

A Short History of eBooks

Marie Lebert

NEF, University of Toronto, 2009

Copyright © 2009 Marie Lebert

All rights reserved

This book is dedicated to all those who kindly answered my questions during ten years, in Europe, in America (the whole continent), in Africa, and in Asia. With many thanks for their time and their friendship.

A short history of ebooks - also called digital books - from the first ebook in 1971 until now, with Project Gutenberg, Amazon, Adobe, Mobipocket, Google Books, the Internet Archive, and many others. This book is based on 100 interviews conducted worldwide and thousands of hours of web surfing during ten years.

This book is also available in French and Spanish, with a longer and different text. All versions can be found online <<http://www.etudes-francaises.net/dossiers/ebook.htm>>.

Unless specified otherwise, quotations are excerpts from NEF interviews <<http://www.etudes-francaises.net/entretiens/>>.

Marie Lebert is a researcher and editor specializing in technology for books, other media, and languages. She is the author of *Technology and Books for All* (in English and French, 2008), *Les mutations du livre* (Mutations of the Book, in French, 2007) and *Le Livre 010101* (The 010101 Book, in French, 2003). Her books are published by NEF (Net des études françaises / Net of French Studies), University of Toronto, Canada, and are freely available online <<http://www.etudes-francaises.net>>.

Table

=== Introduction

1971: Project Gutenberg is the first digital library

1990: The web boosts the internet

1993: The Online Books Page is a list of free ebooks

1994: Some publishers get bold and go digital

1995: Amazon.com is the first main online bookstore

1996: There are more and more texts online

1997: Multimedia convergence and employment

1998: Libraries take over the web

1999: Librarians get digital

2000: Information is available in many languages

2001: Copyright, copyleft and Creative Commons

2002: A web of knowledge

2003: eBooks are sold worldwide

2004: Authors are creative on the net

2005: Google gets interested in ebooks

2006: Towards a world public digital library

2007: We read on various electronic devices

2008: "A common information space in which we communicate"

=== Chronology

=== Acknowledgements

Introduction

The book is no longer what it used to be.

The electronic book (ebook) was born in 1971, with the first steps of Project Gutenberg, a digital library for books from public domain. It is nearly 40 years old, already. But this is a short life compared to the 5-century old print book.

The internet went live in 1974, with the creation of the protocol TCP/IP by Vinton Cerf and Bob Kahn. It began spreading in 1983 as a network for research centers and universities. It got its first boost with the invention of the web by Tim Berners-Lee in 1990, and its second boost with the release of the first browser Mosaic in 1993. From 1994 onwards, the internet quickly spread worldwide.

In Bookland, people were reluctant, curious or passionate.

The internet didn't bring print media, movies, radio or television to an end. It created its own space as a new medium, to get information, access documents, broaden our knowledge and communicate across borders and languages.

Booksellers began selling books online within and outside their home country, offering excerpts on their websites.

Libraries began creating websites as a "virtual" window, as well as digital libraries stemming from their print collections. Librarians helped patrons to surf on the web without being drowned, and to find the information they needed at a time search engines were less accurate. Library catalogs went online. Union catalogs offered a common point for hundreds and then thousands of catalogs.

Newspapers and magazines began being available online, as well as their archives. Some journals became "only" electronic to skip the costs of print publishing, while offering print on demand. Some newsletters, zines and journals started online from scratch, skipping a print version.

Authors began creating websites to self-publish their work or post it while waiting to find a publisher. Communication with readers became easier through email, forums, chat and instant messaging. Some authors explored new ways of writing, called hypertext literature.

More and more books were published with both a print version and a digital version. Some books were "only" digital. Other books were digitized from print versions.

New online bookstores began selling "only" digital books. Aggregators partnered with publishers to produce and sell digital versions of their books.

People no longer needed to run after information and to worry about living in a remote place with no libraries and bookstores. Information was there, by the numbers, available on our

screen, often at no cost.

In 2009, most of us would not be able to work, study, communicate and entertain without connecting with others through the internet.

Here is the “virtual” journey we are going to follow:

- 1971: Project Gutenberg is the first digital library
- 1990: The web boosts the internet
- 1993: The Online Books Page is a list of free ebooks
- 1994: Some publishers get bold and go digital
- 1995: Amazon.com is the first main online bookstore
- 1996: There are more and more texts online
- 1997: Multimedia convergence and employment
- 1998: Libraries take over the web
- 1999: Librarians get digital
- 2000: Information is available in many languages
- 2001: Copyright, copyleft and Creative Commons
- 2002: A web of knowledge
- 2003: eBooks are sold worldwide
- 2004: Authors are creative on the net
- 2005: Google gets interested in ebooks
- 2006: Towards a world public digital library
- 2007: We read on various electronic devices
- 2008: "A common information space in which we communicate"

1971: Project Gutenberg is the first digital library

[Overview]

The first ebook was available in July 1971, as eText #1 of Project Gutenberg, a visionary project launched by Michael Hart to create electronic versions of literary works and disseminate them worldwide. In the 16th century, Gutenberg allowed anyone to have print books for a small cost. In the 21st century, Project Gutenberg would allow anyone to have a digital library at no cost. Its critics long considered Project Gutenberg as impossible on a large scale. But Michael went on keying book after book during many years, with the help of some volunteers. Project Gutenberg got its first boost with the invention of the web in 1990 and its second boost with the creation of Distributed Proofreaders in 2000, to help digitizing books from public domain. In 2008, Project Gutenberg had a production rate of 340 new books each month, 40 mirror sites worldwide, and books being downloaded by the tens of thousands every day. There have been Project Gutenberg websites in the U.S., in Australia, in Europe and in Canada, with more websites to come in other countries.

From 1971 until now

Beginning

As recalled by Michael Hart in January 2009 in an email interview: "On July 4, 1971, while still a freshman at the University of Illinois (UI), I decided to spend the night at the Xerox Sigma V mainframe at the UI Materials Research Lab, rather than walk miles home in the summer heat, only to come back hours later to start another day of school. I stopped on the way to do a little grocery shopping to get through the night, and day, and along with the groceries they put in the faux parchment copy of *The U.S. Declaration of Independence* that became quite literally the cornerstone of Project Gutenberg. That night, as it turned out, I received my first computer account - I had been hitchhiking on my brother's best friend's name, who ran the computer on the night shift. When I got a first look at the huge amount of computer money I was given, I decided I had to do something extremely worthwhile to do justice to what I had been given. This was such a serious, and intense thought process for a college freshman, my first thought was that I had better eat something to get up enough energy to think of something worthwhile enough to repay the cost of all that computer time. As I emptied out groceries, the faux parchment *Declaration of Independence* fell out, and the light literally went on over my head like in the cartoons and comics... I knew what the future of computing, and the internet, was going to be.. 'The Information Age.' The rest, as they say, is history."

Michael decided to search the books from public domain available in our libraries, digitize these books, and store the electronic books (ebooks) in the simplest way, using the low set of ASCII - called Plain Vanilla ASCII - for them to be read on any hardware and software. A book would become a continuous text file instead of a set of pages, with caps for the terms in italic, bold or underlined of the print version. As a text file, a book would be easily copied, indexed, searched, analyzed and compared with other books. (Doing such searches is much harder in various markup formats.)

Project Gutenberg's mission would be the following: to put at everyone's disposal, in electronic versions, as many literary works from public domain as possible for free. Years

later, in August 1998, Michael wrote in an email interview: "We consider etext to be a new medium, with no real relationship to paper, other than presenting the same material, but I don't see how paper can possibly compete once people each find their own comfortable way to etexts, especially in schools."

After keying in *The U.S. Declaration of Independence* in 1971, Michael typed in *The U.S. Bill of Rights* in 1972. A volunteer typed in *The United States Constitution* in 1973.

Persevering

From one year to the next, disk space was getting larger, by the standards of the time - there was no hard disk yet -, making it possible to store larger files. Volunteers began typing in the Bible, with one individual book at a time, and a file for each book. Michael typed in the collected works of Shakespeare, with volunteers, one play at a time, and a file for each play. This edition of Shakespeare was never released, unfortunately, due to changes in copyright law. Shakespeare's works belong to public domain, but comments and notes may be copyrighted, depending on the publication date. Other editions of Shakespeare from public domain were posted a few years later.

10 to 1,000 ebooks

In August 1989, Project Gutenberg completed its 10th ebook, *The King James Bible* (1769), both testaments, and 5M for all files.

In 1990, there were 250,000 internet users. The web was in its infancy. The standard was 360 K disks.

In January 1991, Michael typed in *Alice's Adventures in Wonderland* (1865), by Lewis Carroll. In July 1991, he typed in *Peter Pan* (1904), by James M. Barrie. These two classics of childhood literature each fit on one disk.

The first browser, Mosaic, was released in November 1993. It became easier to circulate etexts and recruit volunteers. From 1991 to 1996, the number of ebooks doubled every year, with one book per month in 1991, two books per month in 1992, four books per month in 1993, and eight books per month in 1994.

In January 1994, Project Gutenberg released *The Complete Works of William Shakespeare* as eBook #100. Shakespeare wrote most works between 1590 and 1613.

The steady growth went on, with an average of 8 books per month in 1994, 16 books per month in 1995, and 32 books per month in 1996.

In June 1997, Project Gutenberg released *The Merry Adventures of Robin Hood* (1883), by Howard Pyle.

Project Gutenberg had 1,000 ebooks in August 1997. eBook #1000 was *La Divina Commedia*, by Dante Alighieri (1321), in Italian, its original language.

As there were more and more ebooks, they got classified in three main sections: (a) "Light Literature", such as *Alice's Adventures in Wonderland*, *Through the Looking-Glass*, *Peter Pan* and *Aesop's Fables*; (b) "Heavy Literature", such as the Bible, Shakespeare's works, *Moby Dick* and *Paradise Lost*; (c) "Reference Literature", such as *Roget's Thesaurus*, almanacs, and a set of encyclopedias and dictionaries. (This classification in three sections was replaced later with a more detailed one.)

"Light Literature" was the main section in number of ebooks. As explained on the website in 1998, "The Light Literature Collection is designed to get persons to the computer in the first place, whether the person may be a pre-schooler or a great-grandparent. We love it when we hear about kids or grandparents taking each other to an etext of *Peter Pan* when they come back from watching Hook at the movies, or when they read *Alice in Wonderland* after seeing it on TV. We have also been told that nearly every Star Trek movie has quoted current Project Gutenberg etext releases (from *Moby Dick* in *The Wrath of Kahn*; a *Peter Pan* quote finishing up the most recent, etc.) not to mention a reference to *Through the Looking-Glass* in JFK. This was a primary concern when we chose the books for our libraries. We want people to be able to look up quotations they heard in conversation, movies, music, other books, easily with a library containing all these quotations in an easy-to-find etext format."

Project Gutenberg has selected books intended for the general public. It has not focused on providing authoritative editions. "We do not write for the reader who cares whether a certain phrase in Shakespeare has a ':' or a ';' between its clauses. We put our sights on a goal to release etexts that are 99.9% accurate in the eyes of the general reader. Given the preferences our proofreaders have, and the general lack of reading ability the public is currently reported to have, we probably exceed those requirements by a significant amount. However, for the person who wants an 'authoritative edition' we will have to wait some time until this becomes more feasible. We do, however, intend to release many editions of Shakespeare and the other classics for comparative study on a scholarly level."

In August 1998, Michael Hart wrote in an email interview: "My own personal goal is to put 10,000 etexts on the net [this goal was reached in October 2003] and if I can get some major support, I would like to expand that to 1,000,000 and to also expand our potential audience for the average etext from 1.x% of the world population to over 10%, thus changing our goal from giving away 1,000,000,000,000 etexts to 1,000 times as many, a trillion and a quadrillion in U.S. terminology."

1,000 to 10,000 ebooks

From 1998 to 2000, the "output" was an average of 36 books per month.

Project Gutenberg reached 2,000 ebooks in May 1999. eBook #2000 was *Don Quijote* (1605), by Cervantes, in Spanish, its original language.

Project Gutenberg reached 3,000 ebooks in December 2000. eBook #3000 was *A l'ombre des jeunes filles en fleurs* (In the Shadow of Young Girls in Flower), vol. 3 (1919), by Marcel Proust, in French, its original language.

Project Gutenberg reached 4,000 ebooks in October 2001. eBook #4000 was *The French Immortals Series* (1905), in English. This book is an anthology of short fictions by authors from the French Academy (Académie française): Emile Souvestre, Pierre Loti, Hector Malot, Charles de Bernard, Alphonse Daudet, and others.

Project Gutenberg reached 5,000 ebooks in April 2002. eBook #5000 was *The Notebooks of Leonardo da Vinci* (early 16th century). Since its release, this ebook has stayed in the Top 100 of downloaded books.

In 1988, Michael Hart chose to type in *Alice's Adventures in Wonderland* and *Peter Pan* because they would each fit on one 360 K disk, the standard of the time. In 2002, the standard disk was 1.44 M and could be compressed as a ZIP file.

A practical file size is about 3 million characters, more than long enough for the average book. The ASCII version of a 300-page novel is 1 M. A bulky book can fit in two ASCII files, that can be downloaded as is or in ZIP format. An average of 50 hours is necessary to get an ebook selected, copyright-cleared, scanned, proofread, formatted and assembled.

A few numbers are reserved for "special" books. For example, eBook #1984 is reserved for George Orwell's classic, published in 1949, and still a long way from falling into public domain.

The "output" in 2001 and 2002 was an average of 100 books per month.

In spring 2002, Project Gutenberg's ebooks represented 25% of all the public domain works freely available on the web, an impressive result if we think of all the pages that were scanned and proofread by thousands of volunteers in several countries.

1,000 ebooks in August 1997, 2,000 ebooks in May 1999, 3,000 ebooks in December 2000, 4,000 ebooks in October 2001, 5,000 ebooks in April 2002, 10,000 ebooks in October 2003. eBook #10000 was *The Magna Carta*, signed in 1215 and known as the first English constitutional text.

From April 2002 to October 2003, in 18 months, the collections doubled, going from 5,000 ebooks to 10,000 ebooks, with a monthly average of 300 new ebooks. The fast growth was the work of Distributed Proofreaders, a website launched in October 2000 by Charles Franks to share the proofreading of books between many volunteers. Volunteers choose one of the books available on the site and proofread a given page. It is recommended they do a page per day if possible.

Books were also copied on CDs and DVDs. As blank CDs and DVDs cost next to nothing, Project Gutenberg began burning and sending a free CD or DVD to anyone asking for it. People were encouraged to make copies for a friend, a library or a school. Released in August 2003, the *Best of Gutenberg* CD contained 600 ebooks. The first Project Gutenberg DVD was released in December 2003 to celebrate the first 10,000 ebooks, with the burning of most titles (9,400 ebooks).

10,000 to 20,000 ebooks

In December 2003, there were 11,000 ebooks, which represented 110 G, in several formats (ASCII, HTML, PDF and others, as is or zipped). In May 2004, there were 12,600 ebooks, with represented 135 G. With more than 300 new books added per month (338 books per month in 2004), the number of gigabytes was expected to double every year.

The Project Gutenberg Consortia Center (PGCC) was affiliated with Project Gutenberg in 2003, and became an official Project Gutenberg site. Since 1997, PGCC had been working on gathering collections of existing ebooks, as a complement to Project Gutenberg focusing on the production of ebooks.

In January 2005, Project Gutenberg had 15,000 ebooks. eBook #15000 was *The Life of Reason* (1906), by George Santayana.

What about languages? There were ebooks in 25 languages in February 2004, and in 42 languages in July 2005, including Sanskrit and the Mayan languages. The seven top languages - with more than 50 books - were English (with 14,548 ebooks on July 27, 2005), French (577 ebooks), German (349 ebooks), Finnish (218 ebooks), Dutch (130 ebooks), Spanish (103 ebooks) and Chinese (69 ebooks). There were ebooks in 50 languages in December 2006. The ten top languages were English (with 17,377 books on December 16, 2006), French (966 books), German (412 books), Finnish (344 books), Dutch (244 books), Spanish (140 books), Italian (102 books), Chinese (69 books), Portuguese (68 books) and Tagalog (51 books).

Project Gutenberg was also spreading worldwide.

In July 2005, Project Gutenberg Australia (launched in 2001) had 500 ebooks.

In Europe, Project Rastko, based in Belgrade, Serbia, launched Distributed Proofreaders Europe (DP Europe) in December 2003 and Project Gutenberg Europe (PG Europe) in January 2004. Project Gutenberg Europe released its first 100 ebooks in June 2005. These books were in several languages, as a reflection of European linguistic diversity, with 100 languages planned for the long term.

New teams were working on launching Project Gutenberg Canada, Project Gutenberg Portugal and Project Gutenberg Philippines.

In December 2006, Project Gutenberg had 20,000 ebooks. eBook #20000 was the audiobook of *Twenty Thousand Leagues Under the Sea* (Vingt mille lieues sous les mers, 1869), by Jules Verne, in its English version.

If 32 years were necessary to digitize the first 10,000 books - between July 1971 and October 2003 -, 3 years and 2 months were necessary to digitize the following 10,000 books - between October 2003 and December 2006.

The section Project Gutenberg PrePrints was set up in January 2006 to collect items submitted to Project Gutenberg which were interesting enough to be available online, but not ready yet to be added to the main Project Gutenberg collections, the reason being missing data, low-quality files, formats which were not handy, etc. This new section had 379 files in December 2006.

Tens of thousands of ebooks

In December 2006, Mike Cook launched Project Gutenberg News as "the news portal for gutenberg.org", a website to complement the existing weekly and monthly newsletters. It has showed for example the weekly, monthly and yearly production stats since 2001.

The weekly production was 24 ebooks in 2001, 47 ebooks in 2002, 79 ebooks in 2003, 78 ebooks in 2004, 58 ebooks in 2005, 80 ebooks in 2006, and 78 ebooks in 2007.

The monthly production was 104 ebooks in 2001, 203 ebooks in 2002, 348 ebooks in 2003, 338 ebooks in 2004, 252 ebooks in 2005, 345 ebooks in 2006, and 338 books in 2007.

The yearly production was 1,244 ebooks in 2001, 2,432 ebooks in 2002, 4,176 ebooks in 2003, 4,058 ebooks in 2004, 3,019 ebooks in 2005, 4,141 ebooks in 2006, and 4,049 ebooks in 2007.

Project Gutenberg Australia reached 1,500 ebooks in April 2007.

Project Gutenberg Canada (PGC) was founded on July 1st, 2007, on Canada Day, by Michael Shepard and David Jones. Distributed Proofreaders Canada (DPC) started production in December 2007. There were 100 ebooks in March 2008, in English, French and Italian.

Project Gutenberg sent out 15 million ebooks via CDs and DVDs by snail mail in 2007. A new DVD released in July 2006 included 17,000 ebooks. CD and DVD files have also been generated as ISO files (since 2005) to be downloaded for burning CDs or DVDs on a CD or DVD writer.

Project Gutenberg reached 25,000 books in April 2008. eBook #25000 was *English Book Collectors* (1902), by William Younger Fletcher.

If Gutenberg allowed everyone to get print books at little cost, Project Gutenberg has allowed everyone to get a library of electronic books at no cost on a cheap device like a USB drive.

In February 2009, there were 32,500 Project Gutenberg (PG) ebooks, including the ebooks at PG Australia (1,750 ebooks), PG Europe (600 ebooks) and PG Canada (250 ebooks), with more Project Gutenberg websites to come in other countries. Ten new ebooks have been added per day.

As explained by Michael Hart: "In addition, there is 'PrePrints' where we put anything we don't know for sure will qualify as a PG ebook. This gets instant exposure, and was created to help keep things flowing. There are 2,020 ebooks available at PrePrints. The Project

Gutenberg Consortia Center (PGCC) has over 75,000 ebooks rendered as PDF files, and some are really quite stunning. The difference? These files were prepared by other eLibraries, not Project Gutenberg, and are using our worldwide distribution network to be seen. Thus, counting these 75,000+ along with our over 32,500 other ebooks, has generated a grand total of over 100,000 ebooks."

From the past to the future

The bet made by Michael Hart in 1971 succeeded. But Project Gutenberg's results are not only measured in numbers. The results can also be measured in the major influence the project has had. As the oldest producer of free books on the internet, Project Gutenberg has inspired many other digital libraries, for example Projekt Runeberg for classic Nordic (Scandinavian) literature and Projekt Gutenberg-DE for classic German literature, to name only two, which started respectively in 1992 and 1994.

Projekt Runeberg was the first Swedish digital library of books from public domain, and a partner of Project Gutenberg. It was initiated in December 1992 by the students' computer club Lysator, in cooperation with Linköping University, as a volunteer project to create and collect free electronic editions of classic Nordic literature and art. Around 200 ebooks were available in full text in 1998. There was also a list of 6,000 Nordic authors as a tool for further collection development.

Projekt Gutenberg-DE was the first German digital library of books from public domain, created in 1994 as a partner of Project Gutenberg. Texts were available for online reading, with one webpage for short texts and with several webpages - one per chapter - for longer works. There was an alphabetic list of authors and titles, and a short biography and bibliography for each author.

Project Gutenberg keeps its administrative and financial structure to the bare minimum. Its motto fits into three words: "Less is more." The minimal rules give much space to volunteers and to new ideas. The goal is to ensure its independence from loans and other funding and from ephemeral cultural priorities, to avoid pressure from politicians and others. The aim is also to ensure respect for the volunteers, who can be confident their work will be used not just for decades but for centuries. Volunteers can network through mailing lists, weekly or monthly newsletters, discussion lists, forums and wikis.

Donations are used to buy equipment and supplies, mostly computers, scanners and blank CDs and DVDs. Founded in 2000, the PGLAF (Project Gutenberg Literary Archive Foundation) has only three part-time employees.

More generally, Michael Hart should be given more credit as the inventor of the electronic book (ebook). If we consider the ebook in its etymological sense - that is to say a book that has been digitized to be distributed as an electronic file - it was born with Project Gutenberg in July 1971. This is a much more comforting paternity than the various commercial launchings in proprietary formats that peppered the early 2000s. There is no reason for the term "ebook" to be the monopoly of Amazon, Barnes & Noble, Gemstar, and others. The non-commercial ebook is a full ebook, and not a "poor" version, just as non-commercial electronic

publishing is a fully-fledged way of publishing, and is as valuable as commercial electronic publishing. Project Gutenberg etexts - the term used originally - have been renamed ebooks, to use the recent terminology in the field.

In July 1971, sending a 5K file to 100 people would have crashed the network of the time. In November 2002, Project Gutenberg could post the 75 files of the *Human Genome Project*, with files of dozens or hundreds of megabytes, shortly after its initial release in February 2001 as a work from public domain. In 2004, a computer hard disk costing US \$140 could potentially hold the entire Library of Congress. And we probably are only a few years away from a USB drive - or an equivalent storage disk - capable of holding all the books on our planet.

What about documents other than text? In September 2003, Project Gutenberg launched Project Gutenberg Audio eBooks, with human-read ebooks. Computer-generated ebooks are "converted" when requested from the existing electronic files in the main collections. Voice-activated requests will be possible in the future. Launched at the same time, the Sheet Music Subproject contains digitized music sheet, as well as a few music recordings. Some still pictures and moving pictures are also available. These collections should take off in the future.

But digitizing books remains the priority, and there is a big demand, as confirmed by the tens of thousands of books that are downloaded every day.

For example, on July 31, 2005, there were 37,532 downloads for the day, 243,808 downloads for the week, and 1,154,765 downloads for the month.

On May 6, 2007, there were 89,841 downloads for the day, 697,818 downloads for the week, and 2,995,436 downloads for the month.

On May 8, 2008, there were 115,138 downloads for the day, 714,323 downloads for the week, and 3,055,327 downloads for the month.

These numbers are the downloads from ibiblio.org (at University of North Carolina, Chapel Hill), the main distribution site, which also hosts the website gutenberg.org. The Internet Archive is the backup distribution site and provides unlimited disk space for storage and processing. Project Gutenberg has 40 mirror sites in many countries and is seeking new ones. It also encourages the use of P2P for sharing its books.

People can choose ebooks from the "Top 100", i.e. the top 100 ebooks and the top 100 authors for the previous day, the last 7 days and the last 30 days.

Project Gutenberg ebooks can also help bridge the "digital divide". They can be read on an outdated computer or a second-hand PDA costing just a few dollars. Solar-powered PDAs offer a good solution in remote regions.

It is hoped machine translation software will be able to convert the books from one to another of 100 languages. In ten years from now (August 2009), machine translation may be judged 99% satisfactory - research is active on that front - allowing for the reading of literary classics in a choice of many languages. Project Gutenberg is also interested in combining translation software and human translators, somewhat as OCR software is now combined with the work of proofreaders.

38 years after the beginning of Project Gutenberg, Michael Hart describes himself as a workaholic who has devoted his entire life to his project. He considers himself a pragmatic and farsighted altruist. For years he was regarded as a nut but now he is respected. He wants to change the world through freely-available ebooks that can be used and copied endlessly, and reading and culture for everyone at minimal cost.

Project Gutenberg's mission can be stated in eight words: "To encourage the creation and distribution of ebooks," by everybody, and by every possible means, while implementing new ideas, new methods and new software.

1990: The web boosts the internet

[Overview]

The internet was born in 1974 with the creation of TCP/IP (Transmission Control Protocol / Internet Protocol) by Vinton Cerf and Bob Kahn. It began spreading in 1983. The internet got its first boost with the invention of the web by Tim Berners-Lee at CERN (European Center for Nuclear Research) in 1989-90, and its second boost with the release of the first browser Mosaic in 1993. The internet could now be used by anyone, and not only by computer literate users. There were 100 million internet users in December 1997, with one million new users per month, and 300 million internet users in December 2000. In summer 2000, the number of non-English-speaking users reached 50%, and went on to increase then. According to Netcraft, the number of websites went from one million (April 1997) to 10 million (February 2000), 20 million (September 2000), 30 million (July 2001), 40 million (April 2003), 50 million (May 2004), 60 million (March 2005), 70 million (August 2005), 80 million (April 2006), 90 million (August 2006) and 100 million (November 2006).

The internet and the web

When Project Gutenberg began in July 1971, the internet was just a glimmer. The pre-internet was created in the U.S. in 1969, as a network set up by the Pentagon. The internet took off in 1974 with the creation of TCP/IP by Vinton Cerf and Bob Kahn. It expanded as a network linking U.S. governmental agencies, universities and research centers.

After the invention of the web in 1989-90 by Tim Berners-Lee at CERN (European Center for Nuclear Research), Geneva, Switzerland, and the release of the first browser, Mosaic (the ancestor of Netscape), in November 1993, the internet began spreading, first in the U.S. because of investments made by the government, then in North America, and then worldwide.

Because the web was easy to use, linking documents and pages with hyperlinks, the internet could now be used by anyone, and not only by computer literate users. There were 100 million internet users in December 1997, with one million new users per month, and 300 million internet users in December 2000.

Why did the internet spread in North America first? The U.S. and Canada were leading the way in computer science and communication technology, and a connection to the internet – mainly through a phone line - was much cheaper than in most countries. In Europe, avid internet users needed to navigate the web at night - when phone rates by the minute were cheaper - to cut their expenses. In 1998, some users in France, Italy and Germany launched a movement to boycott the internet one day per week, for internet providers and phone companies to set up a special monthly rate. This action paid off, and providers began to offer "internet rates".

Christiane Jadelot, a French engineer at INaLF-Nancy (INaLF: National Institute for the French Language), wrote in July 1998: "I began to really use the internet in 1994, with a browser called Mosaic. I found it a very useful way of improving my knowledge of computers,

linguistics, literature... everything. I was finding the best and the worst, but as a discerning user, I had to sort it all out, and make choices. I particularly liked the software for email, file transfers and dial-up connections. At that time, I had problems with a program called Paradox and character sets I couldn't use. I tried my luck and threw out a question in a specialist news group. I got answers from all over the world. Everyone seemed to want to solve my problem!"

The World Wide Web Consortium (W3C) was founded in October 1994 to develop interoperable technologies (specifications, guidelines, software, and tools) for the web, for example specifications for markup languages (HTML, XML, and others), and to act as a forum for information, commerce, communication and collective understanding.

The "Technorealism" movement started on the web in March 1998. Technorealism was "an attempt to assess the social and political implications of technologies so that we might all have more control over the shape of our future. The heart of the technorealist approach involves a continuous critical examination of how technologies - whether cutting-edge or mundane - might help or hinder us in the struggle to improve the quality of our personal lives, our communities, and our economic, social, and political structures" (excerpt from the website). The document *Technorealism Overview* was approved by hundreds of people signing their names. It stated that, "regardless of how advanced our computers become, we should never use them as a substitute for our own basic cognitive skills of awareness, perception, reasoning, and judgment."

The internet and other media

In 1998, people were also wondering whether the print media and the internet would be antagonistic or complementary. Would the internet swallow up the print media? Would the internet get the top place in the hearts of people buying books or subscribing to magazines? The internet was about to change books and other media in a sweeping way, like the printing presses in the past. Authors, booksellers, librarians, printers, publishers and translators were watching the storm, or participating in it in heated debates on copyright issues and distribution control.

In some African countries, the internet meant more information. The number of newspapers was very low compared to the population figures. Each copy was read by at least twenty people. In January 1997, during the Symposium on Multimedia Convergence organized by the International Labor Organization (ILO), Wilfred Kiboro, managing director of Nation Printers and Publishers, in Kenya, expressed the idea of a printing system through a satellite internet connection, instead of carrying newspapers every day by truck all over the country. This printing system would mean cheaper distribution costs, and a drop in the price of newspapers.

Did the internet compete with television and reading? In Quebec, 30.7% of the population was connected to the internet in March 1998. A poll showed that 28.8% of internet users were watching television less than before, but only 12.1% were reading less. As stated by the online magazine *Multimédium* in April 1998, this was "rather encouraging for the department of Culture and Communications which has the double task of furthering the development of

information highways... and reading!"

According to a survey for Online MSNBC in February 1998, the internet – as a new medium – was well liked, matching and sometimes surpassing other media. Merrill Brown, editor-in-chief of Online MSNBC, wrote in *Internet Wire* of February 1998: "The internet news usage behavior pattern is shaping up similar to broadcast television in terms of weekday use, and is used more than cable television, newspapers and magazines during that same period of time. Additionally, on Saturdays, the internet is used more than broadcast television, radio or newspapers, and on a weekly basis has nearly the same hours of use as newspapers." People were spending 2.4 hours per week reading magazines, 3.5 hours surfing the web, 3.6 hours reading newspapers, 4.5 hours listening the radio, 5 hours watching cable TV, and 5.7 hours watching broadcast TV.

Jean-Pierre Cloutier was the editor of *Chroniques de Cybérie*, a weekly French-language online report of internet news. When interviewed in fall 1997 by François Lemelin, chief-editor of *L'Album*, a magazine from Club Macintosh of Quebec, he expressed his views about the internet as a medium: "I think the medium is going to continue being essential, and then give birth to original, precise, specific services, by which time we will have found an economic model of viability. For information cybemedias like *Chroniques de Cybérie* as well as for info-services, community and online public services, electronic commerce, distance learning, the post-modern policy which is going to change the elected representatives / principals, in fact, everything is coming around. (...) Concerning the relationship with other media, I think we need to look backwards. Contrary to the words of alarmists in previous times, radio didn't kill music or the entertainment industry any more than the cinema did. Television didn't kill radio or cinema. Nor did home videos. When a new medium arrives, it makes some room for itself, the others adjust, there is a transition period, then a 'convergence'. What is different with the internet is the interactive dimension of the medium and its possible impact. We are still thinking about that, we are watching to see what happens.

Also, as a medium, the net allows the emergence of new concepts in the field of communication, and on the human level, too – even for non-connected people. I remember when McLuhan arrived, at the end of the sixties, with his concept of 'global village' basing itself on television and telephone, and he was predicting data exchange between computers. There were people, in Africa, without television and telephone, who read and understood McLuhan. And McLuhan changed things in their vision of the world. The internet has the same effect. It gives rise to some thinking on communication, private life, freedom of expression, the values we are attached to, and those we are ready to get rid of, and it is this effect which makes it such a powerful, important medium."

"The dream behind the web"

Tim Berners-Lee invented the web in 1990. Pierre Ruetschi, a journalist for the Swiss daily *Tribune de Genève*, asked him in December 1997: "Seven years later, are you satisfied with the way the web has evolved?". He answered that, if he was pleased with the richness and diversity of information, the web still lacked the power planned in its original design. He would like "the web to be more interactive, and people to be able to create information

together", and not only to be consumers of information. The web was supposed to become a "medium for collaboration, a world of knowledge that we share."

In a short essay posted on his webpage, Tim Berners-Lee wrote in May 1998: "The dream behind the web is of a common information space in which we communicate by sharing information. Its universality is essential: the fact that a hypertext link can point to anything, be it personal, local or global, be it draft or highly polished. There was a second part of the dream, too, dependent on the web being so generally used that it became a realistic mirror (or in fact the primary embodiment) of the ways in which we work and play and socialize. That was that once the state of our interactions was online, we could then use computers to help us analyse it, make sense of what we are doing, where we individually fit in, and how we can better work together." (excerpt from *The World Wide Web: A very short personal history*, available on the W3 website)

1993: The Online Books Page is a list of free ebooks

[Overview]

Founded in 1993 by John Mark Ockerbloom while he was a student at Carnegie Mellon University (in Pittsburgh, Pennsylvania), The Online Books Page is "a website that facilitates access to books that are freely readable over the internet. It also aims to encourage the development of such online books, for the benefit and edification of all." John Mark first maintained this page on the website of the School of Computer Science of Carnegie Mellon University. In 1999, he moved it to its present location at the University of Pennsylvania Library, where he is a digital library planner and researcher. The Online Books Page offered links to 12,000 books in 1999, 20,000 books in 2003 (including 4,000 books published by women), 25,000 books in 2006, and 30,000 books in 2008. The books "have been authored, placed online, and hosted by a wide variety of individuals and groups throughout the world", with 7,000 books from Project Gutenberg. The FAQ also gives copyright information about most countries in the world with links to further reading.

In 1993, the web was still in its infancy, with Mosaic as its first browser. John Mark Ockerbloom was a graduate student at the School of Computer Science (CS) of Carnegie Mellon University (CMU, Pittsburgh, Pennsylvania). He created The Online Books Page as "a website that facilitates access to books that are freely readable over the internet. It also aims to encourage the development of such online books, for the benefit and edification of all" (excerpt from the website).

In September 1998, John Mark wrote in an email interview: "I was the original webmaster here at CMU CS, and started our local web in 1993. The local web included pages pointing to various locally developed resources, and originally The Online Books Page was just one of these pages, containing pointers to some books put online by some of the people in our department. (Robert Stockton had made web versions of some of Project Gutenberg's texts.) After a while, people started asking about books at other sites, and I noticed that a number of sites (not just Gutenberg, but also Wiretap and some other places) had books online, and that it would be useful to have some listing of all of them, so that you could go to one place to download or view books from all over the net. So that's how my index got started. I eventually gave up the webmaster job in 1996, but kept The Online Books Page, since by then I'd gotten very interested in the great potential the net had for making literature available to a wide audience. At this point there are so many books going online that I have a hard time keeping up (and in fact have a large backlog of books to list). But I hope to keep up my online books works in some form or another. I am very excited about the potential of the internet as a mass communication medium in the coming years. I'd also like to stay involved, one way or another, in making books available to a wide audience for free via the net, whether I make this explicitly part of my professional career, or whether I just do it as a spare-time volunteer."

In 1998, there was an index of 7,000 etexts that could be browsed by author, title or subject. There were also pointers to significant directories and archives of online texts, and to special

exhibits. From the main search page, users could search in four types of media: books, music, art, and video.

"Along with books, The Online Books Page is also now listing major archives of serials (such as magazines, published journals, and newspapers) (...). Serials can be at least as important as books in library research. Serials are often the first places that new research and scholarship appear. They are sources for firsthand accounts of contemporary events and commentary. They are also often the first (and sometimes the only) place that quality literature appears. (For those who might still quibble about serials being listed on a 'books page', back issues of serials are often bound and reissued as hardbound 'books'.)" (excerpt from the 1998 website)

In 1999, after graduating from Carnegie Mellon with a Ph.D. in computer science, John Mark moved to work as a digital library planner and researcher at the University of Pennsylvania Library. He also moved The Online Books Page there, kept it as clear and simple, and went on expanding it.

The Online Books Page offered links to 12,000 ebooks in 1999, 20,000 ebooks in 2003 (including 4,000 ebooks published by women), 25,000 ebooks in 2006, and 30,000 ebooks in 2008. The books "have been authored, placed online, and hosted by a wide variety of individuals and groups throughout the world", with 7,000 books from Project Gutenberg. The FAQ lists copyright information about most countries in the world, with links to further reading.

1994: Some publishers get bold and go digital

[Overview]

Some bold publishers decided to use the web as a marketing tool. In the U.S., NAP (National Academy Press) was the first publisher in 1994 to post the full text of some books, for free, with the authors' consent. NAP was followed by MIT Press in 1995. Michael Hart, founder of Project Gutenberg, wrote in 1997: "As university publishers struggle to find the right business model for offering scholarly documents online, some early innovators are finding that making a monograph available electronically can boost sales of hard copies" (excerpt from the Project Gutenberg Newsletter of October 1997). Digital publishing became mainstream in 1997. Digitization accelerated the publication process. Editors, designers and other contributors could all work at the same time on the same book. For educational, academic and scientific publications, digital publishing was a cheaper solution than print books, with regular updates to include the latest information.

Publishers get bold

Some publishers decided to use the web as a marketing tool. In the U.S., NAP (National Academy Press) was the first publisher in 1994 to post the full text of some books, for free, with the authors' consent. NAP was followed by MIT Press (MIT: Massachusetts Institute of Technology) in 1995.

NAP was created by the National Academy of Sciences to publish its own reports and the ones of the National Academy of Engineering, the Institute of Medicine, and the National Research Council. In 1994, NAP was publishing 200 new books a year in science, engineering, and health. The new NAP Reading Room offered 1,000 entire books, available online for free in various formats: "image" format, HTML format and PDF format. Oddly enough, there was no drop in sales - on the contrary, sales increased.

In 1995, MIT Press was publishing 200 new books per year and 40 journals, in science and technology, architecture, social theory, economics, cognitive science, and computational science. MIT Press also decided to put a number of books online for free, as "a long-term commitment to the efficient and creative use of new technologies". Sales of print books with a free online version increased.

Michael Hart, founder of Project Gutenberg, wrote in 1997: "As university publishers struggle to find the right business model for offering scholarly documents online, some early innovators are finding that making a monograph available electronically can boost sales of hard copies. The National Academy Press has already put 1,700 of its books online, and is finding that the electronic versions of some books have boosted sales of the hard copy monographs - often by two to three times the previous level. It's 'great advertising', says the Press's director. The MIT Press is experiencing similar results: 'For each of our electronic books, we've approximately doubled our sales. The plain fact is that no one is going to sit there and read a whole book online. And it costs money and time to download it'." (excerpt from the *Project Gutenberg Newsletter* of October 1997)

Publishers go digital

Digital publishing became mainstream in 1997, as the latest step in the many changes underwent by traditional publishing since the 1970s. Traditional printing was first disrupted by new photocomposition machines, with lower costs. Text and image processing began to be handed over to desktop publishing and graphic art studios. Impression costs went on decreasing with photocopiers, color photocopiers and digital printing. Digitization also accelerated the publication process. Editors, designers and other contributors could all work at the same time on the same book.

For educational, academic and scientific publications, online publishing became a cheaper solution than print books, with regular updates to include the latest information. Readers didn't need any more to wait for a new printed edition, often postponed if not cancelled because of commercial constraints. Some universities began to create their own textbooks online, with chapters selected in an extensive database, as well as papers and comments from professors. For a seminar, a few print copies could be made upon request, with a selection of online articles sent to a printer.

Digital publishing and traditional publishing became complementary. The frontier between the two supports - electronic and paper - began to vanish. Recent print media already stem from an electronic version anyway, on a word processor, a spreadsheet or a database. More and more documents became "only" electronic, and more and more print books were digitized to be included in digital libraries and bookstores.

In the mid-1990s, though, there was no proof that electronic documents would make us paperless in the near future, and save some trees. Many people still needed a print version for easier reading, or for their archives, in the fear the electronic file would be accidentally deleted. We were still in a transition period, from paper to digital.

1995: Amazon.com is the first main online bookstore

[Overview]

The online bookstore Amazon.com was launched by Jeff Bezos in July 1995, in Seattle, on the West coast of the U.S., after a market study which led him to conclude that books were the best "products" to sell on the internet. When Amazon.com started, it had 10 employees and a catalog of 3 million books. Unlike traditional bookstores, Amazon doesn't have windows looking out on the street and books skillfully lined up on shelves or piled upon displays. The "virtual" windows are its webpages, with all transactions made through the internet. Books are stored in huge storage facilities before being put into boxes and sent by mail. In November 2000, Amazon had 7,500 employees, a catalog of 28 million items, 23 million clients worldwide and four subsidiaries in United Kingdom (launched in August 1998), Germany (August 1998), France (August 2000) and Japan (November 2000). A fifth subsidiary opened in Canada in June 2002, and a sixth subsidiary, named Joto, opened in China in September 2004.

Amazon in the U.S.

First steps

The online bookstore Amazon.com was launched by Jeff Bezos in July 1995, in Seattle, on the West coast of the U.S., after a market study which led him to conclude that books were the best products to sell on the internet. When Amazon.com started, it had 10 employees and a catalog of 3 million books. Unlike traditional bookstores, Amazon.com didn't have windows looking out on the street and books skillfully lined up on shelves or piled upon displays. The "virtual" windows are its webpages, with all transactions made through the internet. Books are stored in huge storage facilities before being put into boxes and sent by mail.

What exactly was the idea behind Amazon.com? In Spring 1994, Jeff Bezos drew up a list of twenty products that could be sold online, ranging from clothing to gardening tools, and then researched the top five, which were CDs, videos, computer hardware, computer software, and books.

As recalled by Jeff Bezos in Amazon's press kit (in its 1998 version), "I used a whole bunch of criteria to evaluate the potential of each product, but among the main criteria was the size of the relative markets. Books, I found out, were an \$82 billion market worldwide. The price point was another major criterion: I wanted a low-priced product. I reasoned that since this was the first purchase many people would make online, it had to be non-threatening in size. A third criterion was the range of choice: there were 3 million items in the book category and only a tenth of that in CDs, for example. This was important because the wider the choice, the more the organizing and selection capabilities of the computer could be put in good use."

People could search the online catalog by author, title, subject, date, or ISBN. The website was offering excerpts from books, book reviews, customer reviews, and author interviews. People could "leaf" through extracts and reviews, order some books online, and pay with their credit card. Books arrived within a week at their doorstep. As an online retailer, Amazon.com could offer lower prices than local bookstores, a larger selection, and a wealth

of product information. Customers could subscribe to a mailing list to get reviews of new books by their favorite authors, or new books in their favorite topics, with 44 topics to choose from.

In 1998, there were discounts on 400,000 titles, with 40% on some feature books, 30% on hardcovers, and 20% on paperbacks. Amazon.com was also selling CDs, DVDs, audiobooks and computer games, with 3 million clients in 160 countries, and a catalog with tentimes as many titles as the largest supermarkets' bookstores.

As mentioned by Jeff Bezos in Amazon's press kit: "Businesses can do things on the web that simply cannot be done any other way. We are changing the way people buy books and music. Our leadership position comes from our obsessive focus on customers. (...) Customers want selection, ease of use, and the lowest prices. These are the elements we work hard to provide. We continued to improve our customer experience during the quarter [the second quarter 1998] with the opening of our music store, our easier-to-navigate store layout, and our expansion into the local U.K. and German book markets. These initiatives will continue to require aggressive investment and entail significant execution challenges."

Expansion

People began buying books across borders. What we take for granted now - buy a book in Europe from the U.S. website Amazon.com, or buy a book in the U.S. from the German website Amazon.de - was making big waves at the time. The local online bookstores complained about "unfair competition".

There were also issues about custom taxes. A first outline agreement was reached between the U.S. and the European Union in December 1997. This agreement was followed by an international convention. The internet was decided a free trade area, i.e. without any custom taxes for software, films and digital books bought online. Material goods (books, CDs, DVDs) and services were subject to existing regulations, with collection of VAT (value added tax) for example, but with no additional custom taxes.

On the footsteps of the Internet Bookstore, based in United Kingdom and the largest online bookstore in Europe, Amazon.com launched its Associates Program. As stated in a press release dated June 8, 1998: "The Amazon.com Associates Program allows website owners to easily participate in hassle-free electronic commerce by recommending books on their site and referring visitors to Amazon.com. In return, participants earn referral fees of up to 15 percent of the sales they generate. Amazon.com handles the secure online ordering, customer service, and shipping and sends weekly email sales reports. Enrollment in the program is free, and participants can be up and running the same day. Associates range from large and small businesses to nonprofits, authors, publishers, personal home pages, and more. The popularity of the program is reflected in the range of additions to the Associates Community in the past few months: Adobe, InfoBeat, Kemper Funds, PR Newswire, Travelocity, Virtual Vineyards, and Xoom." There were 60,000 "associates" in June 1998.

Barnes & Noble, a leading U.S. bookseller, entered the world of e-commerce in 1997. Barnes & Noble had 481 stores nationwide in 1997, in 48 states out of 50, as well as 520 bookstores

(B. Dalton stores) in shopping malls, and a catalog of 175,000 titles from 20,000 publishers. Barnes & Noble also published books under its own imprint for exclusive sale through its retail stores and its nationwide mail-order catalogs.

Barnes & Noble first launched its America OnLine (AOL) website in March 1997 - as the exclusive bookseller for the 12 million AOL customers -, before launching its own website, barnesandnoble.com, in May 1997. The site was offering reviews from authors and publishers, with a catalog of 630,000 titles available for immediate shipping, and significant discounts: 30% off all in-stock hardcovers, 20% off all in-stock paperbacks, 40% off select titles, and up to 90% off bargain books. Its Affiliate Network spread quickly, with 12,000 affiliate websites in May 1998, including CNN Interactive, Lycos and ZDNet.

In May 1998, Barnes & Noble.com launched a revamped website with a better design and an Express Lane one-click ordering, improved book search capabilities, and expanded product offerings with a new software "superstore". Jeff Killeen, chief operating officer, stated in a press release dated May 27, 1998: "Through our first year in business we have listened intently to what our customers have asked for and believe we have delivered a vastly superior product based on those requests. (...) Innovation based on customer-focus has been the hallmark of our success and we see our new site as proof-positive of our commitment to be the leader in online bookselling and related products. We're also extremely excited to have Intel, a leader in the technology products category, open its SoftwareForPCs.com site at barnesandnoble.com."

Barnes & Noble.com began a fierce price war with Amazon.com for the best book discounts. Amazon.com came to be known as "Amazon.toast". Jeff Bezos didn't mind the competition. In the magazine *Success* of July 1998, he explained to journalist Lesley Hazleton: "The gap has increased rather than decreased. We went from \$60 million annualized sales revenue in May to \$260 million by the end of the year, and from 340,000 customers to 1.5 million, 58 percent of them repeat customers - all that in the context of 'Amazon.toast'. We're doing more than eight times the sales of Barnes & Noble. And we're not a stationary target. We were blessed with a two-year head start, and our goal is to increase that gap."

Amazon in Europe

The European presence of Amazon began in October 1998, with the creation of two subsidiaries in Germany and in United Kingdom.

In August 2000, Amazon had 1.8 million customers in U.K., 1.2 million customers in Germany, and less than 1 million customers in France. Amazon opened its third subsidiary, Amazon France, with books, music, DVDs and videos - software and video games were added later, in June 2001 - and a 48-hour delivery. At the time, online sales represented only 0.5% of the book market in France, against 5.4% in the United States.

The opening of Amazon France was announced at the last minute, on August 23, 2000, after months of secrecy surrounding the next "American cultural invasion". The French subsidiary opened in Guyancourt, in the suburbs of Paris, with 100 employees - some of them trained in the U.S. headquarters in Seattle - for administration, technical services, and marketing. The

distribution service opened in Boigny-sur-Bionne, near Orleans, a town in the south of Paris. The customer service landed in The Hague, Netherlands, because Amazon was expecting to broaden its European network.

Amazon France had four competitors: Fnac.com, Alapage, Chapitre.com, and BOL.fr.

Fnac.com was the online branch of Fnac , a network of “traditional” bookstores spread throughout France and other European countries, and run by the group Pinault-Printemps-Redoute.

Alapage was an online bookstore founded in 1996 by Patrice Magnard, before being bought by France Telecom in September 1999. Alapage became a subsidiary of Wanadoo, the internet service provider of France Telecom, in July 2000.

Chapitre.com was an independent online bookstore, created in 1997 by Juan Pirlot de Corbion.

BOL.fr was the French subsidiary of BOL.com (BOL: Bertelsmann On Line), launched in August 1999 by Bertelsmann, a German media giant, in partnership with Vivendi, a French multinational company.

Unlike their counterparts in the U.S. and in U.K., where book prices were free, French online bookstores couldn't offer significant bargains. A French law – the Lang law - regulated prices. (Jacques Lang was the ministry of culture who fathered the law to protect independent bookstores.) The 5% discount allowed by law for both traditional and online bookstores was offering little latitude to Amazon.fr, Fnac.com, and the likes, who were nevertheless optimistic about the prospects offered by the French-language international market. A significant number of orders was already coming from abroad, with 10% of orders for Fnac.com as early as 1997.

Interviewed by AFP (Agence France-Presse) on the Lang Law and the meager 5% discount allowed for book prices, Denis Terrien, president of Amazon France (until May 2001), explained in August 2000: "Our experience in Germany, where book prices are also regulated, shows that prices are not the main factor for our customers to purchase books at Amazon. The main factor resides in the additional services we provide. We offer a whole bunch of services, beginning with a large choice in our catalog- we sell all the French cultural products. We have a powerful search engine. As for music, our site offers the only catalog searchable by song title. In addition to the editorial content of our site, which ranges from the one of a traditional bookstore to the one of a magazine, we have a customer service 24h/24 7days/7, something unique in the French market. Finally, an additional specificity of Amazon is our commitment for a fast delivery. We aim to have more than 90% of our products in stock (at our storage facility)."

Amazon's economic model was already admired by many in Europe, but could hardly be considered a model too for staff management, with short-term labor contracts, low wages, and poor working conditions.

Despite the secrecy surrounding the working conditions of the European staff, problems began to filter. In November 2000, the Prewitt Organizing Fund and the French union SUD-PTT Loire Atlantique launched an awareness campaign among the employees of Amazon France, after meeting with a group of 50 employees in the distribution center of Boigny-sur-Bionne. In a statement following the meeting, SUD-PTT denounced "degraded working conditions, flexible schedules, short-term labor contracts in periods of flux, low wages, and minimal social guarantees". Similar action was conducted in Germany and in U.K. Patrick Moran, head of the Prewitt Organizing Fund, founded an employee organization under the name of Alliance of New Economy Workers. In response, Amazon sent internal memos to its employees, stressing the pointlessness of unions within the company.

At the end of January 2001, Amazon, which employed 1,800 people in Europe, announced a 15% reduction of its European staff. It also closed its customer service center in The Hague (Netherlands). Its 240 employees were offered to work in one of the two other European customer service centers, in Slough (United Kingdom) and in Regensburg (Germany).

Amazon worldwide

The second group of foreign clients - after European customers - was in Japan. In July 2000, during an international symposium on information technology in Tokyo, Jeff Bezos announced his intention to launch Amazon Japan in the near future. He insisted on the high potential of the Japanese market, with expensive real estate affecting the prices of goods and services and, as a result, online shopping being more convenient than traditional shopping. High population density would mean easy and cheap home deliveries.

A Japanese call center opened in August 2000 in Sapporo, a city on the Hokkaido island. Amazon Japan opened three months later, in November 2000, as the fourth subsidiary of Amazon and first non-European subsidiary, with a catalog of 1.1 million titles in Japanese and 600,000 titles in English. To reduce delivery times to 24 to 48 hours instead of six weeks for books published in the U.S., a large distribution center (15,800 m²) was created in Ichikawa, a town in the east of Tokyo.

In November 2000, Amazon had 7,500 employees, a catalog of 28 million items, and 23 million clients worldwide. It opened its digital library with 1,000 ebooks, and the promise of many more titles for soon.

Amazon also began focusing on the French-language market in Canada. It hired staff knowing the language and the market, to be able to offer French-language books, music and films (VHS and DVD) in a Canadian subsidiary. Amazon Canada, the fifth subsidiary of the company, was launched in June 2002 with a bilingual (English, French) website.

Surprisingly, even for the marketing of a main online bookstore, paper was not dead. For two consecutive years, in 1999 and 2000, Amazon sent a print catalog to its customers (10 million in 2000) before the holiday season.

2001 marked a turning point for the company, with the need to address the internet bubble affecting the "new" economy and so many companies. Following a deficit for the fourth quarter 2000, Amazon reduced its workforce by 15% in January 2001. 1,300 employees lost their jobs in the U.S. 270 employees lost their jobs in Europe. Jeff Bezos decided to diversify the products sold online, and to sell not only books, videos, CDs and software, but also health care products, toys, electronics, kitchen utensils, and garden tools. In November 2001, cultural products - books, CDs and videos – represented only 58% of sales, the total of which were US \$4 billion, with 29 million customers.

The company was beneficiary for the first time in the third quarter 2003.

In October 2003, Amazon launched a full text search (Search Inside the Book) after scanning the text of 120,000 titles, with many more to come. Amazon also launched its own search engine, A9.com.

A sixth subsidiary - named Joyo - opened in China in September 2004.

The net income of Amazon was US \$588 million for 2004 - 45% of which from its six subsidiaries (Canada, China, France, Germany, Japan, U.K.) -, with a total of \$6.9 billion for sales.

Amazon became a reference for global online commerce.

In July 2005, for its 10-year anniversary, Amazon had 9,000 employees, and 41 million clients enjoying attractive prices for a whole range of products they could get within 48 hours in one of the seven countries with an Amazon platform.

Amazon also sold more and more ebooks. In April 2005, it bought the French company Mobipocket, specializing in readers (software) and ebooks for PDAs.

In November 2007, Amazon launched its own reading device, named Kindle, with a catalog of 80,000 ebooks on Amazon's website. 538,000 Kindle were sold in 2008. A new version of Kindle, named Kindle 2, was launched in February 2009, with a catalog of 230,000 ebooks.

What about small bookstores?

Local bookstores have closed one after the other, or have had a hard time keeping up with the competition of Amazon.com and other online bookstores. Amazon and others are also bad news for specialist bookstores, for example the travel bookstore created in 1971 by Catherine Domain in Paris, France.

According to Catherine, Librairie Ulysse (Ulysses Bookstore) is the oldest travel bookstore in the world. Its 20,000 out-of-print or new books, maps and magazines - in a number of languages and about any country – are all packed up in a tiny space, in the heart of Paris, on Ile Saint-Louis, a small island surrounded by the Seine river.

Catherine has been a traveller since she was a child. She travels every summer - usually sailing on the Mediterranean, the Atlantic or the Pacific - while her boyfriend runs the bookstore. She is also a member of the French National Union of Antiquarian and Modern Bookstores (SLAM: Syndicat national de la librairie ancienne et moderne), the Explorers' Club (Club des explorateurs) and the International Club of Long-Distance Travelers (Club international des grands voyageurs).

Catherine visited 140 countries, and some trips were quite challenging. But her most difficult challenge was to set up a website on her own, from scratch, without knowing anything about computers. In December 1999, she wrote in an email interview: "My site is still pretty basic and under construction. Like my bookstore, it is a place to meet people before being a place of business. The internet is a pain in the neck, takes a lot of my time and I earn hardly any money from it, but that doesn't worry me... I am very pessimistic though, because the internet is killing off specialist bookstores."

Some booksellers decided to run most of their business online, for example Pierre Joppen and his wife Joke Vrijenhoek, the owners of Paulus Swaen Old Maps and Prints, a bookstore founded in 1978 in the Netherlands that relocated in 1996 in Florida. The bookstore offers maps, atlases and globes ranging from the 16th to the 18th century. The maps cover all parts of the world, and were produced by renowned cartographers, such as Ortelius, Mercator, Blaeu, Janssonius, Hondius, Visscher, de Wit, etc. The bookstore has also sold travel books and Medieval manuscripts. It has offered an online internet auction since November 1996, first twice a year, in March and November, and then four times a year, in March, May, September and November.

1996: There are more and more texts online

[Overview]

Created in 1992, the Etext Archives were "home to electronic texts of all kinds". Created in 1993, the E-zine-list was a list of electronic zines around the world. The first electronic versions of print newspapers were available in the early 1990s through commercial services like America Online and CompuServe. In 1996, newspapers and magazines began offering websites with a partial or full version of their latest issue, available freely or through subscription (free or paid), as well as online archives. In United Kingdom, the daily Times and the Sunday Times set up a common website called Times Online. The weekly publication The Economist also went online, as well as the weekly Focus and Der Spiegel in Germany, the daily Le Monde and Libération in France, and the daily El País in Spain. The computer press went logically online as well, first the monthly Wired, "the magazine of the future at the avant-garde of the 21st century", then ZDNet, another leading computer magazine. More and more "only" electronic magazines were also created.

Electronic texts and newsletters

The Etext Archives were founded in 1992 by Paul Southworth, and hosted on the website of the University of Michigan, They were "home to electronic texts of all kinds, from the sacred to the profane, and from the political to the personal". They provided electronic texts without judging their content, in six sections: (a) "E-zines": electronic periodicals from the professional to the personal; (b) "Politics": political zines, essays, and home pages of political groups; (c) "Fiction": publications of amateur authors; (d) "Religion": mainstream and off-beat religious texts; (e) "Poetry": an eclectic mix of mostly amateur poetry; and (f) "Quartz": the archive formerly hosted at quartz.rutgers.edu.

As recalled on the website in 1998: "The web was just a glimmer, gopher was the new hot technology, and FTP was still the standard information retrieval protocol for the vast majority of users. The origin of the project has caused numerous people to associate it with the University of Michigan, although in fact there has never been an official relationship and the project is supported entirely by volunteer labor and contributions. The equipment is wholly owned by the project maintainers. The project was started in response to the lack of organized archiving of political documents, periodicals and discussions disseminated via Usenet on newsgroups such as alt.activism, misc.activism.progressive, and alt.society.anarchy. The alt.politics.radical-left group came later and was also a substantial source of both materials and regular contributors. Not long thereafter, electronic 'zines (e-zines) began their rapid proliferation on the internet, and it was clear that these materials suffered from the same lack of coordinated collection and preservation, not to mention the fact that the lines between e-zines (which at the time were mostly related to hacking, phreaking, and internet anarchism) and political materials on the internet were fuzzy enough that most e-zines fit the original mission of The Etext Archives. One thing led to another, and e-zines of all kinds - many on various cultural topics unrelated to politics - invaded the archives in significant volume."

Another list, the E-zine-list, was launched by John Labovitz in summer 1993 to list e-zines around the world, accessible via FTP, gopher, email, the web, and other services. The list was updated monthly.

What exactly is a zine? John Labovitz explained on his website: "For those of you not acquainted with the zine world, 'zine' is short for either 'fanzine' or 'magazine', depending on your point of view. Zines are generally produced by one person or a small group of people, done often for fun or personal reasons, and tend to be irreverent, bizarre, and/or esoteric. Zines are not 'mainstream' publications - they generally do not contain advertisements (except, sometimes, advertisements for other zines), are not targeted towards a mass audience, and are generally not produced to make a profit. An 'e-zine' is a zine that is distributed partially or solely on electronic networks like the internet."

3,045 zines were listed in November 1998. John wrote on his website: "Now the e-zine world is different. The number of e-zines has increased a hundredfold, crawling out of the FTP and gopher woodworks to declaring themselves worthy of their own domain name, even asking for financial support through advertising. Even the term 'e-zine' has been co-opted by the commercial world, and has come to mean nearly any type of publication distributed electronically. Yet there is still the original, independent fringe, who continue to publish from their heart, or push the boundaries of what we call a 'zine'." After many years of maintaining this list, John passed the torch to others.

Chroniques de Cybérie was launched in November 1994 by Jean-Pierre Cloutier, a journalist living in Montreal, Quebec. As a weekly French-language report of internet news, Jean-Pierre's newsletter was sent by email to its subscribers (free subscription), and available on the web on a dedicated website (from April 1995). Bruno Giussani, journalist, wrote in *The New York Times* of November 25, 1997: "Almost no one in the United States has ever heard of Jean-Pierre Cloutier, yet he is one of the leading figures of the French-speaking internet community. For the last 30 months Cloutier has written one of the most intelligent, passionate and insightful electronic newsletters available on the internet, (...) an original mix of relevant internet news, clear political analysis and no-nonsense personal opinions. It was a publication that gave readers the feeling that they were living week after week in the intimacy of a planetary revolution."

Venezuela Analítica was a Spanish-language electronic magazine conceived as a public forum to exchange ideas on politics, economics, culture, science and technology. Roberto Hernández Montoya, its editor, wrote in September 1998: "The internet has been very important for me personally. It became my main way of life. As an organization it gave us the possibility to communicate with thousands of people, which would have been economically impossible if we had published a paper magazine. I think the internet is going to become the essential means of communication and of information exchange in the coming years."

Print magazines go online

The first electronic versions of print newspapers were available in the early 1990s through commercial services like America Online and CompuServe.

In 1996, newspapers and magazines began offering websites with a partial or full version of their latest issue, available freely or through subscription (free or paid), as well as online archives.

For example, the site of *The New York Times* could be accessed free of charge, with articles of the print daily newspaper, breaking news updated every ten minutes, and original reporting only available online. The site of *The Washington Post* gave the daily news online, with a full database of articles, with images, sound and video.

In United Kingdom, the daily *Times* and the *Sunday Times* set up a common website called Times Online, with a way to create a personalized edition. The weekly publication *The Economist* went online, as well as the daily *Le Monde* and *Libération* in France, the daily *El País* in Spain, and the weekly *Focus* and *Der Spiegel* in Germany.

The computer press went logically online as well, first the monthly *Wired*, created in 1992 in California to cover cyberculture as "the magazine of the future at the avant-garde of the 21st century", then *ZDNet*, as a leading computer online magazine.

"More than 3,600 newspapers now publish on the internet", Eric K. Meyer stated in late 1997 in an essay published on the website of AJR/NewsLink. "A full 43% of all online newspapers now are based outside the United States. A year ago, only 29% of online newspapers were located abroad. Rapid growth, primarily in Canada, the United Kingdom, Norway, Brazil and Germany, has pushed the total number of non-U.S. online newspapers to 1,563. The number of U.S. newspapers online also has grown markedly, from 745 a year ago to 1,290 six months ago to 2,059 today. Outside the United States, the United Kingdom, with 294 online newspapers, and Canada, with 230, lead the way. In Canada, every province or territory now has at least one online newspaper. Ontario leads the way with 91, Alberta has 44, and British Columbia has 43. Elsewhere in North America, Mexico has 51 online newspapers, 23 newspapers are online in Central America and 36 are online in the Caribbean. Europe is the next most wired continent for newspapers, with 728 online newspaper sites. After the United Kingdom, Norway has the next most - 53 - and Germany has 43. Asia (led by India) has 223 online newspapers, South America (led by Bolivia) has 161 and Africa (led by South Africa) has 53. Australia and other islands have 64 online newspapers."

The online versions of these newspapers brought us a wealth of information. The web provided not only news available online, but also a whole encyclopedia to help us understand them. As readers, we could click on hyperlinks to get maps, biographies, official texts, political and economic data, photographs, and audio and video coverage. We could easily access other articles on the same topic with search engines sorting out articles by date, author, title, or subject.

1997: Multimedia convergence and employment

[Overview]

More and more people were using digital technology. Previously distinct information-based industries, such as printing, publishing, graphic design, media, sound recording and film making, were converging into one industry, with information as a common product. This trend was named "multimedia convergence", with a massive loss of jobs, and a serious enough issue to be tackled by the ILO (International Labor Organization) by 1997. The first ILO Symposium on Multimedia Convergence was held in January 1997 at ILO headquarters in Geneva, Switzerland, with employers, unionists, and government representatives from all over the world. Some participants, mostly employers, demonstrated the information society was generating or would generate jobs, whereas other participants, mostly unionists, demonstrated there was a rise in unemployment worldwide, that should be addressed right away through investment, innovation, vocational training, computer literacy, retraining, and fair labor rights, including for teleworkers.

As explained in the introduction of the symposium's proceedings: "With the advent of digitalization, technological convergence has been set into motion. Today all forms of information - whether based in text, sound or images - can be converted into bits and bytes for handling by computer. Digitalization has made it possible to create, record, manipulate, combine, store, retrieve and transmit information and information-based products in ways which magnetic tape, celluloid and paper did not permit. Digitalization thus allows music, cinema and the written word to be recorded and transformed through similar processes and without distinct material supports. Previously dissimilar industries, such as publishing and sound recording, now both produce CD-ROMs, rather than simply books and records. (...)

Multimedia convergence deserves our attention for reasons which go far beyond the entertainment, mass media and telecommunications industries. The technological revolution which has made multimedia convergence possible will continue apace, creating new configurations among an ever-widening range of industries. The digitalization of information processing and delivery is transforming the way financial systems operate, the way enterprises exchange information internally and externally, and the way individuals work in an increasingly electronic environment."

Held in January 1997 at the ILO headquarters in Geneva, Switzerland, the three-day Symposium on Multimedia Convergence intended to discuss the social and labor issues arising from this process. The industry-centred debates focused on three main concerns: (a) the information society: what it means for governments, employers and workers; (b) the convergence process: its impact on employment and work; and (c) labor relations in the information age. The purpose of these debates was "to stimulate reflection on the policies and approaches most apt to prepare our societies and especially our workforces for the turbulent transition towards an information economy."

One of the participants, Peter Leisink, an associate professor of labor studies at the Utrecht University, Netherlands, explained: "A survey of the United Kingdom book publishing industry showed that proofreaders and editors have been externalized and now work as home-based teleworkers. The vast majority of them had entered self-employment, not as a first-choice option, but as a result of industry mergers, relocations and redundancies. These people should actually be regarded as casualized workers, rather than as self-employed, since they have little autonomy and tend to depend on only one publishing house for their work."

Wilfred Kiboro, managing director of Nation Printers and Publishers, Kenya, made the following comments: "In content creation in the multimedia environment, it is very difficult to know who the journalist is, who the editor is, and who the technologist is that will bring it all together. At what point will telecom workers become involved as well as the people in television and other entities that come to create new products? Traditionally in the print media, for instance, we had printers, journalists, sales and marketing staff and so on, but now all of them are working on one floor from one desk."

Formerly, the production staff was keying in the articles, and not the editorial staff. Journalists and editors could now type in their articles online, and these articles went directly from text to layout. In book publishing, digitization speeded up the editorial process, which used to be sequential, by allowing the copyeditor, the image editor and the layout staff to work at the same time on the same book.

Michel Muller, secretary-general of the French Federation of Book, Paper and Communication Industry (Fédération des industries du livre, du papier et de la communication), stated that, in France, jobs in this industry fell from 110,000 to 90,000 in the last decade (1987-1996), with expensive social plans to re-train and re-employ the 20,000 people who lost their jobs.

He also explained that, "if the technological developments really created new jobs, as had been suggested, then it might have been better to invest the money in reliable studies about what jobs were being created and which ones were being lost, rather than in social plans which often created artificial jobs. These studies should highlight the new skills and qualifications in demand as the technological convergence process broke down the barriers between the printing industry, journalism and other vehicles of information. Another problem caused by convergence was the trend towards ownership concentration. A few big groups controlled not only the bulk of the print media, but a wide range of other media, and thus posed a threat to pluralism in expression. Various tax advantages enjoyed by the press today should be re-examined and adapted to the new realities facing the press and multimedia enterprises. Managing all the social and societal issues raised by new technologies required widespread agreement and consensus. Collective agreements were vital, since neither individual negotiations nor the market alone could sufficiently settle these matters."

Quite theoretical compared to the unionists' concerns was the answer of Walter Durling, director of AT&T Global Information Solutions (United States): "Technology would not change the core of human relations. More sophisticated means of communicating, new mechanisms for negotiating, and new types of conflicts would all arise, but the relationships between workers and employers themselves would continue to be the same. When film was invented,

people had been afraid that it could bring theatre to an end. That has not happened. When television was developed, people had feared that it would do away cinemas, but it had not. One should not be afraid of the future. Fear of the future should not lead us to stifle creativity with regulations. Creativity was needed to generate new employment. The spirit of enterprise had to be reinforced with the new technology in order to create jobs for those who had been displaced. Problems should not be anticipated, but tackled when they arose." In short, humanity shouldn't fear technology

In fact, employees were not so much afraid of technology as they were afraid of losing their jobs. In 1996, unemployment was already significant in any field, which was not the case when film and television were invented. What would be the balance between job creation and lay-off in the near future? Unions were struggling worldwide to promote the creation of jobs through investment, innovation, vocational training, computer literacy, retraining for new jobs in digital technology, fair conditions for labor contracts and collective agreements, defense of copyright for the re-use of articles from the print media to the web, protection of workers in the artistic field, and defense of teleworkers as workers having full rights.

The European Commission was expecting 10 million teleworkers in Europe by the year 2000, which would represent 20% of teleworkers worldwide.

Despite unions' efforts, would the situation become as tragic as suggested in a note of the symposium's proceedings? "Some fear a future in which individuals will be forced to struggle for survival in an electronic jungle. And the survival mechanisms which have been developed in recent decades, such as relatively stable employment relations, collective agreements, employee representation, employer-provided job training, and jointly funded social security schemes, may be sorely tested in a world where work crosses borders at the speed of light."

Twelve years later, outsourcing has become a "standard" in information technology, to cut the costs. How many companies care about fair labor conditions for the employees of their outsourcing partners?

1998: Libraries take over the web

[Overview]

The first library website was the one created by the Helsinki City Library in Finland, which went live in February 1994. Four years later, in 1998, more and more traditional libraries had a website as a new "virtual" window for their patrons and beyond. Patrons could check opening hours, browse the online catalog, and surf on a broad selection of websites on various topics. Libraries developed digital libraries alongside their standard collections, for a large audience to be able to access their specialized, old, local and regional collections, including images and sound. Librarians could now fulfill two goals that used to be in contradiction - preservation (on shelves) and communication (on the internet). Library treasures went online, like Beowulf on the website of the British Library. Beowulf is the earliest known narrative poem in English, and one of the most famous works of Anglo-Saxon poetry. The British Library holds the only known manuscript of Beowulf, dated circa 1000, and digitized it for the world to enjoy.

Libraries create websites

Libraries began creating websites as a "virtual" window, as well as digital libraries stemming from their print collections. Thousands of public works, literary and scientific articles, pictures and sound tracks became available on the screen for free.

On the one hand, books were taken out of their shelves only once to be scanned. On the other hand, books could easily be accessed anywhere at any time, without the need to go to the library and struggle through a lengthy process to access the original books, because of reduced opening hours, forms to fill out, safety concerns for rare and fragile books, and shortage of staff. Some researchers still remember the unfailing patience and an out-of-the-ordinary determination they needed to finally get to a given book in some cases. People could now access digital facsimiles, and access the original books only when needed.

Before broadband internet became mainstream, full-screen images were quite long to appear on the screen. After enthusiastically posting large image files, librarians decided to post small images that people could either see as is, or click on to get a larger format.

Some amazing image collections went online, for example American Memory, as "an effort to digitize and deliver electronically the distinctive, historical Americana holdings at Library of Congress, including photographs, manuscripts, rare books, maps, recorded sound, and moving pictures".

SPIRO (Slide and Photograph Image Retrieval Online) was the Visual Online Public Access Catalog (VOPAC) for UC (University of California) Berkeley's Architecture Slide Library (ASL) collection of 200,000 35mm slides.

IMAGES 1 was the database of the Pictorial Collection at the National Library of Australia, with 15,000 historical and contemporary images relating to Australia and its influence in the world, including paintings, drawings, rare prints, objects and photographs.

Librarians also helped patrons to surf on the web without being drowned, and to find the information they needed at a time search engines were less accurate. Library catalogs went online. Some patrons were already hoping that online catalogs would no longer only be a list of bibliographic records, and a prelude to a lengthy process to find the document itself if it didn't belong to their library - forms to fill out for interlibrary loan, fees to pay in some cases, and a long waiting period to finally get the book. They were hoping that, some day, bibliographic catalogs would give instant online access to the full text of books and journals.

Gabriel in Europe

Gabriel - an acronym for "Gateway and Bridge to Europe's National Libraries" - was launched as a trilingual (English, French, German) website by the Conference of European National Librarians (CENL).

As stated on the website in 1998: "Gabriel also recalls Gabriel Naudé, whose *Advis pour dresser une bibliothèque* (Paris, 1627) is one of the earliest theoretical works about libraries in any European language and provides a blueprint for the great modern research library. The name Gabriel is common to many European languages and is derived from the *Old Testament*, where Gabriel appears as one of the archangels or heavenly messengers. He also appears in a similar role in the *New Testament* and the *Qu'ran*."

In 1998, 38 national libraries participated in Gabriel: the ones of Albania, Austria, Belgium, Bulgaria, Croatia, Czech Republic, Denmark, Estonia, Finland, France, Germany, Greece, Hungary, Iceland, Ireland, Italy, Latvia, Liechtenstein, Lithuania, Luxembourg, (Former Yugoslav Republic of) Macedonia, Malta, Netherlands, Norway, Poland, Portugal, Romania, Russia, San Marino, Slovakia, Slovenia, Spain, Sweden, Switzerland, Turkey, United Kingdom and Vatican City.

How did Gabriel begin? During the 1994 CENL meeting in Oslo, Norway, it was suggested that national libraries should set up a common electronic board with updates about their ongoing projects. Representatives from the national libraries of Netherlands (Koninklijke Bibliotheek), United Kingdom (British Library) and Finland (Helsinki University Library) met in March 1995 in The Hague, Netherlands, to launch the pilot Gabriel project. Three other national libraries joined the project, the ones of Germany (Deutsche Bibliothek), France (Bibliothèque nationale de France) and Poland (Biblioteka Narodowa). Gabriel would describe their services and collections, while seeking to attract other national libraries into the project. The original Gabriel website was launched in September 1995. It was maintained by the British Library Network Services and mirrored by the national libraries of Netherlands and Finland.

In November 1995, other national libraries were invited to submit entries describing their services and collections. At the same time, more and more national libraries were launching their own websites and online catalogs. Gabriel also became a common portal for those.

During the 1996 CENL meeting in Lisbon, it was decided that Gabriel would become an official CENL website in January 1997. Gabriel was maintained by the national library in the

Netherlands, and mirrored by four other national libraries, in United Kingdom, Finland, Germany and Slovenia.

Eight years later, in summer 2005, Gabriel merged with the European Library's website, as a common portal for the 43 national libraries in Europe. In March 2006, the European Commission launched the project of a European digital library, after a "call for ideas" from September to December 2005. This European digital library – named Europeana - opened its "virtual" doors in November 2008, with a crash from the server within 24 hours, followed by an experimental period with part of the collections.

In 1998, eight years before launching Europeana, the European Commission was running a Library Program(me) for public libraries, that aimed "to help increase the ready availability of library resources across Europe, and to facilitate their interconnection with the information and communications infrastructure. Its two main orientations will be the development of advanced systems to facilitate user access to library resources, and the interconnection of libraries with other libraries and the developing 'information highway'. Validation tests will be accompanied by measures to promote standards, disseminate results, and raise the awareness of library staff about the possibilities afforded by telematics systems."

In December 1998, according to a document posted on the website of the European Commission, 1,000 public libraries from 26 European countries had their own websites, that ranged from one webpage - with a postal address and opening hours- to several webpages - with full access to the library's OPAC (Online Public Access Catalog) and a variety of services. The leading countries were Finland (247 libraries), Sweden (132 libraries), United Kingdom (112 libraries), Denmark (107 libraries), Germany (102 libraries), Netherlands (72 libraries), Lithuania (51 libraries), Spain (56 libraries) and Norway (45 libraries). Newcomers were the Czech Republic (29 libraries) and Portugal (3 libraries). Russia had a common website for 26 public reference libraries.

Digital libraries

A definition

What exactly is a digital library? The Universal Library Project, hosted by Carnegie Mellon University, defined it in 1998 as "a digital library of digital documents, artifacts, and records. The advantage of having library material available in digital form is threefold: (1) the content occupies less space and can be replicated and made secure electronically; (2) the content can be made immediately available over the internet to anyone, anywhere; and (3) search for content can be automated. The promise of the digital library is the promise of great cost reductions while providing great increases in archive availability and accessibility. (...) There are literally thousands of digital library initiatives of a great many varieties going on in the world today. Digital libraries are being formed of scholarly works, archives of historical figures and events, corporate and governmental records, museum collections and religious collections. Some take the form of scanning and putting documents to the World Wide Web. Still other digital libraries are formed of digitizing paintings, films and music. Work even exists in 3D reconstructive digitization that permits a digital deconstruction, storage, transmission, and reconstruction of solid object."

Since the mid-1990s, libraries were studying how to store an enormous amount of data and make it available on the internet through a reliable search engine. Library 2000 was a project run between 1995 and 1998 by the MIT Laboratory for Computer Science (MIT: Massachusetts Institute of Technology) to explore the implications of large scale online storage, using the digital library of the future as an example. It developed a prototype using the technology and system configurations expected to be economically feasible in 2000.

Another project was the Digital Library Initiative, supported by grants from NSF (National Science Foundation), DARPA (Defense Advanced Research Projects Agency) and NASA (National Aeronautics and Space Administration). As mentioned on its website in 1998: "The Initiative's focus is to dramatically advance the means to collect, store, and organize information in digital forms, and make it available for searching, retrieval, and processing via communication networks - all in user-friendly ways."

The British Library was a pioneer in Europe. Brian Lang, chief executive of the library, explained on its website in 1998: "We do not envisage an exclusively digital library. We are aware that some people feel that digital materials will predominate in libraries of the future. Others anticipate that the impact will be slight. In the context of the British Library, printed books, manuscripts, maps, music, sound recordings and all the other existing materials in the collection will always retain their central importance, and we are committed to continuing to provide, and to improve, access to these in our reading rooms. The importance of digital materials will, however, increase. We recognize that network infrastructure is at present most strongly developed in the higher education sector, but there are signs that similar facilities will also be available elsewhere, particularly in the industrial and commercial sector, and for public libraries. Our vision of network access encompasses all these."

The Digital Library Programme was expected to begin in 1999. "The development of the Digital Library will enable the British Library to embrace the digital information age. Digital technology will be used to preserve and extend the Library's unparalleled collection. Access to the collection will become boundless with users from all over the world, at any time, having simple, fast access to digitized materials using computer networks, particularly the internet."

Another pioneer in Europe was the French National Library (BnF: Bibliothèque nationale de France). The BnF launched its digital library Gallica in October 1997 as an experimental project to offer digitized texts and images from print collections relating to French history, life and culture. When interviewed by Jérôme Strazzulla in the daily *Le Figaro* of June 3, 1998, Jean-Pierre Angremy, president of BnF, stated: "We cannot, we will not be able to digitize everything. In the long term, a digital library will only be one element of the whole library." The first step of the program, a major collection of 19th-century French texts and images, was available online one year later.

Some projects

In Germany, the Bielefeld University Library (Bibliothek der Universität Bielefeld) began posting online versions of German rare prints in 1996. Michael Behrens, in charge of the

digital library project, wrote in September 1998: "To some here, 'digital library' seems to be everything that, even remotely, has to do with the internet. The library started its own web server some time in summer 1995. (...) Before that, it had been offering most of its services via Telnet, which wasn't used much by patrons, although in theory they could have accessed a lot of material from home. But in those days almost nobody really had internet access at home... We started digitizing rare prints from our own library, and some rare prints which were sent in via library loan, in November 1996. (...)

In that first phase of our attempts at digitization, starting November 1996 and ending June 1997, 38 rare prints were scanned as image files and made available via the web. During the same time, there were also a few digital materials prepared as accompanying material for lectures held at the university (image files as excerpts from printed works). These are, for copyright reasons, not available outside of campus. The next step, which is just being completed, is the digitization of the *Berlinische Monatsschrift*, a German periodical from the Enlightenment, comprising 58 volumes, and 2,574 articles on 30,626 pages. A somewhat bigger digitization project of German periodicals from the 18th and early 19th century is planned. The size will be about 1,000,000 pages. These periodicals will be not just from the holdings of this library, but the project would be coordinated here, and some of the technical would be done here, also."

Other digital libraries were created from scratch, with no back up from a traditional library. They were "only" digital. This was the case of Athena in Switzerland, and Progetto Manuzio in Italy.

Athena was founded in 1994 by Pierre Perroud, a Swiss teacher, and hosted on the website of the University of Geneva. Athena was created as a multilingual digital library specializing in philosophy, science, literature, history and economics, either by digitizing documents or by providing links to existing e texts. The Helvetia section provided documents about Switzerland. Geneva being the main city in French-speaking Switzerland, Athena also focused on putting French texts online. A specific page offered an extensive selection of other digital libraries worldwide, with relevant links.

Progetto Manuzio was launched by Liber Liber as a free digital library for texts in Italian. Liber Liber is an Italian cultural association aimed at the promotion of any kind of artistic and intellectual expression. It wanted to link humanities and science by using computer technology in humanities. Progetto Manuzio was named after the famous 16th-century Venetian publisher who improved the printing techniques invented by Gutenberg.

As stated on its website in 1998, Progetto Manuzio wanted "to make a noble idea real: the idea of making culture available to everybody. How? By making books, graduation theses, articles, tales or any other document which could be digitized in a computer available all over the world, at any minute and free of charge. Via modem, or using floppy disks (in this case, by adding the cost of a blank disk and postal fees), it is already possible to get hundreds of books. And Progetto Manuzio needs only a few people to make such a masterpiece as Dante Alighieri's *Divina Commedia* available to millions of people."

Some "only" digital libraries were organized around an author, for example The Marx/Engels Internet Archive (MEIA). MEIA was created in 1996 to offer a chronology of the collected works of Karl Marx and Frederick Engels, and link this chronology to the digital versions of these works "as one work after another is brought online". As explained on the website in 1998: "There's no way to monetarily profit from this project. 'Tis a labor of love undertaken in the purest communitarian sense. The real 'profit' will hopefully manifest in the form of individual enlightenment through easy access to these classic works. Besides, transcribing them is an education in itself... Let us also add that this is not a sectarian/One-Great-Truth effort. Help from any individual or any group is welcome. We have but one slogan: 'Piping Marx & Engels into cyberspace!'"

A search engine was set up for the digital library. "As larger works come online, they will also have small search pages made for them alone - for instance, *Capital* will have a search page for that work alone."

The Biographical Archive gave access to biographies of Marx and Engels, as well as short biographies and photographs of their family members and friends. The Photo Gallery gathered photos of the Marx and Engels clan from 1839 to 1894, and their dwellings from 1818 to 1895, with "many more to come". The section "Others" included a list of works from all Marxist writers, for example James Connolly, Daniel DeLeon and Hal Draper, as well as a short biography. The Non-English Archive listed the works of Marx and Engels freely available online in other languages (Danish, French, German, Greek, Italian, Japanese, Polish, Portuguese, Spanish and Swedish). It seems that the project was later renamed the Marxists Internet Archive.

Library treasures go online

Libraries began digitizing their treasures, and putting the digital versions on the web for the world to enjoy. The British Library was a pioneer in this field. One of the first digitized treasures was *Beowulf*, the earliest known narrative poem in English, and one of the most famous works of Anglo-Saxon poetry. The British Library holds the only known manuscript of *Beowulf*, dated circa 1000. The poem itself is much older than the manuscript - some historians believe it might have been written circa 750. The manuscript was badly damaged by fire in 1731. 18th-century transcripts mentioned hundreds of words and characters which were then visible along the charred edges, and subsequently crumbled away over the years. To halt this process, each leaf was mounted on a paper frame in 1845.

Scholarly discussions on the date of creation and provenance of the poem continue around the world, and researchers regularly require access to the manuscript. Taking *Beowulf* out of its display case for study not only raised conservation issues, it also made it unavailable for the many visitors who were coming to the British Library expecting to see this literary treasure on display. Digitization of the manuscript offered a solution to these problems, as well as providing new opportunities for researchers and readers worldwide.

The Electronic *Beowulf* Project was launched as a database of digital images of the *Beowulf* manuscript, as well as related manuscripts and printed texts. In 1998, the database included: (a) the fiber-optic readings of hidden characters and ultra-violet readings of erased text in the

manuscript; (b) the full electronic facsimiles of the 18th-century transcripts of the manuscript; and (c) selections from the main 19th-century collations, editions and translations.

Major additions to the database were planned for the following years, such as images of contemporary manuscripts, links to the Toronto *Dictionary of Old English* Project, and links to the comprehensive Anglo-Saxon bibliographies of the *Old English Newsletter*.

The database project was developed in partnership with two leading experts in the United States, Kevin Kiernan, from the University of Kentucky, and Paul Szarmach, from the Medieval Institute of Western Michigan University. Professor Kiernan edited the electronic archive and supervised the making of a CD-ROM with the main electronic images.

Brian Lang, chief executive of the British Library, explained on its website in 1998: "The *Beowulf* manuscript is a unique treasure and imposes on the Library a responsibility to scholars throughout the world. Digital photography offered for the first time the possibility of recording text concealed by early repairs, and a less expensive and safer way of recording readings under special light conditions. It also offers the prospect of using image enhancement technology to settle doubtful readings in the text. Network technology has facilitated direct collaboration with American scholars and makes it possible for scholars around the world to share in these discoveries. Curatorial and computing staff learned a great deal which will inform any future programmes of digitization and network service provision the Library may undertake, and our publishing department is considering the publication of an electronic scholarly edition of *Beowulf*. This work has not only advanced scholarship; it has also captured the imagination of a wider public, engaging people (through press reports and the availability over computer networks of selected images and text) in the appreciation of one of the primary artefacts of our shared cultural heritage."

Other treasures of the British Library were available online as well: *Magna Carta*, the first English constitutional text, signed in 1215, with the Great Seal of King John; the *Lindisfarne Gospels*, dated 698; the *Diamond Sutra*, dated 868, sometimes referred to as the world's earliest print book; the *Sforza Hours*, dated 1490-1520, an outstanding Renaissance treasure; the *Codex Arundel*, a notebook from Leonardo Da Vinci, in the late 15th or early 16th century; and the *Tyndale New Testament*, as the first print version in English by Peter Schoeffer in Worms.

New treasures followed. The digitized version of the *Bible of Gutenberg* was available online in November 2000. Gutenberg printed its Bible in 1454 or 1455 in Germany, perhaps printing 180 copies, with 48 copies still available in 2000, and three copies - two full ones and one partial one - at the British Library. The two full copies - a little different from each other - were digitized in March 2000 by Japanese experts from Keio University of Tokyo and NTT (Nippon Telegraph and Telephone Communications). The images were then processed to offer a full digital version on the web a few months later.

1999: Librarians get digital

[Overview]

The job of librarians, that had already changed a lot with computers, went on to change even more with the internet. Electronic mail became commonplace for internal and external communications. Librarians could subscribe to newsletters and participate in newsgroups and discussion forums. In 1999, librarians were running intranets for their organizations, like Peter Raggett at the OECD Library, or they were running library websites, like Bruno Didier at the Institute Pasteur Library. Computers made catalogs much easier to handle, as well as library loans and book orders. Librarians could type in bibliographic records in a computer database that was sorting out book records by alphabetical order, with search engines for queries by author, title, year and subject. By networking computers, the internet gave a boost to union catalogs for a state, a province, a department, a country or a region, and made things simpler for interlibrary loan.

Two experiences

At the OECD

The OECD Library was among the first ones in Europe to set up an extensive intranet for the staff of its organization. What is OECD (Organization for Economic Cooperation and Development)? "The OECD is a club of like-minded countries. It is rich, in that OECD countries produce two thirds of the world's goods and services, but it is not an exclusive club. Essentially, membership is limited only by a country's commitment to a market economy and a pluralistic democracy. The core of original members has expanded from Europe and North America to include Japan, Australia, New Zealand, Finland, Mexico, the Czech Republic, Hungary, Poland and Korea. And there are many more contacts with the rest of the world through programmes with countries in the former Soviet bloc, Asia, Latin America - contacts which, in some cases, may lead to membership." (excerpt from the website in 1999)

The OECD Central Library serves the OECD staff to support their research work, with more than 60,000 monographs and 2,500 periodicals in early 1999, as well as microfilms and CD-ROMs, and subscriptions to databases like Dialog, Lexis-Nexis and UnCover.

Peter Raggett, deputy-head (and then head) of the Central Library, first worked in government libraries in United Kingdom before joining the OECD in 1994. An avid internet user since 1996, Peter wrote in August 1999: "At the OECD Library we have collected together several hundred websites and have put links to them on the OECD intranet. They are sorted by subject and each site has a short annotation giving some information about it. The researcher can then see if it is possible that the site contains the desired information. This is adding value to the site references and in this way the Central Library has built up a virtual reference desk on the OECD network. As well as the annotated links, this virtual reference desk contains pages of references to articles, monographs and websites relevant to several projects currently being researched at the OECD, network access to CD-ROMs, and a monthly list of new acquisitions. The Library catalogue will soon be available for searching on the intranet. The reference staff at the OECD Library uses the internet for a good deal of their work. Often an academic working paper will be on the web and will be available for full-text

downloading. We are currently investigating supplementing our subscriptions to certain of our periodicals with access to the electronic versions on the internet."

What about finding information on the internet? "The internet has provided researchers with a vast database of information. The problem for them is to find what they are seeking. Never has the information overload been so obvious as when one tries to find information on a topic by searching the internet. When one uses a search engine like Lycos or AltaVista or a directory like Yahoo!, it soon becomes clear that it can be very difficult to find valuable sites on a given topic. These search mechanisms work well if one is searching for something very precise, such as information on a person who has an unusual name, but they produce a confusing number of references if one is searching for a topic which can be quite broad. Try and search the web for Russia AND transport to find statistics on the use of trains, planes and buses in Russia. The first references you will find are freight-forwarding firms who have business connections with Russia."

How about the future? "The internet is impinging on many peoples' lives, and information managers are the best people to help researchers around the labyrinth. The internet is just in its infancy and we are all going to be witnesses to its growth and refinement. (...) Information managers have a large role to play in searching and arranging the information on the internet. I expect that there will be an expansion in internet use for education and research. This means that libraries will have to create virtual libraries where students can follow a course offered by an institution at the other side of the world. Personally, I see myself becoming more and more a virtual librarian. My clients may not meet me face-to-face but instead will contact me by email, telephone or fax, and I will do the research and send them the results electronically."

At the Pasteur Institute

"The Pasteur Institutes are exceptional observatories for studying infectious and parasite-borne diseases. They are wedded to the solving of practical public health problems, and hence carry out research programmes which are highly original because of the complementary nature of the investigations carried out: clinical research, epidemiological surveys and basic research work. Just a few examples from the long list of major topics of the Institutes are: malaria, tuberculosis, AIDS, yellow fever, dengue and poliomyelitis." (excerpt from the website in 1999)

Bruno Didier, librarian and webmaster of the library website, explained in August 1999: "The main aim of the Pasteur Institute Library website is to serve the Institute itself and its associated bodies. It supports applications that have become essential in such a big organization: bibliographic databases, cataloging, ordering of documents and of course access to online periodicals (presently more than 100). It is a window for our different departments, at the Institute but also elsewhere in France and abroad. It plays a big part in documentation exchanges with the institutes in the worldwide Pasteur network. I am trying to make it an interlink adapted to our needs for exploration and use of the internet. The website has existed in its present form since 1996 and its audience is steadily increasing. (...) I build and maintain the webpages and monitor them regularly. I am also responsible for training users. The web is an excellent place for training and it is included in most ongoing discussion

about that."

How about the future? "Our relationship with both the information and the users is what changes. We are increasingly becoming mediators, and perhaps to a lesser extent 'curators'. My present activity is typical of this new situation: I am working to provide quick access to information and to create effective means of communication, but I also train people to use these new tools. (...) I think the future of our job is tied to cooperation and use of common resources. It is certainly an old project, but it is really the first time we have had the means to set it up."

Online catalogs

OPACs

The internet boosted library catalogs through cyberspace. OPACs (OPAC: Online Public Access Catalog) were more attractive and user-friendly than the older print and computer catalogs. Some catalogs began to give instant online access to the full text of books and journals, something that would become a major trend ten years later.

The first step was UNIMARC, as a common bibliographic format for library catalogs. The IFLA (International Federation of Library Associations) published the first edition of *UNIMARC: Universal MARC Format* in 1977, followed by a second edition in 1980 and a *UNIMARC Handbook* in 1983.

UNIMARC (Universal Machine Readable Cataloging) was set up as a solution to the 20 existing national MARC (Machine Readable Cataloging) formats. 20 formats meant lack of compatibility and extensive editing when bibliographic records were exchanged. With UNIMARC, catalogers would be able to process records created in any MARC format. Records in one MARC format would first be converted into UNIMARC, and then be converted into another MARC format. UNIMARC would also be promoted as a format on its own.

In May 1997, the British Library launched OPAC 97 to provide free online access to the catalogs of its main collections in London and Boston Spa. It also launched Blaise, an online bibliographic information service (with a small fee), and Inside, a catalog of articles from 20,000 journals and 16,000 conferences. As explained on the website at the time: "The Library's services are based on its outstanding collections, developed over 250 years, of over one hundred and fifty million items representing every age of written civilisation, every written language and every aspect of human thought. At present individual collections have their own separate catalogues, often built up around specific subject areas. Many of the Library's plans for its collections, and for meeting its users' needs, require the development of a single catalogue database. This is being pursued in the Library's Corporate Bibliographic Programme which seeks to address this issue." The "single catalogue database" was fully operational a few years later.

Another leading effort was the one of the Library of Congress with its Experimental Search System (ESS). The ESS was "one of the Library of Congress' first efforts to make selected cataloging and digital library resources available over the World Wide Web by means of a

single, point-and-click interface. The interface consists of several search query pages (Basic, Advanced, Number, and a Browse screen) and several search results pages (an item list of brief displays and an item full display), together with brief help files which link directly from significant words on those pages. By exploiting the powerful synergies of hyperlinking and a relevancy-ranked search engine (InQuery from Sovereign Hill Software), we hope the ESS will provide a new and more intuitive way of searching the traditional OPAC (Online Public Access Catalog)." (excerpt from the website in 1998)

Another interesting - and totally different - initiative was the creation of the Internet Public Library (IPL) by the School of Information and Library Studies at the University of Michigan. The IPL went live in March 1995 as the first U.S. digital public library to serve the internet community, and to catalog websites and webpages. The librarians' task was to choose the best documents available on the web, and process them as library documents for them to be easily accessed from the IPL website, that acted as a portal. The IPL sections were: Reference, Exhibits, Magazines and Serials, Newspapers, Online Texts, and Web Searching. There were also Teen and Youth sections. All items were carefully selected, catalogued and described by the IPL staff. As an experimental library, IPL also listed the best internet projects that were run by librarians, in the section Especially for Librarians. Since then, students from the IPL Consortium, a consortium of colleges and universities with programs in information science, have worked on maintaining and developing the IPL as a public library for the web.

Union catalogs

In 1999, the two main union catalogs were WorldCat, run by OCLC (Online Computer Library Center), and RLIN (Research Library Information Network), run by the Research Libraries Group (RLG).

What exactly is a union catalog? The idea behind a union catalog is to earn time by avoiding the cataloging of the same document by many catalogers worldwide. When catalogers of a member library catalog a new document, they first search the union catalog. If the record is available, they import it into their own library catalog and add the local data. If the record is not available, they create it in their own library catalog and export it into the union catalog. The new record is immediately available to all catalogers of member libraries. Depending on their status, experience and quality of cataloging, member libraries can either import records only, or import and export records.

OCLC (Online Computer Library Center) was created in 1971 as a non-profit organization dedicated to furthering access to the world's information while reducing information costs. The OCLC Online Union Catalog - renamed WorldCat much later - began as the union catalog of the university libraries in the State of Ohio. Over the years, OCLC became a national and then worldwide library cooperative, and WorldCat the largest library catalog in the world. In early 1998, WorldCat had 38 million records in 400 languages - with transliteration for non-Roman languages - and an annual increase of 2 million records. In 1998, 27,000 libraries in 65 countries were using OCLC services (paid subscription) to manage their collections and provide online reference services.

WorldCat has accepted only one bibliographic record per document, unlike RLIN (Research Library Information Network), another union catalog launched by the Research Libraries Group (RLG) in 1980. RLIN accepted several records for the same document, with 88 million records in early 1998.

Members of RLG were mainly research and specialized libraries. RLIN was later renamed the RLG Union Catalog. Its free web version, RedLightGreen, was launched in fall 2003 as a beta version, and in spring 2004 as a full version. This was a major move, not only for library members, but for all internet users, who could also access it for free.

In 2005, WorldCat had 61 million bibliographic records in 400 languages, from 9,000 member libraries in 112 countries. In 2006, 73 million bibliographic records were linking to one billion documents available in these libraries.

In August 2006, WorldCat began to migrate to the web through the beta version of its new website worldcat.org. Member libraries now provided free access to their catalogs and electronic resources: books, audiobooks, abstracts and full-text articles, photos, music CDs and videos. RedLightGreen ended its service in November 2006, and RLG joined OCLC.

2000: Information is available in many languages

[Overview]

2000 was a turning point for a multilingual internet, both for its content and its users. In summer 2000, non-English-speaking users reached 50%. This percentage went on to increase steadily: 52.5% in summer 2001, 57% in December 2001, 59.8% in April 2002, 64.4% in September 2003 - with 34.9% non-English-speaking Europeans and 29.4% Asians - and 64.2% in March 2004 - with 37.9% non-English-speaking Europeans and 33% Asians (source: Global Reach). The internet is a good tool for minority languages, as stated by Caoimhín Ó Donnáil, who teaches computing at the Institute Sabhal Mór Ostaig, located on the Island of Skye, in Scotland. Caoimhín also maintains the college website, which is the main site worldwide with information on Scottish Gaelic, with a bilingual (English, Gaelic) list of European minority languages. He wrote in May 2001: "Students do everything by computer, use Gaelic spell-checking, a Gaelic online terminology database. Gaelic radio (both Scottish and Irish) is now available continuously worldwide via the internet. A major project has been the translation of the Opera web browser into Gaelic - the first software of this size available in Gaelic."

"Language nations"

At first, the internet was nearly 100% English. Born in the United States, it spread in North America before taking over the whole planet. Then people from all continents began connecting to the internet and posting webpages in their own languages. In the 1990s, the percentage of English decreased from nearly 100% to 85% (reached in 1997 or 1998, depending on the sources).

In 1997, Babel - a joint initiative from Alis Technologies (language translation services) and the Internet Society - ran the first major study relating to distribution of languages on the web. The results were published in June 1997 on a webpage named Web Languages Hit Parade. The main languages were English with 82.3%, German with 4.0%, Japanese with 1.6%, French with 1.5%, Spanish with 1.1%, Swedish with 1.1%, and Italian with 1.0%.

In July 1998, according to Global Reach, a company specializing in international online marketing, the fastest growing groups of internet users were non-English-speaking: Spanish-speaking, 22.4%, Japanese-speaking, 12.3%; German-speaking, 14%; and French-speaking, 10% - with 56 million non-English-speaking users. More than 80% of all webpages were still in English, whereas only 6% of the world population spoke English as a native language (16% spoke Spanish).

Randy Hobler was a consultant in internet marketing for Globalink, a company specializing in language translation software and services. He wrote in September 1998: "85% of the content of the web in 1998 is in English and going down. This trend is driven not only by more websites and users in non-English-speaking countries, but by increasing localization of company and organization sites, and increasing use of machine translation to/from various languages to translate websites."

Randy also brought up the concept of "language nations": "Because the internet has no national boundaries, the organization of users is bounded by other criteria driven by the medium itself. In terms of multilingualism, you have virtual communities, for example, of what I call 'Language Nations'...all those people on the internet wherever they may be, for whom a given language is their native language. Thus, the Spanish Language nation includes not only Spanish and Latin American users, but millions of Hispanic users in the U.S., as well as odd places like Spanish-speaking Morocco."

Robert Ware created OneLook Dictionaries in April 1996, as a "fast finder" of words in hundreds of online dictionaries. He wrote about an experience he had in 1994, that showed the internet could promote both a common language and multilingualism: "In 1994, I was working for a college and trying to install a software package on a particular type of computer. I located a person who was working on the same problem and we began exchanging email. Suddenly, it hit me... the software was written only 30 miles away but I was getting help from a person half way around the world. Distance and geography no longer mattered! OK, this is great! But what is it leading to? I am only able to communicate in English but, fortunately, the other person could use English as well as German which was his mother tongue. The internet has removed one barrier (distance) but with that comes the barrier of language. It seems that the internet is moving people in two quite different directions at the same time. The internet (initially based on English) is connecting people all around the world. This is further promoting a common language for people to use for communication. But it is also creating contact between people of different languages and creates a greater interest in multilingualism. A common language is great but in no way replaces this need. So the internet promotes both a common language *and* multilingualism. The good news is that it helps provide solutions. The increased interest and need is creating incentives for people around the world to create improved language courses and other assistance, and the internet is providing fast and inexpensive opportunities to make them available."

The internet could also be a tool to develop a "cultural identity". During the Symposium on Multimedia Convergence organized by the International Labor Organization (ILO) in January 1997, Shinji Matsumoto, general secretary of the Musicians' Union of Japan (MUJ), explained: "Japan is quite receptive to foreign culture and foreign technology (...) Foreign culture is pouring into Japan and, in fact, the domestic market is being dominated by foreign products. Despite this, when it comes to preserving and further developing Japanese culture, there has been insufficient support from the government. (...) With the development of information networks, the earth is getting smaller and it is wonderful to be able to make cultural exchanges across vast distances and to deepen mutual understanding among people. We have to remember to respect national cultures and social systems."

As the internet quickly spread worldwide, more and more people in the U.S. realized that, although English may stay the main international language for exchanges of all kinds, not everyone in the world reads English and, even so, people prefer to read information in their own language. To reach as large an audience as possible, companies and organizations needed to offer bilingual, trilingual, even multilingual websites, while adapting their content to a given audience. Thus the need of both internationalization and localization, which

became a major trend in the following years, not only in the U.S. but in many countries, where foreign companies set up bilingual websites - in their language and in English - to reach a wider audience, and get more clients.

Translation software available on the web was far from perfect, but was helpful, because instantaneous and free, unlike a high-quality professional translation. In December 1997, AltaVista, a leading search engine, was the first to launch such software with Babel Fish - also called AltaVista Translation -, which could translate webpages (up to three pages at the same time) from English into French, German, Italian, Portuguese or Spanish, and viceversa. The software was developed by Systran, a company specializing in machine translation. This initiative was followed by others, with free and/or paid versions on the web, developed by Alis Technologies, Globalink, Lernout & Hauspie, IBM (with the WebSphere Translation Server), Softissimo, Champollion, TMX or Trados.

Brian King, director of the WorldWide Language Institute (WWLI), brought up the concept of "linguistic democracy" in September 1998: "Whereas 'mother-tongue education' was deemed a human right for every child in the world by a UNESCO report in the early '50s, 'mother-tongue surfing' may very well be the Information Age equivalent. If the internet is to truly become the Global Network that it is promoted as being, then all users, regardless of language background, should have access to it. To keep the internet as the preserve of those who, by historical accident, practical necessity, or political privilege, happen to know English, is unfair to those who don't."

Jean-Pierre Cloutier was the editor of *Chroniques de Cybérie*, a weekly French-language online report of internet news. He wrote in August 1999: "We passed a milestone this summer. Now more than half the users of the internet live outside the United States. Next year, more than half of all users will be non English-speaking, compared with only 5% five years ago. Isn't that great?"

The internet did pass this second milestone in summer 2000, with non-English-speaking users reaching 50%. As shown in the statistics of Global Reach, they were 52.5% in summer 2001, 57% in December 2001, 59.8% in April 2002, 64.4% in September 2003 (with 34.9% non-English-speaking Europeans and 29.4% Asians), and 64.2% in March 2004 (with 37.9% non-English-speaking Europeans and 33% Asians).

From ASCII to Unicode

Used since the beginning of computing, ASCII (American Standard Code for Information Interchange) is a 7-bit coded character set for information interchange in English. It was published in 1968 by ANSI (American National Standards Institute), with an update in 1977 and 1986. The 7-bit plain ASCII, also called Plain Vanilla ASCII, is a set of 128 characters with 95 printable unaccented characters (A-Z, a-z, numbers, punctuation and basic symbols), i.e. the ones that are available on the English/American keyboard.

With the use of other European languages, extensions of ASCII (also called ISO-8859 or ISO-Latin) were created as sets of 256 characters to add accented characters as found in French, Spanish and German, for example ISO-8859-1 (ISO-Latin-1) for French.

Yoshi Mikami, who lives in Fujisawa, Japan, launched the bilingual (Japanese, English) website *The Languages of the World by Computers and the Internet*, also known as Logos Home Page or Kotoba Home Page, in late 1995. Yoshi was the co-author (with Kenji Sekine and Nobutoshi Kohara) of *The Multilingual Web Guide* (Japanese edition), a print book published by O'Reilly Japan in August 1997, and translated in 1998 into English, French and German.

Yoshi Mikami explained in December 1998: "My native tongue is Japanese. Because I had my graduate education in the U.S. and worked in the computer business, I became bilingual in Japanese and American English. I was always interested in languages and different cultures, so I learned some Russian, French and Chinese along the way. In late 1995, I created on the web *The Languages of the World by Computers and the Internet* and tried to summarize there the brief history, linguistic and phonetic features, writing system and computer processing aspects for each of the six major languages of the world, in English and Japanese. As I gained more experience, I invited my two associates to help me write a book on viewing, understanding and creating multilingual web pages, which was published in August 1997 as *The Multilingual Web Guide*, in a Japanese edition, the world's first book on such a subject."

Yoshi added in the same email interview: "Thousands of years ago, in Egypt, China and elsewhere, people were more concerned about communicating their laws and thoughts not in just one language, but in several. In our modern world, most nation states have each adopted one language for their own use. I predict greater use of different languages and multilingual pages on the internet, not a simple gravitation to American English, and also more creative use of multilingual computer translation. 99% of the websites created in Japan are written in Japanese."

Brian King, director of the WorldWide Language Institute (WWLI), explained in September 1998: "A pull from non-English-speaking computer users and a push from technology companies competing for global markets has made localization a fast growing area in software and hardware development. This development has not been as fast as it could have been. The first step was for ASCII to become Extended ASCII. This meant that computers could begin to start recognizing the accents and symbols used in variants of the English alphabet - mostly used by European languages. But only one language could be displayed on a page at a time. (...) The most recent development is Unicode. Although still evolving and only just being incorporated into the latest software, this new coding system translates each character into 16 bytes. Whereas 8-byte Extended ASCII could only handle a maximum of 256 characters, Unicode can handle over 65,000 unique characters and therefore potentially accommodate all of the world's writing systems on the computer. So now the tools are more or less in place. They are still not perfect, but at last we can at least surf the web in Chinese, Japanese, Korean, and numerous other languages that don't use the Western alphabet. As the internet spreads to parts of the world where English is rarely used - such as China, for example, it is natural that Chinese, and not English, will be the preferred choice for interacting with it. For the majority of the users in China, their mother tongue will be the only choice."

Ten years later, in 2008, 50% of all the documents available on the internet were encoded in Unicode, with the other 50% encoded in ASCII. ASCII is still very useful, especially the original 7-bit plain ASCII, because it can be read, written, copied and printed by any text editor or word processor, and it is the only format compatible with 99% of all hardware and software.

First published in January 1991, Unicode "provides a unique number for every character, no matter what the platform, no matter what the program, no matter what the language" (excerpt from the website). This double-byte platform-independent encoding provides a basis for the processing, storage and interchange of text data in any language, and any modern software and information technology protocols. Unicode is maintained by the Unicode Consortium, and is a component of the W3C (World Wide Web Consortium) specifications.

Language dictionaries

Logos is an international translation company with headquarters in Modena, Italy. In 1997, Logos had 200 in-house translators in Modena and 2,500 free-lance translators worldwide, who processed around 200 texts per day. The company made a bold move, and decided to put on the web all the linguistic tools used by its translators, for the internet community to freely use them as well. The linguistic tools were the *Logos Dictionary*, a multilingual dictionary with 7 billion words (in fall 1998); the *Logos Wordtheque*, a multilingual library with 300 billion words extracted from translated novels, technical manuals and other texts; the *Logos Linguistic Resources*, a database of 500 glossaries; and the *Logos Universal Conjugator*, a database for verbs in 17 languages.

When interviewed by Annie Kahn on December 7, 1997 for the French daily *Le Monde*, Rodrigo Vergara, head of Logos, explained: "We wanted all our translators to have access to the same translation tools. So we made them available on the internet, and while we were at it we decided to make the site open to the public. This made us extremely popular, and also gave us a lot of exposure. This move has in fact attracted many customers, and also allowed us to widen our network of translators, thanks to contacts made in the wake of the initiative."

In the same article, Annie Kahn wrote: "The Logos site is much more than a mere dictionary or a collection of links to other online dictionaries. The cornerstone is the document search program, which processes a corpus of literary texts available free of charge on the web. If you search for the definition or the translation of a word ('didactique', for example), you get not only the answer sought, but also a quote from one of the literary works containing the word (in our case, an essay by Voltaire). All it takes is a click on the mouse to access the whole text or even to order the book, including in foreign translations, thanks to a partnership agreement with the famous online bookstore Amazon.com. However, if no text containing the required word is found, the program acts as a search engine, sending the user to other web sources containing this word. In the case of certain words, you can even hear the pronunciation. If there is no translation currently available, the system calls on the public to contribute. Everyone can make suggestions, after which Logos translators check the suggested translations they receive."

Robert Beard, a language teacher at Bucknell University (in Lewisburg, Pennsylvania), created the website *A Web of Online Dictionaries* (WOD) in 1995, and included it then in a

larger project, yourDictionary.com, that he co-founded in early 2000. He wrote in January 2000: "The new website is an index of 1,200+ dictionaries in more than 200 languages. Besides the WOD, the new website includes a word-of-the-day-feature, word games, a language chat room, the old *Web of Online Grammars* (now expanded to include additional language resources), the *Web of Linguistic Fun*, multilingual dictionaries; specialized English dictionaries; thesauri and other vocabulary aids; language identifiers and guessers, and other features; dictionary indices. yourDictionary.com will hopefully be the premiere language portal and the largest language resource site on the web. It is now actively acquiring dictionaries and grammars of all languages with a particular focus on endangered languages. It is overseen by a blue ribbon panel of linguistic experts from all over the world."

yourDictionary.com wants to be the premiere portal for all languages without any exception, and as such offers a specific section called *Endangered Language Repository*. Robert Beard explained in the same email interview: "Languages that are endangered are primarily languages without writing systems at all (only 1/3 of the world's 6,000+ languages have writing systems). I still do not see the web contributing to the loss of language identity and still suspect it may, in the long run, contribute to strengthening it. More and more Native Americans, for example, are contacting linguists, asking them to write grammars of their language and help them put up dictionaries. For these people, the web is an affordable boon for cultural expression."

The 6,700 languages of our planet are catalogued in *The Ethnologue: Languages of the World*, an encyclopedia published by SIL International (SIL: Summer Institute of Linguistics). Barbara Grimes was the editor of the 8th to 14th editions, 1971-2000. She wrote in January 2000: "*The Ethnologue* is a catalog of the languages of the world, with information about where they are spoken, an estimate of the number of speakers, what language family they are in, alternate names, names of dialects, other socio-linguistic and demographic information, dates of published Bibles, a name index, a language family index, and language maps." The *Ethnologue* is freely available on the web. The print version or CD-ROM can be bought online.

Minority languages

Caoimhín Ó Donnáille teaches computing - through the Gaelic language - at the Institute Sabhal Mór Ostaig, located on the Island of Skye, in Scotland. He also maintains the bilingual (English, Gaelic) college website, which is the main site worldwide with information on Scottish Gaelic, as well as the webpage *European Minority Languages*, a bilingual list of minority languages by alphabetic order and by language family. He wrote in May 2001: "There has been a great expansion in the use of information technology in our college. Far more computers, more computing staff, flat screens. Students do everything by computer, use Gaelic spell-checking, and a Gaelic online terminology database. There are more hits on our website. There is more use of sound. Gaelic radio (both Scottish and Irish) is now available continuously worldwide via the internet. A major project has been the translation of the Opera web browser into Gaelic - the first software of this size available in Gaelic."

What about the internet and endangered languages? "I would emphasize the point that as regards the future of endangered languages, the internet speeds everything up. If people

don't care about preserving languages, the internet and accompanying globalisation will greatly speed their demise. If people do care about preserving them, the internet will be a tremendous help."

Guy Antoine is the founder of Windows on Haiti, a reference website about Haitian culture. He wrote in November 1999: "In Windows on Haiti, the primary language of the site is English, but one will equally find a center of lively discussion conducted in 'Kreyòl'. In addition, one will find documents related to Haiti in French, in the old colonial Creole, and I am open to publishing others in Spanish and other languages. I do not offer any sort of translation, but multilingualism is alive and well at the site, and I predict that this will increasingly become the norm throughout the web."

Guy added in June 2001: "Kreyòl is the only national language of Haiti, and one of its two official languages, the other being French. It is hardly a minority language in the Caribbean context, since it is spoken by eight to ten million people. (...) I have taken the promotion of Kreyòl as a personal cause, since that language is the strongest of bonds uniting all Haitians, in spite of a small but disproportionately influential Haitian elite's disdainful attitude to adopting standards for the writing of Kreyòl and supporting the publication of books and official communications in that language. For instance, there was recently a two-week book event in Haiti's Capital and it was promoted as 'Livres en folie' ('A mad feast for books'). Some 500 books from Haitian authors were on display, among which one could find perhaps 20 written in Kreyòl. This is within the context of France's major push to celebrate Francophony among its former colonies. This plays rather well in Haiti, but directly at the expense of Creolophony. What I have created in response to those attitudes are two discussion forums on my website, Windows on Haiti, held exclusively in Kreyòl. One is for general discussions on just about everything but obviously more focused on Haiti's current socio-political problems. The other is reserved only to debates of writing standards for Kreyòl. Those debates have been quite spirited and have met with the participation of a number of linguistic experts. The uniqueness of these forums is their non-academic nature."

Translations

Henk Slettenhaar is a professor in communication technologies at Webster University, Geneva, Switzerland. He has regularly insisted on the need of bilingual websites, in the original language and in English. He wrote in December 1998: "I see multilingualism as a very important issue. Local communities that are on the web should principally use the local language for their information. If they want to present it to the world community as well, it should be in English too. I see a real need for bilingual websites. I am delighted there are so many offerings in the original language now. I much prefer to read the original with difficulty than getting a bad translation."

Henk added in August 1999: "There are two main categories of websites in my opinion. The first one is the global outreach for business and information. Here the language is definitely English first, with local versions where appropriate. The second one is local information of all kinds in the most remote places. If the information is meant for people of an ethnic and/or language group, it should be in that language first, with perhaps a summary in English. We have seen lately how important these local websites are - in Kosovo and Turkey, to mention

just the most recent ones. People were able to get information about their relatives through these sites."

Jean-Pierre Cloutier was the editor of *Chroniques de Cybérie*, a weekly French-language online report of internet news. Jean-Pierre wrote in August 1999: "The web is going to grow in non-English-speaking regions. So we have to take into account the technical aspects of the medium if we want to reach these 'new' users. I think it is a pity there are so few translations of important documents and essays published on the web - from English into other languages and vice versa. (...) In the same way, the recent spreading of the internet in new regions raises questions which would be good to read about. When will Spanish-speaking communication theorists and those speaking other languages be translated?"

Marcel Grangier is the head of the French Section of the Swiss Federal Government's Central Linguistic Services, which means he is in charge of organizing translations into French for the Swiss government. He wrote in January 1999: "We can see multilingualism on the internet as a happy and irreversible inevitability. So we have to laugh at the doomsayers who only complain about the supremacy of English. Such supremacy is not wrong in itself, because it is mainly based on statistics (more PCs per inhabitant, more people speaking English, etc.). The answer is not to 'fight' English, much less whine about it, but to build more sites in other languages. As a translation service, we also recommend that websites be multilingual. The increasing number of languages on the internet is inevitable and can only boost multicultural exchanges. For this to happen in the best possible circumstances, we still need to develop tools to improve compatibility. Fully coping with accents and other characters is only one example of what can be done."

2001: Copyright, copyleft and Creative Commons

[Overview]

Creative Commons (CC) was founded in 2001 by Lawrence Lessing, a professor at Stanford Law School, California. As explained on its website, "Creative Commons is a nonprofit corporation dedicated to making it easier for people to share and build upon the work of others, consistent with the rules of copyright. We provide free licenses and other legal tools to mark creative work with the freedom the creator wants it to carry, so others can share, remix, use commercially, or any combination thereof." There were one million Creative Commons licensed works in 2003, 4.7 million licensed works in 2004, 20 million licensed works in 2005, 50 million licensed works in 2006, 90 million licensed works in 2007, and 130 million licensed works in 2008. Science Commons was founded in 2005 to "design strategies and tools for faster, more efficient web-enabled scientific research." ccLearn was founded in 2007 as "a division of Creative Commons dedicated to realizing the full potential of the internet to support open learning and open educational resources."

Copyright on the web

What did people think about copyright on the web, when there were heated debates about print articles and other copyrighted works being posted and re-posted without the consent of their authors? Here are some answers.

Based in San Francisco, California, Jacques Gauchey was a journalist in information technology and a "facilitator" between the United States and Europe. He wrote in July 1999: "Copyright in its traditional context doesn't exist any more. Authors have to get used to a new situation: the total freedom of the flow of information. The original content is like a fingerprint: it can't be copied. So it will survive and flourish."

Guy Antoine is the founder of Windows on Haiti, a reference website about Haitian culture. He wrote in November 1999: "The debate will continue forever, as information becomes more conspicuous than the air that we breathe and more fluid than water. (...) Authors will have to become a lot more creative in terms of how to control the dissemination of their work and profit from it. The best that we can do right now is to promote basic standards of professionalism, and insist at the very least that the source and authorship of any work be duly acknowledged. Technology will have to evolve to support the authorization process."

Alain Bron is a consultant in information systems and a novelist. He wrote in November 1999: "I regard the web today as a public domain. That means in practice the notion of copyright on it disappears: everyone can copy everyone else. Anything original risks being copied at once if copyrights are not formally registered or if works are available without payment facilities. A solution is to make people pay for information, but this is no watertight guarantee against it being copied."

Peter Raggett was the deputy-head (and now the head) of the OECD Central Library (OECD: Organization for Economic and Cooperation Development). He wrote in August 1999: "The copyright question is still very unclear. Publishers naturally want their fees for each article

ordered and librarians and end-users want to be able to download immediately the full text of articles. At the moment, each publisher seems to have its own policy for access to electronic versions and they would benefit from having some kind of homogenous policy, preferably allowing unlimited downloading of their electronic material."

Tim McKenna is an author who thinks and writes about the complexity of truth in a world of flux. He wrote in October 2000: "Copyright is a difficult issue. The owner of the intellectual property thinks that s/he owns what s/he has created. I believe that the consumer purchases the piece of plastic (in the case of a CD) or the bounded pages (in the case of book). The business community has not found a new way to add value to intellectual property. Consumers don't think very abstractly. When they download songs for example, they are simply listening to them, they are not possessing them. The music and publishing industry need to find ways to give consumers tactile vehicles for selling the intellectual property."

Copyright and WIPO

Since the web became mainstream, the posting by the thousands of electronic texts and other documents has been an headache for organizations in charge of applying the rules relating to intellectual property.

The World Intellectual Property Organization (WIPO) is an intergovernmental organization, and one of the 16 specialized agencies of the United Nations. It is responsible for protecting intellectual property throughout the world through cooperation among countries. It is also responsible for implementing various multilateral treaties dealing with the legal and administrative aspects of intellectual property.

Intellectual property comprises industrial property and copyright. Industrial property relates to inventions, trademarks, industrial designs and appellations of origin. Copyright relates to literary, musical, artistic, photographic and audiovisual works. WIPO stated on its website in 1999: "As regards the number of literary and artistic works created worldwide, it is difficult to make a precise estimate. However, the information available indicates that at present around 1,000,000 books/titles are published and some 5,000 feature films are produced in a year, and the number of copies of phonograms sold per year presently is more than 3,000 million."

Copyright protection means that using a copyrighted work is lawful only if we get authorization from the copyright owner. As explained by WIPO on its website in the section *International Protection of Copyright and Neighbouring Rights*, the authorizations granted by the copyright owner can be: "The right to copy or otherwise reproduce any kind of work; the right to distribute copies to the public; the right to rent copies of at least certain categories of works (such as computer programs and audiovisual works); the right to make sound recordings of the performances of literary and musical works; the right to perform in public, particularly musical, dramatic or audiovisual works; the right to communicate to the public by cable or otherwise the performances of such works and, particularly, to broadcast, by radio, television or other wireless means, any kind of work; the right to translate literary works; the right to rent, particularly, audiovisual works, works embodied in phonograms and computer programs; the right to adapt any kind of work and particularly the right to make audiovisual works thereof."

Under some national laws, some of these rights - which together are referred to as "economic rights" - are not exclusive rights of authorization but, in some specific cases, merely rights to remuneration. In addition to economic rights, authors - whether or not they own the economic rights - enjoy "moral rights" on the basis of which authors have the right to claim their authorship and require that their names be indicated on the copies of the work and in connection with other uses, and they have the right to oppose the mutilation or deformation of their works.

Shrinking of public domain

Michael Hart created Project Gutenberg in July 1971 to make electronic versions of literary works and disseminate them for free. In 2009, Project Gutenberg has had tens of thousands of downloads every day. As recalled by Michael in January 2009, "I knew [in July 1971] that the future of computing, and the internet, was going to be... 'The Information Age.' That was also the day I said we would be able to carry quite literally the entire Library of Congress in one hand and the system would certainly make it illegal... too much power to leave in the hands of the masses."

As defined by Project Gutenberg, "public domain is the set of cultural works that are free of copyright, and belong to everyone equally", i.e. for books, the ones that can be digitized and released on the internet for free. But the task of Project Gutenberg hasn't been made any easier by the increasing restrictions to public domain. In former times, 50% of works belonged to public domain, and could be freely used by everybody. A much tougher legislation was set in place over the centuries, step by step, especially during the 20th century, despite our so-called "information society". In 2100, 99% of works might be governed by copyright, with a meager 1% for public domain.

In the *Copyright HowTo* section of its website, Project Gutenberg explains how to confirm the public domain status of books according to U.S. copyright laws. Here is a summary: (a) Works published before 1923 entered the public domain no later than 75 years from the copyright date: all these works belong to public domain; (b) Works published between 1923 and 1977 retain copyright for 95 years: no such works will enter the public domain until 2019; (c) Works created from 1978 on enter the public domain 70 years after the death of the author if the author is a natural person: nothing will enter the public domain until 2049; (d) Works created from 1978 on enter the public domain 95 years after publication or 120 years after creation in case of a corporate author: nothing will enter the public domain until 2074.

Each copyright legislation is more restrictive than the previous one. A major blow for digital libraries was the amendment to the 1976 Copyright Act signed on October 27, 1998. As explained by Michael Hart in July 1999: "Nothing will expire for another 20 years. We used to have to wait 75 years. Now it is 95 years. And it was 28 years (+ a possible 28-year extension, only on request) before that, and 14 years (+ a possible 14-year extension) before that. So, as you can see, this is a serious degrading of the public domain, as a matter of continuing policy."

John Mark Ockerbloom, founder of The Online Books Page in 1993, got also deeply concerned by the 1998 amendment. He wrote in August 1999: "I think it is important for people on the web to understand that copyright is a social contract that is designed for the public good - where the public includes both authors and readers. This means that authors should have the right to exclusive use of their creative works for limited times, as is expressed in current copyright law. But it also means that their readers have the right to copy and reuse the work at will once copyright expires. In the U.S. now, there are various efforts to take rights away from readers, by restricting fair use, lengthening copyright terms (even with some proposals to make them perpetual) and extending intellectual property to cover facts separate from creative works (such as found in the 'database copyright' proposals). There are even proposals to effectively replace copyright law altogether with potentially much more onerous contract law. (...) Stakeholders in this debate have to face reality, and recognize that both producers and consumers of works have legitimate interests in their use. If intellectual property is then negotiated by a balance of principles, rather than as the power play it is too often ends up being ('big money vs. rogue pirates'), we may be able to come up with some reasonable accommodations."

Michael Hart wrote in July 1999: "No one has said more against copyright extensions than I have, but Hollywood and the big publishers have seen to it that our Congress won't even mention it in public. The kind of copyright debate going on is totally impractical. It is run by and for the 'Landed Gentry of the Information Age.' 'Information Age'? For whom?"

Sure enough. We regularly hear about the great "information age" we live in, while seeing the tightening of laws relating to dissemination of information. The contradiction is obvious. This problem has also affected several European countries, where the copyright law switched from "author's life plus 50 years" to "author's life plus 70 years", following pressure from content owners who successfully lobbied for "harmonization" of national copyright laws as a response to "globalization of the market". To regulate the copyright of digital editions in the wake of the relevant WIPO international treaties, the *Digital Millennium Copyright Act* (DMCA) was ratified in October 1998 in the United States, and the *European Union Copyright Directive* (EUCD) was ratified in May 2001 by the European Commission.

According to Michael Hart, and Project Gutenberg CEO Greg Newby, "as of January 2009, the total number of separate public domain books in the world is between 20 and 30 million, and 5 million are already on the internet, and we expect another million per year from now until all the easy-to-find books are done. 10 million or so will be done before people start to think about the facts telling them the rate cannot continue to double as they come up to the point of already having done half. New copyrights lasting virtually for ever in the U.S. will bring the growth process to a screeching halt when the Mickey Mouse copyright laws, literally, copyright laws on Mickey Mouse, and Winnie-the-Pooh, etc., stop all current copyright from expiring for the foreseeable future."

Copyleft and Creative Commons

The term "copyleft" was invented in 1984 by Richard Stallman, a computer scientist at MIT (Massachusetts Institute of Technology), who launched the GNU project to develop a complete Unix-like operating system called the GNU system.

As explained on the GNU website: "Copyleft is a general method for making a program or other work free, and requiring all modified and extended versions of the program to be free as well. (...) Copyleft says that anyone who redistributes the software, with or without changes, must pass along the freedom to further copy and change it. Copyleft guarantees that every user has freedom. (...) Copyleft is a way of using of the copyright on the program. It doesn't mean abandoning the copyright; in fact, doing so would make copyleft impossible. The word 'left' in 'copyleft' is not a reference to the verb 'to leave' — only to the direction which is the inverse of 'right'. (...) The GNU Free Documentation License (FDL) is a form of copyleft intended for use on a manual, textbook or other document to assure everyone the effective freedom to copy and redistribute it, with or without modifications, either commercially or non commercially."

Creative Commons (CC) was founded in 2001 by Lawrence Lessing, a professor at Stanford Law School, California. As explained on its website: "Creative Commons is a nonprofit corporation dedicated to making it easier for people to share and build upon the work of others, consistent with the rules of copyright. We provide free licenses and other legal tools to mark creative work with the freedom the creator wants it to carry, so others can share, remix, use commercially, or any combination thereof."

There were one million Creative Commons licensed works in 2003, 4.7 million licensed works in 2004, 20 million licensed works in 2005, 50 million licensed works in 2006, 90 million licensed works in 2007, and 130 million licensed works in 2008.

Science Commons was founded in 2005. As explained on its website: "Science Commons designs strategies and tools for faster, more efficient web-enabled scientific research. We identify unnecessary barriers to research, craft policy guidelines and legal agreements to lower those barriers, and develop technology to make research, data and materials easier to find and use. Our goal is to speed the translation of data into discovery — unlocking the value of research so more people can benefit from the work scientists are doing."

ccLearn was founded in 2007. As explained on its website: "ccLearn is a division of Creative Commons dedicated to realizing the full potential of the internet to support open learning and open educational resources. Our mission is to minimize legal, technical, and social barriers to sharing and reuse of educational materials."

2002: A web of knowledge

[Overview]

The MIT OpenCourseWare (MIT OCW) is an initiative launched by MIT (Massachusetts Institute of Technology) in 2002 to put its course materials for free on the web, as a way to promote open dissemination of knowledge. In September 2002, a pilot version was available online with 32 course materials. In November 2007, all 1,800 course materials were available, with 200 new and updated courses per year. From 2003 onwards, in the same spirit of free access of knowledge, the Public Library of Science (PLOS) launched several high-quality online periodicals. New kinds of encyclopedias were set up, for the general public to both use available articles and contribute to their writing. Wikipedia, launched in 2001, became the leading online cooperative encyclopedia worldwide, with hundreds and then thousands of contributors writing articles or editing and updating them, leading the way to other initiatives like Citizendium, launched in 2006, and the Encyclopedia of Life, launched in 2007.

New ways of teaching

More and more computers connected to the internet were available in schools and at home in the mid-1990s. Teachers began exploring new ways of teaching. Going from print book culture to digital culture was changing relationship to knowledge, and the way both scholars and students were seeing teaching and learning. Print book culture provided stable information whereas digital culture provided "moving" information. During a conference organized by the International Federation of Information Processing (IFIP) in September 1996, Dale Spender gave a lecture about *Creativity and the Computer Education Industry*, with insightful comments on forthcoming trends.

Here are some excerpts: "Throughout print culture, information has been contained in books - and this has helped to shape our notion of information. For the information in books stays the same - it endures. And this has encouraged us to think of information as stable - as a body of knowledge which can be acquired, taught, passed on, memorised, and tested of course. The very nature of print itself has fostered a sense of truth; truth too is something which stays the same, which endures. And there is no doubt that this stability, this orderliness, has been a major contributor to the huge successes of the industrial age and the scientific revolution. (...)

But the digital revolution changes all this. Suddenly it is not the oldest information - the longest lasting information that is the most reliable and useful. It is the very latest information that we now put the most faith in - and which we will pay the most for. (...) Education will be about participating in the production of the latest information. This is why education will have to be ongoing throughout life and work. Every day there will be something new that we will all have to learn. To keep up. To be in the know. To do our jobs. To be members of the digital community. And far from teaching a body of knowledge that will last for life, the new generation of information professionals will be required to search out, add to, critique, 'play with', and daily update information, and to make available the constant changes that are occurring."

Russon Wooldridge, a professor in the Department of French Studies at the University of Toronto, Canada, wrote in February 2001: "All my teaching makes the most of internet resources (web and email): the two common places for a course are the classroom and the website of the course, where I put all course materials. I have published all my research data of the last 20 years on the web (re-edition of books, articles, texts of old dictionaries as interactive databases, treaties from the 16th century, etc.). I publish proceedings of symposiums, I publish a journal, I collaborate with French colleagues by publishing online in Toronto what they can't publish online at home. In May 2000, I organized an international symposium in Toronto about French studies enhanced by new technologies (Les études françaises valorisées par les nouvelles technologies). (...)

I realize that without the internet I wouldn't have as many activities, or at least they would be very different from the ones I have today. So I don't see the future without them. But it is crucial that those who believe in free dissemination of knowledge make sure that knowledge is not 'eaten' by commercial ventures for them to sell it. What has happened in book publishing in France, in linguistics for example, where you can only find textbooks for schools and exams, should be avoided on the web. You don't go to Amazon.com and the likes to find disinterested science. On my website, I refuse any sponsorship."

A few leading projects

MIT OpenCourseWare

The MIT OpenCourseWare (MIT OCW) is an initiative launched by MIT (Massachusetts Institute of Technology) to put its course materials for free on the web, as a way to promote open dissemination of knowledge. In September 2002, a pilot version was available online with 32 course materials. The website was officially launched in September 2003. 500 course materials were available in March 2004. In May 2006, 1,400 course materials were offered by 34 departments belonging to the five schools of MIT. In November 2007, all 1,800 course materials were available, with 200 new and updated courses per year.

MIT also launched the OpenCourseWare Consortium (OCW Consortium) in November 2005, as a collaboration of educational institutions that were willing to offer free online course materials. One year later, it included the course materials of 100 universities worldwide.

Public Library of Science

With the internet as a powerful medium to disseminate information, it seems quite outrageous that the results of research - original works requesting many years of efforts - are "squatted" by publishers claiming ownership on these works, and selling them at a high price. The work of researchers is often publicly funded, especially in North America. It would therefore seem appropriate that the scientific community and the general public can freely enjoy the results of such research. In science and medicine for example, more than 1,000 new articles reviewed by peers are published daily.

The Public Library of Science (PLoS) was founded in October 2000 by biomedical scientists Harold Varmus, Patrick Brown and Michael Eisen, from Stanford University, Palo Alto, and

University of California, Berkeley. Headquartered in San Francisco, PLoS is a non-profit organization whose mission is to make the world's scientific and medical literature a public resource in free online archives. Instead of information disseminated in millions of reports and thousands of online journals, a single point would give access to the full content of these articles, with a search engine and hyperlinks between articles.

PLoS posted an open letter requesting the articles presently published by journals to be distributed freely in online archives, and asking researchers to promote the publishers willing to support this project. From October 2000 to September 2002, the open letter was signed by 30,000 scientists from 180 countries. The publishers' answer was much less enthusiastic, although a number of publishers agreed for their articles to be distributed freely immediately after publication, or six months after publication in their journals. But even the publishers who initially agreed to support the project made so many objections that it was finally abandoned.

Another objective of PLoS was to become a publisher while creating a new model of online publishing based on free dissemination of knowledge. In early 2003, PLoS created a non-profit scientific and medical publishing venture to provide scientists and physicians with free high-quality, high-profile journals in which to publish their work. The journals were *PLoS Biology* (launched in 2003), *PLoS Medicine* (2004), *PLoS Genetics* (2005), *PLoS Computational Biology* (2005), *PLoS Pathogens* (2005), *PLoS Clinical Trials* (2006) and *PLoS Neglected Tropical Diseases* (2007), the first scientific journal on this topic.

All PLoS articles are freely available online, on the websites of PLoS and in the public archive PubMed Central, run by the National Library of Medicine. The articles can be freely redistributed and reused under a Creative Commons license, including for translations, as long as the author(s) and source are cited. PLoS also launched PLoS ONE, an online forum meant to publish articles on any subject relating to science or medicine.

Three years after the beginning of PLoS as a publisher, *PLoS Biology* and *PLoS Medicine* have had the same reputation for excellence as the leading journals *Nature*, *Science* and *The New England Journal of Medicine*. PLoS has received financial support from several foundations while developing a viable economic model from fees paid by published authors, advertising, sponsorship, and paid activities organized for PLoS members. PLoS also hopes to encourage other publishers to adopt the open access model, or to convert their existing journals to an open access model.

Wikipedia

Wikipedia was launched in January 2001 by Jimmy Wales and Larry Sanger (Larry resigned later on). It has quickly grown into the largest reference website on the internet, financed by donations, with no advertising. Its multilingual content is free and written collaboratively by people worldwide, who contribute under a pseudonym. Its website is a wiki, which means that anyone can edit, correct and improve information throughout the encyclopedia. The articles stay the property of their authors, and can be freely used according to the GFDL (GNU Free Documentation License).

In December 2004, Wikipedia had 1.3 million articles (by 13,000 contributors) in 100 languages. In December 2006, it had 6 million articles in 250 languages. In May 2007, it had 7 million articles in 192 languages, including 1.8 million articles in English, 589,000 articles in German, 500,000 articles in French, 260,000 articles in Portuguese, and 236,000 articles in Spanish.

Wikipedia is hosted by the Wikimedia Foundation, founded in June 2003, which has run a number of other projects, beginning with Wiktionary (launched in December 2002) and Wikibooks (launched in June 2003), followed by Wikiquote, Wikisource (texts from public domain), Wikimedia Commons (multimedia), Wikispecies (animals and plants), Wikinews, Wikiversity (textbooks), and Wiki Search (search engine).

Citizendium

Citizendium was launched in October 2006 as a pilot project to build a new encyclopedia, at the initiative of Larry Sanger, who was the cofounder of Wikipedia (with Jimmy Wales) in January 2001, but resigned later on, over policy and content quality issues. Citizendium - which stands for a "citizen's compendium of everything" - is a wiki project open to public collaboration, but combining "public participation with gentle expert guidance".

The project is experts-led, not experts-only. Contributors use their own names, not anonymous pseudonyms like in Wikipedia, and they are guided by expert editors. As explained by Larry in his essay *Toward a New Compendium of Knowledge*, posted in September 2006: "Editors will be able to make content decisions in their areas of specialization, but otherwise working shoulder-to-shoulder with ordinary authors." There are also constables who make sure the rules are respected.

Citizendium was launched on March 25, 2007, with 1,100 articles, 820 authors and 180 editors. There were 9,800 high-quality articles in January 2009, and 11,800 articles in August 2009. Citizendium also wants to act as a prototype for upcoming large scale knowledge-building projects that would deliver reliable reference, scholarly and educational content.

Encyclopedia of Life

The Encyclopedia of Life was launched in May 2007 as a global scientific effort to document all known species of animals and plants (1.8 million), including endangered species, and expedite the millions of species yet to be discovered and catalogued (about 8 million).

This collaborative effort is led by several main institutions: Field Museum of Natural History, Harvard University, Marine Biological Laboratory, Missouri Botanical Garden, Smithsonian Institution, Biodiversity Heritage Library (BHL). The initial funding came from the MacArthur Foundation (US \$10 million) and the Sloan Foundation (\$2.5 million). A \$100 million funding over ten years will be necessary before self-financing.

The multimedia encyclopedia will gather texts, photos, maps, sound and videos, with a webpage for each species. It will provide a single portal for millions of documents scattered online and offline. As a teaching and learning tool for a better understanding of our planet,

the encyclopedia wants to reach everyone: researchers, teachers, students, pupils, media, policy makers and the general public.

The encyclopedia's honorary chair is Edward Wilson, professor emeritus at Harvard University, who was the first to express the wish for such an encyclopedia, in an essay dated 2002. Five years later, his project could become reality thanks to technology improvements for content aggregators, mash-up, wikis, and large scale content management.

As a consortium of the ten largest life science libraries, the Biodiversity Heritage Library (BHL) started the digitization of 2 million documents from public domain spanning over 200 years. In May 2007, when the project was officially launched, 1.25 million pages were already digitized in London, Boston and Washington DC, and available in the Text Archive section of the Internet Archive.

The Encyclopedia of Life is built on the work of thousands of experts around the globe, in a moderated wiki-style environment, for the general public to be able to contribute. The first pages were available in mid-2008. The encyclopedia should be fully "operational" in 2012 and completed with all known species in 2017. The English version will be translated in several languages by partner organizations. People will be able to use the encyclopedia as a "macroscope" to identify major trends from a considerable stock of information - in the same way they use a microscope for the study of detail.

2003: eBooks are sold worldwide

[Overview]

First, publishers began to sell digital versions of their books online, on their own websites or on the new eBookstores of Amazon.com and Barnes & Noble.com. In 2000, new online bookstores were created to sell "only" digital books (ebooks), like Palm Digital Media (renamed Palm eBook Store), Mobipocket or Numilog. At the same time, publishers were digitizing their books by the hundreds, while the public was getting used to read ebooks on computers, laptops, phones, smartphones and reading devices. 2003 was a turning point in an emerging market. More and more books were published simultaneously as a print book and a digital book, and thousands of new books, beginning with best-sellers, were sold as ebooks in various formats: PDF (to be read on Acrobat Reader, replaced by Adobe Reader), LIT (to be read on Microsoft Reader), PRC (to be read on Mobipocket Reader) and others, with the Open eBook format becoming a standard for ebooks.

Books, from print to digital

The new online bookstores selling "only" digital books were also called aggregators because they were producing and selling ebooks from many publishers. It took them a few years (at least in Europe) to convince publishers that books should have two versions, print and digital, and to wait for the public to be ready to read on an electronic device, be it a computer, a laptop, a PDA, a mobile phone, a smartphone or a reading device. This emerging market took off in 2003, and more and more books were simultaneously published as a print book and a digital book.

In the 1990s, few people believed digital books would be commonplace in the near future. They thought people would still be attached to print books regardless of whatever happened, remembering this sentence of Robert Downs, a librarian who wrote in the 1980s: "My lifelong love affair with books and reading continues unaffected by automation, computers, and all other forms of the twentieth-century gadgetry." (excerpt from *Books in My Life*, Library of Congress, 1985)

In an article published in February 1996 by the Swiss magazine *Informatique-Informations*, Pierre Perroud, founder of the digital library Athena, explained that "electronic texts represent an encouragement to reading and a convivial participation to culture dissemination", particularly for textual research and text study. These texts are "a good complement to the print book, which remains irreplaceable when for 'true' reading. (...) The book remains a mysteriously holy companion with profound symbolism for us: we grip it in our hands, we hold it against us, we look at it with admiration; its small size comforts us and its content impresses us; its fragility contains a density we are fascinated by; like man it fears water and fire, but it has the power to shelter man's thoughts from time."

Roberto Hernández Montoya, an editor of the electronic magazine *Venezuela Analítica*, wrote in September 1998: "The printed text can't be replaced, at least not for the foreseeable future. The paper book is a tremendous 'machine'. We can't leaf through an electronic book in the same way as a paper book. On the other hand electronic use allows us to locate text

chains more quickly. In a certain way we can more intensively read the electronic text, even with the inconvenience of reading on the screen. The electronic book is less expensive and can be more easily distributed worldwide (if we don't count the cost of the computer and the internet connection)."

In the 2000s, while many people still prefer reading a print book, more and more readers enjoy reading their ebooks on their notebook, smartphone or any other electronic device. They buy their ebooks online from Amazon, Barnes & Noble, Yahoo, Palm, Mobipocket or Numilog.

In March 2000, Numilog was founded by Denis Zwim near Paris, France, as a company specializing in the distribution of digital books. Numilog launched in September 2000 an online bookstore that became the main French-speaking aggregator of digital books over the years. Numilog has sold books and audiobooks in partnership with a number of publishers, including Gallimard, POL, Le Dilettante, Le Rocher, La Découverte, De Vive Voix, Eyrolles or Pearson Education France. Numilog was bought in May 2008 by Hachette Livre, a leading publishing group.

Adobe Reader

Adobe launched PDF (Portable Document Format) in June 1993, with Acrobat Reader (free, to read PDF documents) and Adobe Acrobat (for a fee, to make PDF documents). As the "veteran" format, PDF was perfected over the years as a global standard for distribution and viewing of information. It "lets you capture and view robust information from any application, on any computer system and share it with anyone around the world. Individuals, businesses, and government agencies everywhere trust and rely on Adobe PDF to communicate their ideas and vision" (excerpt from the website). Adobe Acrobat gave the tools to create and view PDF files, in several languages and for several platforms (Windows, Mac, Linux).

In August 2000, Adobe bought Glassbook, a company specializing in digital books software for publishers, booksellers, distributors and libraries. Adobe also partnered with Amazon.com and Barnes & Noble.com to offer ebooks for the Acrobat Reader and the Glassbook Reader.

In January 2001, Adobe launched the Acrobat eBook Reader (free) and the Adobe Content Server (for a fee).

The Acrobat eBook Reader was used to read PDF files of copyrighted books, while adding notes and bookmarks, getting the book covers in a personal library, and browsing a dictionary.

The Adobe Content Server was intended for publishers and distributors for the packaging, protection, distribution and sale of copyrighted books in PDF format, while managing their access with DRM (Digital Rights Management), according to instructions given by the copyright holder, for example allowing or not the printing and loan of ebooks. (It was replaced by the Adobe LiveCycle Policy Server in November 2004.)

In April 2001, Adobe partnered with Amazon.com, for the online bookstore to include 2,000 copyrighted books for the Acrobat eBook Reader. These were titles of major publishers, travel guides, and children books.

The same year, the Acrobat Reader was available for PDAs, beginning with the Palm Pilot (May 2001) and the Pocket PC (December 2001).

Between 1993 and 2003, over 500 million copies of Acrobat Reader were downloaded worldwide. In 2003, Acrobat Reader was available in many languages and for many platforms (Windows, Mac, Linux, Palm OS, Pocket PC, Symbian OS, etc.). Approximately 10% of the documents on the internet were available in PDF.

In May 2003, Acrobat Reader (5th version) merged with Acrobat eBook Reader (2nd version) to become Adobe Reader (starting with version 6), which could read both standard PDF files and secure PDF files of copyrighted books.

In late 2003, Adobe opened its own online bookstore, the Digital Media Store, with titles in PDF format from major publishers (HarperCollins, Random House, Simon & Schuster, etc.) as well as electronic versions of newspapers and magazines like *The New York Times*, *Popular Science*, etc. Adobe also launched Adobe eBooks Central as a service to read, publish, sell and lend ebooks, and Adobe eBook Library as a prototype digital library.

Open eBook and ePub

In 1999, there were nearly as many ebook formats as ebooks, with each new company creating its own format for its own ebook reader (software) and its own electronic device, for example the Glassbook Reader, the Peanut Reader, the Rocket eBook Reader (for the Rocket eBook), the Franklin Reader (for the eBookMan), the Cytale ebook reader (for the Cybook), the Gemstar eBook Reader (for the Gemstar eBook), the Palm Reader (for the Palm Pilot), etc.

The digital publishing industry felt the need to work on a common format for ebooks. It released in September 1999 the first version of the Open eBook (OeB) format, based on XML (eXtensible Markup Language) and defined by the Open eBook Publication Structure (OeBPS). The Open eBook Forum was created in January 2000 to develop the OeB format and OeBPS specifications. Since 2000, most ebook formats were derived from - or are compatible with the OeB format, for example the PRC format from Mobipocket or the LIT format from Microsoft.

In April 2005, the Open eBook Forum became the International Digital Publishing Forum (IDPF). The OeB format was replaced with the ePub format, a global standard for ebooks with PDF. The PDF files created with recent versions of Adobe Acrobat are compatible with the ePub format.

Microsoft Reader

Microsoft launched the Microsoft Reader in April 2000, for people to read books in LIT (from "literature") format on its new PDA, the Pocket PC. Four months later, in August 2000, the Microsoft Reader was available for computers, and then for any Windows platform, for example the platforms of Tablets PC launched in November 2002.

Microsoft billed publishers and distributors for the use of its DRM technology through the Microsoft DAS Server, with a commission on each sale. Microsoft also partnered with major online bookstores - Barnes & Noble.com in January 2000 and Amazon.com in August 2000 - for them to offer ebooks for the Microsoft Reader in eBookstores soon to be launched. Barnes & Noble.com opened its eBookstore in August 2000, followed by Amazon in November 2000.

Mobipocket Reader

Mobipocket was founded in March 2000 in Paris, France, by Thierry Brethes and Nathalie Ting, as a company specializing in ebooks for PDAs. The Mobipocket format (PRC, based on the OeB format) and the Mobipocket Reader were "universal" and could be used on any PDA - and also on any computer from April 2002. They quickly became global standards for ebooks on mobile devices.

In October 2001, the Mobipocket Reader received the eBook Technology Award from the International Book Fair in Frankfurt. Mobipocket partnered with Franklin for the Mobipocket Reader to be available on the eBookMan, Franklin's personal assistant, instead of the initially planned Microsoft Reader.

The Mobipocket Web Companion was a software (for a fee) for extracting content from partner news sites. The Mobipocket Publisher was used by individuals (free version for private use, and standard version for a fee) or publishers (professional version for a fee) to create ebooks using the Mobipocket DRM technology for controlling access to copyrighted ebooks. The Mobipocket Publisher could also create ebooks in LIT format for the Microsoft Reader.

In spring 2003, the Mobipocket Reader was available in several languages (French, English, German, Spanish, Italian) and could be used on any PDA and any computer, and on the smartphones of Nokia and Sony Ericsson. 6,000 titles in several languages were available on Mobipocket's website and in partner online bookstores.

Mobipocket was bought by Amazon in April 2005. It now operates within the Amazon brand, with a multilingual catalog of 70,000 books in 2008.

2004: Authors are creative on the net

[Overview]

Some authors have enjoyed creating websites, posting their works and communicating with readers by email. Other authors have begun searching how using hyperlinks could expand their writing towards new directions, while linking it to images and sound. Jean-Paul switched from being a print author to being an hypermedia author, while enjoying the freedom given by online (self-)publishing: "The internet allows me to do without intermediaries such as record companies, publishers and distributors. Most of all, it allows me to crystallize what I have in my head: the print medium only allows me to partly do that. (...) Surfing the web is like radiating in all directions (I am interested in something and I click on all the links on a home page) or like jumping around (from one click to another, as the links appear). You can do this in the written media, of course. But the difference is striking. So the internet changed how I write. You don't write the same way for a website as you do for a script or a play."

The internet as a research tool

Murray Suid is a writer of educational books and material living in Palo Alto, in the heart of Silicon Valley. He has also written books for kids, multimedia scripts and screenplays. How did using the internet change his professional life? He wrote in September 1998: "The internet has become my major research tool, largely - but not entirely - replacing the traditional library and even replacing person-to-person research. Now, instead of phoning people or interviewing them face to face, I do it via email. Because of speed, it has also enabled me to collaborate with people at a distance, particularly on screenplays. (I've worked with two producers in Germany.) Also, digital correspondence is so easy to store and organize, I find that I have easy access to information exchanged this way. Thus, emailing facilitates keeping track of ideas and materials. The internet has increased my correspondence dramatically. Like most people, I find that email works better than snail mail. My geographic range of correspondents has also increased - extending mainly to Europe. In the old days, I hardly ever did transatlantic penpalling. I also find that emailing is so easy, I am able to find more time to assist other writers with their work - a kind of a virtual writing group. This isn't merely altruistic. I gain a lot when I give feedback. But before the internet, doing so was more of an effort."

Murray was among the first authors to add a website to his books - an opportunity that many would soon adopt: "If a book can be web-extended (living partly in cyberspace), then an author can easily update and correct it, whereas otherwise the author would have to wait a long time for the next edition, if indeed a next edition ever came out. (...) I do not know if I will publish books on the web - as opposed to publishing paper books. Probably that will happen when books become multimedia. (I currently am helping develop multimedia learning materials, and it is a form of teaching that I like a lot - blending text, movies, audio, graphics, and - when possible - interactivity)."

He added in August 1999: "In addition to 'web-extending' books, we are now web-extending our multimedia (CD-ROM) products - to update and enrich them."

In October 2000, "our company - EDVantage Software - has become an internet company instead of a multimedia (CD-ROM) company. We deliver educational material online to students and teachers."

The internet as a novel "character"

Alain Bron lives in Paris, France. He is a consultant in information systems and a writer. The internet is one of the "characters" of his second novel, *Sanguine sur toile* (Sanguine on the web), available in print from Editions du Choucas in 1999, and in PDF format from Editions 00h00 in 2000.

Alain wrote in November 1999: "In French, 'toile' means the web as well as the canvas of a painting, and 'sanguine' is the red chalk of a drawing as well as one of the adjectives derived from blood ('sang' in French). But would a love of colors justify a murder? *Sanguine sur toile* is the strange story of an internet surfer caught up in an upheaval inside his own computer, which is being remotely operated by a very mysterious person whose only aim is revenge. I wanted to take the reader into the worlds of painting and enterprise, which intermingle, escaping and meeting up again in the dazzle of software. The reader is invited to try to untangle for himself the threads twisted by passion alone. To penetrate the mystery, he will have to answer many questions. Even with the world at his fingertips, isn't the internet surfer the loneliest person in the world? In view of the competition, what is the greatest degree of violence possible in an enterprise these days? Does painting tend to reflect the world or does it create another one? I also wanted to show that images are not that peaceful. You can use them to take action, even to kill."

What part does the internet play in his novel? "The internet is a character in itself. Instead of being described in its technical complexity, it is depicted as a character that can be either threatening, kind or amusing. Remember the computer screen has a dual role - displaying as well as concealing. This ambivalence is the theme throughout. In such a game, the big winner is of course the one who knows how to free himself from the machine's grip and put humanism and intelligence before everything else."

The web and its hyperlinks

Like many artists, Jean-Paul began searching how hyperlinks could expand his writing towards new directions. He switched from being a print author to being an hypemedia author, and created *Cotres furtifs* (Furtive Cutters) as a website "telling stories in 3D". He enjoyed the freedom given by online (self-)publishing, and wrote in August 1999: "The internet allows me to do without intermediaries, such as record companies, publishers and distributors. Most of all, it allows me to crystallize what I have in my head: the print medium (desktop publishing, in fact) only allows me to partly do that."

He also insisted on the growing interaction between digital literature and technology. "The future of cyber-literature, techno-literature, digital literature or whatever you want to call it, is set by the technology itself. It is now impossible for an author to handle all by himself the words and their movement and sound. A decade ago, you could know well each of Director, Photoshop or Cubase (to cite just the better known software), using the first version of each.

That is not possible any more. Now we have to know how to delegate, find more solid financial partners than Gallimard, and look in the direction of Hachette-Matra, Warner, the Pentagon and Hollywood. At best, the status of multimedia director (?) will be the one of video director, film director, manager of the product. He is the one who receives the golden palms at Cannes, but who would never have been able to earn them just on his own. As twin sister (not a clone) of the cinematograph, cyber-literature (video + the link) will be an industry, with a few isolated craftsmen on the outer edge (and therefore with below-zero copyright)."

Jean-Paul added in June 2004: "Surfing the web is like radiating in all directions (I am interested in something and I click on all the links on a home page) or like jumping around (from one click to another, as the links appear). You can do this in the written media, of course. But the difference is striking. So the internet changed how I write. You don't write the same way for a website as you do for a script or a play. (...)

In fact, it is not the internet which changed how I write, it is the first Mac that I discovered through the self-learning of HyperCard. I still remember how astonished I was during the month when I was learning about buttons, links, surfing by analogies, objects or images. The idea that a simple click on one area of the screen allowed me to open a range of piles of cards, and each card could offer new buttons and each button opened on to a new range, etc. In brief, the learning of everything on the web that today seems really banal, for me it was a revelation (it seems Steve Jobs and his team had the same shock when they discovered the ancestor of the Mac in the laboratories of Rank Xerox). Since then I write directly on the screen: I use the print medium only occasionally, to fix up a text, or to give somebody who is allergic to the screen a kind of photograph, something instantaneous, something approximate. It is only an approximation, because print forces us to have a linear relationship: the text is developing page after page (most of the time), whereas the technique of links allows another relationship to the time and space of imagination. And, for me, it is above all the opportunity to put into practice this reading/writing 'cycle', whereas leafing through a book gives only an idea - which is vague because the book is not conceived for that."

2005: Google gets interested in ebooks

[Overview]

The beta version of Google Print went live in May 2005. In October 2004, Google launched the first part of Google Print as a project aimed at publishers, for internet users to be able to see excerpts from their books and order them online. In December 2004, Google launched the second part of Google Print as a project intended for libraries, to build up a world digital library by digitizing the collections of main partner libraries. In August 2005, Google Print was stopped until further notice because of lawsuits filed by associations of authors and publishers for copyright infringement. The program resumed in August 2006 under the new name of Google Books. Google Books has offered books digitized in the participating libraries (Harvard, Stanford, Michigan, Oxford, California, Virginia, Wisconsin-Madison, Complutense of Madrid and New York Public Library), with either the full text for public domain books or excerpts for copyrighted books. Google settled a lawsuit with associations of authors and publishers in October 2008, with an agreement to be signed in 2009.

Google Print

In October 2004, Google launched the first part of Google Print as a project aimed at publishers, for internet users to be able to see excerpts from their books and order them online. In December 2004, Google launched the second part of Google Print as a project intended for libraries, to build up a digital library of 15 million books by digitizing the collections of main partner libraries, beginning with the universities of Michigan (7 million books), Harvard, Stanford and Oxford, and the New York Public Library. The planned cost in 2004 was an average of US \$10 per book, and a total budget of \$150 to \$200 million for ten years. The beta version of Google Print went live in May 2005. In August 2005, Google Print was stopped until further notice because of lawsuits filed by associations of authors and publishers for copyright infringement.

Google Books

The program resumed in August 2006 under the new name of Google Books. Google Books has offered excerpts from books digitized by Google in the participating libraries - that now included Harvard, Stanford, Michigan, Oxford, California, Virginia, Wisconsin-Madison, Complutense of Madrid and New York Public Library. Google Books provided the full text for public domain books and excerpts for copyrighted books. According to some media buzz, Google was scanning 3,000 books a day.

The inclusion of copyrighted works in Google Books was widely criticized by authors and publishers worldwide. In the U.S., lawsuits were filed by the Authors Guild and the Association of American Publishers (AAP) for alleged copyright infringement. The assumption was that the full scanning and digitizing of copyrighted books infringed copyright laws, even if only snippets were made freely available. Google replied this was "fair use", referring to short excerpts from copyrighted books that could be lawfully quoted in another book or website, as long as the source (author, title, publisher) was mentioned. After three years of conflict, Google reached a settlement with the associations of authors and publishers in October 2008, with an agreement to be signed in 2009.

As of December 2008, Google had 24 library partners, including a Swiss one (University Library of Lausanne), a French one (Lyon Municipal Library), a Belgian one (Ghent University Library), a German one (Bavarian State Library), two Spanish ones (National Library of Catalonia and University Complutense of Madrid) and a Japanese one (Keio University Library). The U.S. partner libraries were, by alphabetical order: Columbia University, Committee on Institutional Cooperation (CIC), Cornell University Library, Harvard University, New York Public Library, Oxford University, Princeton University, Stanford University, University of California, University of Michigan, University of Texas at Austin, University of Virginia and University of Wisconsin-Madison.

2006: Towards a world public digital library

[Overview]

Conceived by the Internet Archive to offer a universal public digital library, the Open Content Alliance (OCA) was launched in October 2005 as a group of cultural, technology, non profit and governmental organizations willing to build a permanent archive of multilingual digitized text and multimedia content. The project took off in 2006, with the digitization of public domain books around the world. Unlike Google Books, the Open Content Alliance (OCA) has made them searchable through any web search engine, and has not scanned copyrighted books, except when the copyright holder has expressly given permission. The first contributors to OCA were the University of California, the University of Toronto, the European Archive, the National Archives in United Kingdom, O'Reilly Media and the Prelinger Archives. The digitized collections are freely available in the Text Archive section of the Internet Archive. In December 2008, one million ebooks were posted under OCA principles by the Internet Archive.

The Internet Archive and Yahoo! conceived the Open Content Alliance (OCA) in early 2005 to offer broad public access to the world culture. The OCA also wanted to address the issues of the Google Book project, with its copyright issues and its availability from one search engine only. The OCA was launched with the goal of digitizing only public domain books and making them searchable and downloadable through any search engine.

What exactly is the Internet Archive? Founded in April 1996 by Brewster Kahle, the Internet Archive is a non-profit organization that has built an "internet library" to offer permanent access to historical collections in digital format for researchers, historians and scholars. An archive of the web is stored every two months or so. In late 1999, the Internet Archive started to include more collections of archived webpages on specific topics. It also became an online digital library of text, audio, software, image and video content. In October 2001, with 30 billion stored webpages, the Internet Archive launched the Wayback Machine, for users to be able to surf the archive of the web by date. In 2004, there were 300 terabytes of data, with a growth of 12 terabytes per month. There were 65 billion pages (from 50 million websites) in 2006 and 85 million pages in 2008. The Internet Archive now defines itself as "a nonprofit digital library dedicated to providing universal access to human knowledge."

In October 2005, the Internet Archive launched the Open Content Alliance (OCA) with other contributors as a collective effort for "building a digital archive of global content for universal access" (subtitle of the OCA home page) that would be a permanent repository of multilingual text and multimedia content.

As explained on its website in 2007, OCA "is a collaborative effort of a group of cultural, technology, nonprofit, and governmental organizations from around the world that helps build a permanent archive of multilingual digitized text and multimedia material. An archive of contributed material is available on the Internet Archive website and through Yahoo! and

other search engines and sites. The OCA encourages access to and reuse of collections in the archive, while respecting the content owners and contributors."

The project aims at digitizing public domain books around the world and make them searchable through any web search engine and downloadable for free. Unlike Google Books, the OCA scans and digitizes only public domain books, except when the copyright holder has expressly given permission. The first contributors to the OCA were the University of California, the University of Toronto, the European Archive, the National Archives in United Kingdom, O'Reilly Media and Prelinger Archives. The digitized collections are freely available in the Text Archive section of the Internet Archive. 100,000 ebooks were publicly available in December 2006 (with 12,000 new ebooks added per month), 200,000 ebooks in May 2007, and one million ebooks in December 2008.

Microsoft has been one of the partners of the OCA, while also developing its own project. The beta version of Live Search Books was released in December 2006, with a search possible by keyword for non copyrighted books digitized by Microsoft in partner libraries. The British Library and the libraries of the universities of California and Toronto were the first ones to join in, followed in January 2007 by the New York Public Library and Cornell University. Books offered full text views and could be downloaded in PDF files. In May 2007, Microsoft announced agreements with several publishers, including Cambridge University Press and McGraw Hill, for their books to be available in Live Search Books. After digitizing 750,000 books and indexing 80 million journal articles, Microsoft ended the Live Search Books program in May 2008, to focus on other activities, and closed the website. These books are available in the OCA collections of the Internet Archive.

A main issue for digital libraries is the lack of proofreading of digitized books, that ensures a better accuracy of the text without any loss from the print version. The only digital library proofreading its books has been Project Gutenberg, with 30,000 high-quality ebooks available in 2008. Good OCR (Optical Character Recognition) software run on image files - obtained from scanning print pages - is said to ensure 99% accuracy. If the step of the proofreading seems essential to Project Gutenberg, whose goal is to reach a 99.99% accuracy for its ebooks - above the 99.95% accuracy set up as a standard for Library of Congress -, this step is skipped by the Internet Archive, OCA, Google and many others. Some R&D teams work on better quality OCR technology, which means that digital libraries would have to go back to the original image files to provide a higher quality book in the future, if they do want to provide digital versions without any loss from the print version.

2007: We read on various electronic devices

[Overview]

Amazon.com launched its own reading device, the Kindle, in November 2007. In the mid-1990s, people read on their desktop computers before reading on their laptops. The Palm Pilot was launched in March 1996 as the first PDA, and people began reading on PDAs. 23 million Palm Pilots were sold between 1996 and 2002. Its main competitors were the Pocket PC (launched by Microsoft in April 2000) and the PDAs of Hewlett-Packard, Sony, Handspring, Toshiba and Casio. People also began reading on the first smartphones launched by Nokia or Sony Ericsson. Some companies launched dedicated reading devices like the Rocket eBook, the SoftBook Reader, the Gemstar eBook and the Cybook, all models that didn't last long. Better reading devices emerged then, like the Cybook (new version) in 2004, the Sony Reader in 2006 and the Kindle in 2007. LCD screens were replaced by screens using the E Ink technology. The next step should be an ultra-thin flexible display called electronic paper (epaper), launched in 2010 by E Ink, Plastic Logic and others.

First reading devices

How about a book-sized electronic reader that could store many books at once? From 1998 onwards, some pioneer companies began working on dedicated reading devices, and launched the Rocket eBook (created by NuvoMedia), the EveryBook (created by EveryBook), the SoftBook (created by SoftBook Press), and the Millennium eBook (created by Librius.com).

The Rocket eBook was launched by NuvoMedia, in Palo Alto, California, as the first dedicated reading device. Founded in 1997, NuvoMedia wanted to become "the electronic book distribution solution, by providing a networking infrastructure for publishers, retailers and end users to publish, distribute, purchase and read electronic content securely and efficiently on the internet." Investors of NuvoMedia were Barnes & Noble and Bertelsmann. The connection between the Rocket eBook and the computer (PC or Macintosh) was made through the Rocket eBook Cradle, which provided power through a wall transformer, and connected to the computer with a serial cable.

EveryBook (EB) was "a living library in a single book". The EveryBook's electronic storage could hold 100 textbooks or 500 novels. The EveryBook used a "hidden" modem to dial into the EveryBook Store, for people to browse, purchase and receive full text books, magazines and sheet music.

SoftBook Press created the SoftBook along with the SoftBook Network, an internet-based content delivery service. With the SoftBook, people could "easily, quickly and securely download a wide selection of books and periodicals using its built-in internet connection" using a machine that, "unlike a computer, was ergonomically designed for the reading of long documents and books." The investors of Softbook Press were Random House and Simon & Schuster.

Librius was a "full-service e-commerce company" that launched a small "low-cost" reading device called the Millennium eBook. The website offered a World Bookstore that delivered

"digital copies of thousands of books via the internet."

The Gemstar eBook was launched in October 2000 by Gemstar-TV Guide International, a company providing digital products and services for the media. Gemstar first bought Nuvomedia (Rocket eBook) and SoftBook Press (SoftBook) in January 2000, as well as the French 00h00.com, a producer of digital books, in September 2000. Two Gemstar eBook were available for sale in the U.S. in November 2000, with a later attempt in Germany to test the European market. The REB 1100 had a black and white screen, like the Rocket eBook. The REB 1200 had a color screen, like the SoftBook Reader. Both were produced by RCA (Thomson Multimedia). New and cheaper models were then launched as GEB 1150 and 2150, produced by Gemstar instead of RCA. But the sales were still far below expectations. The company stopped selling reading devices in June 2003, and digital books the following month.

What people thought of them

In 2000 and 2001, I was interviewing some book professionals about these new reading devices they were so curious about, while wondering how a reading device could ever replace a print book. (As shown in the answers below, people often used the word "ebook" for an ebook reading device.)

Peter Raggett is the head of the Central Library at the OECD (Organization for Economic and Cooperation Development). He wrote in July 2000: "It is interesting to see that the electronic book mimics the traditional book as much as possible except that the paper page is replaced by a screen. I can see that the electronic book will replace some of the present paper products but not all of them. I also hope that electronic books will be waterproof so that I can continue reading in the bath."

Henk Slettenhaar is a professor in communication technologies at Webster University in Geneva, Switzerland. He wrote in August 2000: "I have a hard time believing people would want to read from a screen. I much prefer myself to read and touch a real book."

Randy Hobler is a consultant in internet marketing living in Dobbs Ferry, New York. He wrote in September 2000: "eBooks continue to grow as the display technology improves, and as the hardware becomes more physically flexible and lighter. Plus, among the early adapters will be colleges because of the many advantages for students (ability to download all their reading for the entire semester, inexpensiveness, linking into exams, assignments, need for portability, eliminating need to lug books all over)."

Eduard Hovy is the head of the Natural Language Group at USC/ISI (University of Southern California / Information Sciences Institute). He wrote in September 2000: "eBooks, to me, are a non-starter. More even than seeing a concert live or a film at a cinema, I like the physical experience holding a book in my lap and enjoying its smell and feel and heft. Concerts on TV, films on TV, and ebooks lose some of the experience; and with books particularly it is a loss I do not want to accept. After all, it is much easier and cheaper to get a book in my own purview than a concert or cinema. So I wish the ebook makers well, but I am happy with paper. And I don't think I will end up in the minority anytime soon - I am much less afraid of

books vanishing than I once was of cinemas vanishing."

Tim McKenna is an author who thinks and writes about the complexity of truth in a world of flux. He wrote in October 2000: "I don't think that they have the right appeal for lovers of books. The internet is great for information. Books are not information. People who love books have a relationship with their books. They re-read them, write in them, confer with them. Just as cybersex will never replace the love of a woman, ebooks will never be a vehicle for beautiful prose."

Steven Krauwer is the coordinator of ELSNET (European Network of Excellence in Human Language Technologies). He wrote in June 2001 that "ebooks still had a long way to go before reading from a screen feels as comfortable as reading a book."

Guy Antoine is the founder of Windows on Haiti, a reference website about Haitian culture. He wrote in June 2001: "Sorry, I haven't tried them yet. Perhaps because of this, it still appears to me like a very odd concept, something that the technology made possible, but for which there will not be any wide usage, except perhaps for classic reference texts. High school and college textbooks could be a useful application of the technology, in that there would be much lighter backpacks to carry. But for the sheer pleasure of reading, I can hardly imagine getting cozy with a good ebook."

PDA's

In the 1990s, Jacques Gauchey was a journalist and writer covering information technology in Silicon Valley. He was also a "facilitator" between the U.S. and Europe. Jacques was among the first to buy a Palm Pilot in March 1996, and wrote about it in his free online newsletter. As a side remark, he remembered in July 1999: "In 1996 I published a few issues of a free English newsletter on the internet. It had about ten readers per issue until the day when the electronic version of *Wired Magazine* created a link to it. In one week I got about 100 emails, some from French readers of my book *La vallée du risque - Silicon Valley* [The Valley of Risk: Silicon Valley, published by Plon, Paris, in 1990], who were happy to find me again." He added: "All my clients now are internet companies. All my working tools (my mobile phone, my PDA and my PC) are or will soon be linked to the internet."

Palm stayed the leader, despite fierce competition, with 23 million Palm Pilots sold between 1996 and 2002. In 2002, 36.8% of all PDAs available on the market were Palm Pilots. Its main competitor was Microsoft's Pocket PC. The main platforms were Palm OS (for 55% of PDAs) et Pocket PC (for 25,7%). In 2004, prices began to drop. The leaders were the PDAs of Palm, Sony, and Hewlett-Packard, followed by Handspring, Toshiba, and Casio.

Phones and reading devices

The first smartphone was Nokia 9210, launched as early as 2001. It was followed by Nokia Series 60, Sony Ericsson P800, and the smartphones of Motorola and Siemens. Smartphones took off quickly. In February 2005, Sony stopped selling PDAs. Smartphones represented 3,7% of all cellphones sold in 2004, and 9% of all cellphones sold in 2006, with 90 million smartphones sold for one billion cellphones.

Many people read ebooks on their PDAs, cellphones and smartphones. The favorite readers (software) were Mobipocket Reader (available in March 2000), Microsoft Reader (available in April 2000), Palm Reader (available in March 2001), Acrobat Reader (available in May 2001 for Palm Pilot, and in December 2001 for Pocket PC), and Adobe Reader (available in May 2003 to replace Acrobat Reader).

For cellphones, smartphones and dedicated reading devices, LCD screens have been replaced by screens using the technology developed by E Ink. As explained on the company's website: "Electronic ink is a proprietary material that is processed into a film for integration into electronic displays. Although revolutionary in concept, electronic ink is a straightforward fusion of chemistry, physics and electronics to create this new material. The principal components of electronic ink are millions of tiny microcapsules, about the diameter of a human hair. In one incarnation, each microcapsule contains positively charged white particles and negatively charged black particles suspended in a clear fluid. When a negative electric field is applied, the white particles move to the top of the microcapsule where they become visible to the user. This makes the surface appear white at that spot. At the same time, an opposite electric field pulls the black particles to the bottom of the microcapsules where they are hidden. By reversing this process, the black particles appear at the top of the capsule, which now makes the surface appear dark at that spot. To form an E Ink electronic display, the ink is printed onto a sheet of plastic film that is laminated to a layer of circuitry. The circuitry forms a pattern of pixels that can then be controlled by a display driver. These microcapsules are suspended in a liquid 'carrier medium' allowing them to be printed using existing screen printing processes onto virtually any surface, including glass, plastic, fabric and even paper. Ultimately electronic ink will permit most any surface to become a display, bringing information out of the confines of traditional devices and into the world around us."

Sony launched its first reading device, Librié 1000-EP, in Japan in April 2004, in partnership with Philips and E Ink. Librié was the first reading device to use the E Ink technology, with a 6-inch screen, a 10 M memory, and a 500-ebook capacity. eBooks were downloaded from a computer through a USB port. The Sony Reader was launched in October 2006 in the U.S. for US \$350, followed by cheaper and revamped models.

Amazon.com launched its own reading device, the Kindle, in November 2007. Before launching the Kindle, Amazon.com bought in April 2005 Mobipocket, a French company specializing in ebooks for PDAs, cellphones and smartphones, with a catalog of several thousands of books in several languages to be read on the Mobipocket Reader.

The Kindle was launched with a catalog of 80,000 ebooks - and new releases for US \$9,99 each. The built-in memory and 2G SD card gave plenty of book storage (1.4 G), with a screen using the E Ink technology, and page-turning buttons. Books were directly bought and downloaded via the device's 3G wireless connection, with no need for a computer, unlike the Sony Reader. 580,000 Kindles were sold in 2008. A thinner and revamped Kindle 2 was launched in February 2009, with a storage capacity of 1,500 ebooks, a new text-to-speech feature, and a catalog of 230,000 ebooks on Amazon.com's website.

Can reading devices like Sony Reader and Kindle really compete with cellphones and smartphones? Will people prefer reading on mobile handsets like the iPhone 3G (with its Stanza Reader) or the T-Mobile G1 (with Google's platform Android and its reader), or will they prefer using reading devices to get a larger screen? Is there a market for both smartphones and reading devices? These are some fascinating questions for the next years. I personally dream about a big flat screen on one of my walls, where I could display my friends' interactive PDFs and hypemedia stories, when I won't be on a budget anymore. In the meantime, I enjoy my netbook, including to read ebooks.

The next generation of reading devices - expected for 2010-11 - should display color and multimedia/hypemedia content with a revamped E Ink technology

The company Plastic Logic has become a key player for new products. As explained on its website: "Technology for plastic electronics on thin and flexible plastic substrates was developed at Cambridge University's renowned Cavendish Laboratory in the 1990s. In 2000, Plastic Logic was spun out of Cavendish Laboratory to develop a broad range of products using the plastic electronics technology. (...) Plastic Logic has raised over \$200M in financing from top-tier venture funding sources in Asia, Europe and the U.S. We are using the funds to complete product development in England and the USA, build a specialized, scalable production facility in Germany, and build our go-to-market teams." Plastic Logic intends to launch in 2010 a very thin and flexible 10.7" plastic screen, using proprietary plastic electronics and the E Ink technology.

Reading devices can count on some fierce competition with smartphones. In February 2009, the 1.5 million public-domain books available in Google Books - and 500,000 more outside the U.S. because of variations of copyright law - were accessible via mobile handsets such as the T-Mobile G1, released in October 2008 with Google's platform Android and its reader. Because of the small screens of mobile handsets, the ebooks are in text format, and not in image format. Android is an open source mobile device platform (built on Linux), that was announced in November 2007 along with the creation of the Open Handset Alliance (OHA). Other leading companies - Motorola, Lenovo, Sony Ericsson, Samsung, etc. - are working on smartphones that will run Android in the near future.

The @folio project

The @folio project is a reading device conceived as early as October 1996 by Pierre Schweitzer, an architect-designer living in Strasbourg, France. It is meant to download and read any text and/or illustrations from the web or hard disk, in any format, with no proprietary format and no DRM. Unfortunately, to this day (in August 2009), @folio has stayed a prototype, because of lack of funding and because of the language barrier - one article in English for dozens of articles in French.

The technology of @folio is novel and simple, and very different from other reading devices, past or present. It is inspired from fax and tab file folders. The flash memory is "printed" like Gutenberg printed his books. The facsimile mode is readable as is for any content, from sheet music to mathematical or chemical formulas, with no conversion necessary, whether it is handwritten text, calligraphy, free hand drawing or non-alphabetical writing. All this is

difficult if not impossible on a computer or any existing reading device.

The lightweight prototype is built with high-quality materials. The screen takes 80% of the total surface and has low power consumption. It is surrounded by a translucent and flexible frame that folds to protect the screen when not in use. @folio could be sold for US \$100 for the basic standard version, with various combinations of screen sizes and flash memory to fit the specific needs of architects, illustrators, musicians, specialists in old languages, etc.

Intuitive navigation allows to "turn" pages as easily as in a print book, to classify and search documents as easily as with a tab file folder, and to choose preferences for margins, paragraphs, font selection and character size. No buttons, only a round trackball adorned with the world map in black and white. The trackball can be replaced with a long and narrow tactile pad on either side of the frame.

The flash memory allows the downloading of thousands of hypertext pages, either previously linked before download or linked during the downloading process. @folio provides an instant automatic reformatting of documents, for them to fit the size of the screen. For "text" files, no software is necessary. For "image" files, the reformatting software is called Mot@Mot - Word@Word in French - and could be used with any other device. This software received much attention from the French National Library (BnF: Bibliothèque nationale de France) for a potential use in Gallica, its digital library of 90,000 books, especially for old books (published before 1812) and illustrated manuscripts.

Since its inception, the @folio project has received a warm welcome during guest presentations in various book fairs and symposiums in France and Europe, and a warm welcome from the French-speaking media - press, radio, television and internet. An international patent was filed in April 2001. The French startup iCodex was created in July 2002 to promote, develop and market @folio. A few years later, there is still a warm welcome, but yet no funding. In August 2007, the @folio team began seeking funding worldwide. Pierre's passion for a cheap and beautiful reading device intended for everybody - and not just the few - has no boundaries, except some financial ones.

2008: "A common information space in which we communicate"

[Overview]

Tim Berners-Lee, who invented the web in 1989-90, wrote in May 1998: "The dream behind the web is of a common information space in which we communicate by sharing information. Its universality is essential: the fact that a hypertext link can point to anything, be it personal, local or global, be it draft or highly polished. There was a second part of the dream, too, dependent on the web being so generally used that it became a realistic mirror (or in fact the primary embodiment) of the ways in which we work and play and socialize. That was that once the state of our interactions was on line, we could then use computers to help us analyse it, make sense of what we are doing, where we individually fit in, and how we can better work together" (excerpt from The World Wide Web: A Very Short Personal History, available on the W3C website). In 2008, Tim Berners-Lee's dream and "second part of the dream" have begun to become reality, with many participative projects across borders and languages.

From etexts to ebooks

Michael Hart founded Project Gutenberg in 1971. He wrote in 1998: "We consider etext to be a new medium, with no real relationship to paper, other than presenting the same material, but I don't see how paper can possibly compete once people each find their own comfortable way to etexts, especially in schools."

John Mark Ockerbloom created the Online Books Page in 1993. He wrote in 1998: "I've gotten very interested in the great potential the net has for making literature available to a wide audience. (...) I am very excited about the potential of the internet as a mass communication medium in the coming years. I'd also like to stay involved, one way or another, in making books available to a wide audience for free via the net, whether I make this explicitly part of my professional career, or whether I just do it as a spare-time volunteer."

Ten years later, Peter Schweitzer, inventor of the @folio project, the prototype of a reading device, wrote in an email interview: "The luck we all have is to live here and now this fantastic change. When I was born in 1963, computers didn't have much memory. Today, my music player could hold billions of pages, a true local library. Tomorrow, by the combined effect of the Moore Law and the ubiquity of networks, we will have instant access to works and knowledge. We won't be much interested any more on which device to store information. We will be interested in handy functions and beautiful objects."

Marc Autret, a journalist and graphic designer, wrote around the same time: "I am convinced that the ebook (or "e-book") has a great future in all non-fiction sectors. I refer to the ebook as a software and not as a dedicated physical medium (the conjecture is more uncertain on this point). The [European] publishers of guides, encyclopedias and informative books in general still see the ebook as a very minor variation of the printed book, probably because the business model and secure management don't seem entirely stabilized. But this is a matter of time. Non-commercial ebooks are already emerging everywhere while opening the way to new developments. To my eyes, there are at least two emerging trends: (a) an

increasingly attractive and functional interface for reading/consultation (navigation, research, restructuring on the fly, user annotations, interactive quiz); (b) a multimedia integration (video, sound, animated graphics, database) now strongly coupled to the web. No physical book offers such features. So I imagine the ebook of the future as a kind of wiki crystallized and packaged in a format. How valuable will it be? Its value will be the one of a book: the unity and quality of editorial work!"

Cyberspace and information society

Over the years, I asked people I was interviewing by email how they would define cyberspace and information society. Here are a few answers, to open new perspectives that will happily replace a "conclusion" for this book.

According to Peter Raggett, head of the Center for Documentation & Information at the OECD (Organization for Economic and Cooperation Development): "Cyberspace is that area 'out there' which is on the other end of my PC when I connect to the internet. Any ISP (Internet Service Provider) or webpage provider is in cyberspace as far as his users or customers are concerned." And the information society? "The information society is the society where the most valued product is information. Up to the 20th century, manufactured goods were the most valued products. They have been replaced by information. In fact, people are now talking of the knowledge society where the most valuable economic product is the knowledge inside our heads."

Steven Krauwer is the coordinator of ELSNET (European Network of Excellence in Human Language Technologies). "For me the cyberspace is the part of the universe (including people, machines and information) that I can reach from behind my desk." And the information society? "An information society is a society: (a) where most of the knowledge and information is no longer stored in people's brains or books but on electronic media; (b) where the information repositories are distributed, interconnected via an information infrastructure, and accessible from anywhere; (c) where social processes have become so dependent on this information and the information infrastructure that citizens who are not connected to this information system cannot fully participate in the functioning of the society."

Guy Antoine is the founder of Windows on Haiti, a reference website about Haitian culture. For him, cyberspace is "literally the newest frontier for mankind, a place where everyone can claim his place, and do so with relative ease and a minimum of financial resources, before heavy intergovernmental regulations and taxation finally set in. But then, there will be another."

Henk Slettenhaar is a professor in communication technologies at Webster University in Geneva, Switzerland. For him, cyberspace is "our virtual space. The area of digital information (bits, not atoms). It is a limited space when you think of the spectrum. It has to be administered well so all the earth's people can use it and benefit from it (eliminate the digital divide)." And the information society is "the people who already use cyberspace in their daily lives to such an extent that it is hard to imagine living without it (the other side of the divide)."

Tim McKenna is an author who thinks and writes about the complexity of truth in a world of flux. "Cyberspace to me is the distance that is bridged when individuals use technology to connect, either by sharing information or chatting. To say that one exists in cyberspace is really to say that he has eliminated distance as a barrier to connecting with people and ideas." And the information society? "The information society to me is the tangible form of Jung's collective consciousness. Most of the information resides in the subconsciousness but browsing technology has made the information more retrievable which in turn allows us greater self-knowledge both as individuals and as human beings."

Chronology

[Each line begins with the year or the year/month.]

- 1968: ASCII is the first character set encoding.
- 1971: Project Gutenberg is the first digital library.
- 1974: The internet takes off.
- 1977: UNIMARC is created as a common bibliographic format for library catalogs.
- 1984: Copyleft is a new license for computer software.
- 1990: The web is invented by Tim Berners-Lee.
- 1991/01: Unicode is a universal character set encoding for all languages.
- 1993/01: The Online Books Page is a list of free ebooks on the internet.
- 1993/06: Adobe launches PDF, Acrobat Reader and Adobe Acrobat.
- 1993/11: Mosaic is the first web browser.
- 1994: The first library website goes online.
- 1994: Bold publishers post free digital versions of copyrighted books.
- 1995/07: Amazon.com is the first main online bookstore.
- 1995: Mainstream print newspapers and magazines launch their own websites.
- 1996/03: The Palm Pilot is launched as the first PDA.
- 1996/04: The Internet Archive is founded to archive the web.
- 1996: Teachers explore new ways of teaching using the internet.
- 1997/01: Multimedia convergence is the topic of a symposium.
- 1997/04: E Ink begins developing a technology called electronic ink.
- 1997: Online publishing begins spreading.
- 1997: The Logos Dictionary goes online for free.
- 1998/05: 00h00.com sells books "only" in digital format.
- 1998: Library treasures like *Beowulf* go online.
- 1999/09: The Open eBook (OeB) format is created as a standard for ebooks.
- 1999/12: Britannica.com is available for free on the web (for a short time).
- 1999: Librarians become webmasters.
- 1999: Authors go digital.
- 2000/01: The Million Book Project wants to digitize one million books.
- 2000/02: yourDictionary.com is a major language portal.
- 2000/03: Mobipocket focuses on readers (software) and ebooks for PDAs.
- 2000/07: Non-English-speaking internet users reach 50%.
- 2000/07: Stephen King (self-)publishes a novel "only" on the web.
- 2000/08: Microsoft launches its own reader (software) and LIT format.
- 2000/09: GDT is a main bilingual (English, French) free translation dictionary.
- 2000/09: Numilog is an online bookstore selling "only" digital books.
- 2000/09: Handicapzero is a portal for the visually impaired and blind community.
- 2000/10: The Public Library of Science works on free online journals.
- 2000/10: Distributed Proofreaders helps in digitizing books from public domain.
- 2000/11: The British Library posts the digitized *Bible of Gutenberg*.
- 2001/01: Wikipedia is a main free online cooperative encyclopedia.
- 2001: Creative Commons works on new ways of respecting authors' rights.

2003/09: MIT offers its course materials for free in its OpenCourseWare.
2004/01: Project Gutenberg Europe is launched as a multilingual project.
2004/10: Google launches Google Print to rename it Google Books later on.
2005/04: Amazon.com buys Mobipocket, its software and ebooks.
2005/10: The Open Content Alliance works on a universal public digital library.
2006/08: Google Books has several partner libraries and publishers.
2006/08: The union catalog WorldCat is available for free on the web.
2006/10: Sony launches its new reading device, the Sony Reader.
2006/12: Microsoft launches Live Search Books (and drops the project later on).
2007/03: Citizendium works on a main "reliable" online cooperative encyclopedia.
2007/03: IATE is the new terminological database of the European Community.
2007/05: The Encyclopedia of Life will document all known species of animals and plants.
2007/11: Amazon.com launches Kindle, its own reading device.
2008/05: Hachette Livre buys the digital bookstore Numilog.
2008/10: Google Books settles a lawsuit with associations of authors and publishers.
2008/11: Europeana starts as the European digital library.
2009/02: Amazon.com launches Kindle 2.

Acknowledgements

Many thanks to all those who kindly answered my questions over the years. Most interviews were published by NEF (Net des études françaises / Net of French Studies), University of Toronto, Canada. They are available online <<http://www.etudes-francaises.net/entretiens/index.html>>. Some interviews were directly included in this book.

Many thanks to Nicolas Ancion, Alex Andrachmes, Guy Antoine, Silvaine Arabo, Arlette Attali, Marc Autret, Isabelle Aveline, Jean-Pierre Balpe, Emmanuel Barthe, Robert Beard, Michael Behrens, Michel Benoît, Guy Bertrand, Olivier Bogros, Christian Boitet, Bernard Boudic, Bakayoko Bourahima, Marie-Aude Bourson, Lucie de Boutiny, Anne-Cécile Brandenbourger, Alain Bron, Patrice Cailleaud, Tyler Chambers, Pascal Chartier, Richard Chotin, Alain Clavet, Jean-Pierre Cloutier, Jacques Coubard, Luc Dall'Armellina, Kushal Dave, Cynthia Delisle, Émilie Devriendt, Bruno Didier, Catherine Domain, Helen Dry, Bill Dunlap, Pierre-Noël Favennec, Gérard Fourestier, Pierre François Gagnon, Olivier Gainon, Jacques Gauchey, Raymond Godefroy, Muriel Goiran, Marcel Grangier, Barbara Grimes, Michael Hart, Roberto Hernández Montoya, Randy Hobler, Eduard Hovy, Christiane Jadelot, Gérard Jean-François, Jean-Paul, Anne-Bénédicte Joly, Brian King, Geoffrey Kingscott, Steven Krauwer, Gaëlle Lacaze, Michel Landaret, Hélène Larroche, Pierre Le Loarer, Claire Le Parco, Annie Le Saux, Fabrice Lhomme, Philippe Loubière, Pierre Magnenat, Xavier Malbreil, Alain Marchiset, Maria Victoria Marinetti, Michael Martin, Tim McKenna, Emmanuel Ménard, Yoshi Mikami, Jacky Minier, Jean-Philippe Mouton, Greg Newby, John Mark Ockerbloom, Caoimhín Ó Donnaíle, Jacques Pataillot, Alain Patez, Nicolas Pewny, Marie-Joseph Pierre, Hervé Ponsot, Olivier Pujol, Anissa Racheff, Peter Raggett, Patrick Rebollar, Philippe Renaut, Jean-Baptiste Rey, Philippe Rivière, Blaise Rosnay, Bruno de Sa Moreira, Pierre Schweitzer, Henk Slettenhaar, Murray Suid, June Thompson, Zina Tucsnak, François Vadrot, Christian Vandendorpe, Robert Ware, Russon Wooldridge, and Denis Zwirn.

Many thanks to Greg Chamberlain, Laurie Chamberlain, Kimberly Chung, Mike Cook, Michael Hart and Russon Wooldridge for revising previous versions of some parts. The author, whose mother tongue is French, is responsible for any remaining mistakes.