Documents in transit
Lars Björk
Extended abstract submitted for the LIDA conference 2012

Background
Transmission is an integrated part in the life cycle of texts, and consequently, in the life cycle of documents. Texts have throughout history been transferred from carrier to carrier in order to reach readers and as a method to ensure their survival beyond the life span of the supporting substrate.

The methods used for the transmission of text based information can build on concrete sequences of action such as inked types pressed against the surface of paper – or less tangible processes like the transformation of information from analogue format into binary code, stored via electrically charged sequences in a computer hard drive. But regardless of technological platform, in order to store (and to read) a text you need a physical surface - whether in the format of a page in a book or as the combination of the display and the platter of the hard drive. The digital revolution has not caused the death of the document, but it has altered its defining parameters.

During the digitisation process – the transmission of information from analogue to digital format – a segment of the source document is captured. The representation of this segment, the captured layer, equals the amount of information that will serve as a surrogate for the source document. Thus the digitisation process creates a new document with an informative capacity that more or less resembles the source document. But the process always initiates changes, the relation between the source document and the reproduction is never a 1:1 situation.

'Differences will always be introduced in copying; the trick is to regulate the process sufficiently so that the resulting differences are of little or no consequence and that the properties of greatest consequence are shared'

(Levy 2000:26)

The characteristics of these "properties of greatest consequence" will vary with the purpose of the digitisation process. However, several attempts have been made to determine generalised criteria for defining the informative content of documents in order to facilitate the design and implementation of digitisation processes. The feasibility of such approaches has been discussed. (e.g. Yeo, 2010; Puglia & Rhodes, 2007; CLIR, 2001; Menne-Haritz & Brübach, 2001)

The digital dilemma
Thus the digitisation process establishes and defines the relation between two manifestations of a work; the source document and its reproduction. This relation is conditioned by the premise that the reproduction is always derived from the source document. The reproduction may fulfil various
purposes; it can act as a substitute for, or an enrichment of, a source document. It can also function as a reference, a pointer, or a comment to that document. But this relation is distinguished by a specific bearing – the reproduction traces back to a source document.

Although the characteristics of the digital reproduction primarily can be described as the result of the interaction between technical components in the process there is also another, less tangible dimension that effects the reproduction, viz. the characteristics of the source document. Supporting any digitisation activity there are decisions regarding what constitutes the information capacity of the source document. The prospect of well-informed use of the reproduced document is consequently dependent on the user’s comprehension of the particulars of the process; how the process has shaped the reproduction.

But on what basis is such an evaluation possible?

The technical density and complexity of digital reproduction technology often results in the digitisation process being regarded as a kind of ”black box” where documents are entered and refined to produce pure information. Under such conditions the user’s possibility to assess the characteristics of the reproduction process are limited. Abram & Luther (2004) use the term "Format agnostic" in reference to the growing group of users who regard the web as a primary source of information. For this group of users there are hardly any differences in credibility between print media – accessed via authorising institutions e.g. libraries – and information accessed via the web; "information is information". Paul Conway (2010) summarises this trend when he refers to Google as:

'...a metaphor for the instant gratification expected in information search and retrieval today. For a new generation of users, Google represents anonymous access to information without human mediation.'

(Conway 2010:63)

But concern has also been raised regarding the trustworthiness of digital reproductions and the related question of authenticity (e.g. Narkiss, 2009; Ross, 2004; Bradley, 2005). There are reasons to believe that the increasing reliance on digital surrogates will have an effect on the role and function of material heritage collections.

A conceptual model

The digitisation process is often referred to using the metaphor of a chain (e.g. Cornell, 2002; IFLA, 2002). The merit of the concept of “digitisation chain” is the reference to digitisation as a process that consists of a series of functions, i.e. the actual digital capture is only one of several connected activities. However, a limitation of the chain metaphor is that it implies a linear process with a clearly defined start and end product. This paper will, based on some of the findings in a PhD-project in
progress, present a complementary view where the digitisation process is described as consisting of a network of clustered activities. The result – the digital reproduction – is only one (albeit a primary one) of several informative outcomes in the process.

During the digitisation process the informative capacity of the source document is radically restructured. The reproduction – the re-presentation in digital format – is conditioned by the characteristics of the source document, the particulars of the process, and the nature of the presentation format. The interaction between these factors is complex.

Using the theoretical frameworks presented by Lund (2010, 2004) and Pédaque (2003) as a point of departure, a conceptual model of the digitisation process was developed, as part of the PhD-project. This model is based on a multimodal view on documents and their potential as providers of information and will enable a more complex understanding of the transformative aspects of digitisation. The potential usefulness of this approach for the analysis of the informative capacity of the digital representation, with a special reference to the “properties of greatest consequence”, will be discussed.

References

The web-based sources were accessed 30/1-2012


