

# Technology and Books for All

Marie Lebert

NEF, University of Toronto, 2008

Copyright © 1999-2008 Marie Lebert

From Project Gutenberg in 1971 to the Encyclopedia of Life in 2007, 38 milestones and as many pages, with an overview and an in-depth description for each milestone. This book is also available in French, with a different text. Both versions are available on the NEF <<http://www.etudes-francaises.net/dossiers/technologies.htm>>.

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

Most quotations are excerpts from NEF interviews. With many thanks to all the persons who are quoted here, and who kindly answered my questions over the years. Most interviews are available online at <[www.etudes-francaises.net/entretiens/index.htm](http://www.etudes-francaises.net/entretiens/index.htm)>.

With many thanks to Greg Chamberlain, Laurie Chamberlain, Kimberly Chung, Mike Cook, Michael Hart and Russon Wooldridge, who kindly edited and/or proofread some parts in previous versions. The author, whose mother tongue is French, is responsible for any remaining mistakes.

# Table

## Introduction

1968: ASCII  
1971: Project Gutenberg  
1974: Internet  
1977: UNIMARC  
1984: Copyleft  
1990: Web  
1991: Unicode  
1993: Online Books Page  
1993: PDF  
1994: Library Websites  
1994: Bold Publishers  
1995: Amazon.com  
1995: Online Press  
1996: Palm Pilot  
1996: Internet Archive  
1996: New Ways of Teaching  
1997: Digital Publishing  
1997: Logos Dictionary  
1997: Multimedia Convergence  
1998: Online Beowulf  
1998: Digital Librarians  
1998: Multilingual Web  
1999: Open eBook Format  
1999: Digital Authors  
2000: yourDictionary.com  
2000: Online Bible of Gutenberg  
2000: Distributed Proofreaders  
2000: Public Library of Science  
2001: Wikipedia  
2001: Creative Commons  
2002: MIT OpenCourseWare  
2004: Project Gutenberg Europe  
2004: Google Books  
2005: Open Content Alliance  
2006: Microsoft Live Search Books  
2006: Free WorldCat  
2007: Citizendium  
2007: Encyclopedia of Life

## Websites

## Introduction

Michael Hart, who founded Project Gutenberg in 1971, wrote: "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." (excerpt from a NEF interview, August 1998)

Tim Berners-Lee, who invented the web in 1989-90, wrote: "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*, May 1998)

John Mark Ockerbloom, who created The Online Books Page in 1993, wrote: "I've gotten very interested in the great potential the net had 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." (excerpt from a NEF interview, September 1998)

Here is the journey we are going to follow:

1968: ASCII is a 7-bit coded character set.

1971: Project Gutenberg is the first digital library.

1974: The internet takes off.

1977: UNIMARC is set up as a common bibliographic format.

1984: Copyleft is a new license for computer software.

1990: The web takes off.

1991: Unicode is a universal double-byte character set.

1993: The Online Books Page is a list of free eBooks.

1993: The PDF format is launched by Adobe.

1994: The first library website goes online.

1994: Publishers put some of their books online for free.

1995: Amazon.com is the first main online bookstore.

1995: The mainstream press goes online.

1996: The Palm Pilot is the first PDA.

1996: The Internet Archive is founded to archive the web.

1996: Teachers explore new ways of teaching.

1997: Online publishing begins spreading.

1997: The Logos Dictionary goes online for free.

1997: Multimedia convergence is the topic of an international symposium.

1998: Library treasures like Beowulf go online.  
1999: Librarians become webmasters.  
1998: The web becomes multilingual.  
1999: The Open eBook format is a standard for eBooks.  
1999: Authors go digital.  
2000: yourDictionary.com is a language portal.  
2000: The Bible of Gutenberg goes online.  
2000: Distributed Proofreaders digitizes books from public domain.  
2000: The Public Library of Science (PLoS) works on free online journals.  
2001: Wikipedia is the first main online cooperative encyclopedia.  
2001: Creative Commons works on new ways to respect authors' rights on the web.  
2003: MIT offers its course materials for free in its OpenCourseWare.  
2004: Project Gutenberg Europe is launched as a multilingual project.  
2004: Google launches Google Print to rename it Google Books.  
2005: The Open Content Alliance (OCA) launches a world public digital library.  
2006: Microsoft launches Live Search Books as its own digital library.  
2006: The union catalog WorldCat goes online for free.  
2007: Citizendium is a main online "reliable" cooperative encyclopedia.  
2007: The Encyclopedia of Life will document all species of animals and plants.

[Unless specified otherwise, all quotations are excerpts from NEF interviews. These interviews are available online at <<http://www.etudes-francaises.net>>.]

## **1968: ASCII**

### **[Overview]**

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. Plain Vanilla ASCII can be read, written, copied and printed by any simple text editor or word processor. It is the only format compatible with 99% of all hardware and software. It can be used as it is or to create versions in many other formats. Extensions of ASCII (also called ISO-8859 or ISO-Latin) are sets of 256 characters that include accented characters as found in French, Spanish and German, for example ISO 8859-1 (Latin-1) for French.

### **[In Depth (published in 2005)]**

Whether digitized years ago or now, all Project Gutenberg books are created in 7-bit plain ASCII, called Plain Vanilla ASCII. When 8-bit ASCII (also called ISO-8859 or ISO-Latin) is used for books with accented characters like French or German, Project Gutenberg also produces a 7-bit ASCII version with the accents stripped. (This doesn't apply for languages that are not "convertible" in ASCII, like Chinese, encoded in Big-5.)

Project Gutenberg sees Plain Vanilla ASCII as the best format by far. It is "the lowest common denominator." It can be read, written, copied and printed by any simple text editor or word processor on any electronic device. It is the only format compatible with 99% of hardware and software. It can be used as it is or to create versions in many other formats. It will still be used while other formats will be obsolete (or are already obsolete, like formats of a few short-lived reading devices launched since 1999). It is the assurance collections will never be obsolete, and will survive future technological changes. The goal is to preserve the texts not only over decades but over centuries. There is no other standard as widely used as ASCII right now, even Unicode, a universal double-byte character encoding launched in 1991 to support any language and any platform.

# 1971: Project Gutenberg

## [Overview]

In July 1971, Michael Hart created Project Gutenberg with the goal of making available for free, and electronically, literary works belonging to public domain. A pioneer site in a number of ways, Project Gutenberg was the first information provider on the internet and is the oldest digital library. When the internet became popular in the mid-1990s, the project got a boost and gained an international dimension. The number of electronic books rose from 1,000 (in August 1997) to 5,000 (in April 2002), 10,000 (in October 2003), 15,000 (in January 2005), 20,000 (in December 2006) and 25,000 (in April 2008), with a current production rate of around 340 new books each month. With 55 languages and 40 mirror sites around the world, books are being downloaded by the tens of thousands every day. Project Gutenberg promotes digitization in "text format", meaning that a book can be copied, indexed, searched, analyzed and compared with other books. Contrary to other formats, the files are accessible for low-bandwidth use. The main source of new Project Gutenberg eBooks is Distributed Proofreaders, conceived in October 2000 by Charles Franks to help in the digitizing of books from public domain.

## [In Depth (published in 2005, updated in 2008)]

The electronic book (eBook) is now 37 years old, which is still a short life comparing to the five and a half century print book. eBooks were born with Project Gutenberg, created by Michael Hart in July 1971 to make available for free electronic versions of literary books belonging to public domain. A pioneer site in a number of ways, Project Gutenberg was the first information provider on an embryonic internet and is the oldest digital library. Long considered by its critics as impossible on a large scale, Project Gutenberg had 25,000 books in April 2008, with tens of thousands downloads daily. To this day, nobody has done a better job of putting the world's literature at everyone's disposal, while creating a vast network of volunteers all over the world, without wasting people's skills or energy.

During the first twenty years, Michael Hart himself keyed in the first hundred books, with the occasional help of others. When the internet became popular, in the mid-1990s, the project got a boost and gained an international dimension. Michael still typed and scanned in books, but now coordinated the work of dozens and then hundreds of volunteers across many countries. The number of electronic books rose from 1,000 (in August 1997) to 2,000 (in May 1999), 3,000 (in December 2000) and 4,000 (in October 2001).

37 years after its birth, Project Gutenberg is running at full capacity. It had 5,000 books online in April 2002, 10,000 books in October 2003, 15,000 books in January 2005, 20,000 books in December 2006 and 25,000 books in April 2008, with 340 new books available per month, with 40 mirror sites worldwide, and with books downloaded by the tens of thousands every day.

Whether they were digitized 30 years ago or digitized now, all the books are captured in Plain Vanilla ASCII (the original 7-bit ASCII), with the same formatting rules, so they can be read

easily by any machine, operating system or software, including on a PDA, a cellphone or an eBook reader. Any individual or organization is free to convert them to different formats, without any restriction except respect for copyright laws in the country involved.

In January 2004, Project Gutenberg had spread across the Atlantic with the creation of Project Gutenberg Europe. On top of its original mission, it also became a bridge between languages and cultures, with a number of national and linguistic sections. While adhering to the same principle: books for all and for free, through electronic versions that can be used and reproduced indefinitely. And, as a second step, the digitization of images and sound, in the same spirit.

# 1974: Internet

## [Overview]

When Project Gutenberg began in July 1971, the internet was not even born. On July 4, 1971, on Independence Day, Michael keyed in The United States Declaration of Independence (signed on July 4, 1776) to the mainframe he was using. In uppercase, because there was no lower case yet. But to send a 5K file to the 100 users of the embryonic internet would have crashed the network. So Michael mentioned where the eText was stored (though without a hypertext link, because the web was still 20 years ahead). It was downloaded by six users. 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. It got a boost with the invention of the web in 1990 and of the first browser in 1993. At the end of 1997, there were 90 to 100 million users, with one million new users every month. At the end of 2000, there were over 300 million users.



# 1977: UNIMARC

## [Overview]

In 1977, the IFLA (International Federation of Library Associations) published the first edition of UNIMARC: Universal MARC Format, followed by a second edition in 1980 and a UNIMARC Handbook in 1983. UNIMARC (Universal Machine Readable Cataloging) is a common bibliographic format for library catalogs, as a solution to the 20 existing national MARC (Machine Readable Cataloging) formats, which meant lack of compatibility and extensive editing when bibliographical 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.

## [In Depth (published in 1999)]

At the time, the future of online catalogs was linked to the harmonization of the MARC format. Set up in the early 1970s, MARC is an acronym for Machine Readable Catalogue. This acronym is rather misleading as MARC is neither a kind of catalog nor a method of cataloguing. According to UNIMARC: An Introduction, a document of the Universal Bibliographic Control and International MARC Core Programme, MARC is "a short and convenient term for assigning labels to each part of a catalogue record so that it can be handled by computers. While the MARC format was primarily designed to serve the needs of libraries, the concept has since been embraced by the wider information community as a convenient way of storing and exchanging bibliographic data."

After MARC came MARC II. MARC II established rules to be followed consistently over the years. The MARC communication format intended to be "hospitable to all kinds of library materials; sufficiently flexible for a variety of applications in addition to catalogue production; and usable in a range of automated systems."

Over the years, however, despite cooperation efforts, several versions of MARC emerged, e.g. UKMARC, INTERMARC and USMARC, whose paths diverged because of different national cataloguing practices and requirements. We had an extended family of more than 20 MARC formats. Differences in data content meant some extensive editing was needed before records could be exchanged.

One solution to incompatible data was to create an international MARC format - called UNIMARC - which would accept 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, so that each national bibliographic agency would need to write only two programs - one to convert into UNIMARC and one to convert from UNIMARC - instead of having to write twenty programs for the conversion of each MARC format (e.g. INTERMARC to UKMARC, USMARC to UKMARC etc.).

In 1977, the IFLA (International Federation of Library Associations and Institutions) published UNIMARC: Universal MARC Format, followed by a second edition in 1980 and a UNIMARC

Handbook in 1983. These publications focused primarily on the cataloguing of monographs and serials, while taking into account international efforts towards the standardization of bibliographic information reflected in the ISBDs (International Standard Bibliographic Descriptions).

In the mid-1980s, UNIMARC expanded to cover documents other than monographs and serials. A new UNIMARC Manual was produced in 1987, with an updated description of UNIMARC. By this time UNIMARC had been adopted by several bibliographic agencies as their in-house format.

Developments didn't stop there. A standard for authorities files was set up in 1991, as explained on the website of IFLA in 1998: "Previously agencies had entered an author's name into the bibliographic format as many times as there were documents associated with him or her. With the new system they created a single authoritative form of the name (with references) in the authorities file; the record control number for this name was the only item included in the bibliographic file. The user would still see the name in the bibliographic record, however, as the computer could import it from the authorities file at a convenient time. So in 1991 UNIMARC/Authorities was published."

In 1991 a Permanent UNIMARC Committee was also created to regularly monitor the development of UNIMARC. Users realized that continuous maintenance - and not just the occasional rewriting of manuals - was needed, to make sure all changes were compatible with what already existed.

On top of adopting UNIMARC as a common format, The British Library (using UKMARC), the Library of Congress (using USMARC) and the National Library of Canada (using CAN/MARC) worked on harmonizing their national MARC formats. A three-year program to achieve a common MARC format was agreed on by the three libraries in December 1995.

Other libraries began using SGML (Standard Generalized Markup Language) as a common format for both the bibliographic records and the hypertextual and multimedia documents linked to them. As most publishers were using SGML for book records, librarians and publishers began working on a convergence between MARC and SGML. The Library of Congress worked on a DTD (Definition of Type of Document, which defines its logical structure) for the USMARC format. A DTD for the UNIMARC format was developed by the European Union. Some European libraries chose SGML to encode their bibliographic data. In the Belgian Union Catalog, for example, the use of SGML allowed to add descriptive elements and to facilitate the production of an annual CD-ROM.

## 1984: Copyleft

### [Overview]

The term "copyleft" was invented in 1984 by Richard Stallman, who was a computer scientist at MIT (Massachusetts Institute of Technology). "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." (excerpt from the GNU website)

# 1990: Web

## [Overview]

The internet got its first boost with the invention of the web and its hyperlinks by Tim Berners-Lee at CERN (European Laboratory for Particle Physics) in 1990, and a second boost with the invention of the first browser Mosaic in 1993. The internet could now be used by anyone, and not only by computer literate people. 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 the number of English-speaking users, with a percentage of 50-50. According to Netcraft, an internet services company, 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).

## [In Depth (published in 1999, updated in 2008)]

The World Wide Web -that became the Web or web- was invented by Tim Berners-Lee in 1989-90. In 1998, he stated: "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 analyze 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, May 1998.)

Christiane Jadelot, researcher at INaLF-Nancy (INaLF: National Institute of the French Language) wrote: "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 e-mail, file transfers and dial-up connections. At that time I had problems with a programme called Paradox and character sets that 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!" (July 1998)

The W3C (World Wide Web Consortium) was founded in October 1994 to develop interoperable technologies (specifications, guidelines, software and tools) for the web, as a forum for information, commerce, communication and collective understanding. The W3C develops common protocols to lead the evolution of the web, for example the specifications of HTML (HyperText Markup Language) and XML (eXtensible Markup Language). HTML is used for publishing hypertext on the web. XML was originally designed as a tool for large-scale electronic publishing. It now plays an increasingly important role in the exchange of a wide

variety of data on the web and elsewhere.

According to the network tracking firm Netcraft, there were 100 million websites on November 1st, 2006. Previous milestones in the survey were reached in April 1997 (1 million sites), February 2000 (10 million), September 2000 (20 million), July 2001 (30 million), April 2003 (40 million), May 2004 (50 million), March 2005 (60 million), August 2005 (70 million), April 2006 (80 million ) and August 2006 (90 million).

## **1991: Unicode**

### **[Overview]**

First published in January 1991, Unicode is the universal character encoding maintained by the Unicode Consortium. "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 a component of the W3C (World Wide Web Consortium) specifications.

## **1993: Online Books Page**

### **[Overview]**

Founded in 1993 by John Mark Ockerbloom while he was a student at Carnegie Mellon University, 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." (excerpt from the website) John Ockerbloom 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 listed 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 2007. 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 lists copyright information about most countries in the world with links to further reading.

### **[In Depth (published in 1999)]**

John Mark Ockerbloom first started the website of the School of Computer Science of Carnegie Mellon University (CMU CS), and began maintaining The Online Books Page on it. Web space and computing resources were provided by the School of Computer Science.

Interviewed by email in September 1998, John wrote: "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, The Online Books Page listed more than 7,000 books, which could be browsed by

author, by title or by subject. It also listed significant directories and archives of online texts, and special exhibits. From the main search page, users could search four types of media: books, music, art, and video.

The Online Books Page began listing serials. As stated on the website: "Along with books, The Online Books Page is also now listing major archives of serials (such as magazines, published journals, and newspapers), as of June 1998. 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'.)"

The Online Books Page participated in the Experimental Search System of the Library of Congress. It also worked with The Universal Library Project, hosted at Carnegie Mellon University.

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



## **1993: PDF**

### **[Overview]**

PDF (Portable Document Format) was conceived by Adobe in 1992, launched in June 1993 with Adobe Acrobat software, and perfected over 15 years as the 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 gives the tools to create and view PDF files and is available in many languages and for many platforms (Macintosh, Windows, Unix, etc.). Ten years later, over 500 million copies of PDF-based Adobe Reader (formerly Acrobat Reader, until May 2003) have been downloaded worldwide. Approximately 10% of the documents on the internet are available in PDF.

## **1994: Library Websites**

### **[Overview]**

The first library website was the one created by the Helsinki City Library in Finland, which went live in February 1994. Traditional libraries began using a website as a new virtual window for their patrons and beyond. Patrons could check opening hours, browse the online catalog, or surf on a broad selection of websites on various topics, depending on their needs. Libraries also began developing digital libraries alongside their standard collections, for a large audience to be able to access their specialized, old, local and regional collections. Librarians could now fulfill two goals that used to be in contradiction - book preservation (on shelves) and book communication (on the internet).

### **[In Depth (published in 1999)]**

The first library website was the one created by the Helsinki City Library in Finland, which went live in February 1994. Many libraries began developing a digital library alongside their standard collections. Digital libraries allowed a large audience to have access to documents belonging to specialized, old, local or regional collections. Thanks to their digital libraries, traditional libraries could achieve a long-time dream and fulfill two goals which used to be in contradiction - book preservation and book communication. 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, and read on the screen of a computer, without the need to go to the library and struggle through a lengthy process to have access to the original books, for various reasons: concern for preservation of rare and fragile documents, reduced opening hours, forms to fill out, long waiting period to get the document, and shortage of staff. All these reasons were often hurdles to get over, and often required of the researcher an unflinching patience and an out-of-the-ordinary determination to finally get to the document.

Some virtual libraries were created from scratch, right on the internet from the beginning, with no back up from a traditional library. This was the case of Athena, founded in 1994 by Pierre Perroud, a Swiss teacher, and hosted on the website of the University of Geneva, Switzerland. Athena was created as a multilingual digital library focusing on philosophy, science, classics, literature, history, and economics. As Geneva is in French-speaking Switzerland, it also focused on putting French texts online. The Helvetia section gathered documents about Switzerland. A specific page offered a number of links to other digital libraries in the world.

In an interview dated February 1996, Pierre Perroud explained: "Electronic texts represent an encouragement to reading and a convivial participation to culture dissemination, (...) [and] a good complement to the paper book, which remains irreplaceable for reading (...). [The paper book] remains a mysteriously holy companion with profound symbolism for us: we grip it in our hands, we hold it against our bodies, 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." (excerpt from the Swiss magazine *Informatique-Informations*)

The Internet Public Library (IPL) opened in March 1995 as the first digital public library of and for the internet community. Its different sections were: Reference, Exhibits, Magazines and Serials, Newspapers, Online Texts, and Web Searching. There were also sections for Teen and Youth. All the items of the collections were carefully selected, catalogued and described by the IPL staff. As an experimental library, IPL also listed the most interesting projects run by librarians on the internet, in the section Especially for Librarians.

## **1994: Bold Publishers**

### **[Overview]**

Some publishers decided to use the web as a new 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. 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)

### **[In Depth (published in 1999)]**

The web became a marketing tool for publishers. Some publishers decided to put the full text of some books on the web, for free, with their authors' consent. Oddly enough, there was no drop in sales - on the contrary, sales increased. In the US, NAP was the first publisher to take such a risk in 1994, followed by the MIT Press in 1995, and it worked.

NAP (National Academy Press) 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 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).

In 1995, the MIT Press (MIT: Massachusetts Institute of Technology) was publishing 200 new books a year and 40 journals, first in science and technology, and then in architecture, social theory, economics, cognitive science, and computational science. The MIT Press 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 the print books 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)

# 1995: Amazon.com

## [Overview]

Amazon.com was a "pioneer" online bookstore that created an entirely new economic model. 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 window is its website, 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.com had 7,500 employees, a catalog of 28 million items, 23 million clients worldwide and four subsidiaries in UK (in August 1998), in Germany (in August 1998), in France (August 2000) and in Japan (October 2000). A fifth subsidiary opened in Canada in June 2002. A sixth subsidiary - named Joyo - opened in China in September 2004.

## [In Depth (published in 1999)]

Jeff Bezos launched Amazon.com in July 1995, after a market study which led him to conclude that books were the best products to sell on the internet.

In Spring 1994, he 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.

"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." (excerpt from the Amazon.com press kit)

In 1998, Amazon.com was offering 3 million books, CDs, audio books, DVDs, computer games - more than 14 times as many titles as the large chain superstores - to 3 million people in 160 countries. "Businesses can do things on the web that simply cannot be done any other way", wrote Jeff Bezos. "We are changing the way people buy books and music." Amazon.com quickly became the largest online bookstore, with a catalog of these 3 million items that could be ordered online, authoritative reviews, author interviews, excerpts, customer reviews, and book recommendations. As an internet retailer, Amazon.com could offer more services than traditional retailers: lower prices, larger selection, and a wealth of product information.

Any book lover could post his own reviews of books on Amazon's website, and read others.

He could read many interviews with authors, and a number of blurbs and excerpts from books. He could search for books by author, subject, title, ISBN or publication date. Prices were discounted, with savings of 20-40% on 400,000 titles (40% on selected feature books, 30% on hardcovers, and 20% on paperbacks). The client usually received the books within a week. If he requested it, he could receive an email announcing a new book by a favorite author or a new book on a favorite topic. He could select some book categories (44 listed), to be sent a monthly review of new books by email. All things that were entirely new at the time.

What we take for granted now, i.e. buy a book in Europe from the US site of Amazon.com, or buy a book in the US from the German site of Amazon.de, was making big waves at the time, first as "unfair competition" with the local online bookstores, then for taxation. A first outline agreement was concluded between the US and the European Union in December 1997, and 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 electronic books bought online. Material goods (books, CDs, DVDs, and so on) and services were subject to existing regulations, with collection of the VAT for example, but with no additional custom taxes.

Amazon.com and others had great assets, but there were bad news for small bookstores. Like the small bookstore set up in 1971 by my friend Catherine Domain in central Paris, on the island Ile Saint-Louis, surrounded by the Seine river.

The small Ulysses Bookstore is known as the oldest travel bookstore in the world. It has more than 20,000 books, maps and magazines, out of print and new, in a number of languages, about any country and any kind of travel, all packed up in a tiny space. Catherine has been a traveller since she was a child. She travels every summer - usually sailing - while her boyfriend runs the bookstore. She is also a member of the French National Union of Antiquarian and Modern Bookstores (SLAM), the Explorers' Club and the International Club of Long-Distance Travellers.

Catherine visited 140 countries, where she sometimes had a hard time. But one of her most difficult challenges was to set up a website on her own, from scratch, without knowing anything about computers. Catherine wrote in December 1999: "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..." Nevertheless, despite the internet, she was pessimistic about the future. "I am very pessimistic, because the internet is killing off specialist bookstores."

## **1995: Online Press**

### **[Overview]**

The first electronic versions of print newspapers were available in the early 1990s through commercial services like America Online and CompuServe. In 1995, newspapers and magazines began creating their own websites to offer a partial or full version of their latest issue - available freely or through subscription (free or paid) - with online archives. In Europe, the 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 also went online in UK, as well as the weekly Focus and the weekly Der Spiegel in Germany, the daily Le Monde and daily Libération in France, and the daily El País in Spain. The computer press went logically online as well, like 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", or ZDNet, another leading computer magazine. More and more "only" electronic magazines were also created.

### **[In Depth (published in 1999)]**

The first electronic versions of newspapers were available in the early 1990s through commercial services like America Online or CompuServe. Then, in 1995, newspapers and magazines began to create websites to offer the full version of their latest issue - available freely or through subscription (free or paid) - which was then archived online. There were also heated debates on copyright issues for articles posted on the web. More and more "only" electronic magazines were created.

In 1996, the New York Times site could be accessed free of charge. It included the contents of the daily newspaper, breaking news updates every ten minutes, and original reporting available only online. The Washington Post site provided the daily news online, with a full database of articles including images, sound and video.

In Europe, the Times and the Sunday Times set up a common website called Times Online, with the possibility to create a personalized edition. The respected Economist was also available online, as were the French daily newspapers Le Monde and Libération, the Spanish daily newspaper El País or the German weekly magazines Focus or Der Spiegel.

The computer press went online as well. First the monthly Wired, created in 1992 in California to focus on cyberculture and be the magazine of the future at the avant-garde of the 21st century. Then ZDNet, a main publisher of computer magazines.

Behind the news, the web was providing a whole encyclopedia to help us understand them. The web was providing instant access to a wealth of information (geographical maps, biographical notes, official texts, political and economic data, audiovisual and video data); speed in information dissemination; access to main photographic archives; links to articles, archives and data on the same topic; and a search engine to browse articles by date, author, title, subject, etc.

From the start, there were also all these zines using the internet as a cheap way to get published. John Labovitz launched The E-Zine-List in Summer 1993 to list electronic zines (e-zines) around the world, the ones that were accessible via the web, FTP, gopher, email, 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 on November 29, 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'." John stopped updating his list a few years later.



## **1996: Internet Archive**

### **[Overview]**

Founded in April 1996 by Brewster Kahle, the Internet Archive is a non-profit organization that 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 October 2001, with 30 billion web pages stored, 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. In 2006, there were 65 billion pages from 50 million websites. In late 1999, the Internet Archive also started to include more collections of archived web pages on specific topics. It also became an online digital library of text, audio, software, image and video content. In October 2005, the Internet Archive launched the Open Content Alliance (OCA) with other contributors as a collective effort to build a permanent archive of multilingual digitized text (Text Archive) and multimedia content.

# 1996: New Ways of Teaching

## [Overview]

With more and more computers available in schools and at home, and more and more internet connections, teachers began exploring new ways of teaching. Going from print book culture to digital culture was changing their 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 the September 1996 meeting of IFIP (International Federation of Information Processing), Dale Spender gave a lecture about Creativity and the Computer Education Industry, with insightful comments on forthcoming trends.

## [In Depth (published in 1999)]

Going from print book culture to digital culture began changing our relationship to knowledge. Book culture provided stable information whereas digital culture provided "moving" information. During the September 1996 meeting of the IFIP (International Federation of Information Processing), Dale Spender gave an interesting lecture about Creativity and the Computer Education Industry.

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, memorized, 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."

## **1996: Palm Pilot**

### **[Overview]**

In the 1990s, Jacques Gauchey was a journalist and writer living in Silicon Valley and specializing in IT (information technology). He was also working as a "facilitator" between the United States 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 10 readers per issue until the day (in January 1996) 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* [editor's note: *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." Despite fierce competition, Palm stayed the leader in the PDA market, with 23 million Palm Pilots sold between 1996 and 2002.

# 1997: Digital Publishing

## [Overview]

Digital publishing became mainstream in 1997. This was a new step in the changes underwent by the traditional publishing chain since the 1970s. The traditional printing business was first disrupted by new photocomposition machines, with lower costs. Text and image processing began to be handed over to desktop publishing shops and graphic art studios. Impression costs went on decreasing with desktop publishing, photocopiers, color photocopiers and digital printing equipment. 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 the possibility of regular updates to include the latest information.

## [In Depth (published in 1999)]

Since the 1970s, the traditional publishing chain has drastically changed. The printing work done by pre-press shops was first disrupted by new photocomposition machines. Text and image processing began to be handed over to advertising and graphic art agencies. Impression costs went on decreasing with desktop publishing, copiers, color copiers and digital printing equipment.

In 1997, text and image processing was provided at a low price by desktop publishing shops and graphic art studios. Digitization accelerated the publication process. Editors, designers and other contributors could all work at the same time on the same book.

Digitization also made possible the online publishing of educational and scientific publications, which appeared as a far better solution than print books, because they could be regularly updated with the latest information. Some universities began distributing their own textbooks online, with chapters selected in an extensive database, and articles and commentaries from professors. For a seminar, a small print could be made upon request with a selection of online articles sent to a printer.

Electronic publishing allowed some academic publishers to keep running their business, with lower costs and quick access. This way, small publishers went on publishing specialized books, for which the printing in a small number of copies had become more and more difficult over the years due to budgetary reasons. These books could now be regularly updated and their readers benefit from the latest version. Readers didn't need to wait any more for a new printed edition, often postponed if not cancelled because of commercial constraints.

Electronic publishing and traditional publishing became complementary. The frontier between the two supports - electronic and paper - was vanishing. Most recent print media already stemmed 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 scanned to be included in digital bookstores and libraries.

At the end of the 1990s, there were no reliable statistics yet proving that the large-scale use of computers and electronic documents would make us paperless and save some tress, as hoped by all of us who believe in nature preservation. We were still in a transition period. Many people still needed a print version for easier reading, or to keep track of a document in case the electronic file was accidentally deleted, or to have some paper support for their documentation or archives.

# 1997: Logos Dictionary

## [Overview]

Logos is a leading translation company located 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 at the time, 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.

## [In Depth (published in 1999)]

The Logos Dictionary is a multilingual dictionary with 7,580,560 words (as of December 10, 1998). The Logos Wordtheque is a word-by-word multilingual library with a massive database of 325,916,827 words extracted from multilingual novels, technical literature and translated texts. Logos Linguistic Resources is a database of 553 glossaries. The Logos Universal Conjugator is a database for the conjugation of verbs in 17 languages.

Logos is an international translation company based in Modena, Italy. In 1997, Logos decided to put all the linguistic tools used by its translators on the web for free. Logos had 200 translators on the spot and 2,500 free-lance translators all over the world, who processed around 200 texts per day.

When interviewed by Annie Kahn in the French daily newspaper *Le Monde* of December 7, 1997, Rodrigo Vergara, the 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. The operation has in fact attracted a great number of customers, but also allowed us to widen our network of translators, thanks to the 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. A system cornerstone is the document search software, 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 button to access the whole text or even to order the book, thanks to a partnership agreement with Amazon.com, the famous online bookstore. Foreign translations are also available. However, if no text containing the required word is found, the system acts as a search engine, sending the user to other websites mentioning the term in question. 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 their own suggestions, after which our translators and the company check the forwarded translations."

# 1997: Multimedia Convergence

## [Overview]

As more and more people were using digital technology, previously distinct information-based industries, such as printing and 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. This international symposium was a tripartite meeting with employers, unionists, and government representatives. 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.

## [In Depth (published in 1999)]

The first ILO Symposium on Multimedia Convergence was held in January 1997 at the headquarters of ILO (International Labor Office) in Geneva, Switzerland.

Peter Leisink, 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."

This international symposium was held as a tripartite meeting with employers, unionists and government representatives. Some participants still thought our information society would generate jobs, whereas it was already stated worldwide that multimedia convergence was leading to a massive loss of jobs.

Michel Muller, secretary-general of the French Federation of Book, Paper and Communication Industry, stated that the French graphics industry had lost 20,000 jobs - falling from 110,000 to 90,000 - within the last decade, and that expensive social plans had been necessary to re-employ those people. He explained: "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' interventions, here was the answer of Walter Durling, director of AT&T Global Information Solutions: "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 the future as they were afraid of losing their jobs. In 1997, our society already had a high unemployment rate, which was not the case when film was invented and television developed. During the next years, what would be the balance between job creation and lay-off? Unions were struggling worldwide to promote the creation of jobs through investment, innovation, vocational training, computer literacy, retraining for new jobs, fair conditions for contracts and collective agreements, defense of copyright, protection of workers in the artistic field, and defense of teleworkers as workers having full rights. The European Commission was expecting 10 million European teleworkers in the year 2000, which would represent 20% of teleworkers worldwide.

Despite unions' efforts, would the situation become as tragic as what we read in the report of the symposium? "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."

## **1998: Online Beowulf**

### **[Overview]**

Libraries began putting (digital versions of) their treasures on the web for the world to enjoy. The British Library was a pioneer in this field. Several treasures were online in 1998, including Beowulf, known as the first great English masterpiece. 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. The poem itself is much older than the manuscript - some historians believe it might have been written circa 750. 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 Library expecting to see this literary treasure on display. The digitization of the manuscript offered a solution to these problems, while providing new opportunities for researchers and book lovers worldwide.

### **[In Depth (published in 1999)]**

Libraries began using the web to make their treasures freely available to the world.

Here is the story of Beowulf.

Beowulf is a treasure of the British Library. "It is an Old English heroic epic poem of anonymous authorship. This work of Anglo-Saxon literature dates to between the 8th and the 11th century, the only surviving European manuscript dating to the early 11th century. At 3,183 lines, it is notable for its length." (excerpt from Wikipedia)

The manuscript was badly damaged by fire in 1731. 18th-century transcripts mention hundreds of words and letters 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 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 readers, and for the world to enjoy.

The Electronic Beowulf Project was launched as a huge database of digital images of the Beowulf manuscript and related manuscripts and printed texts. In 1998, the database included fiber-optic readings of hidden letters and ultraviolet readings of erased text in the manuscript; full electronic facsimiles of the 18th-century transcripts of the manuscript; and selections from important 19th-century collations, editions and translations. Major additions were planned, such as images of contemporary manuscripts, and links with the Toronto

Dictionary of Old English Project and with the comprehensive Anglo-Saxon bibliographies of the Old English Newsletter.

The project was developed in partnership with two leading experts, Kevin Kiernan, from the University of Kentucky and Paul Szarmach, from the Medieval Institute, Western Michigan University. Professor Kiernan edited the electronic archive and produced a CD-ROM containing a number of electronic images.

Brian Lang, chief executive of the British Library, explained 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 artifacts of our shared cultural heritage." (excerpt from the 1998 website)

The British Library was a pioneer in Europe. Other treasures of the library were already online: 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, which could be the world's earliest print book; the Sforza Hours, dated 1490-1520, an outstanding Renaissance treasure; the Codex Arundel, a notebook of Leonardo Da Vinci (1452-1519), and the Tyndale New Testament, the first English version of the New Testament, printed by Peter Schoeffer, in Worms.

Brian King also stated the importance of the paper world, and the ongoing commitment of the British Library to its paper collections. He added: "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 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." (excerpt from the website)

Other national libraries started digitizing their collections to offer a free digital library. When interviewed by Jérôme Strazzulla in the daily newspaper *Le Figaro* of June 3, 1998, Jean-Pierre Angremy, president of the French National Library, 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 digital library Gallica went online in 1997 with thousands of texts and images

relating to French history, life and culture. A major collection of 19th-century French texts and images was available one year later.

## **1998: Digital Librarians**

### **[Overview]**

The job of librarians, that had already changed a lot with computers, went on to change even more with the internet. Computers made catalogs much easier to handle. Instead of all these cards to be patiently classified into wood or metal drawers, librarians could type in bibliographic records in a program that was sorting out books by alphabetical, chronological and systematic order. Librarians also began using computer programs to lend books and buy new ones. By networking computers, the internet gave a boost to union catalogs for a state, a country or a region, and furthered interlibrary loan. Electronic mail became commonplace for internal and external communications. Librarians could subscribe to newsletters and participate in newsgroups and discussion forums. A number of librarians became webmasters to run library websites, online catalogs and digital libraries.

### **[In Depth (published in 1999)]**

I interviewed Peter Raggett, a digital librarian at OECD (Organization for Economic Co-operation and Development), and Bruno Didier, a digital librarian at Institute Pasteur. Here are some excerpts.

#### **At the OECD Library**

What is OECD? "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 1998 website)

The Center for Documentation and Information (CDI) of OECD provides information to OECD agents in support of their research work. In 1998, there were 60,000 monographs and 2,500 periodicals. The CDI also provides information in electronic format from databases, CD-ROMs and the internet.

Peter Raggett, head of CDI, has been a professional librarian for nearly twenty years, first working in UK government libraries and then at the OECD since 1994. He has used the internet since 1996. He built up the CDI Intranet pages, which became a main tool for the staff.

Peter wrote in June 1998: "At the OECD Library we have collected together several hundred World Wide Web sites 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."

Peter added: "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."

What 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 e-mail, telephone or fax, and I will do the research and send them the results electronically."

### **At the Institute Pasteur Library**

In 1999, Bruno Didier was the webmaster of the Institute Pasteur Library. "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 1999 website)

In August 1999, Bruno wrote about his work as a webmaster: "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, cataloguing, ordering of documents and of course access to online periodicals (presently more than 100). It is also 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 web pages and monitor them regularly. I am also responsible for training users, which you can see from my pages. The web is an excellent place for training and is included in most ongoing discussions about training."

What about the future of librarians? "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."

## **1998: Multilingual Web**

### **[Overview]**

In 1998, Randy Hobler was a consultant in internet marketing for Globalink, a company specializing in language translation software and services. Randy 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. (...) 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 US, as well as odd places like Spanish-speaking Morocco."

### **[In Depth (published in 2000, updated in 2004)]**

In 1998, other languages than English began spreading on the web. In fact, main non-English languages were present nearly from the start. But most of the web was in English. Then people from all over the world began having access to the internet, and posting pages in their own languages. The percentage of the English language began to slowly decrease from nearly 100% to 90%.

In 1998, Randy Hobler was an internet marketing consultant for Globalink, a company specialized in language translation software and services. Previously, Randy worked as a consultant for IBM, Johnson & Johnson, Burroughs Wellcome, Pepsi, Heublein, and others.

Randy wrote in September 1998: "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 US, as well as odd places like Spanish-speaking Morocco."

In 1999, Jean-Pierre Cloutier was the editor of Chroniques de Cybérie, a weekly report of internet news. Jean-Pierre wrote in August 1999: "The web is going to grow in these 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. (...) The recent introduction of the internet in regions where it is spreading raises questions which would be good to read about. When will Spanish-speaking communications theorists and those speaking other languages be translated?"



In 1999, Marcel Grangier was the head of the French Section of the Swiss Federal Government's Central Linguistic Services, which meant he was in charge of organizing translation matters for the Swiss government. Marcel 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."

In 1998, Henri Slettenhaar was a professor at Webster University, Geneva, Switzerland. He insisted regularly 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."

He added in August 1999: "There are two main categories 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."

He added in August 2000: "Multilingualism has expanded greatly. Many e-commerce websites are multilingual now and there are companies that sell products which make localization possible (adaptation of websites to national markets)."

Non English-speaking users reached 50% in Summer 2000. According to the company Global Reach, they were 52.5% in Summer 2001, 57% in December 2001, 59.8% in April 2002, 64.4% in September 2003 (including 34.9% non-English-speaking Europeans and 29.4% Asians) and 64.2% in March 2004 (including 37.9% non-English-speaking Europeans and 33% Asians).

## **1999: Open eBook Format**

### **[Overview]**

In 1999, there were nearly as many eBook formats as eBooks, with every company and organization creating its own format for its own eBook reader and its own electronic device. The publishing industry felt the need to work on a common format for eBooks and published in September 1999 the first version of the Open eBook (OeB) format, an eBook 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. In April 2005, the Open eBook Forum became the International Digital Publishing Forum (IDPF), and the OeB format became the ePub format. The ePub format is one of the standards for the digital publishing industry.

## **1999: Digital Authors**

### **[Overview]**

Like many artists, Jean-Paul began exploring the internet and searching what hyperlinks could offer to expand his writing towards new directions. He switched from being a print author to being an hypermedia author, and created Cotres furtifs (Furtive Cutters), a website telling stories in 3D. He also 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 added in June 2000: "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 didn't change my life, but it did change how I write. You don't write the same way for a website as you do for a script or a play."

### **[In Depth (published in 2000)]**

I interviewed Murray Suid, a writer of educational books, who was living in Palo Alto, California. Back in Paris, I interviewed Jean-Paul, an hypermedia author, who wrote some interesting comments about digital literature.

### **Educational Books**

In 1998, Murray Suid was living in Palo Alto, in the heart of Silicon Valley. He was writing educational books, books for kids, multimedia scripts and screenplays. He was among the first to choose a solution that many authors would soon adopt. He explained in September 1998: "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)."

Murray 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." A few months later, he added: "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."

### **Hypermedia Writing**

In 1999, Jean-Paul, an hypermedia author, was the webmaster of cotres.net, a site telling stories in 3D. He really enjoyed the freedom given by online publishing. He 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. Then the intermediaries will take over and I will have to look somewhere else, a place where the grass is greener..."

Jean-Paul added in June 2000: "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 print media, of course. But the difference is striking. So the internet didn't change my life, but it did change how I write. You don't write the same way for a website as you do for a script or a play.

But it wasn't exactly the internet that changed my writing, it was the first model of the Mac. I discovered it when I was teaching myself Hypercard. I still remember how astonished I was during my month of learning about buttons and links and about surfing by association, objects and images. Being able, by just clicking on part of the screen, to open piles of cards, with each card offering new buttons and each button opening onto a new series of them. In short, learning everything about the web that today seems really routine was a revelation for me. I hear Steve Jobs and his team had the same kind of shock when they discovered the forerunner of the Mac in the labs of Rank Xerox.

Since then I have been writing directly on the screen. I use a paper print-out only occasionally, to help me fix up an article, or to give somebody who doesn't like screens a rough idea, something immediate. It is only an approximation, because print forces us into a linear relationship: the words scroll out page by page most of the time. But when you have links, you have a different relationship to time and space in your imagination. And for me, it is a great opportunity to use this reading/writing interplay, whereas leafing through a book gives only a suggestion of it -- a vague one because a book is not meant for that."

## **2000: yourDictionary.com**

### **[Overview]**

After founding A Web of Online Dictionaries (WOD) in 1995, Robert Beard included it in a larger project, yourDictionary.com, that he cofounded 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 On-line 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."

### **[In Depth (published in 2001)]**

After creating A Web of Online Dictionaries in 1995, Robert Beard cofounded yourDictionary.com in early 2000. He wrote in January 2000: "A Web of Online Dictionaries (WOD) is now a part of yourDictionary.com (as of February 15, 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 On-line 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."

Answering my question about multilingualism, Robert Beard added in January 2000: "While English still dominates the web, the growth of monolingual non-English websites is gaining strength with the various solutions to the font problems. 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."

Answering the same question, Caoimhín Ó Donnáil wrote in May 2001: "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 globalization will greatly speed their demise. If people do care about preserving them, the internet will be a tremendous help."

Caoimhín Ó Donnáile teaches computing - through the Gaelic language - at the Institute Sabhal Mór Ostaig, located on the Island of Skye, in Scotland. He also maintains the college website, which is the main site worldwide with information on Scottish Gaelic. He also maintains European Minority Languages, a 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 at the Gaelic-medium college here. Far more computers, more computing staff, flat screens. Students do everything by computer, use Gaelic spell-checking, Gaelic online terminology database. More hits on our web site. More use of sound. Gaelic radio (both Scottish and Irish) now available continuously worldwide via the internet. Major project has been translation of the Opera web-browser into Gaelic - the first software of any size available in Gaelic."

Published by SIL International (SIL: Summer Institute of Linguistics), *The Ethnologue: Languages of the World* is a catalogue of more than 6,700 languages. A paper version and a CD-ROM are also available. Barbara Grimes was the editor of the 8th to 14th editions, 1971-2000. She wrote in January 2000: "It 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 sociolinguistic and demographic information, dates of published Bibles, a name index, a language family index, and language maps."

## **2000: Online Bible of Gutenberg**

### **[Overview]**

The Bible of Gutenberg went online in November 2000, on the website of the British Library. As we all know, the Bible of Gutenberg is considered as the first print book. Gutenberg printed it in 1455 in Germany, perhaps printing 180 copies, with 48 copies that would still exist in 2000. Three copies - two full ones and one partial one - belong to the British Library. The two full copies - a little different from each other - were digitized in March 2000 by experts from the Keio University of Tokyo and NTT (Nippon Telegraph and Telephone Communications).

## **2000: Distributed Proofreaders**

### **[Overview]**

Conceived in October 2000 by Charles Franks, Distributed Proofreaders was launched online in March 2001 to help in the digitization of public domain books. The method is to break up the tedious work of checking eBooks for errors into small, manageable chunks. Originally meant to assist Project Gutenberg in the handling of shared proofreading, Distributed Proofreaders has become the main source of Project Gutenberg eBooks. In 2002, Distributed Proofreaders became an official Project Gutenberg site. The number of books processed through Distributed Proofreaders has grown fast. In 2003, about 250-300 people were working each day all over the world producing a daily total of 2,500-3,000 pages, the equivalent of two pages a minute. In 2004, the average was 300-400 proofreaders participating each day and finishing 4,000-7,000 pages per day, the equivalent of four pages a minute. Distributed Proofreaders processed a total of 3,000 books in February 2004, 5,000 books in October 2004, 7,000 books in May 2005, 8,000 books in February 2006 and 10,000 books in March 2007, with the help of 36,000 volunteers.

### **[In Depth (published in 2005, updated in 2008)]**

The main "leap forward" of Project Gutenberg since 2000 is due to Distributed Proofreaders. In 2002, Distributed Proofreaders became an official Project Gutenberg site. In May 2006, Distributed Proofreaders became a separate entity and continues to maintain a strong relationship with Project Gutenberg.

Volunteers don't have a quota to fill, but it is recommended they do a page a day if possible. It doesn't seem much, but with hundreds of volunteers it really adds up. In December 2007, five books were produced per day by thousands of volunteers.

From the website one can access a program that allows several proofreaders to be working on the same book at the same time, each proofreading different pages. This significantly speeds up the proofreading process. Volunteers register and receive detailed instructions. For example, words in bold, italic or underlined, or footnotes are always treated the same way for any book. A discussion forum allows them to ask questions or seek help at any time. A project manager oversees the progress of a particular book through its different steps on the website.

The website gives a full list of the books that are: (a) completed, i.e. processed through the site and posted to Project Gutenberg; (b) in progress, i.e. processed through the site but not yet posted, because currently going through their final proofreading and assembly; (c) being proofread, i.e. currently being processed. On August 3, 2005, 7,639 books were completed, 1,250 books were in progress and 831 books were being proofread. On May 1st, 2008, 13,039 books were completed, 1,840 books were in progress and 1,000 books were being proofread.

Each time a volunteer (proofreader) goes to the website, s/he chooses a book, any book. Then one page of the book appears in two forms side by side: the scanned image of one



page and the text from that image (as produced by OCR software). The proofreader can easily compare both versions, note the differences and fix them. OCR is usually 99% accurate, which makes for about 10 corrections a page. The proofreader saves each page as it is completed and can then either stop work or do another. The books are proofread twice, and the second time only by experienced proofreaders. All the pages of the book are then formatted, combined and assembled by post-processors to make an eBook. The eBook is now ready to be posted with an index entry (title, subtitle, author, eBook number and character set) for the database. Indexers go on with the cataloging process (author's dates of birth and death, Library of Congress classification, etc.) after the release.

Volunteers can also work independently, after contacting Project Gutenberg directly, by keying in a book they particularly like using any text editor or word processor. They can also scan it and convert it into text using OCR software, and then make corrections by comparing it with the original. In each case, someone else will proofread it. They can use ASCII and any other format. Everybody is welcome, whatever the method and whatever the format.

New volunteers are most welcome too at Distributed Proofreaders (DP), Distributed Proofreaders Europe (DP Europe) and Distributed Proofreaders Canada (DPC). Any volunteer anywhere is welcome, for any language. There is a lot to do. As stated on both websites, "Remember that there is no commitment expected on this site. Proofread as often or as seldom as you like, and as many or as few pages as you like. We encourage people to do 'a page a day', but it's entirely up to you! We hope you will join us in our mission of 'preserving the literary history of the world in a freely available form for everyone to use'."

## **2000: Public Library of Science**

### **[Overview]**

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 early 2003, PLoS created a non-profit scientific and medical publishing venture to provide scientists and physicians with high-quality, high-profile journals in which to publish their most important work: PLoS Biology (launched in 2003), PLoS Medicine (2004), PLoS Genetics (2005), PLoS Computational Biology (2005), PLoS Pathogens (2005), PLoS Clinical Trials (2006), PLoS Neglected Tropical Diseases (2007). All PLoS articles are freely available online, and deposited in the free public archive PubMed Central. They can be freely redistributed and reused, including for translations, as long as the author(s) and source are cited. PLoS also hopes to encourage other publishers to adopt the open access model, or to convert their existing journals to an open access model.

## **2001: Wikipedia**

### **[Overview]**

Launched in January 2001 by Jimmy Wales and Larry Sanger (Larry resigned later on), Wikipedia has quickly grown into the largest reference website on the internet. Its multilingual content is free and written collaboratively by people worldwide. 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). Wikipedia is hosted by the Wikimedia Foundation, which runs a number of other projects, for example Wiktionary - launched in December 2002 - followed by Wikibooks, Wikiversity, Wikinews and Wikiquote. In December 2004, Wikipedia had 1.3 million articles from 13,000 contributors in 100 languages. Two years later, in December 2006, it had 6 million articles in 250 languages.

## **2001: Creative Commons**

### **[Overview]**

Creative Commons (CC) was founded in 2001 by Lawrence Lessing, a professor at Stanford Law School, California. As stated 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."

## **2002: MIT OpenCourseWare**

### **[Overview]**

The MIT OpenCourseWare (MIT OCW) is a large-scale, web-based electronic publishing initiative launched by MIT (Massachusetts Institute of Technology) to promote open dissemination of knowledge and information. A pilot version of the MIT OpenCourseWare (MIT OCW) was available online in September 2002, with 32 course materials of MIT. In September 2003, the site was officially launched with several hundred course materials. In March 2004, 500 course materials were available in 33 different topics. 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. In November 2005, the MIT launched the OpenCourseWare Consortium (OCW Consortium) as a collaboration of educational institutions creating a broad body of open educational content using a share model. One year later, the OCW Consortium included the courses of 100 universities worldwide.

## **2004: Project Gutenberg Europe**

### **[Overview]**

In January 2004, Project Gutenberg spread across the Atlantic with the launching of Project Gutenberg Europe (PG Europe) and Distributed Proofreaders Europe (DP Europe) by Project Rastko, a non-governmental cultural and educational project located in Belgrade, Serbia. DP Europe uses the software of the original Distributed Proofreaders. DP Europe is a multilingual website, with its main pages translated into several European languages by volunteer translators. In April 2004, DP Europe was available in 12 languages. The long-term goal is 60 languages and 60 linguistic teams representing all European languages. DP Europe supports Unicode to be able to proofread eBooks in numerous languages. Unicode is an encoding system that gives a unique number for every character in any language. DP Europe finished processing its 100th book in May 2005 and its 500th book in October 2008. DP Europe operates under "life +50" copyright laws. When it gets up to speed, DP Europe will provide eBooks for several national and/or linguistic digital libraries.

### **[In Depth (published in 2005, updated in 2008)]**

In 2004, multilingualism became one of the priorities of Project Gutenberg, like internationalization. Michael Hart went off to Europe, with stops in Paris, Brussels and Belgrade. In Belgrade, he met with the team of Project Rastko, to support the creation of Distributed Proofreaders Europe (launched in December 2003) and Project Gutenberg Europe (launched in January 2004).

The launching of Distributed Proofreaders Europe (DP Europe) by Project Rastko was indeed a very important step. DP Europe uses the software of the original Distributed Proofreaders and is dedicated to the proofreading of books for Project Gutenberg Europe. Since the very beginning, DP Europe has been a multilingual website, with its main pages translated into several European languages by volunteer translators. DP Europe was available in 12 languages in April 2004 and 22 languages in May 2008.

The long-term goal is 60 languages and 60 linguistic teams representing all the European languages. When it gets up to speed, DP Europe will provide books for several national and/or linguistic digital libraries. The goal is for every country to have its own digital library (according to the country copyright limitations), within a continental network (for France, the European network) and a global network (for the whole planet).

A few lines now on Project Rastko, which launched such a difficult and exciting project for Europe, and catalyzed volunteers' energy in both Eastern and Western Europe (and anywhere else: as the internet has no boundaries, there is no need to live in Europe to register). Founded in 1997, Project Rastko is a non-governmental cultural and educational project. One of its goals is the online publishing of Serbian culture. It is part of the Balkans Cultural Network Initiative, a regional cultural network for the Balkan peninsula in south-eastern Europe.

In May 2005, Distributed Proofreaders Europe finished processing its 100th book. In June 2005 Project Gutenberg Europe was launched with these first 100 books. DP Europe supports Unicode to be able to proofread books in numerous languages. Created in 1991 and widely used since 1998, Unicode is an encoding system that gives a unique number for every character in any language, contrary to the much older ASCII that was meant only for English and a few European languages.

On August 3, 2005, 137 books were completed (processed through the site and posted to Project Gutenberg Europe), 418 books were in progress (processed through the site but not yet posted, because currently going through their final proofreading and assembly), and 125 books were being proofread (currently being processed). On May 10, 2008, 496 books were completed, 653 books were in progress and 91 books were being proofread.

## **2004: Google Books**

### **[Overview]**

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. 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. 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. The lawsuit with associations of authors and publishers was settled in October 2008.

### **[In Depth (published in 2008)]**

In October 2004, Google launched the first part of Google Print as a project aimed at publishers, for users to be able to see snippets of their books and order them online. The beta version of Google Print went on line in May 2005. 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 scanning and digitizing the collections of main libraries, beginning with the Universities of Michigan (7 million books), Harvard, Stanford and Oxford, and the New York Public Library. The planned cost was an average of US \$10 per book, and \$150 to \$200 million on ten years. In August 2005, Google Print was stopped until further notice because of lawsuits filed by publishers for copyright infringement. The program resumed in August 2006 under the new name of Google Books.

Google Books was launched in August 2006 to replace the controversial Google Print, stopped in August 2005 because of main copyright concerns. Google Books offers excerpts of books digitized by Google in the participating libraries (Harvard, Stanford, Michigan, Oxford, California, Virginia, Wisconsin-Madison, Complutense of Madrid and New York Public Library). Google scans 3,000 books a day, including copyrighted books. The inclusion of copyrighted books is widely criticized by authors and publishers worldwide. In the US, lawsuits were filed by the Authors Guild and the Association of American Publishers (AAP) for alleged copyright infringement. The assumption is that the full scanning and digitizing of copyrighted books infringes copyright laws, even if only snippets are made freely available on the search engine. To counteract copyright concerns and the problems of a closed platform, the Internet Archive launched the Open Content Alliance (OCA) with the goal of digitizing only public domain books and make them searchable and downloadable through any search engine.



## **2005: Open Content Alliance**

### **[Overview]**

The Open Content Alliance (OCA) was conceived by the Internet Archive in early 2005 to offer broad, public access to the world culture. It 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 aims at digitizing public domain books around the world and make them searchable through any web search engine and downloadable for free. Unlike the Google Print project, the OCA scans and digitizes only public domain 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 the United Kingdom, O'Reilly Media and Prelinger Archives. The digitized collections are freely available in the Text Archive of the Internet Archive. In December 2006, they reached a milestone of 100,000 digitalized books publicly available, with 12,000 new books added per month. Two years later, in December 2008, one million books were "posted under OCA principles or otherwise public domain hosted by the Internet Archive."

## **2006: Microsoft Live Search Books**

### **[Overview]**

Microsoft has also participated in the Open Content Alliance (OCA), launched by the Internet Archive in October 2005. In December 2006, Microsoft released the beta version of Live Search Books. The book search engine performs keyword searches for non copyrighted books digitized by Microsoft from the collections of the British Library, University of California, and University of Toronto, followed in January 2007 by the New York Public Library and Cornell University. Books offer full text views and can be downloaded in PDF files. In the future, Microsoft intends to add copyrighted works with the permission of their publishers. In May 2007, Microsoft announced agreements with several main publishers, including Cambridge University Press and McGraw Hill. After digitizing 750,000 books and indexing 80 million journal articles, Microsoft ended the Live Search Books program in May 2008 and closed the website.

## **2006: Free WorldCat**

### **[Overview]**

WorldCat was created in 1971 by the non-profit OCLC (Online Computer Library Center) as the union catalog of the university libraries in the State of Ohio. Over the years, OCLC became a national and worldwide library cooperative, and WorldCat the largest library catalog in the world. In 2005, WorldCat had 61 million bibliographic records in 400 languages from 9,000 member libraries (paid subscription) in 112 countries. In 2006, 73 million bibliographic records were linking to 1 billion documents available in these libraries. In August 2006, WorldCat began to migrate to the web through the beta version of the new website WorldCat.org. Member libraries now provide free access to their catalogs and electronic resources: books, audio books, abstracts and full-text articles, photos, music CDs and videos. Another pioneer site was RedLightGreen, launched in Spring 2004 (with a beta version in Fall 2003) as the web version of the RLG Union Catalog, another major union catalog created in 1980 by the Research Libraries Group (RLG). RedLightGreen ended its service in November 2006, after a successful 3-year run, and RLG joined OCLC.

### **[In Depth (published in 1999)]**

In 1998, two organizations - OCLC (Online Computer Library Center) and RLIN (Research Library Information Network) - were running international bibliographical databases through the internet.

The OCLC Online Computer Library Center is a non-profit, membership, library computer service and research organization dedicated to furthering access to the world's information and reducing information costs. More than 27,000 libraries in 65 countries were using OCLC services to manage their collections and to provide online reference services. The website was available in English, Chinese, French, German, Portuguese, and Spanish.

OCLC services included: access services; collections and technical services; reference services; resource sharing; Dewey Decimal Classification (published by OCLC Forest Press); and preservation resources. From its headquarters in Dublin, Ohio, OCLC operated one of the world's largest library information networks. Libraries in the US joined OCLC through their OCLC-affiliated regional networks. Libraries outside the US received OCLC services through OCLC Asia Pacific, OCLC Canada, OCLC Europe, OCLC Latin America and the Caribbean, or via international distributors.

OCLC was also running WorldCat - the name of the OCLC Online Union Catalog - which is a merged electronic catalog of library catalogs around the world, and the world's largest bibliographic database with its 38 million records (in early 1998) in 400 languages (with transliteration for non-Roman languages), and an annual increase of 2 million records.

WorldCat stemmed from a concept which is the same for all union catalogs: earn time to avoid the cataloging of the same document by many catalogers worldwide. When they are about to catalog a publication, the catalogers of the member libraries search the OCLC

catalog. If they find the record, they copy it in their own catalog and add some local information. If they don't find the record, they create it in the OCLC catalog, and this new record is immediately available to all the catalogers of the member libraries worldwide.

Unlike RLIN, another main union catalog that accepts several records for the same document (please see below), the OCLC Online Union Catalog accepts only one record per document, and asks its members not to create duplicate records for documents that were already cataloged. The records are created in USMARC format (MARC: Machine Readable Catalog) according to the Anglo-American Cataloguing Rules, 2nd version (AACR2).

What is the history of OCLC? "In 1967, the presidents of the colleges and universities in the state of Ohio founded the Ohio College Library Center (OCLC) to develop a computerized system in which the libraries of Ohio academic institutions could share resources and reduce costs. OCLC's first offices were in the Main Library on the campus of the Ohio State University (OSU), and its first computer room was housed in the OSU Research Center. It was from these academic roots that Frederick G. Kilgour, OCLC's first president, oversaw the growth of OCLC from a regional computer system for 54 Ohio colleges into an international network. In 1977, the Ohio members of OCLC adopted changes in the governance structure that enabled libraries outside Ohio to become members and participate in the election of the Board of Trustees; the Ohio College Library Center became OCLC, Inc. In 1981, the legal name of the corporation became OCLC Online Computer Library Center, Inc. Today, OCLC serves more than 27,000 libraries of all types in the US and 64 other countries and territories." (excerpt from the 1998 website)

In early 1998, WorldCat had 38 million records - with one record per document. RLIN (Research Libraries Information Network) had 88 million records - with several records per document.

RLIN was run by the Research Libraries Group (RLG). The central RLIN database was a union catalog of 88 million items held in main libraries belonging to RLG member institutions, including research and specialized libraries, like law, technical, and corporate libraries.

RLIN included:

(1) records that described works cataloged by the Library of Congress, the National Library of Medicine, the US Government Printing Office, CONSER (Conversion of Serials Project), the British Library, the British National Bibliography, the National Union Catalog of Manuscript Collections, and RLG members and users;

(2) nearly all the books cataloged since 1968 and rapidly expanding coverage for older materials;

(3) information about non-book materials ranging from musical scores, films, videos, serials, maps, and recordings, to archival collections and machine-readable data files;

(4) unique on-line access to special resources, such as the United Nations' DOCFILE and CATFILE records, and the Rigler and Deutsch Index to pre-1950 commercial sound recordings;

(5) international book vendors' in-process records, that were transferred to bibliographers, acquisition services and catalogers, to order records or help them for cataloging items in their own local databases.

RLIN also provided:

(1) A catalog of computer files. Machine-readable data files were useful to a growing number of disciplines. RLIN contained records describing a number of such files, from the full-text French literary works in the ARTFL Database to the statistical data collected by the Inter-university Consortium for Political and Social Research (ICPSR) at the University of Michigan;

(2) A catalog of archives and special collections. The archival and manuscript collections of research libraries, museums, state archives, and historical societies contained essential primary resources, but information about their contents was often elusive. Archivists and curators worked with RLG to create an automated format for these collections. In 1998, there were 500,000 records available in RLIN for archival collections located throughout North America. These records described many collections by personal name, organization, subject, and format.

RLIN also hosted the English Short Title Catalogue (ESTC), an invaluable research tool for scholars in English culture, language, and literature. This file provided extensive descriptions and holdings information for letterpress materials printed in UK or any of its dependencies in any language, from the beginnings of print to 1800 - as well as for materials printed in English anywhere else in the world. Produced by the ESTC editorial offices at the University of California, Riverside, and the British Library, in partnership with the American Antiquarian Society and over 1,600 libraries worldwide, the file was updated and expanded daily. ESTC served as a comprehensive bibliography of the hand-press era and as a census of surviving copies. ESTC included 420,000 records as of June 1998, from the beginnings of print (1473) through the 18th century - including materials ranging from Shakespeare and Greek New Testaments to anonymous ballads, broadsides, songs, advertisements and other ephemera.

## **2007: Citizendium**

### **[Overview]**

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, and they are guided by expert editors. "Editors will be able to make content decisions in their areas of specialization, but otherwise working shoulder-to-shoulder with ordinary authors." (Larry Sanger, *Toward a New Compendium of Knowledge*, September 2006) Constables make sure the rules are respected. Citizendium was launched on March 25, 2007, with 1,100 articles, 820 authors and 180 editors.

## **2007: Encyclopedia of Life**

### **[Overview]**

Launched in May 2007, the Encyclopedia of Life is a global scientific effort to document all known species of animals and plants (1.8 million), and expedite the millions of species yet to be discovered and catalogued (8 to 10 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 comes from the MacArthur Foundation (US \$10 million) and the Sloan Foundation (\$2.5 million). A number of pages will be available by mid-2008. The encyclopedia will be operational in 3-5 years and completed (with all known species) in 10 years. Built on the scientific integrity of thousands of experts around the globe, the Encyclopedia will be a moderated wiki-style environment, freely available to all users everywhere.

## Websites

1968: ASCII: <http://www.asciitable.com/>  
1971: Project Gutenberg: <http://www.gutenberg.org/>  
1974: Internet: <http://www.isoc.org/>  
1977: UNIMARC: <http://www.unimarc.net/>  
1984: Copyleft: <http://www.gnu.org/copyleft/>  
1990: Web: <http://www.w3.org/>  
1991: Unicode: <http://unicode.org/>  
1993: Online Books Page: <http://onlinebooks.library.upenn.edu/>  
1993: PDF: <http://www.adobe.com/products/acrobat/adobepdf.html>  
1994: Library Websites: <http://lists.webjunction.org/libweb/>  
1994: Bold Publishers: <http://www.nap.edu/>  
1995: Amazon.com: <http://www.amazon.com/>  
1995: Online Press: <http://www.ipl.org/div/news/>  
1996: Internet Archive: <http://www.archive.org/>  
1996: New Ways of Teaching: <http://www.ifip.org/>  
1996: Palm Pilot: <http://www.palm.com/>  
1997: Digital Publishing: <http://www.adobe.com/devnet/digitalpublishing/>  
1997: Logos Dictionary: <http://www.logos.it/>  
1997: Multimedia Convergence: <http://www.ilo.org/public/english/dialogue/sector/>  
1998: Online Beowulf: <http://www.bl.uk/onlinegallery/onlineex/englit/beowulf/>  
1998: Digital Librarians: <http://www.ifla.org/>  
1998: Multilingual Web: <http://www.w3.org/International/>  
1999: Digital Authors: <http://www.cotres.net/>  
1999: Open eBook: <http://www.idpf.org/>  
1999: yourDictionary.com: <http://www.yourdictionary.com/>  
2000: Online Bible of Gutenberg: <http://www.bl.uk/treasures/gutenberg/homepage.html>  
2000: Distributed Proofreaders: <http://www.pgdp.net/>  
2000: Public Library of Science: <http://www.plos.org/>  
2001: Wikipedia: <http://www.wikipedia.org/>  
2001: Creative Commons: <http://creativecommons.org/>  
2002: MIT OpenCourseWare: <http://ocw.mit.edu/>  
2004: Project Gutenberg Europe: <http://pge.rastko.net/>  
2004: Google Print / Book Search: <http://books.google.com/>  
2005: Open Content Alliance: <http://www.opencontentalliance.org/>  
2006: Microsoft Live Search Books: <http://blogs.msdn.com/livesearch/>  
2006: Free WorldCat: <http://www.worldcat.org/>  
2007: Citizendium: <http://en.citizendium.org/>  
2007: Encyclopedia of Life: <http://www.eol.org/>