Received 27 April 2012
Revised 1 July 2012
Accepted 4 July 2012

Research libraries and the internet

On the transformative dynamic between institutions and digital media

Niels Ole Finnemann
Department of Aesthetics and Communication, Aarhus University, Aarhus, Denmark

Abstract

Purpose – The purpose of this paper is to improve comprehension of some of the intricate interrelations between research libraries, the role of media and the knowledge production system.

Design/methodology/approach – This paper establishes arguments from a historical analysis of stages in the conceptual development of digital media and stages in the digitization of library functions. The historical approach leads to some discussions and forecasts of the future of research libraries.

Findings – Digital media have a disruptive, revolutionary potential, but path dependency is often a modifying component in the historical development. This is demonstrated in different stages of the development of the interrelationship between digitization, digital media and research libraries. Digital media become disruptive due to the strength of the historical dynamic, rather than as a result of particular agencies. Today the historical dynamic has reached a point where all institutions concerned with knowledge handling will have to redefine themselves. Research libraries are gradually incorporated into a number of new "research infrastructures" which are being built around different kinds of data materials, and each research library may specialize according to some sort of coordinated criteria.

Originality/value – This paper demonstrates new openings to a theoretical and conceptual understanding of the interrelationship between digital media and developments of research libraries.

Keywords Digitization, Research libraries, Digital media, Knowledge production

Paper type Research paper

1. Digitization of library functions as defined within the traditional institutional framework developed around the printed text

Historically, the exponential growth of knowledge has continuously initiated new forms of knowledge production, e.g. the specialization and reconfiguration of relations between disciplines, and it has led to changes in the overall chain of production, distribution, consumption/appropriation and recirculation of knowledge[1]. This process has always included the development of new theories, new methods, new research questions, new disciplines and changing interdisciplinary relations. It has also incorporated new media, whether they were developed specifically to solve particular issues within the knowledge production process or whether they were imported from other spheres and elaborated further.

The exponential growth of knowledge production does not simply manifest itself in the increasing number of scientific and scholarly books and journals, but also includes a global expansion of the institutional system of universities, commercial and

Journal of Documentation
Vol. 70 No. 2, 2014
pp. 202-220

0022-0418
DOI 10.1108/JD-05-2013-0059

non-commercial scientific laboratories, libraries and research libraries as part of the whole system of knowledge production. One might include all types and levels of school in this system.

A main component in the modern knowledge production system has been the building of an ever-increasing number of virtually identical institutions, each serving a particular area, town, region or country. If, in any given place, there is a university or a laboratory, there is also a research library and other types of collections. If there are scientists and scholars, there are also librarians located close to one another. The system functions as a "Luhmann-like", relatively-autonomous, "autopoietic" system, separated from other parts of society and defining its own internal standards, which refer to similar international knowledge communities.

This system was built on spoken, written and printed media. Speech dictates the local and corporeal here and now. Writing and typewriting allowed for the cumulative process of cataloguing the particular local collections of books and other materials. Print of a multitude of identical copies finally allowed for the building of many institutions in parallel, which were spread around the globe, giving access to more-or-less the same knowledge reservoir; the librarian being an important intermediary between the different libraries and between the library and the institutions served. The modern dynamic media – emerging from the mid-nineteenth century –, such as the telegraph, telephone, radio, movie, television, sound and video recorders, did of course add to the knowledge production system, as these media were used both for communication, and the documentation and dissemination of knowledge. They added to the exponential growth of knowledge production, but they did not provide new solutions for the handling of the growing amounts of knowledge, and they were mainly used within the existing institutional framework or in separate institutions, built in accordance with this institutional setup. However, over the years some kinds of specialization and task distribution between the research libraries emerged.

Until recently, it has been possible to treat the institutions within this system as relatively stable and separate units, universities, libraries etc., each type with more-or-less the same, well-defined and delimited set of tasks. In the case of the research library these tasks were centered on collecting, storing, and giving access to the global reservoir of scholarly and scientific knowledge including written, printed and electronic materials. New media were gradually incorporated into the existing, institutional system. Handwritten catalogues were supplemented with typewritten cards and, later, with electronic databases, which facilitated mechanical search. Microfilm was introduced to replace archived newspapers.

Similarly, digital media were for many years developed and incorporated in libraries as additional to or as simple replacements of existing media or manual practices within the existing institutional and conceptual framework. In the libraries the electronic version of the printed catalogues was modeled on the same organizing principles, mainly adding the mechanized search function within the catalogue of the particular institution. This is path dependency in a very strong form and fully in accordance with the classical conceptualization of computers as a means of mechanizing and automating repetitive processes, as they allow mechanical processing of the stored content.

Today a far-reaching break up of this system is taking place. The entire system, which originally developed to deal with the megatrend in the past few centuries, has been overblown by its own triumphant results: the very same exponential growth of books, journals and other print materials, and of scholars, researchers, scientists, and librarians.

The overall process comprises a number of particular trajectories, each of which brings its own dimension to the general process. Globalization of knowledge production comes together with further specialization of knowledge, accompanied by new interdisciplinary, trans-disciplinary and multidisciplinary research challenges. The notions of disciplines and interdisciplinary relations are renegotiated, while the individual scholars and scientists become increasingly unable to overview the production of new knowledge, even within their own, specialized field. The societal position and function of researchers and scholars is challenged by the on-going industrialization of knowledge production and education. Finally, the exponential growth in knowledge production is directly related to the exponential growth in use of material resources, potentially threatening the biosphere.

As a consequence of these more-or-less interrelated tendencies, we are in the midst of a major disruption in fundamental societal structures and institutional relations. The knowledge production system is under pressure, from a variety of needs, to develop new ways of handling the many and complex processes involved. Digital media are developed both within and outside the library institutions as a part of the response to these pressures as it is manifested in the explosive development and spread of an increasing variety of digital media for storing, preserving, searching, retrieving, transmitting, and analyzing digital data corpora of all varieties[2].

In the twentieth century, computer technology was predominantly considered a practical instrument for a more "rational" or cost effective performance of more-or-less the same tasks within in the existing institutional setup. Catalogues were digitized, but still kept separate from the texts catalogued and maintained as catalogues for local collections. Buckland (1992) characterizes this epoch as the "automated paper library", in which information technology is used as a "sustaining technology", which improves the performance of established processes or products by fulfilling the same kind of requirements, just more effectively (Lewis, 2004, p. 69; Christensen, 1997). The cataloguing system was the first thing to be digitized in libraries, but the process continued with digitization of the publications and was then expanded further to include distribution. For Buckland (1992), this was a third phase leading to the electronic library, where both collections and bibliographic control mechanisms are electronic. According to Lewis (2004), the transition from the "automated" to the "electronic" library began in the 1990 s, with the development of full text databases, the internet and the world wide web, and we are still in the early part of this transition, which "is likely to run another decade or two" (Lewis, 2004, p. 70). It remains to be seen whether we will ever rely solely on purely digital materials and processes, or whether older media will continue play a role, or perhaps given new functions. However, nowadays, all sorts of knowledge and symbolic content are primarily stored in digital format.

The story told here is not simply about the spread of digitization into areas within (and outside) the library, it is also a story about changes in the relations between medium and institution; digital media are produced to handle all the dimensions in

knowledge production within the institutions while, at the same time, they remain a challenge and a threat for the existing institutional system. They do so not least because of their flexible and "malleable" character which allow the transition from being primarily sustaining to being, what Lewis and Christensen call, a "disruptive technology", which brings with it new value propositions and enables a game change to take place[3]. The changing concepts of the computer are further discussed in section 2.

Unfortunately, for the existing institutions, Lewis (2004) also argued that "established organizations generally fail when change involves disruptive technologies, and organizations at the periphery or from different sectors often succeed." Perhaps what Lewis had in mind, writing in 2004, was the story of game changing agencies like Wikipedia, Amazon and Google, while Facebook and other social media were still to appear.

Lewis' skepticism towards the existing institutions maybe explains that he overlooks the real dilemma. In spite of his awareness of the difference between sustaining and disruptive technologies, Lewis fails to address the core issue of the relation: what kind of disruption is at stake? And why is it that the electronic library becomes disruptive if it is simply a more complete or extended version of the automated library, i.e. if the digital catalogue is simply supplemented with digitization of the collections and reference systems?

While Lewis fears a complete dissolution of the library functions, it is maybe more appropriate to focus on the reciprocal relations between digital media and institutions. The disruption at hand does not depend on the degree of digitization, as assumed in the idea of a complete replacement, but on the fact that digital media allow the catalogue and collections to be combined and accessed in ways previously not possible, neither in the automated nor in the paper-based library. The possible breakdown of these institutional, functional and physical barriers does not even stop at the walls of the existing libraries. The technology can easily be extended to include the relations between all catalogues and libraries, and a huge variety of other sorts of collections outside the existing system of libraries. Since the internet places any kind of knowledge repository at one's fingertips, it opens up a whole new trajectory in knowledge repository relations, whether they be general or specialized research libraries, or commercial, as well as civilly-driven, knowledge repositories. The internet facilitates the use of repositories in a way which undermines the *raison d'être* of the institutions acting in parallel to serve each their local research communities.

Book and journal collections are rapidly changing into digitized archives and repositories. Traditional site-specific catalogues are supplemented with site-specific, as well as globalized, keyword searching or free-text search, within and beyond the single work. Scientific and scholarly papers are increasingly stored in digital repositories outside the libraries, even if the libraries continue to act as intermediaries, providing access to some of these repositories. The "work" itself is now stored as a file, which can be deliberately split and recombined with elements from other files. It can also be retrieved and delivered anywhere and anytime, within the bounds of privacy, property and copy right restrictions. This also impacts on the notion of authorship and its role in the library, as an increasing number of files will become a type of bricolage, whether machine-generated or deliberately edited. Authorship, editorship and curation could merge into each other[4].

At the same time, digitized and digital formats call for new kinds of bridging between the storage and the user. Bridge building is not simply a matter of creating a user-friendly Graphical User Interfaces (GUI). It is – not least in the area of research libraries – also a matter of continuous development and adaptation of specialized interfaces connecting the archives and the analytical software tools. This will be an on-going process, due to the ever-changing needs and interests of scientists, researchers and scholars from across the globe. The process will also include the incorporation of continuously renewed methods of software-supported analysis.

Digital media undermine historical reasons for the physical location of institutions. The books, the buildings and the postal system can be converted into bits, which can be spread, stored, and retrieved from anywhere. The former reasons for the separation between research, libraries, archives, museums and news media are also undermined as these activities become more intertwined and to a high degree converted into digital processes. Digital media are production media, distribution media, search media and storage media all-in-one thus encompassing all the dimensions of knowledge production.

Today, digital devices, platforms and search methods permeate the knowledge production system, as well as the surrounding society and culture and, centered on the internet, initiate new combinations of communicational and institutional relations in society. At the bottom of this is a long-term transformation of the relation between media and institutions, due to the increasingly central role of the internet in the overall communicative infrastructure of society[5].

The fact that the institutional *raison d'être* evaporates does not mean that the institution dissolves in a postmodern deconstruction, or that all functions and processes will eventually be automated. Neither is it likely that there will be a "revolutionary" replacement of old institutions, such as universities and research libraries, with completely new institutions. It does mean, however, that the existing institutions and their population will have to rethink and redefine their role, identity and functions within a new communicational infrastructure. In the long term, libraries will have to reinvent themselves and see their own history as an antecedent to future agencies in a knowledge production system, which contains an extensive variety of different agencies, often more specialized and more networked than local, full-service research libraries used to be.

Digital media both add to the exponential growth and provide a new, general basis for managing the exponential growth, because they allow for a new communicative infrastructure in society in general. In the case of libraries, they form a media platform for institutional transformations, since they facilitate new types of division of labor, new relations to the wide array of stakeholders (whether other institutions, researchers, corporations or civil society), and lead to a reduction in the number of redundant books stored in various locations. We no longer need thousands of copies of any single book in different places. A few files will do the same job, since digital media can be copied and distributed easily. Digital media also allow printing "just in time" of any appropriate number of copies, in any given place and in the required physical or digital format.

So, what are the appropriate institutional forms for the various services and functions that should be nurtured in this type of future knowledge system?

One would like to be able to answer such a question but, unfortunately, it is impossible to do so. Even if we know that the megatrend, which has been a main driver, will continue, it is not possible to predict the direction beyond the next few years.

There are several reasons for this.

One is that digital media are less stable than older media, and they are more susceptible to future modifications and the projection of new ideas and concepts into the functional architecture. They are, therefore, less predictable than older media. The breakthroughs of the desktop computer/PC in late 1970 s and of the internet in the 1990 s were unforeseen, as were the breakthroughs of the mobile phone and social media early this century. Even if we know that we will see a tremendous development of online services in the years to come, we cannot predict their particular forms and functionalities.

A second reason is that the loosening of the relations between the various parts of the production, distribution and consumption of knowledge releases the internal dynamics in each of these areas. This process opens up the possibility for a variety of new commercial and civic agencies in the fields of production, distribution and consumption of knowledge, while the whole area has become subject to on-going, intensive public management initiatives on the level of both national governments and EU-institutions.

A third reason is that the very notion of knowledge is subject to significant changes, which, among other things, favors usability, pragmatism, collaboration, transactionalism, process-orientation, situated perspectives, and a transition to fragmentary composition of knowledge and scanning perspectives; practices the results of which are difficult to predict, even if they appear to fit nicely with the hyper textual, interactive and multimodal features that form the basic properties of digital media.

In the end, nobody can predict where the institutional reconfigurations will lead.

One major dynamic will be the further development of digital media and software supported methods on all levels of knowledge production. A second will be the impacts of trans-institutional library networks interfering with the parental institutions. These two dynamics are interdependent, but each of them is also influenced by other sources.

2. The historical unfolding of the disruptive properties of digital media

The reader may have noticed that the notion of the computer and of information technology (IT and ICT) is replaced in this article with the notion "digital media". This latter term is preferred, because it signifies the plurality of digital media, and refers to the distinct properties of this particular sort of information technology and not to the whole array of information technologies throughout history. The plurality of digital media includes the plurality of gadgets, dedicated computers, and the plurality of interfaces and semantic regimes. A main reason for speaking of digital media in the plural is that the computer contrary to all formerly known machines comes with a variable functional architecture. At the same time the notion of digital media allows us to reconsider a whole array of implicit ideas, which have been associated with the computer and IT for the past seventy-five years. One of the most basic of these is the implicit idea that IT is a driver of development and not itself driven and given form and function by social, cultural and political needs. The determinism inherent in the

idea of IT as a driver is strange, since the computer is a purely cultural artifact. If compared to former machines and media, it is more open for modifications by projections of new ideas, not only on the level of content. New ideas can also be projected into the functional architecture, thus changing the operational structures of the machine. The variability of the functional architecture even allows that the computer can be made sensitive to the content of any individual message if so wanted.

The concept of the computer plays a decisive role in the development of the technology, but the concept itself and the properties in focus change due to societal and cultural needs and desires. Thus the variability of the functional architecture is not a new property, but it has played only a minor role in previous conceptualizations of the computer.

The history of digital media can be traced back to the 1930 s, if not earlier. In 1936, the English mathematician, Alan M. Turing, wrote a seminal paper on "Computable numbers" (launching the idea of the universal computing machine) (Turing, 1936). This can be considered a crucial landmark in the conceptual history of computers and a point of departure for later conceptualizations, which present the computer as a number cruncher, an electronic brain, an intelligent machine, whether it operates according to logical principles or associative, "neural" processes[6]. Turing touched upon all of these conceptualizations, either in his 1936 paper or in his second seminal paper, "Can Machines Think?" which he wrote in 1951 (launching the idea of artificial intelligence). In his 1936 paper, Turing also introduced an interesting distinction between the automated computer and what he called "the choice machine". The two machines were, in fact, identical, but they differed with respect to the tasks performed. If the task could be completed and automatically processed, the machine functioned as an automatic machine. In that case, the next step was completely defined by the present state. However, if the task could not be completely processed in the machine and new inputs were required, the machine functioned as a choice machine. In that case, the present state did not determine the next step. This is the point of departure for utilizing the variability of the functional architecture as there are in principle no semantic or syntactical restrictions for the future sequencing. For Turing, the relevance of this distinction related to the mathematical question he tried to solve (a subtheme of Gödel's Entscheidungsproblem) and he showed no further interest in the choice machine. This, however, is the more general conceptualization, because it does not rely on the limitations of a particular type of task. It shows that the computer functions independently of any particular task, purpose or meaning, which is also the reason that it can be used for an indefinite array of possible tasks, purposes and meanings. For such reasons, it may be considered not simply as a multimodal machine, but as a machine allowing all sorts of semi-automatic processes on any semiotic level whether the level of expressional units, syntactic structures or semantic regimes, as for instance word processing, optional linking, many different sorts of interactivity and so forth.

The conceptualization of the computer in the mid-twentieth century was partly driven by formalist theories, and partly by specific, practical (not least military) requirements, such as the need to cope with a huge amount of weather data and to calculate missile-trajectories, and, during the war, for code-breaking, and thus reinforcing the idea that the computer is rooted in the handling of numbers. There is no doubt that this idea has played, and still plays, a fundamental role in the development of computers. But the computer does not operate on the basis of numbers, which are

semantic units, but on the basis of bits, which are semantically empty and, hence, open for changing syntactical structures and semantic meanings. This is why the computer is radically different from other machines, as it makes possible the projection of different conceptualizations into the functional architecture.

The gradual expansion of purposes and fields for computer usage – spreading into commercial areas, civic society and home life – has also led to new conceptualizations beyond the number cruncher and beyond the interpretation of the two bits as a logical alternative, "either – or", or Boolean logic. In the 1960 s, logical interpretations led to ideas of artificial intelligence and to the idea of a universal programming language, Algol, also resulting in heated debates on so-called "dirty programming", which uses arbitrary go-to functions instead of logical relations[7]. A somewhat wider concept was developed within various information theories, which acknowledged that there is no simple limitation of the concept of information and there is no simple number system or logic capable of capturing the possible relations between the two bits. A constellation of bits may function as a letter in our Latin alphabet, as part of an image, as part of a sound or program, a mathematical rule or simply as data and content of any expression.

If we consider the computer to be an autonomous, intelligent machine, though artificial, it is conceived within a modern paradigm of autonomous agents, separated from the surrounding natural world. A Cartesian machine put into a Newtonian physical world. Within this framework, computerization would allow everything to be done more rationally and more perfectly, but would leave everything in much the same fundamental order as it was.

The idea of the logical machine, with a focus on formalisms, had its heyday in the era of the mainframe computers, which were in the hands of professional computer experts and specialized departments within universities, corporations and public institutions. Those were the days when it was assumed that, in a country like Denmark, a handful of mainframe computers would do the job for the whole country. During this time, many seeds were sown for a new game change. Ideas for Hypertext, a decentered internet, new graphical interfaces and more intuitive input devices appeared in the 1960 s. With the development of the microprocessor in the early 1970 s, the foundations for a game change were laid.

To the centralized and professionally-controlled mainframe computers, PCs were added, accompanied by new ways of thinking, such as object-oriented programming, the emerging human computer interaction perspective (HCI), Computer Supported Cooperative Work (CSCW), participatory design, and a variety of related ideas, which viewed the computer as a tool or a toolbox, often stressing the flexibility and variability of the box (e.g. Norman and Draper (1986), Bannon and Pylyshyn (1989), Ehn (1989), Bolter (1991)). The idea of the autonomously-operating, automatic machine was abandoned and replaced with the idea of interaction between man and machine, thus aiming to end human alienation towards our machines by making the machine a friend. The machine was softened to a tool with a user-friendly interface.

What started as mainframe-oriented IT in research libraries would soon develop into a new phase, based on the increasing access to catalogues and collections and reference systems, which still continues today. With HCI, the focus is shifted from the internal operations of the machine to the interface between man and machine, but the computer is still considered as a separate entity. The mainframe and PC/HCI

paradigms differ to the extent that non-IT experts are included as possible "users" in the PC/HCI-paradigm.

The PC/HCI or toolbox paradigm opens for new conceptualizations not least by means of graphical user interfaces. It also opens for new kinds of path dependencies as the machinery can now be operated by non-IT experts that bring with them a variety of other sorts of competences and qualifications. But the perspectives are still restricted to individual machines.

With the arrival of the internet, digital media become the most overarching notion. Digital media are moving into society at large, and the technologies are now in the hands of a growing number of different social groups; in the 1980 s, technology was primarily utilized in the workplace, but it soon spread into the family home and thus opened the door for dependency of another path. The fast growing number of digital gadgets which have emerged since then shows that the variability of the delimitation hardware/software, which is unique for computers has become a major trajectory for innovation of digital media.

With the internet, an even richer potential is opened, bringing technologies into the center of society, engaging with all institutions and with public life. Thus doors are opened for new path dependencies. This is also the moment for the birth of large-scale internet archiving strategies.

It is of interest to note, that the process of domesticating digital media took a path very different from the domestication of the electronic mass media and print media as the former process started at work and only later entered the home. This story also explain why digital media are not from the very early on considered as "media" on a par with other media, and why there is still much confusion about what are included in the notion of media.

The diversification of digital media does not mean that older notions of the computer become obsolete, but it does mean that they are no longer "universal" or the most general concepts; instead, they are specialized concepts relevant only for certain kinds of usage. While the mainframe perspective reflects centralized systems controlled by professional IT experts, the toolbox perspective is decentered as it puts the IT expert in the background; however, it remains IT-centrist, because it focuses on the relation between man and machine and does not reflect the particular purposes for the professional and civic usages, nor for the broader social and cultural aspects of media and infrastructure and nor for long-term perspectives of data storage.

Given that hypertext and interactivity are fundamental properties of the computer, they were, of course, also properties of the mainframe systems. However, they were only brought into prominence in the wake of the HCI paradigm and reinforced by the spread of internet connections.

The notion hypertext has not been much used since the breakthrough of the internet, but many particular and still growing variety of subcategories, such as navigational links, URLs, menu links, search fields, news-lists, buttons etc. are constitutive for the functioning of the internet. Narrative linking is less constitutive, but still it provides an important trajectory as in remix culture, in gaming and various forms of virtual reality systems.

Within the two main conceptualizations of standalone machines, it is assumed that computerization will spread into a growing range of fields, but none of these theories considers the relation between technology, institutions and the overall

communicational infrastructures of society, and they do not address the disruptive potentials at this level; or, if they do, they address them one-sidedly, as projections of unspecified "IT" into everything else.

These are the issues that are put onto the agenda by the large-scale breakthrough of the internet in the 1990 s, which call for a third type of conceptualization of the computer based on the concept of a medium. The media concept was first applied to computer theory in the late 1970 s, by Kay and Goldberg (1977), who described "personal media" as a meta-medium capable of containing all other media and symbolic forms of expression. The notion of a meta-medium is somewhat misleading. There are always overlaps between different media, which does not make them into meta-media (writing and print for instance). The computer also has a particular set of properties alongside the possible incorporation of the functionalities of older media. It is a medium in its own right. Since it can be dedicated to a variety of physically and semantically different devices and gadgets, the most convenient notion is "digital media", which refers to media in which a part of the functional architecture, as well as the content, is represented and processed by means of sequences of bits.

The basic ingredients in the concept of digital media are not automatization, the stand-alone relation between a machine and a user, or the graphical user interface, but hypertext, interactivity and multimodality (Bolter (1991), Landow (1992), Bolter and Grusin (1998), Benkler (2006), Bruns (2008), Norman and Draper (1986), Bannon and Pylyshyn (1989), Ehn (1989), Bolter (1991)). Furthermore, the transformative potential is not limited to the possible inclusion of all kinds of human communication. It is also based on the fundamental property of computers; that they integrate the storage capacities of print media with the transmission speed of electronic media, conditioned only by the demand that all expressions should be manifested as sequences of bits. The integration of data and processing rules, of machine architecture and materials, which are processed in the same binary alphabet, provide the device with the variable functional architecture as described above. The internet connection add to this finally, that digital media allow physical distances as well as institutional walls to be bypassed, distances to be overcome and institutions to be integrated via digital interfaces and networks. Using these features purposely is what today makes digital media disruptive in the history of institutions for knowledge production as for all sorts of public, semi-public and private communication (Jenkins (2006), Finnemann (2005)).

Thus they incorporate communication and stored information, facilitating all sorts of mixtures and combinations of archived materials and communicational transactions. Communicational transactions always aim to change the previously-stored materials.

The fundamental relation between information and communications, storage and transaction, is intrinsic in all computers and it is directly related to the fundamental role of search processes. To access stored content, you need to perform a search. However, due to the internet, the reach is now dramatically expanded, since search processes can be performed from anywhere. Coded access barriers are now more important than the physical distance between different machines. The search engine thus brings the stored materials, the transport system and a huge array of research instruments formed around advanced elaborations of search engines onto the same physical level. Digital materials are storable, retrievable and exchangeable between interconnected digital media of all kinds (mobile media, dedicated gadgets, the internet

of things) including also online connected microchip implants in our bodies (be it in the form of online connected pace makers, digitized nerve fibers, small scale digital equipment travelling in the veins). They can be accessed anytime, from anywhere, and allow for endless combinations, selections and recombination between any deliberately-defined units within the system. How then, in the future, do we define the limits of which sort of digital materials should be safeguarded by which sorts of institutions? The space is open.

While the distinction between collection and catalogue is based on the separateness of the two corpora (the latter being a small representation of the former unit for unit, or unit to group of units), digitization implies that this one-to-one relation, or this token-to-type relation, is no longer a constitutive condition for accessing the collection. The fundamental integration of storage and transmission can be applied on any scale, and the storage can be accessed with a rapidly-expanding repertoire of elaborated search engines, with an increasing variety of search criteria. Since individual users can access the collections, they may also be able to define their own particular search criteria. Thus, the internet and the search engine now constitute the infrastructure around which the knowledge production and many other forms of social development takes place, regardless of which resources are stored and how they are institutionalized. But digital materials are still also physical materials that have to be stored and to be institutionalized.

Over the years, a number of conceptualizations of the computer have contributed to a development that ranges from numerous concepts of artificial intelligence (stressing the ambitions of establishing automated semantic processes), concepts of the computer as a tool and toolbox (stressing the instrumental and user-oriented perspectives of human computer interaction-studies), to, in recent years, a variety of digital media concepts that emphasize the hypertextual, interactive and multimodal repertoire, software capacities for mediated social communication, and the potential for institutional reconfigurations, which occurs as a result of the breakdown of locality as a precondition for communicative presence. This growth is not simply a growth in the number of digital processes, but rather a process of increasingly varied and specialized program-software, search methods and types of data collections.

New conceptualizations will follow in years to come. It is simply not possible to delimit the potential set of ideas and practices that can be implemented in digital media.

Since the computer itself is void of structure and meaning, its development depends on the ideas and interpretations that are projected into programs, data structures and interfaces. This is what defines the changes of path dependencies, which manifests itself in the history of conceptualizations. In the mainframe era, only a specialized group of experts were in touch with the computer, and the repertoire of ideas projected was small. With the development of the PC and the graphical user interface, the repertoire was dramatically expanded and less predictable. New social groups entered the scene, manifesting their needs and desires. As opposed to the mainframe era, the PC became a game changer, not least because it also facilitated the projection of more conventional ideas and experiences from other fields as manifested in the Graphical User Interface mimicking the traditional office. Finally – so far – the internet opened a wider range of new paths and a more fundamental change of basic infrastructural relations in society, to which all institutions will have to accommodate.

A main trend in this long-term trajectory is the transition from what can be described as the fourth matrix of media, formed around print and electronic media since the mid-nineteenth century, to a fifth matrix, formed around the internet[8]. However, with respect to the libraries, it is a transition from the third matrix directly to the fifth, as the electronic media (phone, radio, television, etc.) did not have a significant impact the structure of the knowledge chain and the role of research libraries. In the library tradition, the main concepts and organizational principles are developed on the basis of stable print media and closed finite works, which, at least in principle, should be classified systematically and with well-defined forms of authorship. The transitional process unfolds itself both gradually and disruptively over many years. Buckland (1992), and Lewis (2004) described a two-step process of digitization, the first being based on "sustaining" technologies, which are introduced to make more effective existing processes and a second being disruptive and game changing, mainly due to the web-based internet. For Lewis, automated services (delivered by either commercial or civic agencies) and self-services are the main threats to established libraries. He also touches upon the potential for distributed operations and new forms for networked collaboration, but his focus is still based on the individual library seen as a paradigmatic unit in a system of distributed institutions acting in parallel.

The argument presented here goes further in two dimensions: First, that the internet, including the web-based internet, will be used in the future to reorganize the institutional landscape due to various forms for collaboration between differently specialized library institutions which will also be part of close collaborations with specialized research milieus. The institutions will be further mediatized, and hence interconnected in new ways. Second, that society will continue to develop digital media both due to known needs, but also due to less known human desires. We may predict that digitization will be expanded, reaching all areas of culture and society, but we do not know much about the particular forms. We may predict for instance that the number of specialized search engines and other forms for analytical software will grow in the future, but we do not know their capacities.

3. The new media landscape and the library

3.1 From acting in parallel to acting in concert

Books and printed materials can now be replaced by electronic texts. The shelves in one library can be replaced with a server in any given location. The catalogues can be replaced by searchable databases and the large buildings, holding thousands of copies of the same books, can be replaced by internet and broadband access. At the same time, a new division of labor is both enabled and required. While the library and university system was based on the carrying out of the same processes in many places, the new communicative infrastructure evolving around the internet allows for a higher degree of coordinated specialization between the locally oriented institutions acted in parallel. There will be path dependency in some respects, but the dependency may switch from one path to another; the existing institutions will be a point of departure for the development of new forms of division of labor. Each institution will take care of a more limited set of more specialized services, which will be provided to a wider range of users.

The library may still have the preservation task and serve as a repository for particular, locally stored corpora, but these corpora will be made globally available via

the internet. The preservation task has already taken on new dimensions as hardware need to be replaced and software need to be recoded to be preserved. Such processes may also modify the stored materials. As a consequence, new issues concerning historical documentation and evidence is raised in the wake of digitization.

Printed materials may still have a role to play. In such a case, the collections will be connected with distributed print and binding facilities, allowing print on demand of any particular – user defined – set of texts, excerpts, images, videos, graphics and diagrams, sounds, etc.

If identified from the point-of-view of the old system of research libraries, this disruption includes a major change in the relation between the libraries, in the relation between libraries and research/scholarship, and in the relation between libraries and other institutions for knowledge production, be they commercial media-corporations, other sorts of commercial knowledge providers or civic enterprises, ranging from the well-known Wikipedia to a variety of civic knowledge providers, which are utilizing the facilities of the internet and search engines.

If the library was formerly identified by its physical location and local affiliations, it will, in the future, have to be identified by the particular set of services it can contribute to the global knowledge community, since it will have to enter into a new kind of collaboration and division of labor.

One model for this is found in the International Internet Preservation Consortium (IIPC), which counts as members most libraries who are responsible for the official internet archiving, ranging from the founding internet archive library in the US, archive.org, to the major national libraries in developed countries.

The case can be seen as paradigmatic.

The IIPC is a relation between institutions for internet archiving. They may enter into a division of tasks concerning crawler software, presentations software, preservation strategies, standardization and development of digital tools for accessing and retrieving information. However, they cannot easily collaborate on the selection and collection of the materials to be archived. Internet archives cannot be complete and selections need to be made.

Internet archives are born digital and the materials are produced by a huge variety of different agencies. The purpose of these archives is to provide documentation of both the historical development of the public parts of the internet itself and– as is presented in a report from a pilot project for the development of the Danish internet archive – of the societal activities which, perhaps, take place exclusively on the internet:

> The overall purpose in establishing an internet archive is thus to ensure the preservation of this contribution to the cultural heritage and thereby the source materials that will provide the foundation for future research, not only into the internet's own history, but also into all the ever more comprehensive cultural, institutional and business activities that take place especially – and in some cases exclusively – on the internet or in close connection with it (Christensen-Dalsgaard *et al.*, 2003).

As a consequence of this purpose, such collections will have to reflect the heterogeneous character of the activities.

At present the main strategies for web-archiving (Christensen-Dalsgaard *et al.* (2003)) are:

- snapshots of a domain or a set of domains or even of all domains;
- selective archiving based on a limited set of predefined websites, based on either the criteria of value or, possibly, the criteria of updating frequencies to compensate for lagoons in the snapshot; and.
- event-based harvesting of a particular event, which may generate new web pages or content in unexpected places.

These strategies are often combined. The selection strategy may use a variety of criteria. Netarkivet.dk in Denmark for instance selects sites which are frequently updated and non-cumulative, and partly towards experimental sites, which are innovative with respect to, for example, aesthetics or functionalities.

There will be a need for editorial practices that take into account the role of the internet in the particular society in question. In this respect there is still a need for "locally defined" nodes in the network.

Neither the archiving institution nor the researchers are able to define any overarching systematic principle for these kinds of collections. Furthermore, internet materials are produced in such vast quantities, which will always exceed the amount archived. Internet archives are ruins, storing only fractional parts and remains. Thus, the archiving institutions need to elaborate procedures for identifying incompleteness as part of their selection strategies. They need to document what sorts of data are missing and for what reason, at any given time.

Since internet materials are often modified, changed, supplemented or even deleted, they need to be archived on the fly, and as regularly as possible. Time is passing and materials are can change dramatically during, for example, the collection of a national domain.

There is also a need to develop a variety of search methods and formats for the visualization of the, otherwise invisible, materials. This is the case for all sorts of digital materials. Due to the heterogeneous character of internet archives, it is most likely that the retrieval and presentation of internet materials calls for a set of software methods which cannot automatically be used directly online or in accessing other kinds of digital corpora.

The IIPC is a meta-organization formed by traditional library institutions, but it might turn out that such institutions are not capable to provide the solutions concerning basic standards for long term preservation of machines, of programming software and of primary source materials stored in the archive as well as the continuing development of interfaces and support for analytical software necessary for the study of the materials. Meta-organizations are maybe only transitional stages into a new and more fully developed model based on the collaboration between specialized sub-institutions to ensure for instance that the various national archives can be integrated even if they are created due to different "national" criteria. Scholars interested in the internet archives, as for instance media scholars, may also want integration with other sorts of archives, such as digitized radio and television and print media archives.

Thus, research libraries are gradually incorporated into a number of new "research infrastructures" which are being built around different kinds of data materials as international structures – in Europe primarily within the EU and thus also serving as a kind of Union building projects (European Science Foundation (ESF) (2011)).

Each research library may specialize according to some sort of coordinated criteria, such as, e.g. the type of collection, the particular sets of search facilities maintained.

They will have to identify which sort of data materials they will preserve, thus, also adopting a function as curators in the face of the vast array of possible sources and repositories made accessible via the internet.

The book library held both research publications and printed source materials for disciplines, which rely on written and printed sources. As these materials are digitized, they are brought into the same universe as many other digital resources, documenting and making human knowledge available.

3.2 A new relation between the research library and university

Science, research and scholarship is no longer unfolding within a relatively self-regulating, meritocratic system, based on expertise and responsible for more-or-less well-defined contributions to various relatively-clearly separated sectors in society, be it the health sector, the juridical sector, the economy, the manipulation of natural resources, or the maintenance of high level education, etc. These formerly relatively separate activities are, today, more closely integrated and mixed with political, commercial, social and cultural projects of all kinds, and often directed towards more clearly-specified political goals. Academia still has a little word to say in the validation of knowledge, but other criteria are often included.

The modern functional differentiation of "autopoietic" social systems, described by Niklas Luhmann, among others, has been replaced by tangled conglomerates, trying to incorporate heterogeneous goals of a variety of political, economical, social and cultural subsystems.

This is also the context for a changing relation between science/research/scholarship and librarian institutions, which is now transformed into a particular institutional system of "research infrastructures" and electronic resources, connecting researchers and librarians in developing bridging platforms for search, retrieval and analysis of digitized collections and archives, as well as an increasing amount of born digital sources from all areas of society.

The relations between research projects and libraries need to be institutionalized in order to take care of a continuous tailoring of software supported research methods and archive interfaces, according to the (changing) research questions posed, the development of new analytical software, and the technical constraints of the particular collection in question. Tremendous input will have to be made to standardization purposes across the globe.

Digital materials also call for new sorts of preservation efforts, which form a major anchor point for the maintenance of research infrastructures. For each research project, there will be a set of analytical tools, as well as a selected set of data that need to be taken care of or to be abandoned as reproducible resources.

There are tensions in this. While the archiving institutions are obliged to maintain a long-term perspective for preservation and a broad perspective for giving access to many different research groups, researchers are obliged to pursue particular research questions and elaborate their repertoire of software-supported methods. There are also tensions between top down initiatives, which are related to the rational organization of huge data corpora and monitored increasingly on a national or regional scale (e.g. EU initiatives), and bottom up processes of particular researchers and research projects. It is probably no coincidence that the huge datasets of all kinds – so-called big data – appeal to research projects focusing on structural patterns of those corpora.

Conclusions

Digital media are, on the one hand, developed as a means to overcome important limitations of the paper and print-based knowledge production systems, but they also allow for far-reaching transformations, well beyond the existing institutional setup up. They have a disruptive potential, which seems to unfold more clearly as the technology comes of age. The disruptive potential is revolutionary in its logical nature, but only unfolds itself slowly, according to particular needs and desires – as it is the case for those kinds of evolutionary processes which lead into more complex systems, due to the emergence of new, mediated layers built into and upon the existing layers.

Path dependency is often a modifying component in development. However, in the case of digital media the dependency is often moved from one path to another, as was the case for instance with the spread of the PC at work and later again in the homes as well as with the breakthrough of e-mail and web-based internet, and again the breakthrough of social media which opened the web based internet for small talk and other forms of – mostly written – social communication. Path dependency continues to take on disruptive forms, be it the development of new semantic interpretations, usages of computers in new areas, or the development of institutional transformations.

The disruptive potential, which can be associated with digital media, is seldom visible in the rhetoric of IT avant-gardism, but more often in the virtually unforeseen unfolding of new trajectories, which are often only visible in retrospect. Digital media become disruptive due to the strength of the historical dynamic, rather than as a result of particular agencies. Like the IT-development, the knowledge production system is transformed due to forces beyond the reach of any single agency or institution or medium. The IT-developments and the institutional transformations are entangled, and each of them is double-edged. Their developments are, to a large extent, driven by the exponential growth of knowledge production, which they serve to reinforce, while, at the same, they offer new means to handle the dilemmas.

Generally speaking, scientists and scholars now have individual access to libraries all over the world and to a vast array of resources outside the institutional system of libraries.

The fact that things are possible does not mean that they will happen, or tell us how far and fast they will proceed, but it does mean that all institutions concerned with knowledge handling will have to redefine themselves and identify their *raison d'être* according to the new infrastructure, which is emerging around the internet and increasingly extending into all areas, due to mobile media and a growing variety of specialized digital gadgets. This process will also include a need to reconsider all the core notions of the institution, the profession and the division of labor among researchers and librarians.

As for individuals, the identity of institutions will increasingly depend on the constellation of networks in which the institutions participates, even if the identity given by path dependencies continues to play a significant role.

Notes

1. The exponential growth of knowledge production was first documented in De Solla Price (1961) and De Solla Price (1963). For a recent quantitative account, including digital materials (see Lyman and Varian, 2000).

2. The development of digital media is also formed by other megatrends, such as the demands for interactive, mediated entertainment and social communication, due to the lifestyle of the modern, educated middle classes. Knowledge production and entertainment are both influential long-term drivers in the development of modern society, but have often been kept apart. They meet each other today in the many efforts to make learning more playful. Although it could be argued that this is a perspective that should be included in future strategies, it is, unfortunately, not possible to discuss this further in this article.
3. The plasticity and malleable character of computers were only fully recognized in the late 1980 s, e.g. by Ehn (1989), Bolter (1991) and others.
4. The implications of digital hypertext and interactivity for authorship has been discussed among others by Bolter (1991), Landow (1992), Bolter and Grusin (1998), Benkler (2006), Bruns (2008) and a number of concepts each referring to some aspects have been suggested (the writerly reader, co-writing, produsage, user-generated content, remix culture).
5. There are a variety of interpretations of the internet as new communicational infrastructure. For a brief overview of theories of replacement, supplementation, enhancement, evolution, and co-evolution of media see Finnemann (2006). Schultz (2004), Krotz (2009) focus on mediatization of society, Bakardjieva (2005) and Berker *et al.* (2007) focuses on domestication of new media.
6. See, e.g. Metropolis *et al.* (1980) and Williams (1985), for a history of main frame computing. The literature is increasingly divided between history of the machinery, devices and protocols, the history of network developments, and studies of particular practices in administration, business, librarian institutions, in social communication, impact on politics and the mass media, to the study of software genres and content.
7. Dirty programming utilizes the "choice machine" by utilizing random access to the storage, which, again, is the point of departure for more elaborate developments of hypertextual link-modes and more elaborate formats for interactivity.
8. The three former epochs were centred on oral communication, writing and print respectively (cf. Finnemann, 2011).

References

Bakadjieva, M. (2005), *Internet Society: The Internet in Everyday Life*, Sage, London.

Bannon, L. and Pylyshyn, Z. (eds) (1989), *Perspectives on the Computer Revolution*, 2nd ed., Ablex, Norwood, NJ.

Benkler, Y. (2006), *The Wealth of Networks. How Social Production Transforms Markets and Freedom*, Yale University Press, New Haven, CT.

Berker, T., Hartmann, M., Punie, Y. and Ward, K. (eds) (2007), *Domestication of Media and Technology*, Open University Press, Maidenhead.

Bolter, J.D. (1991), *Writing Space. The Computer, Hypertext, and the History of Writing*, Lawrence Earlbaum, Mahwah, NJ.

Bolter, J.D. and Grusin, R. (1998), *Remediation: Understanding New Media*, MIT Press, Cambridge, MA.

Bruns, A. (2008), *Blogs, Wikipedia, Second Life and Beyond: From Production to Produsage*, Peter Lang, New York, NY.

Buckland, M. (1992), *Redesigning Library Service: A Manifesto*, American Library Association, Chicago, IL.

Christensen, C.M. (1997), *The Innovator's Dilemma: When New Technologies Cause Great Firms to Fail*, Harvard Business School Press, Boston, MA.

Christensen-Dalsgaard, B., Fønss-Jørgensen, E., von Hielmcrone, H., Finnemann, N.O., Brügger, N., Henriksen, B. and Vejrup Carlsen, S. (2003), "Experiences and conclusions from a pilot study: web archiving of the District and County Elections 2001", Final Report for The Pilot Project "netarkivet.dk", available at: http://netarchive.dk/publikationer/webark-final-rapport-2003.pdf

De Solla Price, D.J. (1961), *Science Since Babylon*, Yale University Press, New Haven, CT.

De Solla Price, D.J. (1963), *Little Science, Big Science*, Columbia University Press, New York, NY.

Ehn, P. (1989), *Work-oriented Design of Computer Artifacts*, Arbeitslivcentrum, Stockholm.

European Science Foundation (ESF) (2011), *Research Infrastructures in the Digital Humanities*, European Science Foundation, Strasbourg.

Finnemann, N.O. (2005), "The cultural grammar of the internet", in Jensen, K.B. (Ed.), *Interface://Culture – The World Wide Web as Political Resource and Aesthetic Form*, Samfundslitteratur/Nordicom, Copenhagen, pp. 52-71.

Finnemann, N.O. (2006), "Public space and the co-evolution of digital and digitized media", in *Media and Politics*, Joint Edition of Tidsskriftet Politik, Vol. 9 No. 2, and Mediekultur No. 40, Copenhagen.

Finnemann, N.O. (2011), "Mediatization theory and digital media", *Communications*, Vol. 36, pp. 67-89.

Jenkins, H. (2006), *Convergence Culture – Where Old and New Media Collide*, New York University Press, New York, NY.

Kay, A. and Goldberg, A. (1977), "Personal dynamic media", *Computer*, Vol. 10 No. 3, pp. 31-41.

Krotz, F. (2009), "Mediatization: a concept with which to grasp media and societal change", in Lundby, K. (Ed.), *Mediatization: Concept, Changes, Consequences*, Peter Lang, New York, NY, pp. 21-31.

Landow, G.P. (1992), *Hypertext: The Convergence of Contemporary Critical Theory and Technology*, The Johns Hopkins University Press, Baltimore, MD.

Lewis, D.W. (2004), "The innovators dilemma: disruptive change and academic libraries", *Library Administration & Management*, Vol. 18 No. 2, pp. 68-74.

Lyman, P. and Varian, H.V. (2000), *How Much Information*, available at: www.sims.berkeley.edu/how-much-info 06.13 2012.

Metropolis, N., Howlett, J. and Rota, G.C. (1980), *A History of Computing in the Twentieth Century*, Academic Press, New York, NY.

Norman, D. and Draper, S. (eds) (1986), *Centered System Design: New Perspectives on Human-Computer Interaction*, Lawrence Erlbaum Associates, Hillsdale, NJ.

Schultz, W. (2004), "Reconstructing mediatization as an analytical concept", *European Journal of Communication*, Vol. 19 No. 1, pp. 87-101.

Turing, A.M. (1938), "On computable numbers, with an application to the entscheidungsproblem: a correction", *Proceedings of the London Mathematical Society*, Vol. 43, pp. 544-546.

Turing, A.M. (1950), "Computing machinery and intelligence", *Mind 4*, Vol. 59 No. 236, pp. 433-460.

Williams, M.R. (1985), *A History of Computing Technology*, Prentice Hall, London.

Further reading

Association of College and Research Libraries (ACRL) (2010), "Value of academic libraries: a comprehensive research review and report", researched by Megan Oakleaf, Association of College and Research Libraries, Chicago, IL, available at: www.acrl.ala.org/value

Brügger, N. (2005), *Archiving Websites. General Considerations and Strategies*, The Centre for Internet Research, Aarhus.

Chomsky, N. (1957), *Syntactic Structures*, Mouton, Gravenhage.

Jensen, J.F. (1996), *Babel og Tomrum*, Gyldendal, Copenhagen.

Lundby, K. (ed) (2009), *Mediatization, Concepts, Changes, Consequences*, Peter Lang, New York, NY.

Turing, A.M. (1936), "On computable numbers, with an application to the entscheidungsproblem", *Proceedings of the London Mathematical Society*, Vol. 2 No. 42, pp. 230-265.

Corresponding author
Niels Ole Finnemann can be contacted at: finnemann@imv.au.dk