

Bibliothek und Wissenschaft

Herausgegeben von
Claudia Fabian, Michael Knoche,
Monika Linder, Elmar Mittler,
Wolfgang Schmitz und Hellmut Vogeler

44 · 2011

Harrassowitz Verlag · Wiesbaden

Digitale Edition und Forschungsbibliothek

Beiträge der Fachtagung
im Philosophicum der Universität Mainz
am 13. und 14. Januar 2011

Herausgegeben von
Christiane Fritze, Franz Fischer, Patrick Sahle
und Malte Rehbein

Harrassowitz Verlag · Wiesbaden

ISSN 0067-8236
ISBN 978-3-447-06568-9

Inhalt

ABSTRACTS	1
ELMAR MITTLER und MALTE REHBEIN: Edition und Forschungsbibliothek – Chancen und Herausforderungen einer traditionsreichen Partnerschaft im digitalen Zeitalter	9
BODO PLACHTA: Edition und Bibliothek	23
ULRICH JOHANNES SCHNEIDER und ZEKI MUSTAFA DOGAN: Digitaler Humanismus: Das Beispiel des Codex Sinaiticus	37
MICHAEL STOLZ: Benutzerführung in digitalen Editionen. Erfahrungen aus dem Parzival-Projekt	49
STEFAN CRAMME: Editionen in einer bildungshistorischen Forschungsbibliothek	81
MATS DAHLSTRÖM: Editing Libraries	91
THOMAS STÄCKER: <i>Creating the Knowledge Site</i> – elektronische Editionen als Aufgabe einer Forschungsbibliothek	107
ANDREA RAPP: Digitale Edition und Forschungsbibliothek. TextGrid als virtuelle Infrastruktur für digitale Editionen	127
PETER BOOT and JORIS VAN ZUNDERT: The Digital Edition 2.0 and The Digital Library: Services, not Resources	141
REINHARD ALTENHÖNER: Trau, schau, wem – Zur Authentizität und Langzeitverfügbarkeit digitaler Objekte	153
NEUERWERBUNGEN DER BIBLIOTHEKEN	171
ANSCHRIFTEN DER AUTOREN	183
DANK	185

Peter Boot and Joris van Zundert

The Digital Edition 2.0 and The Digital Library: Services, not Resources

Introduction

In this article we want to explore and theorize the future forms, functions and requirements of the digital scholarly edition, the IT architecture that would fit such a digital edition, and the opportunities and problems such digital scholarly editions might pose for institutional libraries in general and digital libraries specifically. Our point of departure is the expectation that future digital editions will be characterized by high volatility of data and functionality during their creation in a research context, and for as long as they are in use for research and reception. We will also argue that such digital editions of the future will be composites of distributed data sources and functionality. These expectations result in a vision of the future digital edition implying changing requirements for digital libraries and other institutions that are in a position to assure their permanency. We suggest that these developments form an opportunity for libraries to establish a new active role of guidance and monitoring in scientific knowledge dissemination.

New possibilities, new methods

Humanities scholars have been keen to pick up on the possibilities of new digital technologies. At first they were primarily driven by the blessings of having remote resources available under a simple mouse click. Recently, the ability to share their resources and research on the web, for others to use and reuse, has become another motivating factor¹. As well as this ›collaborative shift‹ the application of computational methods to humanities research will be another driver behind the creation of digital resources and services². Communities of textual and documentary scholars are moving into a period dominated by the need for digital resources: increasingly they are using sophisticated digital technologies to create tools,

1 Spiro, Lisa: Examples of Collaborative Digital Humanities Projects. Houston 2009. Accessed online 13 October 2010, <<http://digitalscholarship.wordpress.com/2009/06/01/examples-of-collaborative-digital-humanities-projects/>>.

2 See Crane, Gregory: Give us editors! Re-inventing the edition and re-thinking the humanities. In: Online Humanities Scholarship: The Shape of Things to Come, ed. by J. McGann. Houston 2010 S. 81–97. Accessed online 10 January 2011, <<http://cnx.org/content/col11199/1.1/pdf>>. Also KNAW, eHumanities. Amsterdam 2011. Accessed online 7 February 2011, <<http://ehumanities.nl/>>.

conduct collaborative research and publish results. The move is from a desktop perspective of research towards a networked paradigm.

Current ›born digital‹ editions are partly illustrating this shift. Foreshadowing such future editions are multimodal and multimedial digital publications like modern drama editions that include performance video³, editions that include a geographical component⁴, an edition of an account book that permits transfer of the mentioned amounts into a data worksheet⁵, the inclusion of a database component in the digital edition⁶, an edition of a musical that includes sound fragments⁷, and so on.

We can identify five main drivers or characteristics of this shift to networked research in digital research resources (and thus to digital scholarly editions in particular). The first of these is an ever rising need for digital *availability*: researchers want full transparent access to any data, anytime, anywhere. Researchers, when looking for data and information, want access to pre-print, e-books, facsimile and web data alike. Related to this need is the requirement for good tools for the discovery and exploration of such resources. The second characteristic is the need for *reliability*, which comes in two forms: trustworthiness and sustainability. Trustworthiness depends on, among other things, authorization schemes, peer evaluation that works in a new digital era, and the ability to evaluate the state of a particular resource at a certain point in time. Sustainability creates trust that digital resources can be relied upon to be there for the foreseeable future. A third characteristic is the inherent *heterogeneity* of humanities research: humanities researchers use multiple and often eclectic approaches to multiple data types. They depend on discovery, exploration and analysis of data of potentially any form and unit of measurement: visual, textual, temporal, numerical, or geographical. The fourth driver of a more networked research paradigm results from the heterogeneous character of data and research in the humanities: the need for *augmentation*. More and more digital instruments and resources in digital scholarly research are used to comment, annotate, enrich and map data, which increases their scholarly value and analytic potential. The rise of enriched data in its turn drives (but is also driven by) emerging *computational approaches* such as those applied in stylistics, discourse

3 Brome, Richard: Richard Brome Online. London 2010. Accessed online 3 October 2010, <<http://www.hrionline.ac.uk/brome/>>.

4 Jewell, Andrew: Mapping a Writer's World. A Geographic Chronology of Willa Cather's Life. Lincoln 2010. Accessed online 3 October 2010, <<http://cather.unl.edu/geochron/>>.

5 Keating, John G., Teehan, Aja, Gallagher, Damien, O'Connor, Thomas: A Digital Edition of a Spanish 18th Century Account Book: Part 1 – User Driven Digitisation. In: *Jahrbuch für Computerphilologie* 10. Paderborn 2010.

6 Rehbein, Malte: Reconstructing the textual evolution of a medieval manuscript. In: *Literary and Linguistic Computing*, 24 (2009) 3 S. 319–327.

7 MITH: Music Theatre Online. Maryland Institute for Technology in the Humanities. Maryland 2009. Accessed online 18 October 2010, <<http://mith.umd.edu/mto/>>.

analysis and corpus linguistics⁸. Having augmented or curated data digitally available in large quantities is a prerequisite for more quantitative and data intensive analysis in the humanities domain⁹.

New forms: open and distributed scholarly digital editions

If one combines current practice in digital scholarly editions with the characteristics of the shift towards a networked and computational research paradigm for the humanities identified above, it is possible to infer the most probable characteristics of future digital editions.

The ›digital scholarly edition 2.0‹ (as we will dub it, tongue-in-cheek) will fundamentally be open ended. It will not ›freeze in time‹ like a book, although versioning technology will make it possible to retrieve the state – a snapshot if you like – of the edition as it existed at a certain date¹⁰. Open ended editions may be enhanced with information about other copies of the edited text, with annotation¹¹, with commentaries, with links to other relevant material, and so on. They may be enhanced through ›Web 2.0‹ and other technologies that are emerging in the field of textual scholarship; technologies like crowd sourcing, social media and data mining¹².

These editions will be the result of networked research: they will arise from the efforts of many individuals and organizations, with different backgrounds, skill sets and financial

8 Cf. e. g. Clement, Tanya, Steger, Sara, Unsworth, John, Uszkalo, Kirsten: *How Not to Read a Million Books*. s.l. 2010. Accessed online 22 September 2010, <<http://www3.isrl.illinois.edu/~unsworth/hownot2read.html>>; Van Dalen-Oskam, Karina, Van Zundert, Joris: *Delta for Middle Dutch: Author and copyist distinction in »Walewein«*. In: *Literary and Linguistic Computing*, 22 (2007) 3 S. 345–62; Bod, Rens et al.: *Computational Humanities; Workshop at the 10th International Conference on Computational Science*. Amsterdam 2010. Accessed online 22 September 2010, <<http://staff.science.uva.nl/~rens/CompHum2010.htm>>.

9 Hey, Tony, Tansley, Stewart, Tolle, Kristin eds.: *The Fourth Paradigm: Data-Intensive Scientific Discovery*. Redmond 2009.

10 Cf. The way the state of a Wikipedia page at a certain date can always be recovered through its history page. For a more principled approach see Van de Sompel, Herbert, et al.: *Memento: Time Travel for the Web*. Arxiv preprint arXiv:0911.1112. s.l. 2009. Accessed online 13 October 2010, <<http://arxiv.org/abs/0911.1112>>.

11 Boot, Peter: *Mesotext: Digitised Emblems, Modelled Annotations and Humanities Scholarship*. Amsterdam 2009. (PhD Thesis, Utrecht University).

12 Siemens, Ray: *Underpinnings of the Social Edition, »A Narrative, 2004–9, for the Renaissance English Knowledgebase (REKn) and Professional Reading Environment (PReE) Projects«*. In: *Online Humanities Scholarship: The Shape of Things to Come*, ed. by J. McGann. Houston 2010 S. 223–259. Accessed online 10 January 2011, <<http://cnx.org/content/col11199/1.1/pdf>>.

resources. They will consist of networked (or mashup¹³) functionality and content. As the analytic needs of humanities scholars and researchers are highly diverse, the corresponding functionality will be highly diverse as well. Reading, lookup, search, annotation, linguistic analysis, stemma analysis, collation, and authorship attribution are the more elementary ones; the expected rise in computational approaches indicated above will add to these functional possibilities. However, as these functionalities become ever more specialized, a digital edition should not, as a rule, establish its own advanced functionalities. Editions should make use of services (e. g. a collation or indexing engine) that are already published on the web, for reasons of both technological and scientific efficiency and reliability. This implies that software/services will be highly distributed.

The data resources that digital editions will fetch from will be similarly distributed. Examples might be: integrated access to research literature on the edited texts; access to (digital representations of) earlier editions of the edited texts or documents; access to literature documenting the critical reception of the edited texts; dictionary lookup in a customizable dictionary or set of dictionaries; access to server log based information about navigation patterns within the edition; etc.

We thus see a digital scholarly edition 2.0 that is diverse and distributed with respect to data sources, services and functionality. Its borders with virtual research environments, digital archives, knowledge sites¹⁴, thematic research collections¹⁵ and digital libraries will be fuzzy¹⁶.

Some of the components of such editions might be created and curated by libraries. Many digital editions, for example, will at some point need to refer to newspaper articles about the circumstances surrounding the creation or publication of the edited work. A library maintaining a digital newspaper corpus might create the necessary services for displaying these articles and for exploring related articles. Another set of services that a digital library could provide would help the reader find texts that are semantically related to the edited texts.

13 For mashups see Schroth, Christoph, Janner, Till: Web 2.0 and SOA: Converging Concepts Enabling the Internet of Services. In: *IT Professional*, 9 (2007) 3 S. 41. Also Floyd, Ingbert R., et al.: Web mash-ups and patchwork prototyping: User-driven technological innovation with web 2.0 and open source software. In: *HICSS '07 Proceedings of the 40th Annual Hawaii International Conference on System Sciences* 40th, IEEE Computer Society. Washington 2007.

14 Shillingsburg, Peter L.: *From Gutenberg to Google: Electronic Representations of Literary Texts*. Cambridge 2006.

15 Palmer, Carole L.: *Thematic Research Collections*. In: *A Companion to Digital Humanities*, ed. by Susan Schreibman, Ray Siemens, and John Unsworth. Oxford 2004 S. 348–65.

16 Price, Kenneth M.: *Edition, Project, Database, Archive, Thematic Research Collection: What's in a Name?* In: *Digital Humanities Quarterly*, 3 (2009) 3. Accessed online 11 February 2011, <<http://www.digitalhumanities.org/dhq/vol/3/3/000053.html>>.

What these examples illustrate is one of the general points we try to make in this paper: it is not enough for a library to own and keep a resource and to make it accessible through its own web pages: the resources that a library owns should be made accessible in such a way that they can be integrated in other sites and products; that is, as services, not as resources.

Sustainability

Innovations in digital scholarly edition technology may have major epistemological and methodological benefits. But these benefits come at a cost. The functionalities of the digital edition 2.0 are built on the highly dynamic (or unstable if you like) platform of the World Wide Web, which is, so to say, shifting under our feet. This is not an imaginary problem: the creation of a shared digital textual heritage spans just a few decades, but already at this point in time that heritage is arguably more at risk than non-digital cultural heritage. Many digital incunabula are already all but lost due to the virtual world's tendency to reinvent itself. This problem of sustainability applies to tools as much as it applies to digital data. As an example for tools we can point to the venerable Collate, the text collation tool developed by Peter Robinson in the early 1990s. Because of platform migrations it has by now become virtually unusable¹⁷. As for digital editions, one example among many is provided by the CURSUS project, once led by Professor David Chadd. The project intended to employ XML together with XSLT transformations to make data available on the Web from sources of medieval Latin liturgy. Unfortunately the primary investigator passed away in 2006. The digital edition's website lingered on, and has meanwhile become unreachable¹⁸. Though the data and code is still recoverable, it is at risk of degeneration¹⁹.

Textual scholars have been re-implementing and reinventing the digital edition on a permanent basis, from its beginnings on CD-ROM, through static web publication, interactive web publication, and now moving onto ›Web 2.0‹. This continuous application of cutting edge technology means introducing issues of sustainability. At the Huygens Institute, work on *eLaborate*²⁰, a framework for preparing and publishing digital scholarly editions under continuous development, shows that regular revisions to the code of the system are

17 Robinson, Peter: Current issues in making digital editions of medieval texts – or, do electronic scholarly editions have a future? In: *Digital Medievalist*, 1 (2005) 1. Accessed online 10 January 2011, <http://www.digitalmedievalist.org/journal/1.1/robinson/>.

18 Compare <http://replay.waybackmachine.org/20061012155248/http://www.cursus.uea.ac.uk/> (accessed 2011-02-09) and <http://www.cursus.uea.ac.uk/> (accessed 2011-02-09)

19 Cummings, James: Liturgy, Drama, and the Archive: Three conversions from legacy formats to TEI XML. In: *Digital Medievalist*, 2 (2006) 1. Accessed online 9 January 2011, <<http://www.digitalmedievalist.org/journal/2.1/cummings/>>.

20 Cf. <http://www.e-laborate.nl/en>.

necessary to keep it in sync with current IT technology. The maintenance load of such a framework is therefore considerable. The digital scholarly edition may prove to be a very vulnerable digital artifact, and preserving the existing digital editions may become a growing technological and organizational burden.

Sustainable digital scholarly edition infrastructure

The changes in form and function of digital editions are driven by both technological and epistemic (or heuristic) innovations. Continuous innovation implies vulnerability for data resources and software components, among other reasons because different components implement different ideas about system design. A service based architecture²¹, now generally accepted as the best approach for developing software for cooperation and publication in the humanities²², can be an important ingredient in counteracting this vulnerability. Services with publicly defined and stable interfaces create the possibility for service consumers to switch between service providers and for service providers to change implementations without bothering service consumers²³. Combining service orientation with publication of both data resources and services through cloud computing and cloud storage will offer enhanced permanency²⁴. In effect this could leverage a fundamentally distributed character of the digital scholarly edition 2.0 as the very mechanism to enhance its permanence and sustainability.

So what are the ingredients of a cloud based edition platform that enhances our editions' resilience to aging? First, a service oriented architecture that provides independence bet-

21 Gold, Nicholas: Service-Oriented Software in the Humanities: A Software Engineering Perspective. In: *Digital Humanities Quarterly*, 3 (2009) 4. Accessed online 11 February 2011, <<http://digitalhumanities.org/dhq/vol/3/4/000072/000072.html>>.

22 See Dreyer, Malte, et al.: eSciDoc – a Scholarly Information and Communication Platform for the Max Planck Society. s.l. 2007. Accessed online 11 February 2011, <<http://edoc.mpg.de/315471>>. Also Küster, Marc Wilhelm, Ludwig, Christoph, and Aschenbrenner, Andreas (2007): TextGrid as a digital ecosystem. In: *IEEE DEST 2007*, ed. by E. Chang. Cairns 2007. and Váradi, Tamás, et al.: CLARIN: Common language resources and technology infrastructure. In: *Proceedings of the 6th International Conference on Language Resources and Evaluation (LREC 2008)*. s.l. 2008. Accessed online 13 October 2010, <<http://www.lrec-conf.org/proceedings/lrec2008/summaries/317.html>>.

23 In informatics this is also known as the application of a façade design pattern towards web enabled services or applications. A façade is a special kind of interface/api that hides the specifics of the underlying implementation, adapting existing interfaces and apis to common standards or more abstract level interfaces. Cf Gamma, Erich, et al.: *Design Patterns: Elements of Reusable Object-Oriented Software*. Boston etc. 1995.

24 Strauber, Christophe: Cloud Computing: Distributed Power, Remote Storage, and Web Services. In: *More technology for the rest of us: a second primer on computing for the non-IT librarian*, ed. by N. Courtney. Santa Barbara 2010 S. 31–40.

ween the editions' components; second, a move towards generic services working on specific data; third, redundancy in storage, functionality and even service implementations; fourth, monitor and broker services, meta-services that observe the status of services and route services invocation based on their current statuses; fifth, alerting services, resources that notify the editions' stakeholders about non-availability of services or data resources; sixth, versioning services, which can fallback to earlier versions of an edition that is currently non-functional; and seventh, migration services that can help adapt to the changes that will, however good the rest of the architecture, still be inevitable. We can here discuss only some of these components.

At the most abstract level, the edition consists of data and services. Among the data are transcriptions, annotations, and facsimile images. The services are invoked to process the data and present the edition's user interface to the reader/user. They function at multiple levels: one service may generate the edition's table of contents, another one fetches the annotations belonging to a single line of poetry and yet another computes a manuscript stemma based on collation data. The basic idea of ensuring permanence of the edition through cloud technology is that (i) data can be stored redundantly, with network facilitated fallback to redundant copies if a master copy fails, and (ii) services can (read: should) have multiple implementations, and service invocations should be handled by broker services that are aware of the current status of these services. The network monitors the services' availability and knows the functionality that services implement. On the basis of that information, the broker services can decide which services to invoke.

One condition for this idea to work is that we need to move in the direction of reducing the amount of custom developed software. At present, the digital edition often consists of one or more source files, usually encoded in TEI, and a software layer that ›knows‹ what the user interface looks like, ›knows‹ where to fetch the relevant image files, and creates HTML pages based on that information. Even though the software layer probably includes a number of more or less standard components (publication frameworks such as Cocoon or XTE, standardised UI and Javascript libraries, etc.), currently the software layer as a whole is custom developed for each edition, and it is therefore unlikely that there will be alternative implementations of the service that it provides.

Rather than custom built software, what we should have is generic logic, implementing generic functionality, and data files that specify the specifics of the edition. The edition's user interface should probably be specified in a user interface markup language such as XUL, mapping between the user interface and the edition content probably in a language still to be developed²⁵. Generic services can then create the user interface from the edition's content, based on the user interface and mapping definitions.

25 What we are suggesting here is that we need a vocabulary to describe declaratively which part of the edition content should go in which ›slots‹ in the edition's user interface. Describing this procedurally (i. e. programming it) implies developing edition-specific logic.

The same approach should be followed wherever possible. The edition data files should declare which facsimile images are relevant to which part of the text; a generic service can then fetch the images corresponding to the text that is being shown. Again, data files should point to relevant texts elsewhere (letters, other editions, newspaper articles); generic services should fetch these texts when necessary.

Once this separation of data and functionality is accomplished, the cloud platform should provide the services mentioned above, such as monitoring, brokering, messaging and versioning. It is important to notice that versioning should not just be application level versioning (showing the edition as it looked five years ago) but should include platform level versioning (restoring or simulating the operating system and other infrastructural components of five years ago), because it is quite possible that the edition's services depend on that. Migration services may be one way of avoiding that necessity, and should include not just the already much discussed format migration but also migration of software to newer environments.

While not covering all eventualities, the redundant publishing and mirroring of such resources means that digital editions as interfaces on those services will remain in operation, even if one of the resources would be unavailable at a certain time. The need for human intervention remains, however, as even in the case of multiple service implementations one day the system may run out of fallback options.

New technologies, new problems

The architecture as described above is a theoretical ideal. Realizing such a triple redundancy based network of resources and services would be a considerably ambitious venture even with major industry partners involved. Moreover it would mean a significant change in the strategies currently deployed for long term archiving, raising in passing much of the challenge from institutional to inter-institutional level. But as long-term availability of the resources for any digital edition is an essential issue, we should nevertheless consider it seriously.

Assuming the primary mission of the library is to ensure long term access, for the scholars of today and tomorrow, to the output of scholarly work, how can the library help address some of the following issues in the longevity and accessibility of this envisioned distributed, cross medial, open edition of the future? The potential challenge is extensive and more important: real. Because even if the redundancy solution is mere utopia for now, the advent of the distributed digital scholarly edition is a reality²⁶. Thus digital libraries and

26 See e. g. the work of Peter Robinson on digital edition referencing schemes, Robinson, Peter: *Electronic Editions for Everyone*. In: *Text and Genre in Reconstruction: Effects of Digitalization on Ideas, Behaviours, Products and Institutions*, ed. by Willard McCarty. Cambridge 2010 S. 145–63.

their host institutions will be confronted by problems of digital sustainability pertaining to digital scholarly editions in the near future.

The list of potential problems is long, and we can identify only a few here. We already identified the problem of the vulnerability of digital resources. Digital resources and services may become unavailable, either temporarily (service is down) or permanently (service is discontinued). Hosting institutions can be closed down or face changing objectives and priorities. This risk to permanency is aggravated by future migrations and upgrades to baseline services (such as authentication services), platforms and code libraries. Migrations of which we can only be certain they will occur, but of which the impact is in principle unpredictable. But also non-baseline services (like facsimile serving, annotation linking services, and analytical services) are sure to undergo upgrading, changing their capabilities and complexity, thus changing the very ›fabric‹ of which distributed digital editions will be made.

Replication and redundancy of data resources, services and implementation would, as suggested, preempt part of these problems. Providing permanency and long term access would then imply libraries adding the replication and mirroring of data and processing services to their repertoire of activities.

Another key problem is presented by the open ended, dynamic, and evolving nature of the future digital scholarly edition. Open editions will, after their initial creation, be enhanced beyond the work of the original makers, with new texts, new witnesses, new commentary, new (bodies of) annotations, new hyperlinks, etc. And, quite possibly, these ›secondary‹ editors will not just add new texts but also modify existing ones, the texts provided by the primary editors. As an open, collaborative and evolving approach, this is, in principle, to be welcomed. But even if we assume the emergence of some trusted networked peer evaluation system, do we really accept an edition will change if one of its components or data resources elsewhere is modified? If, for example, we edit a writer's correspondence and a historian elsewhere decides a letter must have been written at another date and changes a database to reflect that decision, do we really want the edition that refers to that database silently to import that changed date? Or do we accept possibly resulting internal inconsistencies? Catching these problems would require some form of editorial control to be embedded in the workflow. Similar problems arise from the possibilities for well intentioned but non-expert contributions to degrade a site's overall reputation and trustworthiness, and a fortiori from malicious contributors and spammers.

But the foregoing is not to say the only possible solution is ›freezing‹ the edition, allowing for no change. If we accept the distributed nature of future editions, this will be impossible

Also the work on Canonical Text Services by Neel Smith: The Canonical Text Services Protocol. Cambridge, MA 2004–2008. Accessed online 30 September 2010, <<http://chs75.chs.harvard.edu/projects/diginc/techpub/cts>>.

in any case. But moreover, probably we do not want to lose the potential the evolving nature offers to keep changing, improving and enhancing an edition²⁷.

An advanced version/state management system should be able to retrieve the state of the composite of resources and services that makes up a certain digital edition for any point in time in the past. This would permit change while still being able to sign a certain state as authoritative. It would then be possible to signal the additions and changes added later (and also to hide them of course). One could even imagine forking paths in the edition's development: a scholarly society or a later respected scholar creating an amalgam out of the original edition and a number of later components, signing the result as his preferred edition, while perhaps a community edition would include a larger number of crowd contributed components. But more probably a certain amount of editorial control would ensure the edition's authoritative state was updated at some point and would guide the development of the digital edition.

Does this essential fluidity of the future digital scholarly edition imply that libraries will have to take responsibility for safeguarding that dynamic nature of editions? This presents us with a paradox: the role of the library in future is to provide permanent access to editions that shift under our feet, to provide the permanency of change?

The ›digital scholarly edition 2.0‹ and the role of the library

If we accept that this paradox is upon us, and if we accept that the primary role of libraries is to safeguard long term access, how can libraries approach that challenge of perpetuating access to digital open ended scholarly editions that might be permanently ›in flux‹? We argued that digital scholarly editions will be composites of three types of digital web services: data sources, processing services, and interfaces, and that redundant implementation of such components combined with automated replication and versioning, would in essence provide for long term accessibility of these editions. Should it therefore be the role of future (digital) libraries to provide for such redundant availability? If so, there is a plethora of actual information technology tasks for the libraries to tend to. Should the libraries be reduced to maintainers of server grids, so that enough redundant cloud storage and real time computing capacity will permanently be accessible? Should they make themselves responsible for digital information network processes like mirroring, replicating, load balancing, etc.? These are indeed essential tasks associated with providing any permanency of access to volatile digital humanities research information and publications. But such tasks are also the core business of academic computing centers and large scale IT industry.

27 And of course the opposite might happen as well: it just might be that researchers will not be interested in enhancing existing sites as a more or less anonymous contributor, preferring instead to create (and be credited with) new editions or new interpretations.

Such institutions might be more efficient than libraries in supporting specific ›low level‹ infrastructure technology.

But on the other hand: humanities research information quality and process is not the core expertise of computing centers and IT industry. Google provides a notorious example. The IT giant's mass digitization efforts and the analytic tools it provides to scholars and general public alike are undoubtedly useful. However, its perception of form and function of digital documentary sources is still a diamond in the rough at best, as was witnessed by the recent backlash against its ngram viewer and the possibility for ›culturomics‹²⁸. Ideally we would see the coverage and scale of the IT might combined with the precision and quality of textual scholar's approach. It is in striving for this hybrid vigor that the libraries may have an important role.

Libraries have an unparalleled experience and knowledge of compiling, safeguarding and providing access to humanities research information. They also have centuries of experience in managing the quality of the associated processes. Libraries therefore should step in as curators, guaranteeing long term access to volatile digital humanities research output. Though IT partners may provide the critical technology, and researchers and research institutions may provide initial form, flow and content, it is the libraries that are able to define, manage and maintain the processes, workflows and quality controls that can assure the edition's long term availability in the digital realm.

The trends and developments we described present an opportunity for libraries to venture into establishing a role of guidance and monitoring. Recognizing the reported trends can help libraries establish the processes that assure long term accessibility of volatile digital humanities research information and ensure their active role in scientific knowledge dissemination. We are in dire need of custodians for the processes that safeguard the services, interfaces and data that the digital humanities produce. This curation is not ›just‹ an IT task. It is an intrinsic and crucial part of the process of long term humanities knowledge diffusion that has been and should be primarily in the care of the library institution.

In essence the trends we described shift the focus of maintenance, archiving and sustaining digital editions from the digital edition as a complete resource to its associated volatile digital data resources, its functional services, and its interfaces. As the ›product‹ becomes volatile, the focus of providing permanency shifts to the processes. These trends are tied to shifting research methodology and changing modes of knowledge output, dissemination and reception. Knowledge representation becomes dynamic and volatile.

Pioneering knowledge creation and publication within this changing context is the task of researchers and editors. Providing and maintaining possible emerging larger scale digital infrastructure could probably be left best to computing centers and IT industry. But capturing, standardizing and monitoring the resulting processes of knowledge flow

28 Michel, Jean-Baptiste et al. (2011): Quantitative Analysis of Culture Using Millions of Digitized Books. In: *Science*, 331 (6014) S. 171–182.

and representation, to be able to guide technology providers on the one hand and support volatile digital editions on the other, is a major challenge for the libraries of the future.