there develops a class of people whose jobs are tied to infrastructure investment.

This paper will thus explore what is needed to foster an acceptance of digital practices in the humanities beyond the creation of pure infrastructure, specifically in terms ofunderstanding and technically modelling traditional scholarly research within a digital medium while enabling new modes of scholarly work that could only be carried out within a digitally-mediated environment.

In the latter case, this means moving beyond the emulation of traditional methods of scholarship tied to the page (albeit with linking metaphors as in the first generation document-centric WWW), to new ways anchored in the web of Linked Data which might be viewed as a combination of notebook and Memex proposed by Schraefel (2007), who is, of course, channelling Bush (1949).

In order to model this, we need to better understand how scholars undertake their research now and in the past, and how their functional framework might adequately translate into a digital context in order to attract them to new working modes. Furthermore, this kind of activity needs to be an integral part of a research infrastructure, otherwise the infrastructure runs the risk of becoming a static environment rather than a

dynamically evolving one that corresponds to ongoing and dynamic research needs.

John Unsworth (2000) conceptualized "scholarly primitives" as basic functions which are common to any scholarly activity in the humanities independent of discipline, theoretical orientation, or era. He suggested seven recursive and interrelated scholarly primitives --discovering, annotating, comparing, referring, sampling, illustrating, and representing -- which he saw as the basis for tool-building enterprises for the Digital Humanities. Since then, Unsworth's scholarly primitives have been often utilized and further revised.

As John Unsworth (2011) acknowledged in an interview almost a decade later, his list of scholarly primitives is not definitive. Subsequent research shows that there is no agreement on the exact definition or scope of scholarly primitives. However, the approach of using scholarly primitives or similar concepts proved to be a valuable and accepted means of structuring and conceptualizing the scholarly domain or aspects of it. Therefore we decided to use Unsworth's conceptualization of scholarly primitives as a starting point for our own Scholarly Domain Model. In our model, however, the scholarly primitives represent some of the most generic humanistic functions which are further broken down into more granular sub-functions which resemble scholarly activities.

Our research is part of web of scholarship currently being carried out within research infrastructure projects to link researchers' processes closer to the development of services and techniques (examples include Europeana Research Cloud and DARIAH's VCC2). Our contribution is a systematic investigation into how we can model primary research activities embracing the assumption that understanding what John Unsworth had originally proposed in terms of "scholarly primitives" more than a decade ago is central to any such approach at modelling the digital scholarly domain.

This paper will examine how deeper modelling of research processes in the humanities could inform the development of tools to enhance and augment scholarship. In particular we will focus on models of how students and scholars conduct research can be used to inform tool development, particularly in the area of text-based scholarship (both of primary texts and

metadata), focusing primarily on transcription, translation, annotation and curation.

Furthermore, our models will be enriched by ontological models which enable scholarly functions. Therefore, the aim is not only to provide a framework for categorizing and assessing tools for the Digital Humanities but also to formalize the model into a computational model in order to capture research activity and thereby also validate our Scholarly Domain Model.

Our Scholarly Domain model will go beyond *categorizing* tools to create a formal model of interrelated research primitives and functions in order to implement operational scenarios in DM2E (see below). But, the one fundamental difference is the fact that our model is explicitly geared towards a web context, to linked data environments, as the future platforms of scholarly communication and collaboration. As a consequence it uses RDF, RDFS and OWL as "glue" in an effort to ontologically formalize the primitives and their attributes as well as the relations that can be established in such an environment.

Our research is being carried out within the EU-funded DM2E[1] project and its sister projects, which includes the development of a digital humanities collaboration environment, and the development of best of breed semantic sampling and annotation tools such as Korbo[2] and Pundit[3] (originating from the SemLib project[4]). We will also share results of the JISC-funded TEXTUS[5] project which has objectives similar to those of DM2E, but extends the semantic annotation functionality into a shared citation and referencing system. And lastly, we will include the perspective of the "Virtual and Real Architecture of Knowledge"[6] activity within the "Image, Knowledge, Gestaltung" excellency cluster funded by DFG.

Among other objectives, one of the main goals of the DM2E project is to "work with digital humanities scholars and specialized application developers to explore usage scenarios of the content provided to Europeana in a

[1] http://dm2e.eu/
[2] http://dm2e.eu/
[3] http://thepund.it/
[4] http://www.semlibproject.eu/
[5] http://textusproject.org/
[6] http://www.interdisciplinary-laboratory.hu-berlin.de/en/Virtual-and-Real-Architecture-of-Knowledge

specialised environment for humanities research generating digital heuristics and making data as well as heuristics available to specialisedvisualisation or reasoning environments". The results of DM2E are intended to contribute to emerging distributed, interactive production and processing environments that go well beyond traditional working paradigms in the scholarly culture of the humanities.

Bibliography

- Anderson, S., Blanke, T,& Dunn, S. (2010): Methodological commons: arts and humanities e-Science fundamentals. In: *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences* 368 (1925)..
- Atkins, D.. E., et al. (2003) Revolutionizing Science and Engineering Through Cyberinfrastructure. Report of the National Science Foundation Blue-Ribbon Advisory Panel on Cyberinfrastructure.http://www.nsf.gov/od/oci/reports/atkins.pdf
- Bamboo (2010): Project Bamboo Scholarly Practice Report. Retrieved from http://www.projectbamboo.org/wp-content/uploads/Project-Bamboo-Scholarly-Practices-Report.pdf
- Benardou, A., Constantopoulos, P., Dallas, C.,&Gavrilis, D. (2010): A Conceptual Model for Scholarly Research Activity. iConference 2010. Retrieved from https://www.ideals.illinois.edu/handle/2142/14945
- Blanke, T., & Hedges, M. (2011a). Scholarly primitives: Building institutional infrastructure for humanities e-Science. *Future Generation Computer Systems*. doi:10.1016/j.future.2011.06.006
- Blanke, T., Bryant, M., Hedges, M., Aschenbrenner, A., &Priddy, M. (2011b). Preparing DARIAH. In IEEE 7th International Conference on e-Science, 2011.
- Bush, Vannever. 'As We May Think'. *Atlantic Magazine* (July 1945). http://www.theatlantic.com/magazine/archive/1945/07/as-we-may-think/303881/.
- Gradmann, S., & Meister, J. C. (2008). Digital document and interpretation: re-thinking "text" and scholarship in electronic settings. *Poiesis Praxis*, *5*(2), doi:10.1007/s10202-007-0042-y
- Rockwell, G. (2010, May 14). As Transparent as Infrastructure: On the research of cyberinfrastructure in the humanities. Retrieved from the Connexions Web site:http://cnx.org/content/m34315/1.2/
- schraefel, m. c. (2007). What is an Analogue for the Semantic Web and Why is Having One Important? Manchester: ACM Hypertext 2007. Retrieved from http://eprints.soton.ac.uk/264274/1/schraefelSWAnalogueHT07pre.pdf

- Unsworth, J. (2000). Scholarly Primitives: what methods do humanities researchers have in common, and how might our tools reflect this? *Symposium on Humanities Computing formal methods experimental practice*. Retrieved fromhttp://www3.isrl.illinois.edu/~unsworth/Kings.5-00/primitives.html
- Unsworth, J., et al. (2006).Our Cultural Commonwealth. Report of the American Council of Learned Societies Commission on yberinfrastructure for the Humanities and Social Sciences, Retrieved from http://www.acls.org/cyberinfrastructure/cyber.htm
- Unsworth, J.,&Tupman, C. (2011): Interview with John Unsworth, April 2011, carried out and transcribed by Charlotte Tupman. In: Marilyn Deegan und Willard McCarty (Hg.): Collaborative research in the digital humanities. Farnham: Ashgate.