The Autological Constitution of Digital Cultural Artefacts

An Analysis of the Implications of ICT on Memory Organizations

Marton, Attila

The Autological Constitution of Digital Cultural Artefacts

An Analysis of the Implications of ICT on Memory Organizations

Attila MARTON, The London School of Economics and Political Science
a.marton@lse.ac.uk

Paper presented at the 3rd Latin American and European Meeting on Organization Studies (LAEMOS), 7–10 April 2010, Buenos Aires.
Sub-theme 4: Digital media, information and organizations

Abstract
The societal shift from writing to printing to information and communication technologies has been accompanied by a shift in the structure of social memory that seems to threaten our capability to remember. Within this context, a preliminary analysis is offered on the impact of the digitization of cultural heritage on the ways social memory is being organized by memory institutions (archives, libraries and museums) attempting to bring their repositories online. Informed by the work of Niklas Luhmann and Elena Esposito, the paper addresses the problem of an ICT driven organization of cultural heritage transforming information objects into autological, self-describing digital information objects. The research aims to contribute the notion of memory as a counter-concept to the discussion on information and its technologies in the information systems field and related domains such as organization studies and the social study of ICT. It also advocates the necessity to focus more on the implications of ICT on the ways social memory is structured.

Keywords: social memory, digital libraries, digital information objects, ICT

Introduction

In the 1980ies, marking the 900th anniversary of the historical Domesday Book – an exhaustive survey of medieval England commissioned by William the Conqueror in 1086 – the BBC developed a digitized multimedia version stored on highly resilient discs and run on an Acorn Microcomputer system. Only 15 years after the digitization and a total cost of £2.5 million, the data was inaccessible because the file formats used to store the data were unreadable by contemporary computer systems. In one of its reports from that time, the British newspaper The Guardian quoted an expert on this matter saying: "*It is ironic, but the 15-year-old version is unreadable, while the ancient one is still perfectly usable* […]. *We're lucky Shakespeare didn't write on an old PC*" (McKie and Thorpe 2002). It took the combined efforts of experts from Leeds University and University of Michigan to program an emulator which made the digital Domesday Book accessible again (BBC 2002).

Despite the happy ending, the digital Domesday Project exemplifies a pressing concern related to the increasing degree of cultural artefacts being mediated in a binary-based digital format. Information and communication technologies are designed for speedy data processing and efficient data transmission, not for long term preservation and persistent accessibility. At the core lies the ephemeral nature of the binary-based medium of 0s and 1s depending on being interpreted by the right soft- and hardware. The challenge is not to keep the 0s and 1s themselves or to copy them from a degrading to a new storage medium but to keep them informative – i.e. to maintain their accessibility for software to be read and processed as well as to be presented in a format suitable for a user. With digitized artefacts, this matter is less problematic since one can always fall back to the original for reference once the digitized version turns out to be inaccessible. However, the problem becomes quite urgent for the, by now, incomprehensible amount of cultural artefacts and documents that are born digital and mediated via the internet and the services it affords. Lacking an institutionalized, trusted and dedicated caretaker, librarian and archival experts are warning from a digital dark age – an era bereft of persistent documents bearing testimony to the historical condition of humanity (Kuny and Cleveland 1996; Russell, Weinberger et al. 1999; Bennett 2001; Baker 2006). It seems ironic that despite the immense storage capabilities at our disposal, in the end, most of the stored data may be forgotten rather sooner than later dissolving into an ocean of incomprehensible noise (Brindley 2009).

An analysis of the implications of preserving online, digital-born content on archival practices is discussed elsewhere (Kallinikos, Aaltonen et al. 2010). This paper focuses on cultural artefacts being digitized, preserved and made accessible online by, so called, memory institutions (libraries, archives and museums) and the change these artefacts go through from an object that is being described to an autological object that describes itself. A preliminary interpretation will be presented based on a single revelatory case study (Yin 2003) on the Europeana initiative[1]. The case study is part of The Internet And Information Growth Research project (TIGAIR)[2] on the implications of technological information on various social domains and institutions (Kallinikos 2006a). Launched in July 2007 by the Conference of European National Librarians (CENL) in cooperation with the European Commission (EC) and EU member states, Europeana brings together archives, libraries and museums from all over Europe under a single framework in order to make their digitized content accessible via the WWW. In November 2008, Europeana launched the first prototype as a proof of concept. Since then, the project is developing a sustainable, operational service with a focus on increasing the number of content providers, achieving a critical mass of 10 million digitized and fully accessible items, implementing multilinguality features for the official EU languages, providing discovery services based on semantic web technologies and supporting member organizations in their digitization efforts, to name a few (Purday 2010). Some goals have already been achieved, especially multilinguality and semantic technologies based services, however, are still in their very early stages of planning, research and development. Though initially launched as the European Digital Library (EDL), it became obvious relatively early during the process that the project was not about librarian services but rather a unifying effort of the traditionally separate domains of archives, libraries and museums into a single *memory organization*.

Embedded within Niklas Luhmann's Theory of Social Systems (Luhmann 1997; 1998), the paper will start with an introduction into 1) the notion of social memory as an operation of forgetting (Esposito 2002; 2008) and 2) the research domain of digital libraries. Followed by 3) an analysis of the autological nature of digital objects and 4) the way Europeana turns them into information objects, the paper will conclude with 5) a more abstract interpretation of the role of memory

[1] www.europeana.eu
[2] www.tigair.info

organizations and the relation forgetting establishes with information and data in the online world. The research aims to contribute the notion of memory as a counter-concept to the discussion on information and its technologies in the information systems field and related domains such as organization studies and the social study of ICT. It also advocates the necessity to focus more on the implications of ICT on the ways social memory is structured.

The ICT-Turn in Social Memory

With the increasing importance of computational information and data in basically all domains of human existence (Kallinikos 2006a), it may seem counter-intuitive for research within the domain of the social study of ICT to focus on a concept that, at a first glance, seems to be quite the opposite to information – memory. While information is often associated with surprises, novelty and learning, memory is more often than not seen as a passive recorder of events stored in an archive-like fashion (Borgmann 1999; Kallinikos 2006b; Schmidt 2008). Consequently, the so called memory institutions – libraries, archives and museums – are described as warehouses storing the collective memory of a social entity – be it groups, civilizations, nation states and so forth (Landheer 1957:91). As this paper will argue, this notion of memory is too simplistic and, therefore, overlooks the active role memory plays in the construction of reality and information.

From a social scientific perspective, the first noteworthy contribution to the topic was made by Marcel Halbwachs during the early decades of the 20th century. In his concept of collective memory, he discusses the formation of a group memory shared by its members. Remembering is a collective practice of reconstructing the past based on the present social framework the group finds itself in (Halbwachs and Coser 1992). The collective memory is located in or, if you will, dispersed among the personal memories of individuals. For Halbwachs, it was the individual that remembered in unison with other members of the same collective (Olick and Robbins 1998). As Esposito (2008) points out, collective memory has become more and more limited with the increasing complexity of society. She suggests that a social memory emerged that is based only on social operations without the involvement of mental memory processes for its own sustainment. Due to developments in communication technologies from writing and printing to relatively recent innovations in telemedia and ICT, social interaction has been enabled to be mediated in increasingly

decontextualized ways leading to a high degree of variability in terms of who is communicating with whom about what topic over what period of time. For instance, writing frees the communicating persons from the necessity of being at the same place at the same time. Mass media constructs its message in a way that is understood by an anonymous audience. Finally, the internet or rather many of the services built on top of it enable the kind of many-to-many communication where creators and audience are one and the same. This process goes hand in hand with an increasingly abstract structure of social memory (Esposito 2002; Boyden 2003).

A defining innovation in terms of memory was the mass production of print media that made the oral tradition of repetition obsolete. Modern libraries emerged as an autonomous organizational form dealing with copies of mass produced books, newspapers and so forth rather than with unique documents or artefacts, as it is still the case with archives and museums today (Marton 2009). The immense increase in production and the popularization of reading due to the availability of cheap books and other print material from roughly the end of the 18th century on was accompanied by another innovation, that is the organization of the librarian collection by means of a 2nd order classification system – the card catalogue (Thompson 1982; Brown and Duguid 2000:95; Weinberger 2007). Ancient and medieval libraries offered walkable information spaces by means of ordering the items themselves. This arrangement can still be found in some public and research libraries today. Closed shelf libraries, on the other hand, construct an information space in their catalogues based on representations, namely descriptive metadata such as the author's name, title of publication or keywords used for indexing purposes. The catalogue exemplifies the shift in terms of social memory from remembering (mnemotechné, ars memoriae) to forgetting (Luhmann 1997; Esposito 2002).

Memory eliminates the unique features of an event constructing sameness into difference that leads to a stabilized set of categories. Consequently, memory does not store each and every event but rather selects what is remarkable and forgets the rest (Coyle 2008; Esposito 2008). It actively enables an observation to distinguish between what is already known or old and what is not known or new (Marton 2009). This notion can already be applied to language – oral, written or printed. Words or terms categorize the unique objects and singularities they refer to in a communicative, hence, social sense. For instance, the term "table" signifies all tables or rather the parameters that makes a table a table and not, for example, a chair irrespective of the

individual features of each and every table there is. Being a surprise or novelty, information, on the other hand, can only occur if an event is comparable with what is expected – a variety of what is remembered (Kallinikos 2006a:103). Consequently, remembering is not the retrieval of a stored event but rather the activation of a set of instructions to reconstruct the event. Hence, what is stored "*in archives* [as well as libraries and museums] *is not facts, but disaggregated classifications that can at will be reassembled to take the form of facts about the world*" (Bowker 2005:18). What is remembered are not actual tables but reconstructions of tables based on the parameters that define the category "table". Equivalently, a catalogue does not remember books but rather classifications and instructions on how to find items in the repository.

In this sense, information and memory are two sides of the same coin. Memory can be described as the organization of observing information as it constructs sameness into difference and, as a result, recursivity and identity (Esposito 2002). Without memory, everything would be new and surprising, every event would be observed as a singularity in all its details. In other words, literally anything would be informative (which is the same as nothing) since the capability to ignore noise (by distinguishing it from information) would not exist. In terms of communication technologies and therefore social memory, the function of forgetting has emerged into the organizational form of libraries, archives and museums. Their practices of collecting a very selective area of cultural heritage, of documenting and cataloguing based on an ex-ante classification system as well as preserving the material integrity of the selected items allow for a persistent findability and accessibility. Within the domain of online communication, however, the professional categorization of communication media is replaced by search engine algorithms and emerging folksonomies based on social tagging (Weinberger 2007). The focus is shifting from packaged media (books, newspapers, CD-ROMs) to the dynamic and momentary rendition of information. It is not only the content that is in constant flux but also the search engine results page helping users to navigate the ever growing online information landscape (Kallinikos 2006a; Kallinikos, Aaltonen et al. 2010).

In opposition to the catalogue of modern memory institutions, the classification system is, so to speak, constructed on the fly in an ex-post fashion by users themselves or by search engine algorithms leading to, what Weinberger (2007) calls, the 3rd order of ordering things. While the catalogue allows for the forgetting of

the actual collected cultural artefacts by remembering only an abstraction of it (the catalogue card), search engines allow for the forgetting of the catalogue by remembering only the algorithms to perform a search. Going back to Bowker's quote above, it is not only the social facts that are being reassembled but also the classification system according to which those facts are being ordered. The static model of information retrieval is replaced by a performative model of information construction (Esposito 2002:357). It is this environment of momentary ordering rather than persistent order memory institutions are stepping into.

Digital Libraries

The phenomenon of interest can be broadly positioned within the discourse on digital libraries. Library and information science (LIS) has been discussing the possibilities for taking advantage of developments in information and communication technologies for quite some time now (Thompson 1982; Agre 2003). In the early nineties, it was the digitization of the card catalogue made accessible via computer terminals on site, followed by projects to digitize collected items themselves for reasons of preservation (e.g. disintegrating newspapers) and accessibility via internet services (e.g. e-books) roughly from the turn of the century on (Petschar 2002). Projects like Project Gutenberg[3], the World Digital Library[4], The European Library[5] and Europeana as well as projects launched by commercial enterprises such as Google Books[6] or the Open Content Alliance[7] bear witness to the immense efforts put into digitizing millions of cultural artefacts and into the sophisticated services making these artefacts available online. Complementing these, projects of preserving online as well as offline born-digital documents are also on the way attempting to bring persistence into an ephemeral medium without an incremental memory function or dedicated archival trustee – the Internet Archive[8] being the most prominent one (Kallinikos, Aaltonen et al. 2010).

The theme of digital libraries encompasses a whole range of very diverse topics ranging from 1) accessibility issues in terms of multi-dimensional search functionalities, new collaborative environments or the usability of face-to-screen

[3] www.gutenberg.org
[4] www.wdl.org
[5] www.theeuropeanlibrary.org
[6] books.google.com
[7] www.opencontentalliance.org
[8] www.archive.org

interfaces and the readability of texts on screen (e.g. Greene, Marchionini et al. 2000; Thong, Hong et al. 2002), 2) information system design for data management, storage, retrieval and search result ranking (e.g. Marcum 2003; Tuominen, Talja et al. 2003) to 3) interoperability and metadata standardization (e.g. Suleman and Fox 2001) and 4) copyright and digital rights management (e.g. Russell, Weinberger et al. 1999; Bearman and Trant 2005). Given the diversity of the discussed topics, it comes as no surprise that the term "digital library" itself is very vague without a clear definition. As a consequence, Oppenheim and Smithson (1999) suggest that contemporary efforts from parts of the librarian world should be referred to as "hybrid libraries" since the digital aspect complements rather than eradicates the paper-based aspect of librarian services.

It is also questionable whether digital repositories and online services – though possibly initiated by librarians - are in fact libraries in a digital format. Binary based digital media renders the distinction between original and copy useless. Consequently, the distinction between the specialized domains of, on the one hand, archives and museums traditionally focusing on unique documents and artefacts and, on the other hand, libraries focusing on mass produced communication media seems to be of little help in the digital world. Be it born-digital or digitized media, a document or artefact does not appear as either unique or as one copy of many. Taking Europeana as an example, it is more appropriate to see some of the projects of digitizing cultural artefacts and providing online access as a unifying process resulting in a single type of *memory organization* dealing with digital media.

The Autology of Information Objects

The traditional way of memory institutions to provide for findability is based on very specific channels for discovery. In a library, for instance, the paper-based card catalogue usually allows only a search according to author names or keywords linked to the actual *information object* (a book, a newspaper issue, a CD-ROM) by a shelf mark.

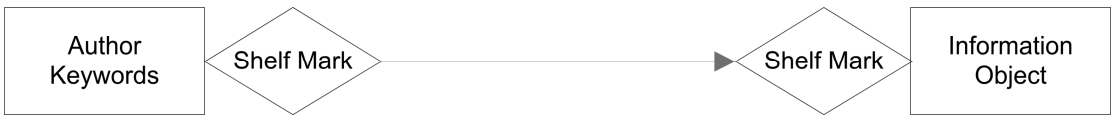

Figure 1: Catalogue Discovery System

The basic set-up shown in Figure 1 can be easily translated into a data-based model as it is the case with Open Public Access Catalogues (OPAC) by copying the metadata from the card to a database. Digital information objects fit into this paradigm by being treated like books or newspapers. A persistent URL, linking to those digital objects, works as a functional equivalent to the shelf-mark. The crucial point is that descriptive metadata and information objects are separated. The user navigates through the index in order to discover information objects. Although database technology allows for additional functionalities such as searching for titles or the usage of Boolean operators, the underlying concept remains the same. Hence, digital libraries can turn out to be nothing more than online portals allowing users to search through the catalogue databases similar to searching through a card catalogue.

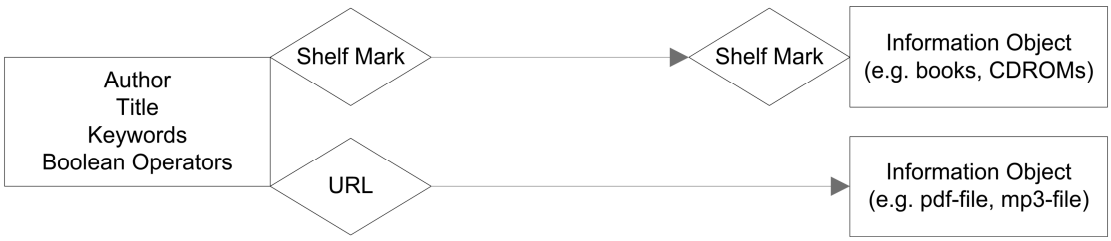

Figure 2: Hybrid Library Discovery System

However, as the example of the Domesday Project showed, digital information objects bring their own set of problems (Kallinikos, Aaltonen et al. 2010). In order to provide for persistent accessibility a memory organization does not only have to tend for the usability of its discovery system but also for the integrity and, above all, authenticity of its information objects. In terms of digital information objects, accessibility depends just as much on the software standards used to create a document as on the integrity of the data stored. In order for a computer file to be displayed in the correct way, instructions on how the binary code needs to be interpreted by software are embedded into the file itself. This is not the case with, say, books that can be read as they are. These instructions blur the initially clear distinction between metadata and data. A case in point is the digitization of complex information objects. The following example is from a metadata enrichment project at the Heidelberg University Library using an xml based metadata standard called

METS (Metadata Encoding and Transmission Standard)[9]. It shows the structural metadata used to bind the various aspects of a digitized manuscript into a complex information object.

The digitization of the manuscript resulted in image files of various qualities for each scanned page. Usually, the high resolution images are used for preservatory reasons while the low resolution versions are used for access via the internet or as thumbnails for navigation. In this example, five different quality levels of the scanned pages are provided: 1) MIN (minimal for work), 2) MINplus, 3) DEFAULT (standard quality), 4) DEFAULTplus and 5) THUMB (thumbnail images). As a first step, the scans are grouped according to their quality.

```
<mets:fileSec>
    <mets:fileGrp USE="MIN">
        <mets:file ID="filemin00001" MIMETYPE="image/jpg">
            <mets:FLocat LOCTYPE="URL" xlink:href="http://diglit.ub.uniheidelberg.
            de/diglitData/image/cpg108/1/000_A_Vorderdeckel.jpg"/>
        </mets:file>
        <mets:file ID="filemin00002" MIMETYPE="image/jpg">
            <mets:FLocat LOCTYPE="URL" xlink:href="http://diglit.ub.uniheidelberg.
            de/diglitData/image/cpg108/1/000_A_Vorderspiegel.jpg"/>
        </mets:file>
        <mets:file ID="filemin00003" MIMETYPE="image/jpg">
            <mets:FLocat LOCTYPE="URL" xlink:href="http://diglit.ub.uniheidelberg.
            de/diglitData/image/cpg108/1/000_B_1ar.jpg"/>
        </mets:file>
.....
.....
.....
</mets:fileGrp>
</mets:fileSec>
```

Figure 3: File Grouping of Scanned Images according to Image Quality

In Figure 3, a group is set for the lowest quality scans (USE=”MIN”). Each image file is assigned a unique identifier starting with the cover of the manuscript (Vorderdeckel) as ID=”filemin00001” accompanied by the definition of the file type as a .jpg compressed image file. Finally, the ID is linked to the actual computer file via a URL. This is done for all the minimum quality scans forming a File Group. Medium and high quality scans as well as thumbnails are grouped into their respective File Groups accordingly.

As a second step, a so called Physical Structural Map is applied that basically reflects the material make-up of the artefact, in this case the page sequence of the

[9] The report is available at
http://enrich.manuscriptorium.com/files/enrich/ENRICH_WP5_D_5_2_final.pdf

manuscript (see Figure 4). The page, used as the basic unit of the manuscript, is merely a conceptual container holding the various versions of the scanned imagery and ordering them based on the sequence of the pages of the original. Again starting with the manuscript cover (Vorderdeckel), all the digitized versions are linked to the cover as the first page (ID=”phys00001”) based on the File Group IDs assigned as shown in Figure 3. In this case, there are five different versions per page starting with the lowest quality scans (filemin and fileminplus) to the standard (filedefault), above standard (filedefaultplus) and thumbnail quality scans (filethumb). This is repeated for every page in the order of the physical manuscript’s page sequence.

```
<mets:structMap TYPE="PHYSICAL">
    <mets:div ID="phys0" TYPE="pageSequence">
        <mets:div ID="phys00001" ORDER="00001" TYPE="page"
        ORDERLABEL="Vorderdeckel">
                <mets:fptr FILEID="filemin00001"/>
                <mets:fptr FILEID="fileminplus00001"/>
                <mets:fptr FILEID="filedefault00001"/>
                <mets:fptr FILEID="filedefaultplus00001"/>
                <mets:fptr FILEID="filethumb00001"/>
        </mets:div>
        <mets:div ID="phys00002" ORDER="00002" TYPE="page"
        ORDERLABEL="Vorderspiegel">
                <mets:fptr FILEID="filemin00002"/>
                <mets:fptr FILEID="fileminplus00002"/>
                <mets:fptr FILEID="filedefault00002"/>
                <mets:fptr FILEID="filedefaultplus00002"/>
                <mets:fptr FILEID="filethumb00002"/>
        </mets:div>
.....
.....
.....
</mets:structMap>
```

Figure 4: Physical Structural Map

In addition to the order of the pages, however, there is also a logical structure in books based, for instance, on chapters. This is reflected in Logical Maps (see Figure 5) breaking the content of the manuscript down into meaningful units rather than into data-files as it was the case in the steps above.

```
<mets:structMap TYPE="LOGICAL">
    <mets:div ID="log0" DMDID="dmd0" TYPE="Monograph"
    ADMID="amdSec_complete_01">
        <mets:div ID="log00192" DMDID="dmd00192" LABEL="Einband vorne"
        TYPE="chapter"></mets:div>
        <mets:div ID="log00193" DMDID="dmd00193" LABEL="1r Revelatio
        nova itineris et passionis undecim milium virginum, Lib. I, dt."
        TYPE="chapter"></mets:div>
        <mets:div ID="log00194" DMDID="dmd00194" LABEL="48r Cordula-
        Legende" TYPE="chapter"></mets:div>
        <mets:div ID="log00195" DMDID="dmd00195" LABEL="49v Elisabeth
        Schonaugiensis, Liber revelationum de sacro exercitu virginum coloniensem,
        I 1-21, dt." TYPE="chapter">
    </mets:div>
.....
.....
.....
</mets:structMap>
```

Figure 5: Logical Map

Each chapter of the manuscript is assigned a unique ID and labelled, starting with the front cover (Einband vorne) that is treated as if it was a chapter of its own.

Finally, the physical and logical maps are linked together in a Structural Links Map. Figure 6 shows how the pages 1-6 (phys00001-phys00006), each of them linked to the actual versions of the scanned page, are linked to the logical unit front cover (log00192 is the ID for “Einband vorne” – the front cover) defined in Figure 5.

```
<mets:structLink>
    <mets:smLink xlink:to="phys0" xlink:from="log0"/>
        <mets:smLink xlink:to="phys00001" xlink:from="log00192"/>
        <mets:smLink xlink:to="phys00002" xlink:from="log00192"/>
        <mets:smLink xlink:to="phys00003" xlink:from="log00192"/>
        <mets:smLink xlink:to="phys00004" xlink:from="log00192"/>
        <mets:smLink xlink:to="phys00005" xlink:from="log00192"/>
        <mets:smLink xlink:to="phys00006" xlink:from="log00192"/>
.....
.....
.....
</mets:structLink>
```

Figure 6: Structural Links Map

All these various maps combined result in the structural metadata of the manuscript. In more abstract terms, the assemblage of the digitized manuscript information object based on the structural metadata can be depicted in the following way:

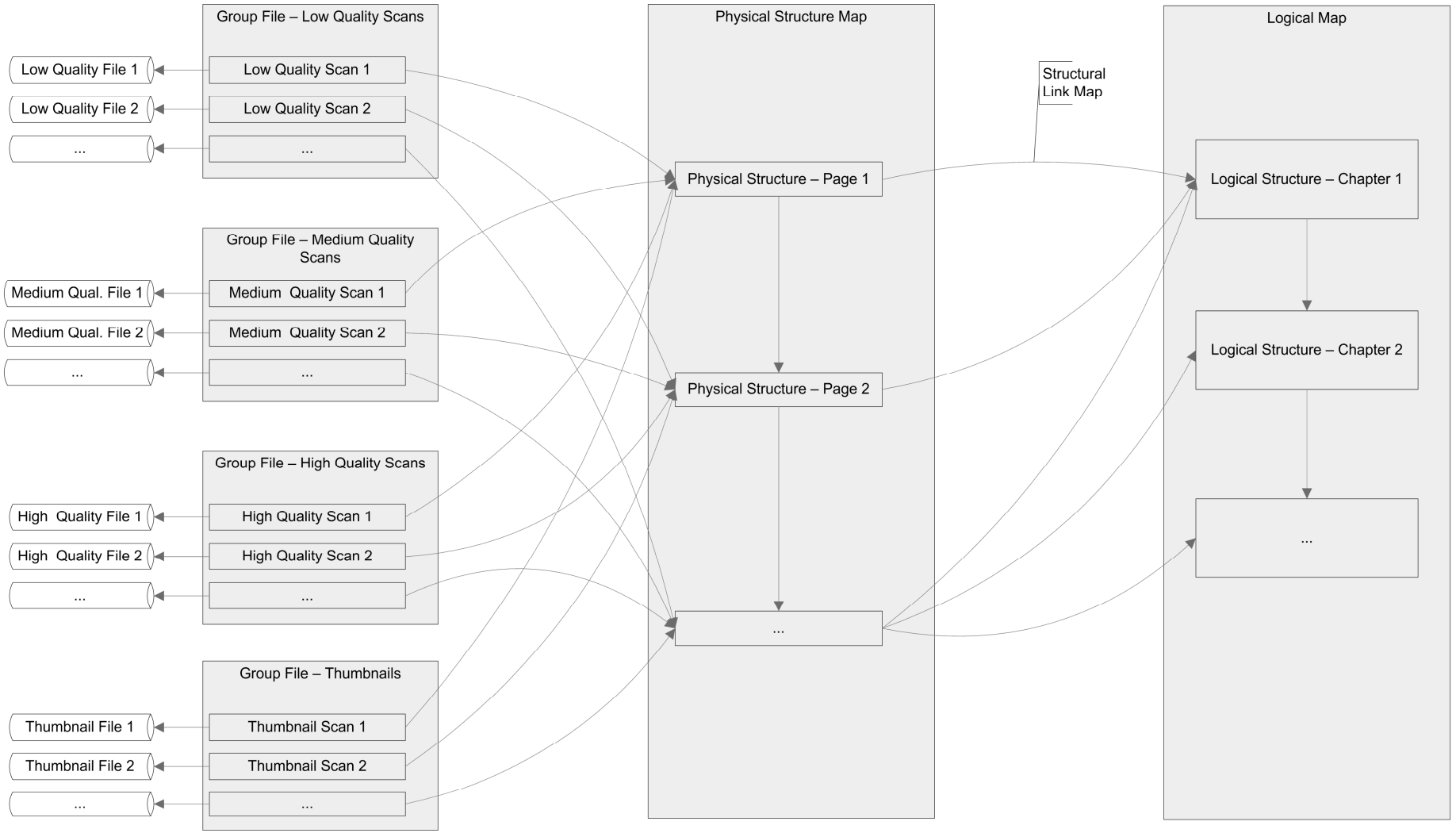

Figure 7: Schematic Depiction of the Structural Metadata of a Complex Information Object

In fact, Figure 7 shows the functional equivalent to the binding of the manuscript – the glue that holds the pages together. The book as a packaged entity actually does not exist anymore but rather becomes a logical entity defined by a set of instructions on how the various parts are to be assembled the moment a user accesses the item[10]. The information object that was given by the material make-up of the physical artefact has to be emulated computationally in order to be usable for a human being. Metadata, traditionally used to describe and to make an item findable by means of a catalogue, is in addition now used to actually assemble the item. The elementary unit used to composite an information object can vary to some degree. While an image scanned manuscript lends itself to be decomposed into single pages, an OCR scanned print-book, for instance, could be decomposed into paragraphs, sentences or words. The textual elements could be stored in files separate from the pictorial elements. The structural metadata would then need to contain information related to lay-outing, page breaks and so forth.

The information object, the item to be organized by a memory organization, is only a set of instructions on how it is supposed to be assembled and displayed. In other words, the information object describes its own construction. It is, what linguists would call, autological (Hughes and Brecht 1978:14; Esposito 1996). An example for an autological term is the word "English" which, as an adjective, signifies itself. In a similar fashion, a digital information object – be it a composition of various parts or a simple computer file - describes itself as well. The structural metadata *is* the manuscript. It defines the logical entity "manuscript" by instructing software applications on the emulation of itself. As a consequence, a distinction can be made between descriptive metadata used for the documentation and discovery of the information object and structural metadata used for the assemblage of the information object. The crucial point, however, is that it is not only the descriptive metadata that needs to be preserved but also the structural metadata. If one finds an item in the catalogue, without the structural metadata intact, it will not be accessible, hence the manuscript ceases to exist.

[10] In principle, the same can be said about the scanned images that form the elementary unit – the page – of this complex information object. Image files – in this case .jpg files – contain metadata instructions as well that tell a software application what to do with the 0s and 1s it processes.

The Europeana Information Space

While the previous section showed the disaggregation of traditional packaged communication media into elementary units, the Europeana project goes one step further and disaggregates the fundamental information system of any memory institution – the catalogue. Be it in a medieval book catalogue, a modern card catalogue or an IT database, descriptive metadata is usually separated from the items it describes to be used as an index for enhanced performance in terms of search and retrieval. In case of Europeana, however, the descriptive metadata becomes part of the information object. One of the goals of the Europeana project is to create "*a network of inter-operating surrogates enabling semantics based object discovery and use*" which is to become an integral part of the overall information architecture of the WWW (Europeana documentation). This is to say that Europeana is planned to become more than just an online portal users visit in order to search through the content providers' repositories but rather opens up the repositories for online services. It will be, for instance, possible for search engines to crawl Europeana. Hence, if a user searches for a certain topic, related items from Europeana will be displayed in the search engine results page. In addition to legal and storage space issues, this is the reason why Europeana works only with, so called, surrogates of the actual digital objects rather than with the digital object itself managed by one of the content providers.

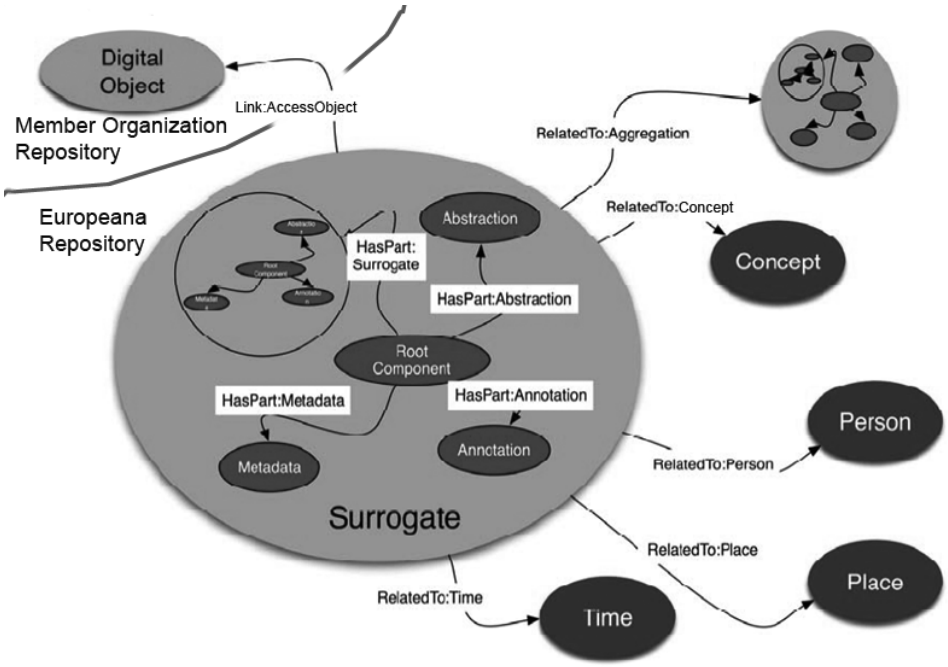

Figure 8: Logical Structure of the Surrogate Model (Source: Europeana)

At its very core, a surrogate is actually only a link to a digital object. The link itself – called the Root Component – functions as a node connecting the digital objects descriptive metadata with annotations and abstractions (e.g. table of content). All these parts form a simple surrogate which, as a whole, is connected to other surrogates. This logical entity would be created for every elementary unit. In case of the example of the manuscript discussed above, that means a surrogate for each page. Therefore, surrogates can also be part of a complex parent surrogate aggregation which, again, contains of a link to the logical entity being the manuscript, descriptive metadata, annotations and abstractions and is linked to other complex surrogate aggregations. In this sense, there is a one-to-one correspondence between each entity at the provider's site and a surrogate at Europeana's site. Complex surrogates reflect the structural relationship not the content of the represented logical object entities. They are the equivalent to the structural metadata of the source.

The Europeana information space ends up being a network of surrogates simple and complex which are in their core functionality nothing more than links qualified by descriptive metadata. Today, users have the possibility to navigate through the Europeana surrogate information landscape by means of a simple, search-engine-like as well as a more specific, advanced search functionality.

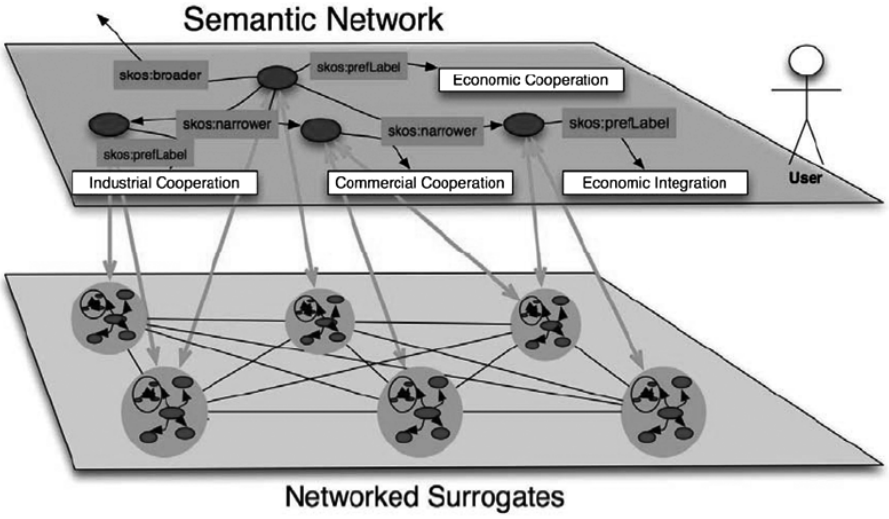

Figure 9: Europeana's Semantic and Surrogate Layer (Source: Europeana)

In the future, users should also be able to browse through the repository by means of a semantic network on top of the surrogate network allowing for concept driven rather than metadata driven discovery. In opposition to the traditional model of user-catalogue interaction (see Figures 1 and 2), "[t]*he user now primarily interacts*

with the semantic network to explore the Europeana surrogate space which now has the metadata as parts of the surrogate and surrogate aggregations" (Europeana documentation). As Figure 9 shows, the user will be able to browse through the Europeana semantic network layer (very much like walking through an open-shelf public library) in addition to the more specific search functionalities usually provided by card-catalogues in closed-shelf settings and data-based OPAC services. The disaggregation of the catalogue and fusion of the descriptive metadata and the root component into an information object allows for an integration of the Europeana information space into the WWW. Specifically, so called "landing pages" are generated presenting the key information about specific items in a webpage format which can be crawled by search engines. Those landing pages are representations of the surrogates which in turn are again representations of the actual digital object.

The key observation at this point is that Europeana is not a unified catalogue (a meta-catalogue of the provider's catalogues) but rather presents a new way of organizing digitized cultural heritage. Europeana constructs an information space based on logical entities being even more abstract than the traditional catalogue cards. The surrogate becomes an information object only in an autological way through the self-description encoded in descriptive metadata being a part of the surrogate it describes. It defines itself rather than being defined by a catalogue. Descriptive metadata used to be the index, now it is being indexed. The information object – that which is being organized – is not an entity or an item itself but rather a network that is actively entified into a delimited information object. It only exists as a logical entity insofar as Europeana preserves not only its key compositional elements but also the links between them. Given the arguments of this paper, those basic units are *operations* encoded as metadata (structural and descriptive) combined with references to the sources (the root component of the surrogate) providing for findability and, ultimately, accessibility (Kallinikos, Aaltonen et al. 2010). This notion is becoming more important with the degree of granularity of the information object. The pages of the digitized manuscript, as an example for low granularity, are still meaningful without structural metadata assembling them into a single information object. If the manuscript were disaggregated into words and imagery, the structural metadata would be key for the existence of the manuscript as a whole. Lose that and the stored words and images turn into meaningless noise.

Analysis

The societal shift of the focus in terms of communication technologies from writing to printing to information and communication technologies has been accompanied by a shift in terms of social memory from remembering to forgetting structured in increasingly abstract forms and classification systems. In fact, communication contains an aspect or rather an event of being informed and of being understood (Luhmann 1997). Information – being a surprise or novelty – only occurs in light of what is remembered or rather what is not forgotten. Hence, social memory – the operation of filtering singular details as noise based on a classification system – is strongly connected to the medium into which communication is embedded (Olick and Robbins 1998; Esposito 2002). The preservation of a book, for instance, is nothing more than the preservation of the potentiality of that communication technology to inform over a long period of time. Communication technologies, after all, do not only have the capabilities to distanciate space but also time (Giddens 1990). However, what is to be preserved undergoes a selection process of communication technologies that fit into the categorization system of the dedicated memory institution.

In line with mass media print, the card catalogue emerged as a second order of ordering things (Weinberger 2007). In opposition to its predecessor – the book catalogue – the card catalogue is able to include metadata, at least in principle, from an unlimited number of items. The organization of the information objects shifts from an organization of the objects themselves to an organization of the representations of the objects – the metadata that fit on a single card. In simpler terms, the library becomes the catalogue. Exceptions are, for instance, open-shelf public libraries. The catalogue exemplifies an increased level of abstraction in terms of the operation of selecting what is remembered and forgetting the rest. The information objects themselves are forgotten that is they are not arranged in a way to allow for their findability but rather to save storage space. What is being remembered is what fits on a catalogue card forming an abstraction of the repository with very specific avenues for discovery.

The WWW, being on the verge of becoming the next primary communication technology, is memorized in an even more abstract way. It's navigability is not provided by catalogues but rather by search engines and increasingly by social tagging (Weinberger 2007). Especially with search engines one can witness a restructuring of social memory. The navigation through the online information space

is guided by the search results page which is created, based on algorithmic calculations, for a specific user every time a search query is processed. Hence, with search engines social memory does forget fixed categorizations and the selective, persistent avenues paved for information discovery. Instead, it is enough if the algorithms are remembered (Esposito 2002). The abstraction of the repository based on representational properties of a fixed order is exchanged by a higher degree of abstraction based on performative order*ing* only to be forgotten the moment the ordering is abandoned – the moment the results page is closed.

This is not to say that the ordering is lost but rather it is stored as data. In terms of memory, the contemporary challenge is to make petabytes of data informative by means of second order technologies (Gantz, Chute et al. 2008). Data-mining tools or online search engine services are, in fact, technologies of remembering. They reconstruct events stripped of their singularity by being categorized and stored as data based on the very parameters according to which they were collected and stored in the first place. Information is forgotten as data. By digitizing their repositories and making them available online, memory institutions basically add their data to the already existing sea of ephemeral data. However, that contradicts the object oriented nature of the services provided by memory institutions. A book comes as an object with physical structures and borders enabling its usability by means of its material make-up. Digitized books are only logical constructs, their borders and structure in need of being created and maintained by means of information technology. In more abstract terms, information objects are actively constructed and entified only when accessed. Europeana is a case in point as it also disassembles the catalogue by making descriptive metadata part of its information object – the surrogate. Hence it can organize the immense amount of information objects stored at the providers' site by simply filtering most of their attributes as noise

Europeana radicalized this notion by making the descriptive metadata of an information object an integral part of the very same information object it describes. Metadata – be it descriptive metadata linked to the Root Component or structural metadata of complex digital information objects represented by compound parent surrogates – is therefore forgotten as data as well. A part of the surrogate does not have any meaning on its own but rather gains its functionality due to its relationship to the other parts. Being a logical entity of its own rather than a copy of another information object, the surrogate only identifies certain aspects such as the location of

what it represents (the Root Component's URL) qualified by a limited set of categories (descriptive metadata) and forgets all the rest.

This notion very much fits into the performative and momentary ordering via online search engines described above. It is not the content of a webpage that matters but rather the algorithms of indexing the webpage based on certain rules and procedures which index certain aspects of a webpage while ignoring others. The instructions that actually make up a webpage are left for browsers to be interpreted and displayed to a user. In a similar fashion, the information object is made up of instructions that need to be processed by software first in order to be presented to a user. However, there are also differences. After all, it is the provision of *persistent* findability and accessibility that distinguishes memory institutions from recently emerged information service providers such as search engines. Given the autological nature of digital information objects, the memory organization needs to take appropriate steps in order to guarantee the integrity of its data but also of its logical objects it is dedicated to preserve. What is actually preserved by Europeana is, first of all, the surrogate model (see Figure 8) that is the blueprint of how the various parts of the surrogate are to be linked. It is this schematic model that provides the parameters according to which surrogates are being constructed out of data and, therefore, remembered. Second, Europeana applies a standardized set of categories for its descriptive metadata. Third, the Root Component of the surrogate consists of a persistent link to the actual digital object. The persistence of the digital objects themselves is managed by the providers and are, therefore, only of peripheral concern. In this sense, Europeana actually provides a service for persistent referencing.

Conclusion

The emergence of social memory as a distinct social (not mental) operation began with the rise of print as a communication technology initializing a shift from remembering to forgetting finalized by the mass production of new printed material (e.g. newspapers, novels) in contrast to the repetition or copying of a canonized set of texts. With mass media becoming the prime communication technologies, societal communication became too complex in its focus on novelty and variety for a collective memory to handle but rather needed to be organized like any other social domain. Organized social forgetting (instead of collective remembering) finally led to the differentiation of a memory institution into libraries, archives and museums as

distinct organizational systems that are able to increase their capacities to remember by forgetting more. Defining social memory as an operation of forgetting, as the construction of sameness into singularities by filtering their unique details, libraries, archives and museums can be seen as organizations of forgetting. Remembering, being the exception, is the reconstruction of events based on a classification system according to which those events were categorized in the first place. Information, on the other hand, is simply a variation of what is remembered.

This paper made an argument about the dynamics between information and memory being two sides of the same coin. The structure of social memory was described as reaching higher levels of abstraction by referring to the ways information has been primarily organized – in other words, how forgetting and remembering has been structured - since the rise of mass print media. Generally, the order of the things themselves has been replaced by an order of abstractions of the things (the catalogue card) followed by a performative ordering. Search engines, being the prime example for this new paradigm, construct a catalogue every time a user makes a search query. Instead of being represented by a fixed catalogue, information is constructed through algorithmic calculations which are, in fact, an abstraction of an abstraction (the catalogue card) of an information object (a book, webpage, etc.). As this paper has argued, this has some wide ranging implications for traditional memory institutions trying to step into the online world.

In terms of their artefacts, digitization already results in an abstraction of the physical item. The digitized information object becomes a logical entity encoded into structural metadata describing itself. The Europeana project introduces another step of abstraction – the surrogate – that basically consists of 1) a link to the digitized information object and 2) descriptive metadata. The card catalogue as an index is dissolved and each of the cards becomes part of what it describes. Including the search functionality, a chain of abstractions of abstractions emerge that finally ends with an information object. Social memory, reflected in the way its dedicated organizations manage cultural heritage, is increasingly based on an autological constitution of digital cultural artefacts that are being constructed by means of information technology. Given these arguments, the problem of accessing the digitized version of the Domesday Book may, in fact, be just one example of many still to come.

Literature

Agre, P. (2003). "Information and Institutional Change: The Case of Digital Libraries", in: A. P. Bishop, N. A. Van House and B. P. Buttenfield (Eds.), Digital library use : social practice in design and evaluation. Cambridge, Mass., MIT Press: 219-40.

Baker, D. (2006). "Digital library futures: a UK HE and FE perspective." Interlending & Document Supply **34**(1): 4-8.

BBC (2002). Digital Domesday book unlocked. **2010** http://news.bbc.co.uk/1/hi/technology/2534391.stm

Bearman, D. and J. Trant (2005). "Converting Scanned Images of the Print History of the World to Knowledge: A Reference Model and Research Strategy." Russian Digital Libraries Journal **8**(5). http://www.elbib.ru/index.phtml?page=elbib/eng/journal/2005/part5/BT

Bennett, S. (2001). "The Golden Age of Libraries." The Journal of Academic Librarianship **27**(4): 256-59.

Borgmann, A. (1999). Holding on to reality : the nature of information at the turn of the millennium. Chicago, University of Chicago Press.

Bowker, G. C. (2005). Memory practices in the sciences. Cambridge, Mass., MIT Press.

Boyden, M. (2003). "The Rhetoric of Forgetting: Elena Esposito on Social Memory." Image [&] Narrative - Online Magazine of the Visual Narrative **6**. http://www.imageandnarrative.be/mediumtheory/michaelboyden.htm

Brindley, L. (2009). We're in danger of losing our memories. The Observer. London http://www.guardian.co.uk/technology/2009/jan/25/internet-heritage

Brown, J. S. and P. Duguid (2000). The social life of information. Boston, Harvard Business School Press.

Coyle, K. (2008). "Managing Sameness." The Journal of Academic Librarianship **34**(5): 452-53.

Esposito, E. (1996). "From self-reference to autology: how to operationalize a circular approach." Social Science Information **35**(2): 269-81.

Esposito, E. (2002). Soziales Vergessen: Formen und Medien des Gedächtnisses der Gesellschaft. Frankfurt am Main, Suhrkamp.

Esposito, E. (2008). "Social Forgetting: A Systems-Theory Approach", in: A. Erll and A. Nünning (Eds.), Cultural memory studies : an international and interdisciplinary handbook. Berlin ; New York, Walter de Gruyter: 181-89.

Gantz, J. F., C. Chute, et al. (2008). The Diverse and Exploding Digital Universe: An Updated Forecast of Worldwide Information Growth Through 2011. Framingham, MA, IDC http://www.emc.com/collateral/analyst-reports/diverse-exploding-idc-exec-summary.pdf

Giddens, A. (1990). The consequences of modernity. Cambridge, Polity.

Greene, S., G. Marchionini, et al. (2000). "Previews and Overviews in Digital Libraries: Designing Surrogates to Support Visual Information Seeking." Journal of the American Society for Information Science **51**(4): 380-93.

Halbwachs, M. and L. A. Coser (1992). On collective memory. Chicago, University of Chicago Press.

Hughes, P. and G. Brecht (1978). Vicious circles and infinity : a panoply of paradoxes. Garden City, N.Y., Doubleday.

Kallinikos, J. (2006a). The consequences of information : institutional implications of technological change. Northampton, MA, Edward Elgar.

Kallinikos, J. (2006b). "Information out of information: On the self-referential dynamics of information growth." Information Technology and People **19**(1): 98-115.

Kallinikos, J., A. Aaltonen, et al. (2010). "A Theory of Digital Objects." First Monday(forthcoming).

Kuny, T. and G. Cleveland (1996). The Digital Library: Myths and Challenges. IFLA General Conference, Beijing, China. http://www.ifla.org/IV/ifla62/62-kuny.pdf

Landheer, B. (1957). Social functions of libraries. New York, Scarecrow Press.

Luhmann, N. (1997). "The control of intransparency." Systems Research and Behavioral Science **14**(6): 359-71.

Luhmann, N. (1998). Die Gesellschaft der Gesellschaft. Frankfurt a.M., Suhrkamp.

Marcum, D. (2003). "Requirements for the future digital library." The Journal of Academic Librarianship **29**(5): 276-79.

Marton, A. (2009). Digital Libraries as Information Organizations. The Re-Unfolding of the Memory/Information Paradox. ECIS, Verona, Italy. http://is2.lse.ac.uk/asp/aspecis/20090224.pdf

McKie, R. and V. Thorpe (2002). Digital Domesday Book lasts 15 years not 1000. The Observer http://www.guardian.co.uk/uk/2002/mar/03/research.elearning

Olick, J. K. and J. Robbins (1998). "Social Memory Studies: From "Collective Memory" to the Historical Sociology of Mnemonic Practices." Annual Review of Sociology **24**(1): 105-40.
Oppenheim, C. and D. Smithson (1999). "What is the hybrid library?" Journal of Information Science **25**(2): 97-112.
Petschar, H. (2002). "Vom Zettel- zum Bildkatalog: das Projekt "Digitales Bildarchiv" an der Österreichischen Nationalbibliothek - ein Werkstattbericht." Zeitschrift für Bibliothekswesen und Bibliographie **49**(2): 59-68.
Purday, J. (2010). Europeana Annual Report: 1 February 2009 – 31 January 2010. The Hague, Europeana http://version1.europeana.eu/c/document_library/get_file?uuid=df49b382-aa16-497f-8405-948b532dc0d7&groupId=10602
Russell, K., E. Weinberger, et al. (1999). "Preserving digital scholarship: The future is now." Learned Publishing **12**(4): 271-80.
Schmidt, S. J. (2008). "Memory and Remembrance: A Constructivist Approach", in: A. Erll and A. Nünning (Eds.), Cultural memory studies : an international and interdisciplinary handbook. Berlin ; New York, Walter de Gruyter: 191-201.
Suleman, H. and E. A. Fox (2001). "A Framework for Building Open Digital Libraries." D-Lib Magazine **7**(12). http://www.dlib.org/dlib/december01/suleman/12suleman.html
Thompson, J. (1982). The end of libraries. London, C. Bingley.
Thong, J. Y. L., W. Hong, et al. (2002). "Understanding user acceptance of digital libraries: what are the roles of interface characteristics, organizational context, and individual differences?" International Journal of Human-Computer Studies **57**(3): 215-42.
Tuominen, K., S. Talja, et al. (2003). "Multiperspective Digital Libraries: The Implications of Constructionism for the Development of Digital Libraries." Journal of the American Society for Information Science and Technology **54**(6): 561-69.
Weinberger, D. (2007). Everything is miscellaneous : the power of the new digital disorder. New York, Times Books.
Yin, R. K. (2003). Case study research : design and methods. Thousand Oaks, Calif., Sage Publications.