

Mats Dahlström

Editing libraries

The title of this essay is made up of two words, editing and libraries, both with a range of varying connotations. Here, however, »editing« is shorthand for scholarly editing based on textual criticism and bibliographical acumen. »Library« denotes many things of course, but the two perhaps most common definitions are, firstly, a collection of documents amassed, organized and managed according to one or several consistent principles, and secondly, the institution that ammasses, organizes and manages such a collection. The title therefore opens up for three avenues within the essay:

1. the editing of collections of documents,
2. libraries as institutions performing such editing tasks, and
3. the relations between these two activities

Critical mass

Recent reconfigurations of the web have subjected these areas to fundamental changes. There is as of yet no sign that these reconfigurations are coming to an end. Through their efforts to digitize cultural heritage material, libraries and similar memory institutions are increasingly becoming editing and publishing agents that not only deliver the material to the general public, but indeed shape the very commons of cultural heritage. Likewise, through digitization these institutions not only face new opportunities to support and enhance education and research, but indeed produce research data themselves and embed scholarly work in their digitized collections. Steadily, cultural heritage digitization efforts are approaching a critical mass, which should allow the memory institutions performing them to develop conscious strategies for such an editing and publishing responsibility.

Libraries have however not yet fully embraced the consequences of this new role. As a case in point, the very idea of seeing themselves as publishers, not to mention as editors, seems strange to most librarians. I think this has some grave consequences. Many libraries simply do not take the publishing task in the digital realm seriously. In many respects, an analog mode is simply transferred to the digital realm, with respect to e. g. exhibiting and making material available (or not). Further, most digitizing libraries seem to take too lightly the issue of whether the products are not only usable but indeed used at all, not only accessible but indeed accessed at all.

Small-scale examples

There will be option to return to these issues later in the essay. For the time being, however, let us have a brief look at a couple of student digitization projects. They stem from a course in cultural heritage digitization, directed at employees within the memory institution sector, which my home department has been running for some years¹. As their major assignment, students perform a digitization project, all the way from managing the physical source documents to creating a web site where the digitized material is presented and commented. The students manage most of the steps in the process themselves. The course has been focussing:

- high resolution image digitization and image editing
- thorough TEI encoding, with XHTML generation through XSLT
- careful selection of source documents and an attention to issues of textual criticism and material bibliography
- accessibility (students are required to publish their material in an open access and open source manner)
- critical documentation (students document their work and compose critical reports about the work they have done)

In the first example, the students produced high-resolution digital facsimiles of some autograph letters by the Swedish poet Gustaf Fröding (1860–1911). In figure 1, the facsimile images can be reached by clicking on thumbnail images. In addition, the students manually transcribed the textual contents of the letters (including envelope paratexts) and encoded the transcription using rich TEI markup, with e. g. tags indicating irregular vs normalized, or original vs modernized spelling. They then wrote XSL style sheets to have the TEI file transformed into various XHTML outputs to serve different user groups, while also making both the TEI file and the style sheet files available open source (in figure 1 this is indicated by live links to the transcribed (diplomatic) text, to the modernized text, and the TEI+XSLT source files).

Another student project digitized a Swedish classic children's book by author and artist Ivar Arosenius, *Kattresan* (*The cat's journey*), originally published in 1909. The students experimented with several approaches to (re)present the digitized book. Not only did they produce a straightforward, linear-mode digital facsimile along with a text transcription marked up with TEI and then transformed to XHTML. They also wrote XSL style sheets to transform the TEI source file into a type facsimile representation (striving for a graphic simulation of the facsimile version), a diplomatic text version, a normalized text version,

1 The course is described in more detail and with critical comment in Dahlström, Mats & Doracic, Alen: Digitization education: courses taken and lessons learned. In: D-Lib Magazine 15, 3/4 (2009). <<http://www.dlib.org/dlib/march09/dahlstrom/03dahlstrom.html>> (accessed at 2011-02-11).

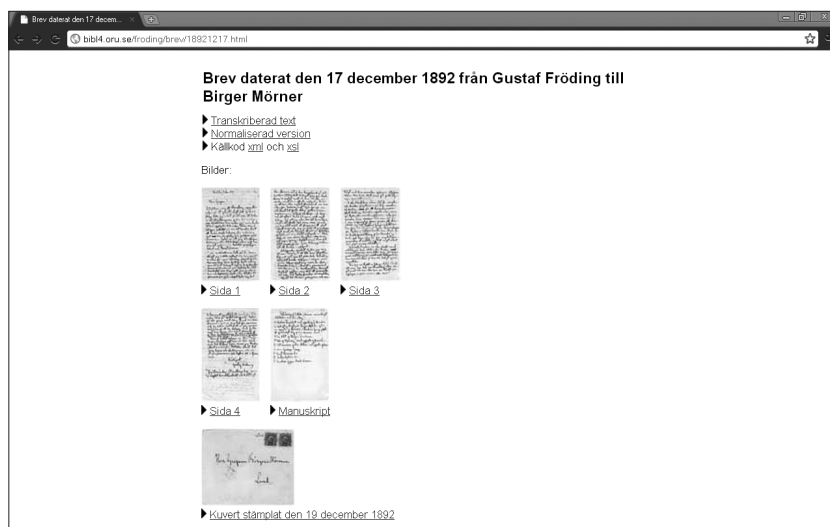


Figure 1. Screenshot of a student project on the Gustaf Fröding letters.

different versions highlighting either tropes and rhetorical figures, or grammatical components, or the mentioning of people and animals, etc. The students even went to some length identifying an extant library first edition copy whose condition would provide them with the best opportunities to digitize. In so doing, they also came across later editions and began to notice significant differences between editions and copies. This inspired them to write ancillary pages to the project web site illustrating such differences with details and commenting them. Although figure 2 is in Swedish, it might convey some idea of the resources the students created: the horizontal menu provides links to different aspects of the digitized work for different user groups such as school pupils, students and scholars of art, literary studies, or linguistics. The vertical left-hand menu lets the user choose between diplomatic and modernized text versions, versions highlighting archaic idioms, word classes, anaphors and epiphors, or humans/animals, biographical information, and finally the bibliographical comparative study (chosen in the screen dump of figure 2). Diligent students, indeed, but they had been encouraged to play with technology, and they were obviously swept away with the joy of working with the material during the project.

Projects such as these are obviously small-scale. Their degree of graphical and technical sophistication must be assessed proportionally to what limited resources the students possessed with respect to time, equipment, and technical skills. My point in mentioning them, however, is that in all their modesty, they display some of the challenges, tasks, and opportunities that libraries and archives tend to come across when embarking on projects to digitize older written materials – challenges and tasks we recognize from scholarly editing.



Figure 2. Screenshot of a student project on *Kattresan*.

Mass and critical

In many respects, the student project examples display a mode of digitization which is as far as you can possibly come from the currently most talked-of mode, mass digitization, with Google Book Search (GBS) as the paradigmatic example². In mass digitization, huge amounts of documents are digitized by automated means during a short period of time (i. e. the amount of time devoted to digitizing each object is minimized). It operates on an industrial scale and with as many steps in the process as possible fully automated. This mass mode systematically digitizes whole collections with no particular means of discrimination. To be at all feasible, it needs to track down and eradicate any time-consuming steps in the process. Therefore, it minimizes manual and labour-intensive work and cannot include intellectual aspects such as careful selection, interpretation, contextual descriptions, descriptive text encoding, or manual proof reading. It cannot afford to produce new metadata and other bibliographic information about the source document. Here, digitization flattens out into a linear streamlined affair. Hence, there has been no shortage of criticism brought fore against massdigitization projects such as GBS³. Critics have found scanning

- 2 See Coyle, Karen: Mass digitization of books. In: *Journal of Academic Librarianship*, 32, 6 (2006) p. 641–645 for an introduction the concept, and Deegan, Marilyn & Sutherland, Kathryn: *Transferred illusions: Digital technology and the forms of print*. Aldershot 2009 Ch. 5 for a thorough critical discussion.
- 3 E. g. by Duguid, Paul: Inheritance and loss? A brief survey of Google Books. In: *First Monday* 12, 8 (2007). <<http://firstmonday.org/htbin/cgiwrap/bin/ojs/index.php/fm/article/view/1972/1847>> (accessed at 2011-02-11).

and transcriptions to be poor or even useless. The metadata are few and often erroneous. Search facilities other than simple, raw free text search are practically non-existent. There is scarce or no bibliographical information about the source documents (e. g. little if any enlightenment about what copy or even what imprint the digitizing agent used as source). The contracts and terms, finally, are non-transparent with respect to e. g. copyright, terms for re-use, future maintenance, and long-term preservation. By sheer serendipity when using it to locate references for a lecture I was delivering last year, I happened to come across the two phenomena in figures 3 and 4.

So yes, it is easy to come across errors, unreadable transcription, and missing pages in GBS, and thorough bibliographical information is a rare commodity in the project. Perhaps I should say *all too easy* to come across: this is of course a state of affairs anticipated by GBS. The primary aim for mass digitization simply is not depth and perfection. It has its value in scale – a considerable value at that: a gigantic, growing bank of digital texts that can be free-text searched, used as localizing tool, and – perhaps more importantly – form the technical base for many kinds of future software optimizations. One of the logics driving mass digitization project states: let us not spend costly manual labour on transcription corrections, scanning reviewing, production of metadata, historical and bibliographical information, registering, or indexing. Better to go on digitizing as much and as fast as the machines will allow us, then make the scanning files and machine produced transcriptions available as much as possible, *however incorrect* some of them currently may be. Because a) users will gladly come to help (and as a document gets accessed and used, chance is it will be corrected), and b) soon enough, machines will come to help: forceful algorithms, recognition software (for patterns, genres, texts, or images), intelligent search, fuzzy search, text mining, and other technologies will aid machines in such a way that manual hands-on can be minimized or even completely avoided. At least, so goes the *rationale* behind the projects.

Be that as it may, mass digitization does not fit all kinds of libraries, all kinds of resources, or all kinds of documents. There are loads of digitization projects that cannot be designed along these streamlined principles, that are anything *but* ›mass‹, where most of the work is and has to be manual and the technology tailored – what do we label such projects?

For a while now I have been toying with the term ›critical digitization‹ to designate some of the strategies that GBS and other mass digitization projects choose not to embrace⁴. Let us keep in mind that at every step in the long digitization process, one can make choices, deselect and interpret. Usually by necessity, mass digitization ignores most of these choices by designing a uniform straitjacket to impose on every document in the collection to be

4 Dahlström, Mats: Kritisk digitalisering. In: Fra samling til sampling: digital formidling af kultur-arv, ed. by Niels D. Lund et al., Copenhagen 2009 p. 171–191; Dahlström, Mats: Critical editing and critical digitization. In: Text comparison and digital creativity: the production of presence and meaning in digital text scholarship, ed. by Ernst Thoutenhoofd, Adriaan van der Weel & Wido Th. van Peursen, Amsterdam 2010 p. 79–97.

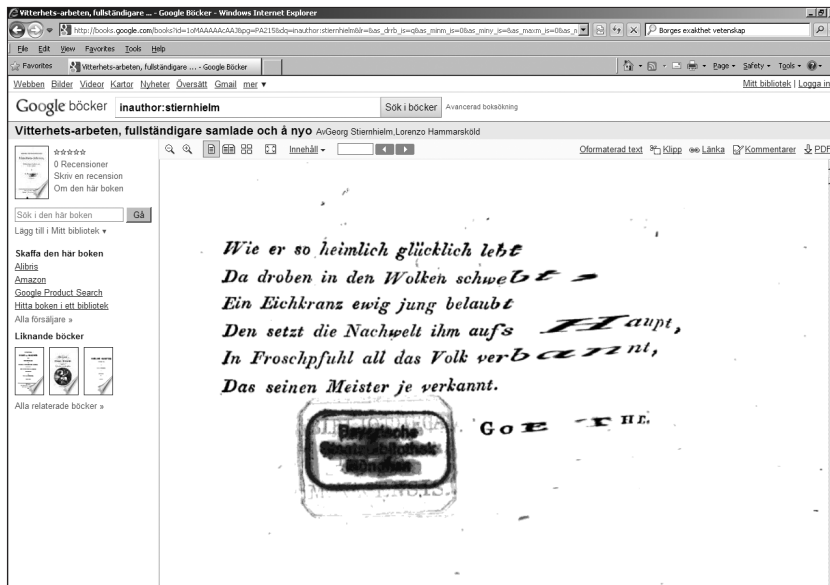


Figure 3. A skewed scanned title page of a digitized book (containing works by the Swedish author George Stjernhielm, 1598–1672). From Google Book Search, April 2010.

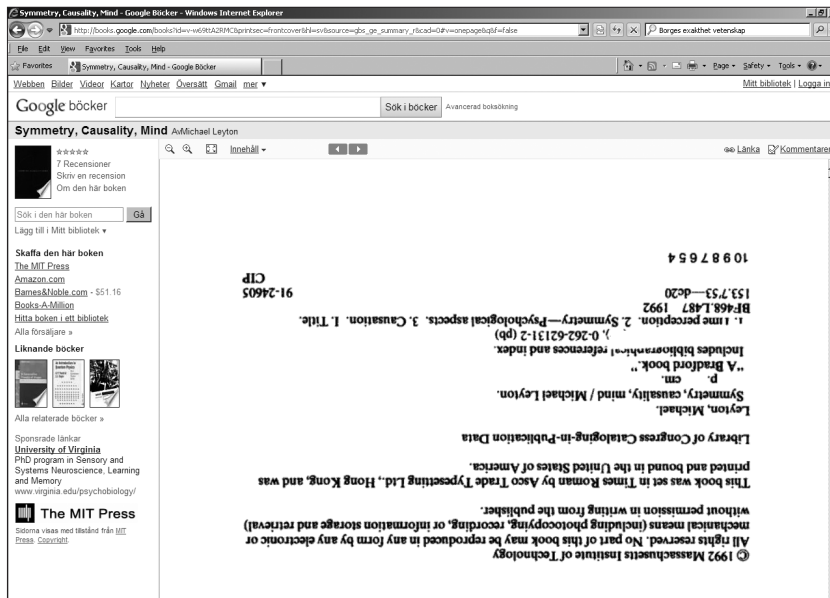


Figure 4. A book turned upside-down in the digitization process. From Google Book Search, April 2010.

digitized. Critical digitization however acknowledges the choices and alternatives, and makes active use of them. A digitizing agent may e. g. need to perform a deliberate and strategic – critical – selection from a number of possible source documents. Perhaps the contents of the document are difficult to decipher. The text or, more often, the image may need to be edited and manipulated to make sense or context. Intellectual editing, preparation and emendation of the contents of the digitized documents thus often occur during the process. Perhaps the project needs to take great care the source documents are not at peril during the digitization process (as is often the case within mass digitization), but on the contrary are subjected to careful preservational or conservational measures. The digitizing agent may aim for a manually and critically produced representation that is as exhaustive and as faithful as possible to the source document – not only with respect to its text, but to its visual and graphical qualities, perhaps even to its artefactual materiality. The digital object may need to be provided with large amounts of metadata, indexing, descriptive encoding, paratexts and bibliographical information. If so, bibliographical and other scholarly research is embedded in the objects. Finally, perhaps fragments from different sources are critically selected to form an eclectic virtual representation in the process.

Therefore, critical digitization is qualitative in the sense that it concentrates on what is unique and contingent in the documents, whereas mass digitization is quantitative in its design to capture what are common, regular, foreseeable traits in large amounts of documents. In consequence then, critical digitization normally has to develop project contingent procedures and tools and tailor them to the nature of the documents in the particular collection. In mass digitization, the individual documents in the digitized collection are on the contrary subordinated (tailored, if you will) to more general, perhaps even universal procedures and tools. Table 1 provides a comparison of some features of the two modes of digitization.

Critical digitization	Mass digitization
Primarily manual	Primarily automated
Depth	Breadth
Maximized editing and metadata	Minimized editing and metadata
Interpretation	Clone
Critical selection	Random and exhaustiveness
Document	Text
Qualitative	Quantitative
Idiographic	Nomothetic

Table 1. Critical digitization versus mass digitization – some features
(based on Dahlström, *Critical* p. 91, see note 4).

As would be evident by now, the student examples exhibited at the outset of this essay can be designated, to some extent, as critical digitization. More to the point, most of us have come across several impressive examples of such digitization endeavours. A prime example would be the digitization of the Codex Sinaiticus⁵. The project's careful and thorough bibliographic investigation of the codex object and its textual and material history is as impressive as its application of cutting-edge technology to dig as deeply as possible into the heretofore invisible layers of the object, or its commendable success at displaying the virtual object and helping the user navigate both horizontally and vertically in it. And there are many other magnificent examples of high quality critical digitization endeavours. Projects working with the Archimedes palimpsest, ancient papyri, or the Vindolanda wax tablets have demonstrated the benefits of applying new technology to old documents. Their methods include artefact scanning in 3D, X-Ray, and multi-spectral imaging technology (taking several shots of a document carrying illegible text and capturing different wavelengths of light, whereby one is able to distinguish ink from paper in a manner prohibited to the naked eye). By stretching the muscles of current digital humanities, projects such as these display how high-quality digitization promises to enhance new scholarship and research, and to enable us to see familiar objects through new lenses.

A prominent Swedish example would be the digitization of the Codex Gigas (the Devil's Bible). The Codex Gigas was taken as booty by the Swedish army as a war trophy in Prague during the Thirty Year's War. During the last decade, the Czech Republic made claims to have the codex returned to Prague – a type of claim that has potentially difficult administrative and cultural ramifications. As a compromise, the National Library of Sweden decided to digitize the whole manuscript book in minute detail and provided its Czech counterpart with the resulting digital files. Digitizing the largest manuscript book in the world proved a considerable task, involving creative adjustment of equipment, and indeed innovative methodological development⁶.

Critical mode criticism

I confess to not being perfectly content with the distinction between the mass and the critical mode of digitization. It might suggest that a digitization project is either a mass

- 5 Presented in one of the adjacent articles in this *Bibliothek und Wissenschaft* issue. It is also presented and discussed in Parker, David: *Ancient scribes and modern encodings: the digital Codex Sinaiticus*. In: *Text comparison and digital creativity: the production of presence and meaning in digital text scholarship*, ed. by Ernst Thoutenhoofd, Adriaan van der Weel & Wido Th. van Peursen, Amsterdam 2010 p. 173–188.
- 6 The digitized Codex Gigas is browsable at <<http://www.kb.se/codex-gigas/>> (accessed at 2011-02-11).

or a critical mode project. Instead, I would like to think of the two modes as poles at each end of a scale – most digitization activities contain features from both strategies and place themselves somewhere along the scale. A more problematic aspect, one that suggests the »horizontal« scale should be supplemented with a vertical scale along more nuanced parameters of e. g. collection size and project scale, is that the two labels might lend themselves to ideas of mutual exclusivity. That is, to suggesting that a digitization project cannot comprise huge quantities of objects while still being performed with critical inquiry and a high level of standards and quality. Current development within technology points to increased automation, but *also* to maintaining high standards and requirements for e. g. preparation and management of source documents, or high quality digital facsimile production. Many memory institutions obviously are interested in finding ways to satisfy both these demands⁷. A final potential problem with the distinction, or rather the description of it that I provided above, is that it might suggest that the one is objectively superior to the other.

But relative to the number of digitized objects and in comparison to mass digitization projects, the critical mode of digitization is expensive, slow, and often attracts few and small user communities. To perform a thorough critical digitization project also requires considerable skills in e. g. textual or image scholarship. As for the scholarly value of critical digitization, it often emphasizes the digital facsimile work to the extent that the production of machine-readable transcriptions is neglected. In some similarity to many qualitative research projects within the humanities, critical digitization projects tend not to really meet thorough requirements of documentation, transparency, and repeatability. As for the scholarly value, some users might find it problematic that the digitized material has already been deliberately and explicitly interpreted and narrowly formed for a particular purpose. On the one hand, critical digitization enriches its objects with an intellectual added value and applies some kind of quality seal as regards selection, textual quality, resolution, proofreading, comments and bibliographical information. On the other hand, not all scholars may be interested in that particular metainformation and added value, but are in need of quite different aspects than those that happened to catch the interest of the digitizing institution. Further, much of the discourse framing mass digitization can be criticized for trivializing the image and text capture phases into a simple »conduit« view of transmission, where the objects and their contents run through the operation unaffected, and where the digital collection thus is a complete surrogate for the source collection. One might equally well claim that critical digitization risks falling for another kind of cloning ideal: as long as the most powerful reproduction technologies available are implemented

7 For instance, the Münchener Digitalisierungszentrum explicitly states on its web site that it does no distinction »between high quality digitisation and mass digitisation. Both is possible due to 10 years of expertise and the presence of a high quality scanning equipment.« <<http://www.digital-collections.de/index.html?c=digitalisierung-policy&l=en>>; accessed at 2011-02-11. I am grateful to Lars Björk for referring the example.

in the project, and if rich enough metainformation is provided to each object, there will be no need for any future digitization of the object – we would be facing the definitive digital representation. To this line of reasoning, mass digitization has an advantage because it does not select nor provide interpretive metadata and explicit text encoding, but rather provides »neutral« banks that scholars ideally can use, reuse and enrich the way it suits them best. Again, we recognize a pattern from the tension in scholarly editing between Alexandrian and Pergamian ideals.

Critical transmission: digitization, editing, and textual scholarship

In fact, as would be evident, much of what goes on in critical digitization reminds us of what scholarly editors have been doing for centuries – and that includes both problems and solutions. This is of course no coincidence: they are both instances of what I would call critical transmission. In that sense, library digitization belongs to the same transmission tradition as 20th century microfilming or the transcribing of manuscripts in ancient libraries and medieval monasteries. The bibliographers of antiquity were engaged in critical full-text transmission⁸. Their practice of transmitting, commenting and correcting texts gradually evolved into textual criticism and scholarly editing, while libraries increasingly devoted their time to producing bibliographical labels for documents rather than reproducing the full documents. The Gutenberg era significantly changed the extent and nature of this full-text transmission in libraries. With digital reproduction technologies and current digitization projects however, libraries are drawing a historical circle and are yet again dedicating much energy and attention to such transmission. In so doing, they face tasks that go beyond trivial and hence enter a critical transmission mode.

When performing such transmission, cultural memory institutions as well as scholarly editors need to select and discriminate materials. They both face the task of having to interpret and sometimes compare and collate variant source documents, perhaps even to identify common – sometimes even ideal – contents within such a range of variant documents. They share a concern for authenticity and faithfulness between the sources and the representations. They edit, comment and produce metatexts. They are practices legitimized by prestige institutions, and based on stabilized principles. As such, critical digitization is, similar to scholarly editing, an exclusive affair – in more senses than one. Efforts such as the Codex Gigas and the Codex Sinaiticus digitization projects increase the symbolic sig-

8 Good introductions to the seeds of textual criticism and critical bibliography planted by bibliographers and librarians in e. g. the library of Alexandria are provided by Reynolds, Leighton D. & Wilson, Nigel G.: *Scribes and scholars: a guide to the transmission of Greek and Latin literature*. Oxford 1991 and by Dickey, Eleanor: *Ancient Greek scholarship: a guide to finding, reading, and understanding scholia, commentaries, lexica, and grammatical treatises, from their beginnings to the Byzantine period*. Oxford 2007.

nificance of the digitized documents. Critical digitization and editing share a canonizing function; by putting a quality seal on a digitized object they honor particular works and documents by raising »cathedrals« in their honor, turning documents into monuments. In general, I would say that library digitization and scholarly editing display a similar internal pattern of conflicting scholarly ideals⁹. I would even say that in the case of *critical* digitization, libraries are in effect engaged in what comes close to textual criticism. Librarians and other employees involved in current digitization projects might feel awkward with such a label of textual criticism or even bibliography put on their work, but the historical connection cannot be denied.

Might we go so far as to say that scholarly editing and critical digitization are one and the same activity? No. The institutional context aside, there are some important differences between them. John Lavagnino recently addressed this issue, formulated as the differences between digital (scholarly) editions and digital libraries (in the sense of collections of objects digitized by memory institutions such as libraries)¹⁰. The former, Lavagnino argues, displays a more profound analytical understanding of the extant source situation, its problems and challenges, for instance by emending textual errors or making arguments about authorship:

It is a perfectly respectable thing to provide an accurate transcription or photographic reproduction of one copy of a book, and in many cases those tasks require a considerable amount of scholarly expertise and judgment, but the potential achievements of these tasks are different from those that are possible in a scholarly edition¹¹.

A library digitizing a collection, on the other hand, »does not involve any analysis; it is devoted to reproducing existing books, but not to any critical or bibliographical analysis.« Producing it is largely a clerical task, and it »can be created by workers who have no special knowledge of the material.«¹²

This is an accurate description of most mass digitization projects performed by libraries, archives, and partner agents such as Google. But if we broaden our view of digitization in libraries, we see a more complex spectrum of analytical, critical levels in the work performed. In high-quality digital imaging projects for instance, teams of conservators, technicians and photographic experts constantly base a seemingly endless row of decisions that need to be informed precisely by critical and bibliographical analysis and by a highly specialized knowledge of the bibliographical, graphical, bibliological, historical, and other research aspects of the material to be digitized. Libraries further produce, with or without the aid of subject specialist scholars, electronic thematic research collections, a task that requires a considerable degree of scholarly, critical skill and deliberation.

9 This pattern is discussed at more length in Dahlström, *Critical* (see note 4).

10 Lavagnino, John: Access. In: *Literary and Linguistic Computing* 24, 1 (2009) p. 63–76.

11 Ibid. p. 63.

12 Ibid.

I agree that the major distinction one might define between scholarly editing and library digitization concerns the textual level: library digitization does not seek, as scholarly editing does, to establish a text (a copy-text as it were) that perhaps never existed previously and which cannot be literally transmitted from a single source document. But this distinction is not as sharp as one might at first expect. On the one hand, movements such as versioning, documentary editing and un-editing push scholarly editions towards becoming documentary, flat archives. There, the products of scholarly editing are broken down into small, modular fragments in the manner of a database, or »library«. Some within the scholarly editing discourse also advocate minimizing editorial interventions and instead providing fulltext and »raw« versions, refraining from highlighting or constructing a single uniform established base text. Scholarly editions are also increasingly based on image-oriented editing, and devote more of their space to harboring digital facsimiles of the source documents. On the other hand, critical digitization approaches the domain of scholarly editing and textual criticism, for example in image-oriented digitization of documents where we have more than one source copy to choose. There, the critical comparison of several source candidates based on e. g. their condition and completeness results either in the deliberate selection of one candidate or in an eclectic amalgam of fragments from several candidates (such as eclectic facsimiles à la Charlton Hinman). In the latter case – at the very least – we are facing, if not the ambition of textual criticism to establish an ideal text, then at least a kind of *document criticism* that seeks to establish an ideal document. This is in fact the primary aspect where critical digitization and scholarly editing are both different and similar. The former devotes its analytical faculties to the critical transmission of graphical and material documents, and has been criticized for not paying too much critical attention to their texts. The latter devotes its analytical faculties to the critical transmission of text, and has been criticized for not paying too much critical attention to the graphical and material documents carrying those texts. Understood this way the two activities not only differ but support and complement each other.

And frankly, how do we label projects such as the ones digitizing the Codex Sinaiticus¹⁵ or the Codex Gigas – are they primarily digital editions or digitization projects within a memory institutional context?¹⁴ What about the digital William Blake or Mark Twain *archives*?¹⁵ That kind of undecidedness is perhaps a symptom that the two fields of scholarly editing and library digitization are brought much closer together in the digital realm. On a practical level, there is of course much tangible cooperation between the two fields. Libraries normally house the source documents of interest to scholarly editors to begin

15 See Schneider, Ulrich & Dogan, Mustafa in this volume.

14 For instance, David Parker refers to the digital Codex Sinaiticus as a »de luxe edition« (Parker, Ancient. p. 179, see note 5).

15 The web site of the William Blake Archive: <<http://www.blakearchive.org/blake/>> (accessed at 2011-02-11); the web site of the Mark Twain Papers and Project: <<http://bancroft.berkeley.edu/MTP>> (accessed at 2011-02-11).

with, and often perform the technical digitization of them to serve large editing projects. Libraries are also arguably best suited to be responsible for the long-term technical and bibliographical maintenance and preservation of the digital files. They are, further, in a good position to coordinate and manage the intricate web of intellectual property rights (IPR) within large editing projects in a way that other agents, including scholarly editors, simply cannot do. This is particularly the case with image-oriented projects, where libraries and archives often are the very IPR holders themselves.

Notwithstanding the high-quality, critical digitization mode discussed above, even quite modest library digitization has significant benefits for textual scholarship. Large amounts of material is made available for both people and search software, at least as facsimiles, and can thus become the object of text capture and turned into machine-readable text to be used and re-used in research and editing projects. Unpublished and sometimes unknown manuscript material is identified, registered and made available. Even the gigantic banks of transcription files produced from scanned books by Google Book Search are beginning to be interesting as research data to textual scholars around the world. We have already seen several fruitful examples of text mining and other technologies to detect patterns over very large collections of texts and images¹⁶. Scholarly editors, textual scholars and digitizing agents alike begin to think about how to coordinate their projects with what agents such as Google are already doing anyway, e. g. by superimposing new collections or patterns on Google's bank (to the extent that is allowed and technically feasible).

Scholarly editing and library digitization thus seem to share an increasing amount of features and concerns, and might occasionally even merge on the project level. Overall, there are ample reasons for being optimistic about cooperative opportunities between memory institutions and scholarly editing endeavours. Conversely, there are of course a number of potential problems and obstacles on the way to such smooth cooperation. The present essay – what is left of it – does not have the space or the task of providing a comprehensive map of such problems. I will settle for pointing to one.

The look-but-don't-touch problem

Many digital text editing projects depend on high-resolution digital facsimiles, carefully prepared and proof-read transcriptions, and XML encodings of the highest quality. Such material can ideally be delivered by the libraries and archives with which the editing projects

16 Crane, Gregory, Seales, Brent & Terras, Melissa: Cyberinfrastructure for classical philology. In: *Digital Humanities Quarterly* 3, 1 (2009). <<http://www.digitalhumanities.org/dhq/vol/003/1/000023.html>> (accessed at 2011-02-11); Flood, Alison: Study claims Agatha Christie had Alzheimer's. In: *The Guardian*, April 3 2009. <<http://www.guardian.co.uk/books/2009/apr/03/agatha-christie-alzheimers-research>> (accessed at 2011-02-11).

have partnered. Not only do the editing teams need to have access to the underlying data, the text encodings, the rich metadata, the high resolution images, the scripts, style sheets etc., but they also need to be able to exchange, use, reuse, integrate and remodule them. Such needs are obviously better met by critical digitization than by mass digitization, which produces information-poor files.

There are indeed many encouraging examples of such intimate cooperation between libraries and scholarly editors¹⁷. These kinds of projects would not be feasible if the partners did not embrace an open source ideal, at least to some extent. Due to legal, economical, or administrative reasons however, a number of libraries, archives and other digitizing agents choose to restrict access to/usage of the information-rich files. In these cases, only the delivery files are made freely available, although it is very often the rich archival files that are needed if research and particularly re-usability is to be realized. Keep in mind that we are talking about publicly funded libraries and about works that have entered the public domain. This poses a major problem. Not a few scholarly editing projects have failed to come to fruition because the institution holding the source documents either claimed copyright to its particular copy of the public domain work or posed a forbidding charge for access to the high quality master file¹⁸.

It is not my intention in this article to discuss whether libraries are right or wrong in implementing this policy – I have done that elsewhere¹⁹. Let me just comment on one out of the many arguments put forth by libraries to rationalize their behaviour: that practically all »end users« usually settle for viewing and browsing lightweight delivery files. This discloses an old familiar view on collections and users. In the traditional, analogue domain of libraries and archives, the institution has a physical object, which the audience can view when visiting the institution (that is pretty much what the audience *can* do): look but don't touch. That perspective on objects and audiences is then carried into the digital domain. The institution digitizes, and the users are expected to view, browse, but not much else. Interestingly enough, Peter Shillingsburg also criticizes editors for designing their digital editions as finished products »to be seen and not touched«²⁰. This attitude is however even more prominent in the world of libraries and archives – a circumstance I think is largely

17 The previously mentioned William Blake Archive is one good example, Schopenhauersource (<<http://www.schopenhauersource.org/>>, accessed at 2011-02-11) another.

18 Cf Robinson, Peter: Electronic editions for everyone. In: Text and genre in reconstruction: effects of digitalization on ideas, behaviours, products and institutions, ed. by W. McCarty, Cambridge 2010 p. 145–163, esp. p. 150–151.

19 E. g. in the presentation entitled »Digitized library collections – an open source approach«, delivered at the conference The Dilemmas of Digitisation: Thinking about the Past, Planning the Future, the University of Oxford, May 22–24 2008.

20 Shillingsburg, Peter: How literary works exist: implied, represented, and interpreted. In: Text and genre in reconstruction: effects of digitalization on ideas, behaviours, products and institutions, ed. by W. McCarty, Cambridge 2010 p. 165–182.

due to differences in institutional culture. In scholarly editing, there is at least a discussion on what to label the results of e. g. archival file transformation: viewer, browser, user, or reuser mode? This is a component of a larger discussion (of which Shillingsburg's article is an example) in the scholarly editing community of the preferred mode and forms in which users are to access the results of digital editing endeavours. Many scholarly editors tend to regard the outcomes of digital editing as modular resources and tools, to be used by others and in new products. This is evident through many concept proposals and discussions in editorial literature, e. g. dynamism²¹, reproductivity²², personal/impersonal perspective and ergodic mode²³, and ›knowledge site‹²⁴. These kinds of discussion are strikingly absent from the world of archives and libraries, who largely lack the tradition of seeing themselves as research data producers, and for whom there seems to be a large distance between digitizer and audience (as viewer and browser). When we digitize however, we produce a new object, which can be more or less powerful and more or less apt to work with and reuse, not merely to view and browse.

Symptomatically, the result of a digitization project (often also a scholarly editing project) is routinely thought of – and referred to – as the ›end product‹, a self-contained entity handed over to the ›end user‹. Libraries – like museums – do showcases and exhibit closed artifact entities. What if these producers started thinking about the products not as end products but as start products or tools? Digitizers and editors need to become much more focused on showing people what they can do with the digitized resources, rather than just amassing a pile of monolith documents and storing them somewhere on the web where not enough people find them, and where the ones that do, are left to themselves to find out how to use them, what to do with them, and what value to assign to them.

The editing community is fond of culinary metaphors (the TEI ›pizza model‹ to mention but one). Alois Pichler and Odd Einar Haugen drew a distinction some years ago between combined editions (such as the Bergen Electronic Edition of Wittgenstein's *Nachlaß*) and dynamic ones²⁵. Where the combined are à la carte menus, enabling the user to choose between fixed sets of dishes, the dynamic might be regarded as smorgasbords, where the

- 21 Siemens, Ray G.: Shakespearean apparatus: explicit textual structures and the implicit navigation of accumulated knowledge. In: *Surfaces* 8 (1999). <<http://www.pum.umontreal.ca/revues/surfaces>> (accessed at 2011-02-11).
- 22 Ore, Espen Smith: Monkey business – or What is an edition? In: *Literary and Linguistic Computing* 19, 1 (2004) p. 35–44; Dahlström, Mats: How reproductive is a scholarly edition? In: *Literary and Linguistic Computing* 19, 1 (2004) p. 17–35.
- 23 Vanhoutte, Edward: Defining electronic editions: a historical and functional perspective. In: *Text and genre in reconstruction: effects of digitalization on ideas, behaviours, products and institutions*, ed. by W. McCarty, Cambridge 2010 p. 1119–144, esp. p. 141–143.
- 24 Shillingsburg, Peter: *From Gutenberg to Google: electronic representations of literary texts*. Cambridge 2006.
- 25 Pichler, Alois & Haugen, Odd Einar: *Fra kombinerte utgaver til dynamisk utgivelse: erfaringer fra edisjonsfilologisk arbeid med Wittgensteins filosofiske skrifter og nordiske middelaldertekster*.

user can combine ingredients according to her own taste and needs, in what sequence she prefers. Sticking to the culinary metaphor, one might equally think of further types: the Dutch treat, where users contribute by bringing their own ingredients or dishes; or the marketplace, where you go to find raw materials of the finest quality, such as TIFF and TEI master files of primary sources; or the kitchen, to which users can bring ingredients, recipes and ideas, and where they can find stoves, sauce-pans, and food processors with which to prepare their dishes²⁶. The same culinary variety would be possible within library digitization as well. But if scholarly editors are beginning to let users have a taste of smorgasbords, libraries and archives are still very much comfortable in the fancy à la carte restaurant. Still, not enough people seem to come to these culinary feasts anyway, whether cooked up by digitizing libraries or by scholarly editors in the digital humanities. Surveys in fact indicate that digitized collections are not at all being used and re-used to the extent one might have hoped for. One study for instance observed that about a third of the large digital humanities collections are rarely or never accessed at all²⁷. Perhaps it is time to pick up the phone and call Gordon Ramsay?

In: *Læsemåder: udgavetyper og målgrupper*, ed. by Per Dahl, Johnny Kondrup & Karsten Kynde, Copenhagen 2005 p. 178–249.

26 Thanks to Maria Berggren at the university library of Uppsala for suggesting the kitchen metaphor.

27 Warwick, Claire et al.: *Evaluating digital humanities resources: the LAIRAH project checklist and the Internet Shakespeare Editions project*. In: *Openness in digital publishing: awareness, discovery and access* (ELPUB 2007), ed. by Leslie Chan & Bob Martens, Vienna 2007 p. 297–306, esp. p. 298.