Journal of Digital Information, Vol 6, No 1 (2005)

Digital Text Cycles: From Medieval Manuscripts to Modern Markup

Terje Hillesund

University of Stavanger, Stavanger, Norway

Email: terje.hillesund@uis.no

Abstract

The paper argues that the current implementation of digital publishing is a minor step in a long development of digital text cycles. Rather than being a revolution, the digital transformation of text is an evolutionary process heavily influenced by social and cultural factors. The paper introduces the concept of a "text cycle". An examination of basic features of paper-based text cycles and features of digital text cycles demonstrates that digital technology has a potential for change that far exceeds that of the "Gutenberg revolution". However, by applying a historical perspective, I will try to show how the deep and enduring cultural heritage of print is impeding the radical potential of digital texts.

1 Introduction

Publishers around the world are implementing new production workflows based on structural markup and cross-media publication. The shift towards eXensible Markup Language (XML) and multi-use of text content is reflected in a new term: *digital publishing*.

In retrospect, it now seems obvious that *electronic publishing* was a digitalization process mainly based on principles of print. By contrast, *digital publishing* represents an effort to use basic digital principles in a more flexible way of producing and distributing verbal texts. By introducing digital publishing, writers and publishers want to enable their texts to be distributed in print, as e-books or e-journals, on the Web, through e-learning applications or in digital encyclopaedias.

In this paper I will introduce the concept of a *text cycle* in order to analyze this new stage in the digital transformation of text. A text cycle, as defined in the paper, consists of several interrelated elements or *phases*, such as *writing*, *distribution* and *reading*. I will examine how digital technology affects the phases of the text cycle. I will further show how digitalization of the phases establishes a range of qualitatively new text cycles: *digital text cycles*. I will pay particular attention to the text cycle of e-books. I will also compare digital text cycles with *written* and *printed text cycles*. This comparison will illustrate the fundamental characteristics of digital text cycles, as well as those of writing and print.

Application of the text cycle concept will provide a basis for discussion of whether digital publishing is as revolutionary as is often claimed. Kasdorf (2003) uses the word "revolution" repeatedly. But is digital publishing really a revolution, or is it a mere evolutionary step in the more than 5000 year long process of technologizing the word?

2 Text cycle concept

2.1 Text

In linguistics the term *text* is used for both written and spoken texts, but is sometimes used to refer to written texts only. In publishing, illustrations and photographic reproductions are often regarded as parts of the overall text, and in semiotics the term "text" covers all kinds of representations: speech, pictures, music, videos and computer games. In this paper, I will deal with texts produced and represented in written forms.

A *text* is usually regarded as a product, rather than a process; the text is the product of a process of text production. In this sense, the text has a physical existence of its own, independent of its sender and receiver (O'Sullivan 1983).

But the word *text* can also be used in an abstract sense, signifying the verbal structure, or wordings, underlying the physical representations of the text. Taken in this sense, a text can be given different representations, as when an ancient manuscript is reproduced in modern print, on the Web or as an e-book. It is still the same text. It is the latter meaning of the word we use, when -- in copyright terms -- we speak of an originator's "ownership" of a text. It is obviously not the text in its many physical representations (the products) that the originator owns: it is rather the wording or the structure of words (the abstract text) that constitutes the intellectual property (Asbj¬ørnsen 2002).

The two interpretations of the word *text* are not necessarily contradictory. They rather describe interrelated aspects of the rich phenomenon "text". A definition can encompass both meanings: a text is a visual representation of verbal information.

2.2 Text cycle

In addition to being regarded as both a product and a structure, a text is also part of several processes such as writing, distributing and reading. The text is part of a *text cycle* (Hillesund 2002b).

I prefer the term *text cycle* to *text system* in describing the phases involved in the overall text process because the latter usually focuses on the text and its internal structures. It covers the signs and codes and the syntactical and semantic relationships -- the elements and structures -- that constitute the meaning of a text. Such elements and structures are essential for a text to be able to represent verbal information; but my choice of the term "text cycle" is intended to indicate that my focus is elsewhere.

The term *text cycle* covers text production, circulation processes and dialogical processes. A text is very often a response to other texts in an ongoing cyclic movement. This dialogic aspect of texts is referred to by King *et al.* (1981), referred to in Tenopir and King (2000), in their model: "The life cycle of scientific information through scholarly journal system functions". In dialectic processes scientists read articles, do research and write their own scientific articles. The journal article life cycle model shows the functions and different roles of authors (researchers), publishers, libraries and readers (researchers) in the creation and spread of scientific information.

The life cycle metaphor is obviously useful in the context of scholarly articles. However, in the context of this paper, it is the "cyclic" part of the metaphor that is most interesting. The model of a text cycle is more abstract than a model of a scientific information cycle (Tenopir and King 2004). The general model draws attention to the fact that all texts -- books, pamphlets, letters, emails *and* scientific articles -- are parts of a text cycle.

A text cycle, I will argue, consists of the basic *phases* of *writing*, *producing*, *storing*, *representing*, *distributing* and *reading*. The characteristics of these phases are very different and they cannot be defined by a simple set of criteria. There is a substantial difference between the physical process of storing a text and the mental act of reading. At an empirical level the characteristics of the same elements will also differ, depending on the technology used in the specific text cycle. The processes involved in writing depend on whether you use a chisel, a quill, a typewriter or a computer keyboard. Furthermore, in handwriting, the phases of writing and producing a text are executed in one process, whereas in print, writing and production involve several processes. Despite these differences, a text must in some way be written, produced and distributed, before it can be read.

My use of the generic term *phase*, to refer to elements of the text cycle, is intended to imply some kind of succession. In many text cycles, one phase is clearly followed by another. A scientific article is written and produced before it is distributed and read. In other cases, the sequential order of the phases is less obvious. When you write on a blackboard using chalk, you write, represent, and store your words at the same time. This is all done in one simultaneous operation. By applying chalk to the board, your text is instantly made readable to your audience (who hopefully read the text as you write). In this case the phases are interwoven and interrelated. They overlap and occur simultaneously. At the same time one process makes up several phases of the text cycle. In spite of this apparent inconsistency, I will use the concept *phase* when discussing elements of the text cycle. Even when phases overlap, they nevertheless have a logical order. A text must be written before it can be read.

In the balance of this paper, I will use the concept of a text cycle and its phases in examining the digital text cycles that are currently emerging in publishing, or in digital publishing. In these new text cycles fewer principles are inherited from preceding printed text cycles and more principles are founded on genuine digital technology. This presents an interesting new situation in the technologizing of the word (Ong 1982).

3 Written text cycles

To draw attention to similarities and differences between traditional text cycles and digital text cycles, it will be necessary to describe the basic features of written and printed text cycles.

Stated simply, we may say that the *text cycle* of a *written* language consists of phases or elements of *writing, storing* and *reading.* When you pick up a pen and write on a piece of paper, you simultaneously store the text and make it readable for yourself and others.

What is needed to accomplish this is a pen, ink, a sheet of paper and a visual sign system capable of *representing* verbal information. The essence of the whole operation is that, by these means, writers are able to physically store verbal information outside the body. Once written and stored, the text can be read -- or it can be transported and read far from the location where it was written.

In this simple and ingenious way written communication exceeds serious limitations of oral communication. Since the voice has a short range and the sound of the voice quickly fades away, the participants of oral communication must be in the same place at the same time to communicate. To be passed on to others the oral message must be stored inside the body of the messenger, in his memory or brain.

Written texts are far more durable, portable, and exact than oral texts. By writing and storing a text on a portable medium, such as paper, parchment or clay, the exact text can be carried to distant places. Written communication thus overcomes both the time and space restrictions inherent in oral communication. A written text cycle including *distribution* represents a very effective system of communication (Innis 1972).

Written text cycles move verbal communication from the realm of voice and ear to the realm of hands and eye. In writing, verbal communication becomes essentially visual. This is obviously the key to its success. Ever since humans started to draw or carve visual patterns signifying verbal entities on to the surface of suitable materials, powerful combined applications for both *storing* and *representing* verbal information have been developed and utilized.

This has been the case since 5000 years ago, when the Sumerians started to press a stylus into clay forming their cuneiform characters in the first written language. Since then, writing and reading systems have evolved and flourished all over the world, making use of all kinds of materials, tools and symbols, such as stone, paper, knives, typing machines, hieroglyphics, Chinese signs and alphabetic characters. The development of the art of writing is one of mankind's greatest achievements, and as Ong (1982) and many others have observed, wherever writing was introduced it had a major impact on culture and society.

Common to the various types of writing systems is that, though different, the very same physical means are used in all phases of the text cycle. When, in our example, you *write* a note, you create a visual pattern of marks on the surface of the paper, using ink. Thus you *produce* a text. Marks applied to paper represent the *storage* system of the text. The set of ink-marks is the visual *representation* of the stored text. The paper marked with ink can be carried around or *distributed*, and the same marks of ink (on the paper) can finally be seen and interpreted by the *reader* of the note. Ink on paper is the technology

throughout this text cycle, as are carvings on wood and stylus marks in clay in other cycles. It is especially important to note that, in written text cycles, the same medium is used to store and represent verbal information.

4 Printed text cycles

A great leap in the development of the text cycle was Gutenberg's invention of loose type and his improved use of the printing press (in the middle of the 15th century). The most significant aspect of this invention was the separation of *writing* and *production* into two different processes or phases.

In the medieval manuscript tradition, the production of books was a time-consuming and labor-intensive activity. Every copy of a book had to be written by hand. For centuries after the start of the print area, manuscripts continued to be written by hand. This process amounted to a text cycle on its own. However, this manuscript was only the starting point of a new process, the printing of books. In this process the manuscript was edited and marked, and the building blocks of the text -- the characters or types -- were set or composed in a case. By use of a press and ink the types were printed on sheets of paper, which were then cut, sawn and bound. Through these operations the text in a single handwritten manuscript was reproduced in many printed books.

Gutenberg's invention introduced a new and highly efficient text duplicating process. The process in the handwritten text cycle, simply called "writing", is split into two processes in the printed text cycle: writing and printing, or, in our terminology, writing and *production*. The printed text cycle thus consists of the phases or elements of *writing, production, storing, representation, distribution* and *reading*.

Widespread use of the printing press led to an enormous increase in production and distribution of books in Europe. Eisenstein (1983) and others have convincingly documented that the art of printing and the spread of printed books (and journals and newspapers) had a profound influence on European history and Western culture.

Diffusion of print technology has deeply influenced communication structures in Europe. The production of printed texts was organised in new (capitalistic) ways. New distribution channels emerged, and new markets evolved for all kinds of printed matter. Literacy spread, and reading became a central part of modern life. Over the centuries improvements in typography and print technology made print production steadily more efficient. The printed book, in particular, has been refined and developed into a highly sophisticated reading technology.

In spite of the many improvements, the fundamentals of book production have been remarkably stable and so have the features of printed books. A comparison of medieval manuscripts and modern printed books reveals striking similarities, which suggest that handwritten and printed text cycles are fundamentally rather similar and that the difference between the two technologies is mostly a matter of productivity.

If we compare the phases of the two cycles, the *writing* is more or less the same in both: with some tool -- a quill, a ball pen or a typewriter -- ink is applied in some pattern directly to the surface of the paper, thus creating a text. In a written text cycle *writing* does the job. When written, the message or text is *stored.* It can then be *distributed* and *read*.

In printed text cycles *writing* (constituting a text cycle in itself) is the starting point of quite a different process, the *production* of texts, as we have seen. The result of this process, the printed paper, is very similar to a handwritten piece of paper. Print, too, is a process of applying ink in certain patterns to the surface of paper. Through this process the text is *represented* and *stored* and in this physical shape, as ink on paper, it is *distributed* and *read*.

The explanation of the similarities between writing and printing is this: writing and printing use the same method of representing and storing verbal information. In both cases visual patterns representing verbal information are made on the surface of suitable, portable material. Through this process the text is fixed and stored in a combined representation and storage medium.

5 Digital text cycles

5.1 Separation of storage and representation

Digitizing of the word and text was the next great leap in the development of the text cycle. The most significant aspect of the digital shift is the separation of *storage* and *representation* in two different phases. This separation has had important consequences for all parts of the text cycle.

This section compares the fundamentals of digital text cycles with those of paper-based cycles (written and printed text cycles). In digital cycles, texts are produced, distributed and read with the aid of computers, networks and monitors in a predominantly digital environment. The transition from paper-based to digital text started in the second half of the 20th century, and is still in progress.

5.2 Digital writing

Let us start by comparing *typewriting* and *digital writing* (writing using a computer system and a text editing application). Both ways of writing constitute complete text cycles and, at first glance typewriting and digital writing seem to be very similar: by striking keys on a keyboard, visual representations of the text are created either on paper or on screen. These similarities are superficial.

Typewriting (like most handwriting) is a process of applying ink to paper and making the text readable -- in more or less one operation (by pressing down keys, letters are punched on to the paper). It probably makes sense to say that you must write before you can read, but clearly the *storing* and making a *representation* of the text are performed in one inseparable operation; by applying ink on paper.

In digital writing, by comparison, writing is performed by the execution of a series of discrete steps. By touching the key of a

keyboard, signals are sent to the computer. Here the signaled information is converted and handled by the central processing unit and temporarily *stored* in main memory. In the computer, new signals are created and transported to the display unit. On screen the text is *represented* in a visual, readable way. In this digital cycle *storing* and making a *representation* of the text are performed in two different operations.

The important and extremely consequential point of the digital shift is that the storing of the text is separated from its visual representation. In a digital text cycle, units of the text are stored as bit patterns (usually in the codes of ASCII or Unicode) in a file. When writing is finished, the whole text is stored as a collection of bit patterns in a mass storage system, often a hard disk, completely separate from any representation of the text.

This way of storing a text is totally different from storing in a written text cycle, where the stored text is physically tied to the surface of some material, often paper. The digitally stored text can easily be fetched from memory and loaded back into the computer. In the digital form, the text can be reedited, rearranged and improved.

The text manipulative powers of computers were utilized in various word processing programs early on in the digital age. These applications rapidly improved as monitors and graphical user interfaces improved, and with the introduction of the mouse. Indeed, one of the main reasons computers supplanted typewriters in homes and offices in the 1980s was that they lent themselves to text editing. Digital writing soon became the dominant way of writing.

Word processing applications as we know them, however, have primarily been designed to facilitate paper text production. Computers are connected to printers, and word processing programs make it possible to edit the typography and the layout of texts, as well as the content. The design of printed pages has set the standards for the layout principles used in most text editors. Thus digital text-files often contain a mix of information on the units of the text (letters, numbers, punctuation marks) and the printing of the text (fonts, sizes and layout), the latter in form of presentational markup. Even when digital text files are separately stored, they carry information on their own representation in print.

As soon as an edited text is printed, it leaves the digital text cycle. Printed on paper the text is physically fixed and stored -- and in this shape (as ink on paper) it is physically carried around and distributed, before it is read. Printed, the (digital) text enters a paper-based text cycle.

5.3 Digital production

Since the 1980s, mass print productions have changed dramatically. The long lasting lead typesetting is now obsolete. Offset, desktop publishing and other digital technologies have completely transformed printing (Kasdorf 2003).

In this process most of the established graphical and typographical practice and conventions have been built into digital applications. Type handling, the making of graphics and illustrations, photo editing, make-up of pages and prepress preparations are all done in software programs. Digital texts, along with all the other elements of print publications are treated in sophisticated desktop publishing applications, resulting in files containing a close weave of information on content and presentation.

Digitalization has made print production far more efficient (and has lead to many crises in the graphical and printing industry). In the process print technology has improved. Huge quantities of highly readable and beautifully designed and illustrated books and magazines now flood the markets. The result of this development is that printed text cycles are dominated by two fundamentally different basic principles. In the writing and production phases digital technologies dominate. In these phases *separate* storage and representation techniques are used. After they have been printed, publications are distributed and read in paper-based text cycles, using *combined* storage and representation techniques. Print production and distribution have one foot in a digital text cycle and the other in a traditional paper-based text cycle.

In spite of digitalization, organization of the phases of the printed text cycle is more or less the same as before the digital shift. If we look at the book cycle: authors still write, publishers still edit, printers still print, and the book is distributed by bookstores, book clubs and libraries. In the end, we read a book in much the same way as we always have, by turning the pages.

Writing and production has been digitized -- and everything is still the same, apparently. However, what was developed as utility tools for written and printed text cycles was also the development of a qualitatively new and independent kind of text cycle: the digital text cycle.

After being printed, a digitally produced text file has obviously not disappeared. On the contrary, it is still digitally stored in the memory of the computer system. From there it can be transported and represented over and over again; on our own computer screen, in print and, more importantly, via networks the text can be represented on any other computer screen.

5.4 Digital distribution

As we have seen, digitalization has split storage and representation into two separate phases of the text cycle. The most important aspect of this separation of storage and representation is the way it has changed the *distribution* phase of the cycle. In much the same way as stored bit patterns can be transported within computers, digital texts can be transported between computers over networks, of which the Internet is obviously the most important. Compared to the distribution of physical books or journals, this capability of digital texts has completely changed text distribution.

The history of computer networks is familiar to most computer users. It started in the United States in the 1960s when computers at some universities were connected, making possible an exchange of digital information. Different local and wide area networks were established. In the mid-1970s the Internet was launched as a collection of protocols regulating transportation of signals and information between different networks and computer systems. The Internet was soon made international, and in the 1980s the Net grew rapidly (Leiner *et al.* 2003). Users formed discussion groups, exchanged electronic mail and established systems for downloading digital documents. The Internet thus formed a completely new way of

exchanging information, vastly increasing the *distribution* potential of digital texts.

The greatest expansion of the Internet came after the introduction of a globally unique identifier to digital information (URL), the development of a hypertext markup language (HTML) and the compilation of a hypertext transmission protocol (HTTP), forming the World Wide Web in the early 1990s (Berner-Lee and Fischetti 1999). In this distribution system documents are linked in a hyper-textual fashion forming an intricate web of digital documents. Once linked up and placed on the Web, a digital text can in principle be accessed by any computer connected to the Internet. In our terminology we would say that the Internet and the Web make up the distribution infrastructure of a global digital text cycle.

Digital network communication is not restricted, most of the time, by space and material limitations inherent in written and printed communication. In written and printed text cycles the physical object containing the text (the book or magazine) must be carried around using different means of transport (airplanes, cars or the human body). In digital text cycles the stored bit patterns of digital texts are converted into electromagnetic pulses (in the telephone system), light signals (in fibre optics) or radio waves (in wireless networks) and transported around the world in seconds.

When a printed piece of paper or book is taken from the place where it is stored, it is removed in its entirety. A downloaded digital text is retained in the memory of the home computer, unchanged. From the storage systems of computers digital texts can be copied (or cloned), over and over again, without loss of data or quality. From new positions in the network they can be duplicated and repeatedly redistributed. It is obvious that a digital text cycle including *network distribution* represents an extremely effective system of communication.

5.5 Digital reading

The point of writing is reading. By the act of reading we receive messages, gain information and enjoy stories. Reading is the final goal of writing and constitutes the essence of all text cycles.

The term "digital reading" refers to reading on display units connected to computer systems, such as desktop computer screens and screens on handheld computers (Hillesund 2002b). After monitors and keyboards became the user interface between computers and humans, it has been possible to perform all phases of the text cycle on a computer system. In a digital text cycle words of texts are visually represented on screens. Computers enable us to produce, store, represent and, finally, *read* a variety of texts.

Digital reading is immensely widespread, being part of most administrative and communicative activities in modern society. The spread of networks placed digital reading at the center of a number of communication forms, such as email, discussion groups, chat, e-learning, electronic publishing and the exchange of all kinds of digital documents.

In spite of this, computer screens fall short when it comes to sustained reading, which is reading of longer texts such as journal articles and books. Even for longer emails and Web articles most people tend to find the print button before they read the text. The computer and the computer screen cannot yet compete with printed paper as a medium for reading (Hill 2001).

This weakness of computer screens is related both to hardware and software issues. The heavy and stationary screens of desktop and laptop computers give static and tiring reading positions. Low resolution and poor type representation causes eye strain. In addition most applications and text formats are designed for the *production* and *distribution* of texts (word processors, Web browsers) with little or no concern for the typography of *screen reading*. It is probably true to say that lengthy digital texts are usually made to be printed.

While computer screens are legible, ordinary screens and applications do not have the readability required for sustained reading. As long as, for longer texts, the reading phase is dominated by printed paper, we cannot yet speak of a fully developed digital text cycle. It is rather the case that most of what is called electronic publishing still has one foot in the printed text cycle. The text is digitally written, stored and distributed, but when the text reaches its destination it is read in a printed version.

Nevertheless, the development of digital reading technologies has been slow and steady. In the near future, LCD screens of handheld computers seem to be the likeliest platform for sustained digital reading, whereas electronic ink and electronic paper may catch up at a later point. Further improvements with respect to weight, cost and resolution of wirelessly connected handheld computer screens are likely to make these a viable alternative medium for sustained reading. This especially is the case after leading software manufacturers (Palm, Adobe, Microsoft) have developed specialized reading applications (e-book readers) for sustained screen reading. Thus, despite the fact that sustained digital reading has caught on rather slowly, the situation may change rapidly once it has started.

The reason for this is apparent when we look at what has happened to printed paper. Printed books (and journals) have over the centuries been developed into highly sophisticated reading technologies, but even if paper-based text cycles still dominate publishing, the role and function of paper has declined considerably.

It is the reading phase of the text cycle that keeps paper going. In principle, we no longer need paper for writing, storing or distributing texts. In print production all pre-press processes are digital. This reduced role of paper makes it vulnerable. When, as is likely, digital reading technologies achieve a competitive level of performance, they will oust printed paper and pure digital text cycles will become the norm.

At this point, some of the advantages of digital texts, such as e-books, will become more obvious. Most e-books have re-flow capabilities, which means that the layout of the pages adjusts to different font sizes and to varying screen sizes. This is generally useful, but especially so to the visually impaired. Most e-book applications can be set to read a text aloud. Furthermore, e-books are easy to store and easy to carry in numbers. They are easy to search, annotate and bookmark. E-books will also soon have multimedia functionality and e-books will have hypertext facilities with the possibility of intelligent linking to all kinds of internal and external resources, such as dictionaries and encyclopaedias. E-books, e-journals and other digital texts will be available whenever one is connected to a network. Along with improved readability these advantages will make digital reading, also of journals and books, competitive in a growing number of situations.

5.6 Digital publishing

Digital reading and new distribution channels represent considerable challenges for publishers. They have digitized their print production workflow and once again they have to redesign their workflow to satisfy both print and digital text cycles. By implementing new production patterns publishers are struggling to face a new world of cross-media publishing.

Cross-media publishing is made possible by the digital division of storage and representation of texts. A consequence of the division is that a digitally stored text can be represented in various media, such as paper, screens (of different sizes), Braille or read aloud by automatic reading applications.

This is not as straightforward as it may sound. Printed media and digital media have different presentational principles, which profoundly affect production workflows of print and digital media respectively. At the present time, these workflows are not compatible.

In print the *page* is the composing unit of visual presentation. Once composed the page is fixed. Not so in screen rendering and digital text cycles, which are governed by a *re-flow* principle. In screen presentation, the layout of the page will differ with screen sizes and readers' preferred font sizes. In contrast to print layout, which is page-oriented and fixed, the layout in digital media is flexible. There is no such thing as a fixed page in digital reading applications.

As we have seen, most digital text production is still carried out according to principles of print. Page composition is a central activity, and at the heart of the process is a digital document containing a mix of information on presentation (typography and layout) and verbal content. Basic features of print, workflows and documents are closely interrelated.

In printed text cycles, as we have repeatedly stressed, the *storage* and *representation* phases are intrinsically interwoven. The same visual patterns (on paper) are used to store and represent text. This is reflected in the way print production has been digitized, and in the way digital text documents convey text information. In digital print workflows, text documents very often contain a mix of verbal information and exact descriptions on how to print signs and graphical elements on pages of fixed sizes, Adobe's Portable Document Format (PDF) being a typical example.

The fact that these documents are productive in print (and utilized in electronic publishing) does not mean that they are suited to text production aimed at cross-media publishing and digital reading. Since re-flow is a key feature in digital reading applications, content information and presentation control must be separated. The same applies when a text is recomposed and reused in different channels and media, i.e. on the Web, in e-learning, in print-on-demand, in e-books and in ordinary printed books.

XML (eXtensible Markup Language) has been developed by the World Wide Web Consortium as a platform-independent meta-language. The aim of XML is to overcome the shortcomings of HTML and to extend the possibilities of the Internet. In XML content structure and presentation are kept separate. In XML-documents (verbal) text content is marked up in a logical and hierarchical way while presentational information is stored separately in a style sheet document.

From the point of view of publishing, XML is an effort by computer scientists to make text production and text distribution compliant with basic features of digital technology. In digital text cycles, *storage* and *representation* are separated into two different phases. XML is developed to make the most of this separation, which is the reason why text content and text presentation are separated. This separation gives both intelligent storing and extended flexibility in text representation; the same content can be given different presentations. Text formats based on XML are thus suitable for re-flow (digital reading) and for cross-media publishing.

The general spread and many promising features of XML are the main reasons why XML-based production processes are permeating publishing. Publishers want to be present in all markets and (in the long run) increase their revenues and profits. For publishers this means another change in the previously digitized writing and production processes of publishing. In short it means adjusting to a production workflow based on structured markup, re-use of content and cross-media publishing, or digital publishing as this production mode is called (Kasdorf 2003).

Digital publishing is yet another transition in the overall digital transformation of the text. In digital publishing the combined storage and representational principle of print is no longer the guiding production principle, as it was in electronic publishing.

Step by step, digital technology is altering all the phases of the text cycle; the writing and production of texts, the storage and representation of texts and the distribution and reading of texts.

6 Is digital publishing a revolution?

6.1 Continued print dominance

Digitalization of the word obviously represents a transformation. It changes all parts of the text cycle and it even changes the text itself. Compared to changes brought by print, changes brought by digital technology are far more fundamental and important.

We have already seen how the invention of printing separated the process of writing from that of production, making the latter much more effective. Digital production, in comparison, changes all phases of the text cycle. In written and printed text cycles, storage and representation of texts are performed in one combined operation using ink and paper. In digital cycles, texts are stored as bit patterns and visually represented (on screens) in two separate processes. This separation makes possible completely new ways of distributing and reading texts, and has totally changed the processes of writing and producing texts.

The changes brought to the text cycle by digital technology are extensive, and as such they deserve to be called revolutionary. However, there is nothing revolutionary in the pace of the changes. The digital transformation of text has been going on for more than half a century, and it is still in its early stages. Rather than being a quick revolution the digital

transformation of the text seems to be somewhat slow and gradual.

More than two decades after text digitalization started in the 1950s, and ASCII was adapted in the 1960s, word processors and layout applications were developed in the 1980s as effective tools in print production. In the 1990s the Web arrived, and there was an explosion in electronic publishing, persistently dominated by documents made for print. In the first decade of the 21st century publishers are moving towards digital publishing, but, as we shall see, digital publishing too is still dominated by print principles, and this will probably be the case for decades to come. The legacy of print seems to be deep and long lasting, slowing down the radical potential of digital text cycles.

There are several reasons why print dominates digital publishing in the same way as it dominated electronic publishing. First, reading, especially sustained reading, is still dominated by printed paper. Handheld digital reading devices are too small, too heavy or too expensive -- and in any case, too crude -- to compete with magazines and books as reading technologies. The development of a digital text cycle is only partly completed. For this simple reason publishers earn most of their money in paper-based text cycles. Instead of a radical conversion to digital publishing, most publishers prefer a slow adjustment of their current print production to the principles of structured markup and cross-publishing. Text versions prepared for print in page-layout applications are prime sources for converted structured documents and later cross-media text publications. Through this workflow, print dominance continues.

Paradoxically the XML-related principle of "single sourcing" or "one input - many outputs" also strengthens the dominance of print in digital publishing. In a single source workflow, texts are written and edited once and stored as XML, before they are published in a variety of printed and digital formats: one input - many outputs. As long as the source documents in this workflow are made for print it is obvious that content structures of printed texts will pervade the central XML documents and consequently all the produced texts in the flow. As long as print dominates publishing, the principle of single source will prolong this domination.

6.2 Misunderstandings about XML

In terms of re-use, or cross-media publishing, the idea of single sourcing promises more than a plain reformatting of (printed) text documents to fit different media. The idea is that you can fetch bits and pieces of texts from a repository (a digital asset management system), reformat and rearrange the text items, and use the content for different purposes in print, e-learning, on the Web, in digital encyclopaedias and in e-books. Such re-use and cross-media publishing is indeed technologically possible, but the whole multi-use notion of single sourcing is a gross overestimation of the possibilities of XML (Hillesund 2002a).

First, single sourcing is contradicted by theoretical insights of linguistics and by practical knowledge of writers and publishers. A text always conforms (more or less) to the norms of a text genre. At macro levels genre norms define the overall organization of themes and issues, and at meso levels genre norms prescribe ways to dramatize, describe, make an argument or tell a story. At micro levels, authors, depending on genre, use different words, metaphors, expressions and technical terms, creating all kinds of language styles. A text is thus affected by genre norms at all levels. Bits and pieces of texts cannot be suitably rearranged and reused in an automatic creation of new texts belonging to other genres.

Secondly, text genres have historically been developed to meet socially determined communicative objectives. In addition most genres are media-specific. By utilizing possibilities of certain media (manuscripts, print or Web) and formats (book, journal, magazine or newspaper) writing communities have developed many different text genres in order to perform a variety of communicative acts. Thus, most verbal texts belong to media-specific genres and are best communicated through the medium they were created for.

The single source ideology is built on a misunderstanding of the scope of the content-presentation separation of XML. The belief that a text can be rearranged and moulded by technological means, presupposes that content and presentation can be treated independently, as logically distinct features: it rests on the false supposition that any kind of written or verbal content can be presented at will in any medium and for whatever purpose.

This is not the case. A text is part of a cycle and in different text cycles all phases are related and interconnected. In written and printed text cycles, storage and representation are done by the same means (ink on paper) in one combined operation. The stored text and the text representation are inseparably tied together. Despite the fact that storage and representation are performed as two separate operations in digital text cycles, they too are interrelated phases of overall text cycles. It makes only limited sense to refer to the bit pattern of a digital text as "text". A text is a visual representation of verbal information. To be a text in the full sense of the word, the bits storing the content have to be visually represented in a way that people can read.

The last point is crucial: a characteristic of writing systems is that verbal information is visually represented. All reading systems make use of a surface area on which written signs are made visible. Such visual representational systems have many shapes and formats: hieroglyphics in stone, letters on parchment, type on paper or computer screens. Printed paper is found in many formats: books, journals, magazines and newspapers, all in different sizes. Similarly, computer screens vary in size, from small PDA displays to big stationary screens.

Each medium and format has specific ways of presenting and structuring verbal information. From medieval manuscripts and handwritten books, print-inherited fonts, fixed pages, margins, columns and paragraphs. During the first century after Gutenberg, printers invented several new ways of visualizing and structuring verbal information: tables, indices, section breaks, page numbers, running heads, footnotes and regular provision of titles and subtitles. These (and many more) typographical features manifest structures of verbal information conveyed in printed texts.

When XML experts analyze existing (printed) text genres, they extract structures consisting of element types like <title>, <paragraph> and <table>. These element structures obviously describe some sort of content structure of the texts. The point is that these content structures are communicated by the layout and typography of the texts, by visual means. The elements extracted are combined content/presentation elements. The structures of the elements, as described in document type

definitions, are abstract descriptions of the way content of texts are visually presented (as titles, paragraphs and tables).

In printed texts, as in all texts, visual presentation and (many of the) content structures are intrinsically interwoven and indistinguishable. While this interdependence between form and content is important in publishing, it is hardly discussed in the XML literature (Hillesund 2002a). On the contrary, experts tend to hypostatize the XML elements and structures and treat them as parts of logical content structures of texts, separated from any presentation (Walsh 2002). These text structures do not capture important argumentative, narrative and other semantic content structures of texts. Another objection is that visually manifested text structures (based on principles of print) cannot be treated as logical content structures of texts in general, independent of media and format.

This promotion of XML structures will have the paradoxical (and obviously unintended) consequence that conventions of print will dominate digital publishing for a long time, especially the parts based on cross-media publishing and single sourcing. These production workflows will lead to print-based content structures being contorted to fit new media, while new genres which exploit the potential of digital media will not be developed.

New innovative digital text genres will clearly draw on many of the most readable typographical features of print. But the presentation and structure of verbal information will be shaped by the formats and communicative potential of digital media, rather than those of print media. We have earlier seen that digital formats and capabilities are based on re-flow, inclusion of multimedia and interactivity, hypertexts and linking to specifically defined (and updated) local or remote information resources. These features will give rise to new text genres with new linguistic styles, element types and element structures, many of which are incompatible with print. The new genres will demonstrate single sourcing to be useless in many circumstances.

However, for the time being it seems that the dependency of publishers on print production for commercial survival will prolong the experts' determination to reconcile the irreconcilable by applying the constraints of printing to innovative digital solutions.

Such a mix of technological innovation and cultural conservatism is no new phenomenon. It was certainly evident in the early European years of writing and printing. Centuries after Homer, narrative and metric structures of oral tales still dominated writing. Gutenberg and the first generations of printers put a great deal of effort into their attempts to replicate the rich and beautiful appearance of contemporary handwritten manuscripts. For these reasons we should not be surprised if digital publishing remains hidebound by print conventions for decades, and that the digital transformation of the text is, in consequence, a slow process indeed.

6.3 Social influence

As we have seen, commercial constraints on publishers, and the tenacity of perceptions regarding the nature of text, affect the pace and direction of its digital transformation. So do other social factors, such as economic competition, the establishment and nature of digital rights management systems and readers' behaviour. Experience with e-books may illustrate how these factors work.

At the present time (2004) one of the leading American online e-book stores (Fictionwise) offers e-books in eight unsecured and four "secure" formats. These are all proprietary formats connected to different reading applications, reading devices and digital rights management systems. Readers can hardly be expected to accept such a chaotic plethora of formats, and it is not likely that any of them will survive.

What most readers want is a simple system that allows them to purchase e-books from any retailer or publisher and read them on whatever device, operating system or reading software they choose (Lee et al. 2002, OpenReader). Instead of agreeing to such a standard end-user format, big companies like Adobe, Palm and Microsoft seem to be willing to fight (a futile fight) over markets. This policy will certainly not speed up e-book interest among readers.

Besides the chaos surrounding formats, there are many different online payment systems and several incompatible DRM systems. The latter, in particular, are a major impediment to e-book diffusion.

In digital environments, texts can be copied and distributed endlessly, without loss of quality. To ensure the income of writers and publishers some kind of copy protection is needed. However, recent US and EU legislation has disturbed a time-tested balance in favour of the "content owners", i.e. publishers, such as Random House and Penguin, and multinational media companies, such as Disney and Bertelsmann. In current DRM systems, the contents of e-books are strictly protected by technological obstructions, which deprive readers and buyers of privacy protection and undermine established owner rights to lend or sell books or to take copies for private use or safekeeping. Many scholars and politicians claim that the new DRM regimes also violate democratic values connected to freedom of speech and free float of information.

In the area of intellectual property and copy protection there are many conflicting interests between readers, writers, scholars, libraries, universities, publishers, software producers, hardware producers and multinational media conglomerates. Around the world there are economic, social and cultural struggles going on regarding intellectual property and digital rights management. Achieving a satisfactory balance between the interests of all those affected will call for a lot of rethinking, leading to new concepts, practices and legislation (Lynch 2001). This, in turn, will profoundly affect the organization of e-book cycles.

Besides format and DRM issues the development of e-book markets is heavily dependent on readers' behaviour. It is not easy to predict what use the general public will make of e-books, what kinds of e-books readers will buy or what books the majority of readers will prefer in printed versions. Nor can we be certain how teachers and students will react to e-books.

Unlike radio, television and the Internet, which were altogether new media, e-books compete with a highly valued existing medium, the printed book. Over the centuries, the printed book has developed into a very sophisticated reading technology. Its distribution system is well established (Hillesund 2001). In addition, printed books are highly valued artefacts, associated with some of the most fundamental values of civilized society: knowledge, education, understanding, development, democracy, literature and culture. To many people, shelves full of books are convincing status symbols that indicate the

owner's cultivation and learning.

To compete with printed books, e-books must improve in readability, price, interoperability, rights of use and cultural status. It is likely that the advantages of e-books (re-flow, linking, hypertext, multimedia, interactivity, storage capacities and accessibility) must be improved, and that e-books must develop into a qualitatively new medium, with new genres and new uses, before they can seriously challenge the domination of paper books.

6.4 Revolution or evolution

The electronic book is by no means a new concept. Before the abbreviation 'e-book' came into use in the middle of the 1990s, there had been a lot of talk about electronic books in the 1980s and even some discussion in the 1970s. The development and diffusion of e-books clearly illustrates how slow the process of text digitalization really is.

With its emphasis on reading, e-book technology is in many ways a completion of the digital text cycle, making all parts of the cycle digital. A few years ago, at the turn of the century, many expected a media revolution. It was widely believed that e-books would inundate the world and quickly change book distribution and reading. For reasons outlined above, these changes have not materialised. Far from being revolutionary, the introduction of e-books, and the digitalization of text in general, are slow processes which are themselves largely subject to parameters imposed by society (Winston 1998).

Seen in this perspective, the current trend towards digital publishing is not particularly revolutionary. Most readers still prefer printed versions of books, journal articles and longer texts. For economic reasons, publishers prefer a gradual adjustment of their current production to the principles of digital publishing, favouring print. The single source workflows also import principles of print, and even the philosophy of XML is pervaded with text concepts inherited from print. Rather than being a revolution, digital publishing is a step in the long evolution of digital text cycles.

Nevertheless, there is nothing insignificant about this step in the long evolution of digital text cycles. Digital publishing can go beyond simply being an auxiliary tool in printing. By use of structured markup, digital publishing takes advantage of the storage and representation separation typical of digital text cycles. Digital publishing makes possible a flexible use of text material in many formats and sizes.

Sooner or later the shortcomings of screen display technology will also be solved. E-books serve as good examples of the changes that digital text cycles bring about. E-books eliminate the need for paper, they are read on screens (mostly handheld) and will soon be read on other electronic substitutes for paper. Furthermore, e-books are stored digitally and distributed over networks, having no use of the traditional chain of paper producers, printers, book distributors, bookstores and bookshelves. E-books are also produced differently (in the workflows of digital publishing) and eventually, as new e-genres evolve, authors will write them differently.

All the phases of an e-book cycle differ substantially from the corresponding phases of a printed book cycle. The same is true for phases and cycles of electronic magazines, online newspapers, e-journals, corporate publishing and e-learning. When fully developed, all digital text cycles will do away with paper. In interplay with cultural and social factors, digital texts will change publishing, news mediation, science, business and education.

Even if the digital transformation of the text is a long and slow process, the development of digital text cycles will make a difference in the end. It will change writing, distribution and reading of texts, it will alter our understanding of texts. Ultimately digital text cycles will affect the organization of social institutions and society. But, as I have argued here, all theses changes will be influenced by economic, cultural and social factors.

7 Summary and conclusion

This paper has introduced the concept of the *text cycle*, which consists of phases such as writing, distribution and reading. An examination of written and printed text cycles, and of digital text cycles, has revealed significant differences.

In written text cycles the phases of writing, storing and representation of a text are, to all intents and purposes, performed in one operation: by applying ink on paper (using hand and pen). Written texts overcome many of the limitations of oral communication. The most important aspect of print is the separation of writing and production into two phases, making print an effective way of duplicating texts (by mechanical means).

Both written and printed text cycles use the same physical means -- ink on paper -- to store and represent texts. In digital text cycles, by comparison, storage and representation are performed in two distinct operations. The text is normally stored on a disc and represented on a display unit. The separation of storage and representation has important consequences because it allows computers to be used as effective tools for text manipulation and, thereby, opens up new ways of writing and producing texts. More importantly, as a consequence of the separation, digital texts are not subject to the physical constraints of printed paper. Over networks, like the Internet, digital texts can instantly be distributed and read on a variety of visual display devices.

This examination of the basic features of text cycles shows that digital text technology has the potential to bring about fundamental changes. While print technology substantially increased the efficiency of text production by splitting the creation process into two distinct phases, all phases are changed in digital text cycles, including the distribution and reading of texts. This gives digital texts a potential for change that far exceeds that of the "Gutenberg revolution".

However, the actual history of the digitalization of text, as shown in this paper, has been, and is, a gradual process bearing little resemblance to a revolution. Rather than being a quick change determined by technology, the digital transformation of text is a rather slow evolutionary process heavily influenced by social and cultural factors. This paper has discussed economic factors, readers' behaviour and especially the enduring cultural impact of print on our understanding of text.

The implementation of digital publishing, as discussed, is a step in this long evolution of digital text cycles, but it is an important step. Built on structured markup, digital publishing directs text production from deep-rooted concepts of print to

basic principles of digital text technology (the storage and representation separation). When acceptable reading technologies are developed (and sustained reading included in digital text cycles), digital publishing will be built on principles that make effective use of this new potential.

The future social organization of digital text cycles (for e-magazines, e-journals and e-books) is uncertain, depending, as it does, on the outcome of economic and cultural struggles. One of the most important of these is currently being fought between large media corporations, who seek content control, and readers and writers, who want open access. Whatever the results of these struggles, in the longer term cultural changes following the digital impact on text cycles will certainly be significant.

References

Asbj¬ørnsen, D. (2002) "Eb¬øker: marked og rettigheter". *Tidvise Skrifter*, No. 48 http://www1.his.no/ebok/ebokinor/rapport2.htm

Berners-Lee, T. and M. Fischetti (1999) *Weaving the Web: the Past, Present and Future of World Wide Web by its Inventors* (London: Orion Business Books)

Eisenstein, E. (1983) *The Printing Revolution in Early Modern Europe* (Cambridge, UK: Cambridge University Press)

Kasdorf, W.E. (ed.) (2003) *The Columbia Guide to Digital Publishing* (New York: Columbia University Press)

Hill, B. (2001) *The Magic of Reading* (Redmond, WA: Microsoft Corporation) http://www.microsoft.com/reader/includes/TheMagicofReading.lit (to view this text requires Microsoft Reader, which is available at http://www.microsoft.com/reader/downloads/pc.asp)

Hillesund, T. (2001) "Will E-books Change the World?" *First Monday*, Vol. 6, No. 10, October 1 http://www.firstmonday.dk/issues/issue6_10/hillesund/

Hillesund, T. (2002a) "Many Outputs -- Many Inputs: XML for Publishers and E-book Designers". *Journal of Digital Information*, Vol. 3, No. 1, Article No. 101, August 6 http://jodi.tamu.edu/Articles/v03/i01/Hillesund/

Hillesund, T. (2002b) "Digital lesing". *Tidvise Skrifter*, No. 49 http://www1.his.no/ebok/ebokinor/rapport3.htm

Innis, H.A. (1972) *Empire and Communication* (Toronto: University of Toronto Press)

King, D.W., D.D. Donald and N.K. Roderer (1981) *Scientific Journals in the United States: Their Production, Use and Economics (*New York: Academic Press)

Lee, K., N. Guttenberg and V. McCrary (2002) "Standardization aspects of eBook content formats". *Computer Standards and Interfaces*, Vol. 24, No. 3, July, 227-239

Leiner, B.M., V.G. Cerf, D.D. Clark, R.E. Kahn, L.K. Kleinrock, D.C. Lynch, J. Postel, L.G. Roberts and S. Wolff (2003) "A Brief History of the Internet". Internet Society, version 3.32, revised 10 December http://www.isoc.org/internet/history/brief.shtml

Lynch, C. (2001) The Battle to Define the Future of the Book in the Digital World". *First Monday*, Vol. 6, No. 6, June 4 http://www.firstmonday.dk/issues/issue6_6/lynch/

Ong, W.J. (1982) *Orality and Literacy: The Technologizing of the Word (*London: Methuen)

O'Sullivan, T., J. Hartley, D. Saunders and J. Fiske (1983) *Key Concepts in Communication* (London: Methuen)

Tenopir, C. and D.W. King (2000) *Towards Electronic Journals: Realities for Scientists, Librarians and Publishers* (Washington: Special Libraries Association)

Tenopir, C. and D.W. King (2004) *Communication Patterns of Engineers* (Piscataway, NJ: IEEE Press)

Walsh, N. (2002) "XML: One Input -- Many Outputs: a response to Hillesund". *Journal of Digital Information* , Vol. 3, No. 1, September 12 http://jodi.tamu.edu/Articles/v03/i01/Walsh/

Winston, B. (1998) *Media Technology and Society* (London: Routledge)

Links

Fictionwise http://www.fictionwise.com

OpenReader http://openreader.org/

Extensible Markup Language (XML) http://www.w3.org/XML/