

Critical editing and critical digitization

Mats Dahlström

Abstract

This article discusses scholarly editing based on textual criticism and its transmission mechanisms, science ideals, and socio-cultural functions. Primarily, its pattern of conflicts is mapped—historically and presently. It is suggested that this pattern is similar to that of other transmission activities within socio-cultural institutions. The article then proceeds to deal with library digitization as a transmission activity. Two digitization approaches are discussed: the currently much debated mass digitization, and an approach that bears many similarities to textual criticism and that this article therefore suggests we refer to as *critical* digitization. The relation between these two digitization strategies is analyzed. Finally, the article returns to scholarly editing and analyzes its relation to library digitization, with a particular focus on comparing the similarities and differences between these two transmission activities.

Scholarly editing and transmission ideals

Scholarly editing based on textual criticism is a bibliographical, referential activity. It examines a bulk of documents, compares their texts, normally clusters these around the abstract notion of a work, arranges them in a web of relations, and attempts to embody this web in the scholarly edition. The edition, then, becomes a surrogate purporting to represent, tag and comment upon the edited work. In a sense, the editor reproduces existing documents by making a new document that also embodies a documentation of the textual history and the editorial process.

At the same time, scholarly editions are hermeneutical documents and subjective interpretations, in two senses: they carry with them an ideological and hermeneutical heritage, and they also exert an interpretative influence over the objects they are designed to embody and represent. Nevertheless editions by tradition pretend to convey a sense of value-free objectivity, a mere recording of facts.

This difference between thinking of scholarly editing as either subjective interpretation or as the objective reporting of scientific facts is one of many interesting tensions and potential conflicts within scholarly editing. In fact, when studying the history of editing and textual criticism, one can easily come up with a long list of such tensions and conflicts. I list a few of them in table 1, and discuss them briefly in the text that follows.

Tensions in scholarly editing

critical	non-critical
interventionist	non-interventionist
interpretative	factual

facts <i>as</i> interpretation	facts separable from interpretation
ambiguous	disambiguable
idiographic	nomothetic
contingent tools	universal tools
material document	abstract text
the one text: discriminatory	the many texts: comprehensive

Table 1: Tensions in scholarly editing

A long established distinction, firstly, is the one between scholarly editing that is *critical* and scholarly editing that is *non-critical*—the latter exemplified by documentary editing or facsimile versioning.¹ At times non-critical editing is even looked upon as more or less mechanical and trivial transmission. One might however argue that digital scholarly archives displaying full-text versions in parallel, with no single established text in the centre, threaten to break down this distinction. In creating such a digital archive, the editors open up the doors to their editorial ‘lab’, so to speak, and turn it into an archive that might be cumulative and be run jointly with other editors and scholars and allow them to access source document representations for new editorial endeavours.² A much discussed feature in such archives is the idea of abandoning the established critical text as a privileged gateway to the complex of versions, and handing over this task to the individual user. What is really left of the concept of the scholarly edition, if the critically established base text is removed?

As will be argued later in this article, the boundary is further blurred between critical editions and some of the digital facsimiles that libraries produce and that might arguably be considered as ‘critical’. Image management (such as capture and subsequent editing) is perhaps the digital editing phase where the presumed distance between objectivity and subjectivity is at its largest. On the one hand, image transmission in general and facsimile production in particular has traditionally been regarded by textual critics as non-critical activities, where the editor supposedly recedes into the background, and where the user is brought closer to the source documents by having ‘direct’ access, as it were, to the originals. On the other hand, digitization and the subsequent editing of images has perhaps more than any other editing phase made us attentive to the fact that virtually all parameters

¹ See e.g. D. Greetham, 'Textual scholarship' in *Introduction to scholarship in modern languages and literatures*, J. Gibaldi (ed), New York, MLA, 1992, pp. 103–137. See also A. Renear, 'Literal transcription: can the text ontologist help?', in *New media and the humanities: research and applications*, D. Fiormonte & J. Usher (eds), Oxford, Oxford university, 2001, pp. 23–30.

² G. Bodard & J. Garcés, 'Open source critical editions: a rationale', in *Text editing, print, and the digital world*, M. Deegan & K. Sutherland (eds), Aldershot, Ashgate, 2009, pp. 83–98.

in the process (image size, colour, granularity, bleed-through, contrast, layers, resolution etc.) require intellectual, critical choices, interpretation, and manipulation.

A related distinction is therefore that between the acknowledged presence and the presumed absence of the editor. This is the fundamental issue of *intervening* or *not intervening* in the text, something Greetham referred to as the Alexandrian and the Pergamian editorial ideals.³ The former accepts and even presupposes intervention and corrections, laying the ground for eclectic editing, while to the latter interventions and corrections are theoretically awkward (and even come close to heresy), making way for the school of facsimile and best-text editing. If in academic discourse scholars appear to want to ‘hide’ their role as narrative writers,⁴ then digital archives based entirely on diplomatic and facsimile editions promise the disappearance of editors altogether, inviting readers to step in and fill the creative, authoritative editorial function.

Then there are as well tensions between different scholarly and scientific *ideals*. The edition can on the one hand be thought of as a scientific, value-neutral and objective report, the ‘calculus’ if you will. On the other hand, there are powerful arguments in favour of viewing the edition as a scholarly, interpretative, subjective statement. In the one case, the edition’s text is primarily recognized as a scientific fact, and in the other as an interpretation. Similarly, there is a tension between the notion that facts *are* interpretations and the idea that we can *separate* facts from interpretation, such as in xml encoding (separating between accidental form and substantive content), in stand-off markup, in other means of producing descriptions and separating them from the objects they describe, as well as in synoptic full versions of facsimiles that represent no tampering and provide ‘raw’ material. So there is a split between thinking that the objects (and tools) of editing *can* or *cannot* be subjected to universally agreed disambiguation.⁵ And if the content can be decontextualized and disambiguated, it can also cumulatively form building blocks in other and different types of editorial endeavours with little or no problem.⁶ In consequence, editions are in this respect either stores of scholarly raw material that support future reusability by other editors and scholars, or argumentative and context-bound statements.

³ D. Greetham, *Theories of the text*, Oxford, Oxford University Press, 1999, pp. 50–51.

⁴ Charles Bazerman (*Shaping written knowledge: The genre and activity of the experimental article in science*, Madison, WI, University of Wisconsin Press, 1988, p. 14) makes the following observation on the scientific article as written genre: ‘[T]o write science is commonly thought not to write at all, just simply to record the natural facts. Even widely published scientists, responsible for the production of many texts over many years, often do not see themselves as accomplished writers, nor do they recognize any self-conscious control of their texts.’ This seems to be a legacy in the scholarly editing community as well.

⁵ For instance, the notion of digital do-it-yourself-editions implies that the digital documents produced (the ‘target’ documents) are in all respects equivalent to the source (or ‘departure’) documents, and that the user, granted with direct access to the same source documents as the editor had, can tread different paths and make different choices from those of the editor. It also assumes that the transmission process has been able to be disambiguated, leaving the target documents therefore unaffected by any significant distortion.

⁶ This discussion is taken further in M. Dahlström, ‘The complete edition’, in *Text editing, print, and the digital world*, M. Deegan & K. Sutherland (eds), Aldershot, Ashgate, 2009, pp. 27–44.

Another tension between scholarly ideals is the one between viewing editing as a primarily *nomothetical* or *idiographical* affair. The former maps patterns of common, regular and predictable traits in large amounts of texts, while the latter rather wants to highlight the unique, the different, the contingent. This creates a subordinated tension between attempting to design *universal, project-general* or *contingent, project-specific* tools.⁷ It is also somewhat related to another, much recognized distinction in scholarly editing, namely between different conceptions of the very empirical object of editing: either the text as reduced to linguistic sign sequences or as a meaning conveyed by those linguistic sequences in conjunction with layout, typeface, colour and the rest of the graphical and material appearance that the document provides. This is in other words a difference in perspective between *text* and *document*, manifested in the distinction between text oriented and image oriented editing.⁸ Text oriented editing works mostly with text transcriptions, image oriented editing mostly with facsimile images. To much text oriented editing of e.g. intentionalist descent, text is an immaterial, abstract, ideal, copy-independent phenomenon, while to much image oriented editing in e.g. sociology of texts or material philology, text is rather a material, physical, concrete, copy-dependent phenomenon. Depending on which ideal you subscribe to, the editing, its tools and the resulting edition will turn out to be very different indeed.

Finally, a significant tension has emerged between displaying the uniformity or multiformity of the edited work—what Peter Robinson has referred to as the *one text* or *the many texts*.⁹ The former ideal strives for choosing or constructing single copy-texts whereas the latter ideal turns the edition rather into an archive. This is a difference between on the one hand selection and discrimination and on the other hand more or less total exhaustiveness.

Will the ideals and tasks of scholarly editing change with new media?

Some of these conflicts have been around for the entire history of scholarly editing, while others have emerged during the last decades. It is sometimes claimed that these conflicts and tensions are largely a result of the Gutenberg paradigm and that new media and the web will turn things topsy-turvy and impose a new paradigm of ideals. I think this is presumptuous. In fact, digital editing does not seem to do away with this pattern of conflicts at all, but rather accentuates some of them.

For instance, the strive for totality and more or less complete exhaustiveness within scholarly editing is considerably strengthened by digital editing. The inclusive potential of digital editions and archives such as the ability to house full-text representation of many or indeed all versions of the

⁷ G. Rockwell, 'What is text analysis, really?' *Literary and linguistic computing*, 18(2), 2003, p. 215.

⁸ The distinction between text based and image based editing is treated in e.g. the thematic issue on «image-based humanities computing», *Computers and the humanities* 36(1), 2002. See also the sharp essay on the topic by G. T. Tanselle, 'Reproductions and scholarship', *Studies in Bibliography*, 42, 1989, pp. 25–34.

⁹ P. Robinson, 'The one text and the many texts', *Literary and linguistic computing*, 15(1), 2000, pp. 5–14.

edited work and to support the modularization of documents into movable fragments *across varying contexts*, seems to boost the idealist strand in editing. This trend is even further supported by text encoding, where form is separated from content, and where fact is quite often conceived of as separable from interpretation.

But these notions have forerunners within printed scholarly editing as well. Printed definitive editions attempted to be matter of fact, exhaustive and final. And printed parallel and synoptic editions are attempts to accommodate versionality and inclusivity within the covers of the codex book, and Kanzog's Archiv-Ausgabe of the work of Heinrich von Kleist takes this idea even further.¹⁰ Admittedly, those forerunners are different in scale, and digital editions such as *The William Blake Archive*, *The Wife of Bath's Prologue*, or *The Rossetti Hypermedia Archive* can partly be seen as attempts to not only embody but prolong these notions and make them come real.

It is worth remembering that bibliographical collections of documents derive much of their strength not only from their inclusive but also from their exclusive mechanisms.¹¹ Scholarly editions in fact gain much power and status by their discriminating task and in the way they define and constitute works by excluding and being precise. This task as a constitutive statement is in some contrast to the notion of total archives and do-it-yourself editions. One might however note that in a number of current digital editing projects, an 'edition' can designate a temporary, editorially argumentative selection from a more general-purpose and comprehensive storage archive. This suggests a possible distinction between edition and archive, where the former but not the latter explicitly takes a stand. At the same time there is a counter-reaction against the 'archival' trend, pleading for a return to the editorial prerogative and to primarily monoversional edition forms.

Some researchers have observed that scholars remain faithful to printed editions and that we might even be witnessing a decline in digital editing. Steding, who devotes his dissertation on scholarly editing to making a plea for the medially superior qualities of digital scholarly editions when compared to the qualities of printed editions, does not really provide a good answer as to why digital editions have not already superseded or even outdated their printed counterparts.¹² Steding and others who have commented on this circumstance concentrate their analytical efforts on technical and material qualities to explain or even predict medial evolution. If we on the other hand put more emphasis on editions as biased arguments with certain meritocratic values, we can include as one of the tasks of a scholarly edition to be a report of an accomplished scholarly labour and a

¹⁰ K. Kanzog, *Prolegomena zu einer historisch-kritischen Ausgabe der Werke Heinrich von Kleists: Theorie und Praxis einer modernen Klassiker-Edition*, München, Hanser, 1970.

¹¹ The scholarly edition is here considered as a bibliographical, referential genre. It is in that sense related to such genres as the catalogue, the reference database, the library collection, the archive, or the enumerative bibliography.

¹² S. A. Steding, *Computer-based scholarly editions: context, concept, creation, clientele*, Berlin, Logos, 2002.

status-carrying constitution. In that sense, the edition shares some of the discursive and rhetorical characteristics that historiographical and genre studies have ascribed to the scholarly journal article genre: their objectivity ideal, their aim to avoid aesthetical style, their task of both constituting and defining what is (and what is not) knowledge alongside the more recognized task of presenting tools for new knowledge. To Frohmann, the flora of scholarly journal articles is primarily a reward system in the form of a gift economy. It forms an archive of accepted and constitutive statements rather than an arena of current and cutting-edge research. He therefore is sceptical as to whether ‘scientific articles contribute information useful in the derivation of new results.’¹³ He further claims scientific ‘facts’ are by no means absolute. They are rather statements, and a chief function of the journal article is to stabilize these statements by piling up armies of footnotes in their defense. There is, I think we would agree, a similar, stabilizing socio-cultural function within the scholarly edition. Not only within the community of textual scholarship, but also within literary culture in general. The fact that a work has been the object of scholarly editing is a seal that it has been raised to literary nobility and invited into the inner rooms of the literary salons. Burman is on the mark when he refers to scholarly editions as the ‘cathedrals’ of literature.¹⁴ Frohmann dresses up a similar thought in a sophism: ‘A text does not belong to the scriptures because its content is holy; rather, its content is holy because it belongs to the scriptures.’¹⁵

These functions of being a scientific report, an interpretative statement, a constitutive and canonizing tool, even a monument, provides the scholarly edition with a particular tension. On the one hand, it is supposed to be dynamic and as a research tool quickly reflect new findings and scholarly development. On the other hand, there are arenas where the scholarly edition is supposed to be conservative, static and confirmative. We see this two-faced character in the way scholarly editing is marked by both being prone to change, experiment, question and discussion while at the same time being highly conservative and traditional. There is a welcoming and there is a resistance. Granted, many editors have proven eager to try out new technologies and media, but equally many—perhaps more—editors have proven unwilling to experiment and change.

So in essence, we might perhaps better understand the development of, and relation between, printed and digital editions if we emphasize this multiple task of the scholarly edition as scientific report, as meritocratic constitution and as monument, and where it is plausible that the digital edition has not yet been able to accommodate the structure of meritocratic, social and symbolic values that surround the fixed and stable printed scholarly edition.

¹³ B. Frohmann, *Deflating information: From science studies to documentation*, Toronto, University of Toronto Press, 2004, p. 46.

¹⁴ L. Burman, 'Det enkla valet: konsekvenserna av en oproblematisks textsituation', in *Vid texternas vägskäl: Textkritiska uppsatser*, L. Burman & B. Ståhle Sjönell (eds), Stockholm, Svenska vitterhetssamfundet, 1999, p. 85.

¹⁵ B. Frohmann, *Deflating information*, p. 153.

It is further possible to assume an interplay between edition types and their users according to parameters such as economical, intellectual and social class, in the sense that a typology of editions based on e.g. format to some extent reflects a social distribution among user groups. Genre studies indicate that already the graphical and textual pattern of a document signals genre and that different medial and material edition types might be linked to different user group adaptations. From such a perspective, it would seem probable that a division of labour would come about between printed and digital scholarly editions. The former would then be assigned the task of constitution and of a concentrated and lucid presentation of the material in a manageable format. The latter would increasingly be thought of as the archival and referential material that has made the digital end product possible or as an extension and continuation of the constantly ongoing editorial work. And this is of course to some extent what we are witnessing within many current editing projects, e.g. in large Scandinavian editing projects.¹⁶ I would argue that this is because new media types do not necessarily replace older ones, but rather bring about a change of roles, a displacement of tasks and functions for the media types concerned.

Library digitization and transmission ideals

Let us however return to the pattern of conflicts within scholarly editing depicted in table 1 and discussed in the text that followed. I would argue that this pattern is primarily not media specific to either printed editing or digital editing. It is rather *a general trait of textual transmission* as a cultural phenomenon. We might therefore expect to detect more or less the same pattern in other cases where textual transmission has been similarly stabilized by institutionalisation. The discussion below goes more deeply into a particular case of institutionalised transmission.

One of the things new and different with digital editing is the division of labour and media that surrounds the field of editing and connects it with neighbouring activities. I am thinking in particular of the changing relation between scholarly editing and the ongoing digitization within libraries. And interestingly enough, digitization within libraries is developing a pattern of conflicts and tensions between ideals and transmission strategies that *is* in many ways similar to the pattern within scholarly editing.

Libraries and other so-called memory institutions have throughout history developed a range of methods and tools for transmitting full texts between material carriers and across media family borders. In this sense, library digitization belongs to the same tradition as 20th century microfilming and the transcribing of manuscripts performed by ancient libraries and medieval copyists. The Gutenberg era marked a sharp decline in this full-text transmitting business, and libraries devoted

¹⁶ M. Dahlström & E. S. Ore, 'Electronic text editing in Scandinavia', in *Bausteine zur Geschichte der Edition: Skandinavische Editionen*, P. Henrikson & C. Janss (Hrsgg), Tübingen: Niemeyer. Forthcoming.

their time to producing bibliographical meta-labels for documents rather than reproducing the full documents themselves. With digital reproduction technologies however, libraries have drawn an historical circle. They are yet again dedicating much energy and attention to the full-text transmission they largely abandoned at the dawn of the print age. In so doing, they take on a much more explicit role of producing and shaping the digitized cultural heritage in addition to its accustomed role of preserving it and making it available.

Digitization strategies

Let us bear in mind that digitization within libraries is much more than the mere technical capture of some content in analogue documents. It is rather a large and quite complex chain of affairs, from e.g. planning, budgeting and selection, via content capture, metadata production and publishing, over to documentation, marketing and archival maintenance. The links in the chain overlap, cooperate and support one another. In principle, every link is a factor that might affect and delimit the nature and quality of the final digital resource. This includes to what extent, at what level and in what form the users are granted access to the resource. How the different links are implemented and work together is of course dependant on the overall strategy for the digitization project.

For instance, library digitization works with two modal strategies: *text digitization* and *image digitization*—similarly to scholarly editing, as we recall. In text digitization, documents are primarily interesting as carriers of the linguistic text rather than as graphical and material artefacts. So the task is to create a machine-readable and (usually xml-compatible) encoded transcription of the text. Image digitization on the other hand wants to capture the source documents as two-dimensional images (digital facsimiles), using scanning or digital camera. Needless to say perhaps, the two approaches are often combined.

There are other distinctions of digitization approaches around as well, such as that between proactive (i.e. just-in-case) and reactive (i.e. just-in-time) digitization, or between conservational (non-tampering) and restaurational (tampering) digitization. One will quite quickly see similar distinctions within scholarly editing as well.

Mass digitization

But the currently most spoken-of strategy is, no doubt, mass digitization. It aims to digitize massive amounts of documents (thus an all-inclusive strategy) using automated means, in a relatively short period of time,¹⁷ such as Google Book Search, Europeana, Open Content Alliance, the Norwegian DigitALT or the late Microsoft Book Search. It operates on an industrial scale and with as many

¹⁷ K. Coyle, 'Mass digitization of books', *Journal of academic librarianship*, 32(6), 2006, pp. 641–645.

chain links as possible fully automated. Mass digitization systematically digitizes whole large collections, document by document, with no particular means of discrimination. The projects might assume more or less ambitious totality claims: from projects limited—by copyright, politics or administration—to a particular subset of a collection to the grand supercollection schemas mentioned above: the idea is to digitize ‘everything’ within the collection or sets of collections. For practical reasons, mass digitization has to minimize manual and labour-intensive work and cannot include intellectual aspects such as textual ambiguity, interpretation, descriptive text encoding and manual proof reading. Neither can it afford to have too much metadata and information about the source document accompanying the digital representation. In mass digitization, transmission has been flattened out into a linear streamlined affair. Mass digitization has its ambitions and its value in *scale*, not in depth. Vast amounts of books are made available and their texts searchable. Projects such as these have been met with no shortage of critical remarks. From a bibliographical and archival standpoint e.g., Google Book Search has been criticized by Duguid on the grounds that:¹⁸

- It produces representations with poor textual and graphical quality;
- It supplies scarce and occasionally erroneous bibliographical information about the source documents as well as the digital files;
- It does not seem to pay particular attention as to which edition or version to use as source, and,
- It is not particularly transparent what will happen with the digital material in regard to, for example,
 - Intellectual property rights (such as who will in the future be considered as holder of copyright and the right to edit the digital files?);
 - Economy (for which services will the future end-user likely have to pay a fee?);
 - Preservation (who will assume administrative and technical responsibility for the long-term curation of the digital material?).

One might also question whether the libraries involved in Google Book Search will be able to turn to other digitizing agents with the same source documents in the future, should Google for any reason change its activity or cease altogether as an enterprise.

In spite of these critical remarks one can definitely see strengths and positive effects of mass digitization projects. They combine on the one hand commercial agents who are strong in financial resources but in need of contents, with on the other hand public libraries who are reversely strong in contents but in need of financial resources. *A marriage made in heaven*, it would seem. The result: a gigantic, growing bank of digital texts that can be searched free-text, used as localizing tool, and—perhaps more importantly—form the technical base for many kinds of future software development and implementation.

¹⁸ P. Duguid, 'Inheritance and loss? A brief survey of Google Books', *First Monday*, 12(8), 2007, viewed on 14 January 2009, <<http://firstmonday.org/htbin/cgiwrap/bin/ojs/index.php/fm/article/view/1972/1847>>.

But mass digitization is not the only strategy around. Only a limited number of large libraries have the interest, competence and resources to implement it. It further suits some objects and collections and not others, such as fragile books, manuscripts and other unique objects, and documents whose texts are difficult to read or interpret. Cases such as those require considerably resource consuming and manual labour, and can not reasonably be referred to as mass digitization. We find in fact a number of digitization projects in libraries world wide that are anything but ‘mass’.¹⁹ What then do we call them? I suggest we refer to them as *critical* digitization.

Critical digitization

Critical digitization implements several of the links in the long digitization chain in a manual, intellectual, critical way. At every step one can make choices, deselect and interpret. Mass digitization turns a blind eye to most of these choices, whereas critical digitization acknowledges and makes active use of them. The project might e.g. focus on a single document or a limited set of documents. It may need to perform a deliberate and strategic selection from a number of possible and more or less complete source documents. Perhaps the source document has text that is difficult to decipher and decode. The text or image may need to be edited and manipulated to make sense or context. Perhaps it is vital not to destroy the source document during the digitization process (as mass digitization does) but rather subject it to careful preservational or conservational measures. The project may wish to manually and critically produce a representation that is as faithful and exhaustive *as possible* in relation to the source document and its text, its graphics and perhaps artefactual materiality. The digital object may need to be provided with large amounts of metadata, indexing, descriptive encoding, paratexts and bibliographical information, i.e. bibliographical and other scholarly research may need to be *embedded* in the objects themselves. Critical digitization is qualitative (or idiographic) 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 (i.e. nomothetic). 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 single documents in the digitized collection are on the contrary subordinated (or ‘tailored’) to more general, perhaps even universal procedures and tools.

Critical digitization is in other words a more exclusive strategy—in more senses than one. Let us briefly sum up the differences between the two approaches using Table 2 below (and bearing in

¹⁹ Two Scandinavian examples are the Codex gigas (<http://www.kb.se/codex-gigas/eng/>) at the National Library in Sweden and the Dirik family scrapbook at the National Library in Norway (<http://www.nb.no/baser/diriks/virtuellealbum/index.html>).

mind also Table 1).

Critical digitization	Mass digitization
is primarily manual	is primarily automated
critically recognizes the distortion digitization brings about	in effect treats digitization as a cloning process
undertakes a well-informed selective analysis of source copies	normally picks whatever source copy praxis or chance happens to present
maximizes interpretation and metadata	minimizes interpretation and metadata
is idiographic	is nomothetic
uses contingent tools	uses universal tools
treats graphical and material document as artefact	treats linguistic text as fact
discriminates	is exhaustive
works in depth	works on scale

Table 2: Tensions in library digitization: critical versus mass digitization

I am aware that table 2 might be interpreted as painting a flattering picture of critical digitization, while mass digitization is attributed a negative role. That is however not the aim here. The advantages as well as the disadvantages of mass digitization can be reversed in the case of critical digitization. Critical digitization is slow and very costly in relation to the number of produced objects. It addresses a small audience. It may require rare skills in e.g. textual and bibliographical scholarship. It often has an image-oriented ideal where the facsimiles are left without accompanying machine-readable transcriptions. The result risks quickly being more or less forgotten after the project is completed, and many digitizers neglect to inform about and market the project in proportion to the labour invested. Manual labour such as interpretations and tailored technical solutions are seldom properly documented (if at all), but run the risk of becoming silent knowledge locked in the minds of the digitizing persons, and therefore available to the institution only as long as the persons remain employed by it. Mass digitization on the other hand requires such an industrial scale that its strategies, practices and technologies need to be documented in order to be properly implemented by many different people and machines over long periods of time.

There is also another consequence of the way critical digitization intervenes in the documents. Someone visiting a critically digitized collection faces a material that is in a sense already encoded,

manipulated, labelled and explicitly interpreted. The more this has been done explicitly, the more it turns into a sort of *comment* on the source document which might work counter to how flexibly the material can be reused in new contexts. Mass digitization on the other hand conveys an aura of objectivity around the objects and a lack of manipulation—but that is of course due to the cloning ideal of mass digitization (an objectivity that is in effect nothing more than a chimera).

So opinions as to how usable and reusable the products of different digitization projects really are to scholars and scientists can radically differ. 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, proof-reading, 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.

One can even turn it all around and claim that critical digitization risks falling for another kind of cloning ideal than the one previously identified as typical for mass digitization. That is to say, this other cloning ideal might work on the assumption that as long as the digitization process inscribes in the digital representations large and deep enough metainformation, we will obviate any future need of new digitization efforts, since all material and all possible aspects already exist in the digital archive we have created. We would in other words be facing the ‘definitive’ digital representation, *once and for all*. Mass digitization on the contrary might be thought of as more advantageous precisely because it does *not* select, provide metadata and explicit text encoding and interpretation, but rather constructs reservoirs of source documents that scholars ideally can use, reuse and enrich the way it suits them best. We obviously recognize this pattern from scholarly editing and its tension between Alexandrian and Pergamian ideals. Again, however, we need to remember that the products of mass digitization can as well be thought of as dependent on interpretations and selections—but that these are in effect ignored and silenced, which leaves the user helplessly dependent on the unknown choices that praxis forced upon the mass digitizing institution. We should also readily admit that the lack of textual quality and proof-reading in e.g. Google Book Search hardly makes any scholar particularly happy, regardless of his or her disciplinary affiliation.

It would seem clear that both strategies have their advantages and disadvantages to different audiences. Whether a digitizing library adheres to an ideal that is closer to critical digitization or to mass digitization, however, should take into account which kind of digital material is being produced, what meta information and added values should be attached to the material, to what extent it should be able to be used and reused, and to which user group it should prove to be of interest. In that way the chosen digitization strategy legitimizes certain kinds and levels of material

at the expense of others, and favours certain user groups over others—all a question of symbolic power. There are of course mechanisms in digitization strategies that might be designated as constitutive and perhaps canonizing. Again, we recognize this tension from our previous discussion about the socio-cultural functions of scholarly editing. All in all, however, the library community in general is increasingly favouring the ideals of mass digitization and its pragmatic notion of transmission as a relatively simple, linear, content-capturing affair. The critical digitization activity is much smaller and is in many cases currently threatened to become extinct as an ineffective, over-costly luxury undertaking.

And the library community probably *is* more suitable to be engaged in mass digitization. Kjellman observes that whereas museums discriminate, select and deselect as part of their joint collecting task, national libraries in particular display an ambition of comprehensiveness in their collecting activity.²⁰ Historically this has created an ideal of objectivity within the national library communities that tends to hide the discriminating mechanisms of the institution. Such an ideal is certainly expressed and fuelled by mass digitization. The library collections, furthermore, are largely made up of printed documents that are mass produced to begin with. Given that the many copies of a published book are normally thought of as identical, it is consequently thought of as more or less indifferent whether the one copy or the other is picked (rather than selected) as source document, i.e. as ‘ideal copy’. Archives and museums on the contrary manage unique objects to a much higher degree, and their digitization in consequence regularly concerns single document artefacts rather than the text as a presumed commonality in e.g. a book edition.

Scholarly editing and library digitization

So to some extent, the area of library digitization seems to develop into a map of tensions and conflicting ideals (table 2) that bear considerable intellectual similarities to that of scholarly editing (table 1). As scholarly editing does, library digitization can make deliberate selections and discriminations, interpret, analytically compare source document candidates and seek to establish something common, sometimes even ideal, in such a collection of candidates. They both edit, optimize, document, comment and produce metatexts. They are both examples of transmission practices that are stabilized and legitimized by socio-cultural prestige institutions and that are based on agreed upon, publicly declared and fairly documented principles. In their exclusive character, they both constitute, consecrate and perhaps even canonize works of culture. If scholarly editions are cathedrals of literature, one might certainly argue that ambitious critical digitization projects within libraries turn the selected documents into national monuments—or even testaments.

²⁰ U. Kjellman, *Från kungaporträtt till läsketikett: en domänanalytisk studie över Kungl. bibliotekets bildsamling med särskild inriktning mot katalogiserings- och indexeringsfrågor*, Uppsala, Uppsala university, 2006, p. 239 f.

I would contend 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.

The major *distinction* one might arguably define between scholarly editing and library digitization concerns the establishing of the text: 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 argument can be countered—in two ways.

Firstly, scholarly editions are to some extent approaching the documentary archive by minimizing editorial interventions and instead providing fulltext and ‘raw’ versions, at times even refraining from highlighting or constructing a single uniform established base text. Secondly, critical digitization approaches scholarly editing, for example in image-oriented digitization of documents where we have more than one copy to choose from. There, the critical comparison of several source candidates based on their condition, completeness etc. results either in the deliberate selection of one candidate or in an eclectic amalgam of fragments from several candidates (such as eclectic facsimiles). 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.

Besides, there is a quite tangible cooperation between the two fields. Libraries normally house the source documents of interest to scholarly editors to begin with, and often perform the technical digitization of them to serve large editing projects. And libraries are arguably best suited to be responsible for the long-term technical and bibliographical maintenance and preservation of the digital files. They are also in a good position to coordinate and manage the intricate web of IPR interests within large editing projects in a way 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.

Increasingly, furthermore, scholarly editors are implementing practices, systems and tools that were developed by the library community for managing, relating and describing large collections of documents, such as the Metadata Encoding and Transmission Standard (METS) for metadata and the Functional Requirements for Bibliographic Records (FRBR) for descriptive cataloguing. Library digitization projects are also valuable to textual scholars and scholarly editors in the way they make large amounts of material available, if nothing else as facsimiles that can to an increasing degree be subjected to OCR and whose text therefore can be turned machine-readable and thus

reusable in scholarly editing projects. Not to mention the many digitized manuscripts that have previously been unpublished and not subjected to research but that are now becoming identified, catalogued and made available. Again, the library's policy for copyright, accessibility and technical quality becomes an absolutely crucial factor in determining the usability and reusability of the material for scholarly research.

So scholarly editing and library digitization are perhaps approaching a point where the two not only meet but even merge, at least on the project level. What happens in effect when scholarly editions and archives more and more turn into digital libraries, and what happens when critically digitized collections within libraries become increasingly granulated, technically and architecturally sophisticated, thus increasingly take on the form of 'editions'?

Scholarly editors will likely need to have access to documents that are information dense, carefully prepared and proof-read, enriched with bibliographical information, and that are accessible and editable in high resolution or deep encoding. Such needs are obviously better met by critical digitization than by mass digitization. Libraries on the other hand might be expected to either only produce information-poor files (in the mass digitization mode) or else (in the critical digitization mode) restrict access to or usage of the information-rich files—which is in fact increasingly the case (due to legal, economical, or administrative reasons) and which poses a major problem for future research and reusability.

At the same time, if the current trend within libraries towards mass digitization turns paradigmatic, and if scholarly editing increasingly will cooperate with those libraries, we will likely see a boosting of the idealistic, positivistic strand, where source documents and their texts are treated as facts rather than as expressions of subjective statements. Whether this is a good or a bad thing remains to be seen.

Bibliography

Bazerman, C. *Shaping written knowledge: The genre and activity of the experimental article in science*. Madison, WI, University of Wisconsin Press, 1988.

Bodard, G. & J. Garcés. 'Open source critical editions: A rationale'. In *Text editing, print, and the digital world*, M. Deegan & K. Sutherland (eds). Aldershot, Ashgate, 2009, pp. 83–98.

Burman, L. 'Det enkla valet: Konsekvenserna av en oproblematiserad textsituation'. In *Vid texternas vägskaal: Textkritiska uppsatser*, L. Burman & B. Ståhle Sjönell (eds). Stockholm, Svenska vitterhetssamfundet, 1999, pp. 83–99.

Coyle, K. 'Mass digitization of books'. *Journal of academic librarianship*, 32(6), 2006, pp. 641–

645.

Dahlström, M.. 'The compleat edition'. In *Text editing, print, and the digital world*, M. Deegan & K. Sutherland (eds). Aldershot, Ashgate, 2009, pp. 27–44.

Dahlström, M. & E. S. Ore. 'Electronic text editing in Scandinavia'. In *Bausteine zur Geschichte der Edition: Skandinavische Editionen*, P. Henrikson & C. Janss (Hrsgg). Tübingen: Niemeyer. Forthcoming.

Duguid, P. 'Inheritance and loss? A brief survey of Google Books'. *First Monday*, 12(8), 2007, viewed on 14 January 2009,

<<http://firstmonday.org/htbin/cgiwrap/bin/ojs/index.php/fm/article/view/1972/1847>>.

Frohmann, B. *Deflating information: From science studies to documentation*. Toronto, University of Toronto Press, 2004.

Greetham, D. 'Textual scholarship'. In *Introduction to scholarship in modern languages and literatures*, J. Gibaldi (ed). New York, MLA, 1992, pp. 103–137.

Greetham, D. *Theories of the text*. Oxford, Oxford University Press, 1999.

Kanzog, K. *Prolegomena zu einer historisch-kritischen Ausgabe der Werke Heinrich von Kleists: Theorie und Praxis einer modernen Klassiker-Edition*. München, Hanser, 1970.

Kjellman, U. *Från kungaporträtt till läsketikett: En domänanalytisk studie över Kungl. bibliotekets bildsamling med särskild inriktning mot katalogiserings- och indexeringsfrågor*. Uppsala, Uppsala university, 2006.

Renear, A. 'Literal transcription: Can the text ontologist help?'. In *New media and the humanities: Research and applications*, D. Fiormonte & J. Usher (eds). Oxford, Oxford university, 2001, pp. 23–30.

Robinson, P. 'The one text and the many texts'. *Literary and linguistic computing*, 15(1), 2000, pp. 5–14.

Rockwell, G. 'What is text analysis, really?'. *Literary and linguistic computing*, 18(2), 2003, pp. 209–219.

Steding, S. A. *Computer-based scholarly editions: Context, concept, creation, clientele*. Berlin, Logos, 2002.

Tanselle, G. T. 'Reproductions and scholarship'. *Studies in Bibliography*, 42, 1989, pp. 25–34.