PROJECT GUTENBERG (1971-2005)

MARIE LEBERT

NEF, University of Toronto, 2005

Copyright © 2005 Marie Lebert

Dated August 15, 2005, this long article (following a short version published in June 2004 [and copied at the end of this file]) is a paper for the third International Colloquium on ICT-enhanced French Studies: Dialogues across languages and cultures, October 2005, York University, Toronto, Canada. This article is dedicated to all Project Gutenberg and Distributed Proofreaders volunteers on the five continents, who offer us a free library of 16,000 high-quality eBooks, mainly classics of world literature, with a goal of one million eBooks in ten years.

With many thanks to Russon Wooldridge, who kindly edited this long article. The original version is available on the NEF, University of Toronto: http://www.etudes-francaises.net/dossiers/gutenberg_eng.htm

The French version is: Le Projet Gutenberg (1971-2005). The updated English version is: Project Gutenberg (1971-2008).

TABLE

1. SUMMARY

My fascination for Project Gutenberg is not new, but it doesn't wane. Nobody has done a better job of putting the world's literature at everyone's disposal. And to create a vast network of volunteers all over the world, without wasting people's skills or energy.

Here is the story in a few lines.

In July 1971, Michael Hart created Project Gutenberg with the goal of making available for free, and electronically, literary works belonging to the public domain. A project that has long been considered by its critics as impossible on a large scale. A pioneer site in a number of ways, Project Gutenberg was the first information provider on the internet and is the oldest digital library. Michael himself keyed in the first hundred books.

When the internet became popular, in the mid-1990s, the project got a boost and an international dimension. Michael still typed and scanned in books, but now coordinated the work of dozens and then hundreds of volunteers in 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).

30 years after its birth, Project Gutenberg is running at full capacity. It had 5,000 books online in April 2002, 10,000 books online in October 2003, and 15,000 books online in January 2005, with 400 new books available per month, 40 mirror sites in a number of countries, and books downloaded by the tens of thousands every day.

Whether they were digitized 20 years ago or they are 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 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 goal of one million eBooks in 2015, and 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.

2. HISTORY, FROM THE ORIGINS TO TODAY

= The Beginnings in 1971

Let us get back to the beginnings of the project. When he was a student at the University of Illinois (USA), Michael Hart was given $100,000,000 of computer time at the Materials Research Lab of his university. 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 upper case, because there was no lower case yet. But to send a 5 K 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. Project Gutenberg was born.

Michael decided to use this huge amount of computer time to search the public domain books that

were stored in our libraries, and to digitize these books. He also decided to store the electronic texts (eTexts) in the simplest way, using the plain text format called Plain Vanilla ASCII, so they can be read easily by any machine, operating system or software. A book would become a continuous text file instead of a set of pages, with caps for the terms in italic, bold or underlined of the print version.

Soon afterwards he defined Project Gutenberg's mission: to put at everyone's disposal, in electronic versions, as many literary works of the public domain as possible for free. As he stated years later, in August 1998, "We consider eText to be a new medium, with no real relationship to paper, other than presenting the same material, but I don't see how paper can possibly compete once people each find their own comfortable way to eTexts, especially in schools."

= Persevering from 1972 to 1989

After he keyed in The United States Declaration of Independence in 1971, Michael went on in 1972 and typed in a longer text, The United States Bill of Rights, that includes the ten first amendments added in 1789 to the Constitution (dated 1787) and defining the individual rights of the citizens and the distinct powers ot the Federal Government and the States. In 1973, Michael typed in the full text of The United States Constitution.

From one year to the next, disk space was getting larger, by the standards of the time (there was no hard disk yet), so it was possible to plan bigger files. Michael began typing in the Bible, because the individual books of the Bible could be processed separately as different files. He also worked on the collected works of Shakespeare, with one play at a time, and a file for each play. That edition of Shakespeare was never released, due to copyright changes. If Shakespeare's works belong to the public domain, the comments and notes may be copyrighted, depending on the publication date. But other editions belonging to the public domain were posted a few years later.

In parallel, the internet, which was still embryonic in 1971, was born in 1974 with the launching of TCP/IP (Transmission Control Protocol / Internet Protocol). Its rapid expansion started in 1983.

In August 1989, Project Gutenberg celebrated the completion of its 10th eText,
The King James Bible.

= 10 to 1,000 eBooks from 1990 to 1996

In 1990, there were 250,000 internet users, and the standard was 360 K disks. In
January 1991, Michael keyed in Alice's Adventures in Wonderland, by Lewis
Carroll (published in 1865). In July 1991, he typed in Peter Pan, by James M.
Barrie (published in 1904). These two worldwide classics of childhood literature
each fitted on one disk.

1991 was also the year the web became operational. The first browser, Mosaic, was released in November 1993. As the web was becoming a popular medium, it became easier to circulate eTexts and recruit volunteers. Project Gutenberg gradually got into its stride, with the digitization of one eText per month in 1991, two eTexts per month in 1992, four eTexts per month in 1993 and eight eTexts per month in 1994. In January 1994, Project Gutenberg celebrated its 100th eText by releasing The Complete Works of William Shakespeare. The steady growth went on, with an average of 8 eTexts per month in 1994, 16 eTexts per month in 1995, and 32 eTexts per month in 1996.

As we can see, from 1991 to 1996, the "output" doubled every year. While continuing to digitize books, Michael was also coordinating the work of dozens of volunteers. At the end of 1993, Project Gutenberg's eTexts were organized into three main sections: a) "Light Literature", such as Alice's Adventures in Wonderland, Peter Pan or Aesop's Fables; b) "Heavy Literature", such as the Bible, Shakespeare's works or Moby Dick; c) "Reference Literature", such as Roget's Thesaurus, and a set of encyclopaedias and dictionaries.

Project Gutenberg's goal is to be "universal" both for the literary works that are chosen and the audience who reads them. The goal is to put literature at everyone's disposal. With a focus on books that many people would use frequently, and not only students and teachers. For example, the "Light Literature" section is intended for pre-schoolers as well as their grandparents. The aim is that they will want to look up the eText of Peter Pan when they come back from watching Hook at the movies. Or that they will read the eText of Alice's Adventures in Wonderland after seeing it on TV. Or that they will look for the context of a quotation after hearing it in one of the Star Trek episodes; nearly every episode of Star Trek quotes from books which are in the Project Gutenberg collections.

The idea is that, whether they were avid readers of print books or not in the past, people should easily be able to look up quotations they hear in conversations, movies, music, or they read in books, newspapers and magazines, within a library containing all these quotations in an easy-to-use format.

eTexts don't take up much space in ASCII format. They can be easily downloaded with a standard phone line. Searching a word or a phrase is simple too. People can easily search an entire eText by using the plain "search" menu available in any program."

= 1,000 eBooks in August 1997

In 1997, the "output" was still an average of 32 eTexts per month. In June 1997, Project Gutenberg released The Merry Adventures of Robin Hood, by Howard Pyle (published in 1883). In August 1997, it released its 1000th eText, La Divina Commedia di Dante (published in 1321), in Italian, its original language.

In August 1998, Michael wrote: "My own personal goal is to put 10,000 eTexts on the Net [editor's note: his goal was reached in October 2003] and if I can get some major support, I would like to expand that to 1,000,000 and to also expand our potential audience for the average eText from 1.x% of the world population to over 10%, thus changing our goal from giving away 1,000,000,000,000 eTexts to 1,000 times as many, a trillion and a quadrillion in US terminology."

= 1,000 to 5,000 eBooks from 1998 to 2002

From 1998 to 2000, there was a steadfast average of 36 new eTexts per month. In May 1999, there were 2,000 eTexts. The 2000th eText was Don Quijote, by Cervantes (published in 1605), in Spanish, its original language.

Around 40 eTexts per month were released during the 1st semester 2001, and 50 eTexts during the 2nd semester. Released in December 2000, the 3000th eText was the third volume of A l'ombre des jeunes filles en fleurs (In the Shadow of Young Girls in Flower), by Marcel Proust (published in 1919), in French, its original language.

Released in October 2001, the 4000th eText was The French Immortals Series, in English. Published in 1905 by Maison Mazarin, Paris, this book is an anthology of short fictions by authors belonging to the renowned French Academy (Académie française), notably Emile Souvestre, Pierre Loti, Hector Malot, Charles de Bernard and Alphonse Daudet.

Available in April 2002, the 5000th eText was The Notebooks of Leonardo da Vinci, which he wrote at the beginning of the 16th century. A text that is still in the Top 100 of downloaded texts in 2005.

In 1988, Michael Hart chose to digitize Alice's Adventures in Wonderland and Peter Pan because they each fitted on one 360 K disk, the standard of the time. Fifteen years later, in 2002, 1.44 M is the standard disk and ZIP is the standard compression. The practical file size is about 3 million characters, more than long enough for the average book. The digitized ASCII version of a 300-page novel is 1 M. A bulky book can fit in two ASCII files, that can be downloaded as is or in ZIP format.

An average of 50 hours is necessary to get an eText selected, copyright-cleared, scanned, proofread, formatted and assembled.

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

In 2002, around 100 eTexts were released per month. In Spring 2002, Project Gutenberg's eTexts represented 1/4 of all the public domain works freely available on the web and listed nearly exhaustively by The Internet Public Library (IPL). An impressive result thanks to the relentless work of 1,000 volunteers in several countries.

= 10,000 eBooks in October 2003

1,000 eTexts in August 1997, 2,000 eTexts in May 1999, 3,000 eTexts in December 2000, 4,000 eTexts in October 2001, 5,000 eTexts in April 2002, 10,000 eTexts in October 2003. eText number 10000 is The Magna Carta, the first English constitutional text, signed at the beginning of the 13th century.

From April 2002 to October 2003, in 18 months, the number of eTexts doubled, going from 5,000 to 10,000, with a monthly average of 300 new digitized books. In December 2003, most of the titles (9,400 eBooks) were also burned on a DVD to celebrate the landmark of 10,000 eTexts, renamed as eBooks, according to the latest terminology in the field. A few months before, in August 2003, a "Best of Gutenberg" CD was made available containing 600 eBooks (as a follow-up to other CDs in the past). People could request the CD and DVD for free, and were then encouraged to make copies for a friend, a library or a school. (In 2005, CD and DVD files are also periodically generated as ISO files. When downloaded, they can be used to make a CD or DVD using a CD or DVD writer.)

10,000 eBooks. An impressive number if we think about all the scanned and proofread pages this number represents. A fast growth thanks to Distributed Proofreaders, a website designed in 2000 by Charles Franks to share the proofreading of eBooks between many volunteers. Volunteers choose one of the eBooks listed on the site and proofread a given page. They don't have any quota to fulfill, but it is recommended they do a page per day if possible. It doesn't seem much, but with hundreds of volunteers it really adds up.

In December 2003, there were 11,000 eBooks digitized in several formats, most of them in ASCII, and some of them in HTML or XML. This represented 46,000 files, and 110 G. On 13 February 2004, the day of Michael Hart's presentation at UNESCO, in Paris (see below), there were exactly 11,340 eBooks in 25 languages. In May 2004, the 12,581 eBooks represented 100,000 files in 20 different formats, and 135 gigabytes. With 400 new eBooks added per month (and more in the years to come), the number of gigabytes is expected to double every year.

= 15,000 eBooks in January 2005

In January 2005, Project Gutenberg had 15,000 eBooks. eBook number 15000 is The Life of Reason, by George Santayana (published in 1906). On June 16, 2005 there were 16,481 eBooks in 42 languages. On August 3, 2005, besides English (14,590 eBooks), the six main languages were French (578 eBooks), German (349 eBooks), Finnish (225 eBooks), Dutch (130 eBooks), Spanish (105 eBooks) and Chinese (69 eBooks).

Michael hopes to reach 1,000,000 eBooks by 2015. Each email he sends includes the current number, and the next significant goal to reach. As of July 2005, the next goal is 20,000 eBooks. This goal should be reached in July 2006, for the 35th anniversary of Project Gutenberg.

Conceived in January 2004, at the same time as the launching of Distributed Proofreaders Europe (DP Europe) by Project Rastko, Project Gutenberg Europe went online in June 2005 and released the 100 first eBooks processed by DP Europe over the past several months. These eBooks are in several languages, a reflection of European linguistic diversity. 100 languages are planned for the long term.

In July 2005, Project Gutenberg of Australia (launched in 2001) reached 500 eBooks, and Project Gutenberg of Canada took its first steps (see the PGCanada List). Project Gutenberg Portugal and Project Gutenberg Philippines will be next. (For the latest news, check the News and Events of Project Gutenberg.)

3. THE PUBLIC DOMAIN, AN ENDLESS TOPIC

Despite the enthusiasm and the persistence of its hundreds of volunteers, the task of Project Gutenberg isn't made any easier by the increasing restrictions to the public domain. As stated in the FAQ, "the public domain is the set of cultural works that are free of copyright, and belong to everyone equally." In former times, 50% of works belonged to the public domain, and could be freely used by everybody. Nowadays, 99% of works are governed by copyright, and some people would like this percentage to reach 100%.

In the Copyright HowTo section, Project Gutenberg presents its own rules for confirming the public domain status of eBooks according to US copyright laws. Here is a summary. Works published before 1923 entered the public domain no later than 75 years from the copyright date. (All these works are now in the public domain.) Works published between 1923 and 1977 retain copyright for 95 years. (No such works will enter the public domain until 2019.) Works created from 1978 on enter the public domain 70 years after the death of the author if the author is a natural person. (Nothing will enter the public domain until 2049.) Works created from 1978 on enter the public domain 95 years after publication (or 120 years after creation) if the author is a corporate one. (Nothing will enter the public domain until 2074.) Other rules apply too.

Much more restrictive than the previous one, the current legislation became effective after the promulgation of amendments to the 1976 Copyright Act, dated October 27th, 1998. As explained by Michael Hart in July 1999: "Nothing will expire for another 20 years. We used to have to wait 75 years. Now it is 95 years. And it was 28 years (+ a possible 28 year extension, only on request) before that, and 14 years (+ a possible 14 year extension) before that. So, as you can see, this is a serious degrading of the public domain, as a matter of continuing policy."

The dates mentioned by Michael are: a) 1790, date of the stranglehold of the Stationers' Guild (the publishers of the time) on the Gutenberg printing press (hence the 14-year copyright); b) 1909, date of the copyright reinforcement to counter the re-publishing of large collections of the public domain by reprint houses using steam and electric presses (hence the 28-year copyright); c) 1976, date of a new

tightening of the copyright following the introduction of the Xerox photocopying machine (hence the 50-year copyright after the author's life); d) 1998, date of a further tightening of the copyright following the development of the internet (hence the 70-year copyright after the author's life). These are only the main lines. The Copyright Act has been amended 11 times in the last 40 years.

As stated by Tom W. Bell in Trend of Maximum U.S. General Copyright Term (with a very useful chart): "The first federal copyright legislation, the 1790 Copyright Act, set the maximum term at fourteen years plus a renewal term of fourteen years. The 1831 Copyright Act doubled the initial term and retained the conditional renewal term, allowing a total of up to forty-two years of protection. Lawmakers doubled the renewal term in 1909, letting copyrights run for up to fifty-six years. The interim renewal acts of 1962 through 1974 ensured that the copyright in any work in its second term as of September 19, 1962, would not expire before Dec. 31, 1976. The 1976 Copyright Act changed the measure of the default copyright term to life of the author plus fifty years. Recent amendments to the Copyright Act [the ones in 1998] expanded the term yet again, letting it run for the life of the author plus seventy years."

The amendments of the Copyright Act, dated October 27, 1998, were a major blow for digital libraries and deeply shocked their founders, beginning with Michael Hart and John Mark Ockerbloom, founder of The Online Books Page. But how were they to measure up to the major publishing companies? Michael wrote in July 1999: "No one has said more against copyright extensions than I have, but Hollywood and the big publishers have seen to it that our Congress won't even mention it in public. The kind of copyright debate going on is totally impractical. It is run by and for the 'Landed Gentry of the Information Age.' 'Information Age'? For whom?"

True enough. The political authorities continually speak about an information age while tightening the laws relating to the dissemination of information. The contradiction is obvious. This problem has also affected Australia (forcing Project Gutenberg of Australia to withdraw dozens of books from its collections) and several European countries. In a number of countries, the rule is now life of the author plus 70 years, instead of life plus 50 years, following pressure from content owners, with the subsequent "harmonization" of national copyright laws as a response to the "globalization of the market". (The Online Books Page gives a summary of the various copyright regimes, with a number of useful links.)

Now, from the volunteer point of view, the wisest thing to do is to choose a book published before 1923. It is also required that copyright clearance be confirmed prior to working on any eBook by sending a photocopy of the title page and verso page (even if the latter is blank) to Michael. The pages should be sent as scans to be uploaded on the website. For people who cannot create scans, it is possible to send photocopies by postal mail. The pages will then be filed, either on paper or electronically, so that the proof will be available in the future, to demonstrate if necessary that the book is in the public domain under the US law. Project Gutenberg doesn't release any eBook until the book's copyright status has been confirmed.

There is nevertheless hope for some books published after 1923. According to Greg Newby, director of PGLAF (Project Gutenberg Literary Archive Foundation), one million books published between 1923 and 1964 could also belong to the public domain, because only 10% of copyrights were actually renewed. Project Gutenberg tries to locate these books. In April 2004, with the help of hundreds of volunteers at Distributed Proofreaders, all Copyright Renewal records were posted for books from 1950 through 1977. So, if a given book published during this period is not on the list, it means the copyright was not renewed, and the book fell into the public domain.

4. THE METHOD ADOPTED BY PROJECT GUTENBERG

Whether digitized years ago or now, all the books are digitized in 7-bit plain
ASCII (American Standard Code for Information Interchange), called Plain Vanilla
ASCII. Used since the beginnings of computing, it is the set of unaccented
characters present on a standard English-language keyboard (A-Z, a-z, numbers,
punctuation and other basic symbols). 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.)

Plain Vanilla ASCII is 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 every computer in the world. 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 between 1999 and 2003). 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" encoding system created in 1991.

Project Gutenberg also publishes eBooks in well-known formats like HTML, XML or RTF. There are Unicode files too. Any other format provided by volunteers (PDF, LIT, TeX and many others) is usually accepted, as long as they also supply an ASCII version where possible.

But a large scale conversion into other formats is handed over to other organizations. For example Blackmask Online, which uses Project Gutenberg's collections to offer thousands of free eBooks in eight different formats based on the Open eBook (OeB) format. Or Manybooks.net, which converts Project Gutenberg's eBooks into formats readable on PDAs. Or Bookshare.org, the main digital library for the visual impaired community in the US, which converts books from Project Gutenberg into Braille format and DAISY (Digital Audio Information System) format.

What is entailed exactly, once copyright clearance is received? Digitization is done by scanning the book page after page to get "image" files. Then volunteers run an OCR (Optical Character Recognition) software to convert "image" files into text files. Then each text file is proofread (i.e. re-read and corrected) by comparing it to the "image" file or the original page of the print version. There is an average of 10 mistakes per page for a good OCR package and... many more mistakes if the quality of the scanner and the OCR package is not great.

The book is proofread twice on the computer screen by two different people, who make any corrections necessary. When the original is in poor condition, as with very old books, it is keyed in manually, word by word. Some volunteers themselves prefer to type short texts, or works they particularly like. But most books are scanned, "OCRized" and proofread.

Digitization in "text format" means a book can be copied, indexed, searched, analyzed and compared with other books. It is possible to search the content of the book with the "Find" button available in any browser and any software, without a specific search engine. Project Gutenberg provides a "Nearly Full Text" search (on the first 100 K of each file) using Google, with a database updated approximately monthly. It also provides a search of book metadata (author, title, brief description, keywords) as a participant in Yahoo!'s Content Acquisition Program, with a database updated weekly. (Please see the bottom of the Online Book Catalog.) In the Advanced Search, several fields can be filled: author, title, subject, language, category (any, audio book, music, pictures), LoCC (Library of Congress Catalog classification), filetype (text, PDF, HTML, XML, JPEG, etc.), and eText/eBook No. A field "Full Text" was recently added as an experimental feature.

The assets of digitization in "text format" are numerous. It makes a smaller and more easily sendable computer file, unlike digitization in "image format", which produces a bulky "photo" file. Contrary to other formats, the files are accessible for low-bandwidth use. They can be copied as much as needed to produce new digital or print versions for free. The typos pointed out after the text is released can be fixed at any time. Readers can change the font and size of characters, the margins or the number of lines per page. Visually impaired readers can increase the letter size. Blind readers can use speech recognition software. All this is very difficult, if not impossible, with many other formats.

If the eBooks released are 99.9% accurate in the eyes of the general reader, the goal is not to create authoritative editions, and to argue with a picky reader whether a certain sentence should have a colon instead of a semi-colon between its clauses.

Project Gutenberg is convinced that proofreading by human beings is a very important step, and that this step makes all the difference. The use of scanned books as is —converted to text format by OCR software with no proofreading— gives a much lower quality result. After running OCR software, the text is 99% reliable, in the best of cases. After proofreading, the text becomes 99.95% reliable (a high percentage which is also the standard at the Library of Congress).

For this reason, Project Gutenberg's perspective is rather different from that of the Million Book Project, another project launched by several professors from Carnegie Mellon University, and whose collections (10,611 books on June 1st, 2005) are hosted by the Internet Archive (the Internet Archive is also the backup distribution site of Project Gutenberg). In the case of the Million Book Project, books are scanned and "OCRized", but they are not proofread. The main formats used are XML, TIF and DjVu.

On Project Gutenberg's website, a File Recode Service allows users to convert books in one format (ASCII, ISO-8859, Unicode and Big-5) into another, and vice versa. A much more powerful conversion

program may be launched in the future, with a conversion into still more formats (XML, HTML, PDF, TeX, RTF), including Braille and voice. It will then also be possible to choose the font and size of characters and the background color. Another eagerly expected conversion is that of a book from one language to another by machine translation software. This may be possible in a few years, when machine translation is accurate to 99%.

5. DISTRIBUTED PROOFREADERS, TO HANDLE SHARED PROOFREADING

The main "leap forward" of Project Gutenberg in the last few years is due to Distributed Proofreaders.

Distributed Proofreaders was conceived in 2000 by Charles Franks to help in the digitizing of public domain books. Originally meant to assist Project Gutenberg in the handling of shared proofreading, Distributed Proofreaders became the main source of Project Gutenberg eBooks. In 2002, Distributed Proofreaders became an official Project Gutenberg site.

The number of eBooks that have been processed through Distributed Proofreaders has grown fast, with a total of 3,000 eBooks in February 2004, 5,000 eBooks in October 2004 and 7,000 eBooks in May 2005. On August 3, 2005, 7,639 books were complete (processed through the site and posted to Project Gutenberg), 1,250 books were in progress (processed through the site but not yet posted, because currently going through their final proofreading and assembly), and 831 books were being proofread (currently being processed).

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 on 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 eBook. 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.

Each time proofreaders go to the website, they choose the book they want. 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. (For more detailed information, check the FAQ Central.) 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 cataloguing process (author's dates of birth and death, Library of Congress classification, etc.) after the release.

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 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.

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-INT) and Distributed Proofreaders Europe (DP Europe). 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'."

6. EBOOKS IN MORE AND MORE LANGUAGES

What about languages?

Initially, the eBooks were mostly in English. As Project Gutenberg is based in the United States, it first focused on the English-speaking community in the country and worldwide.

In October 1997, Michael Hart expressed his intention to expand the publishing of eBooks in other languages. At the beginning of 1998, the catalog had a few titles in French (10 titles), German, Italian, Spanish and Latin. In July 1999, Michael wrote: "I am publishing in one new language per month right now, and will continue as long as possible."

In early 2004, there were works in 25 languages. In July 2005, there were works in 42 languages, including Iroquoian, Sanskrit and the Mayan languages. The seven "main" languages were: English (with 14,548 books on July 27, 2005), French (577 books), German (349 books), Finnish (218 books), Dutch (130 books), Spanish (103 books) and Chinese (69 books).

Let us take French as an example. On February 13, 2004, there were 181 eBooks in French (out of a total of 11,340 eBooks). On May 16, 2005, there were 547 eBooks in French (out of 15,505 Books). The number tripled in 15 months. This number should rise significantly during the next few years, notably with Project Gutenberg Europe (launched in June 2005).

What were the first eBooks posted in French? They were six novels by Stendhal and two novels by Jules Verne, all released in early 1997. The six novels by Stendhal were: L'Abbesse de Castro, Les Cenci, La Chartreuse de Parme, La Duchesse de Palliano, Le Rouge et le Noir and Vittoria Accoramboni. The two novels by Jules Verne were: De la terre à la lune and Le tour du monde en quatre-vingts jours. In early 1997, whereas Project Gutenberg offered no English version of any of Stendhal's writings (yet), three of Jules Verne's novels were available in English: 20,000 Leagues Under the Seas (original title: Vingt mille lieues sous les mers), posted in September 1994; Around the World in 80 Days (original title: Le tour du monde en quatre-vingts jours), posted in January 1994 and From the Earth to the Moon(original title: De la terre à la lune), posted in September 1993. Stendhal and Jules Verne were followed by Edmond Rostand with Cyrano de Bergerac, posted in March 1998.

In late 1999, the "Top 20" —the 20 most downloaded authors— included Jules Verne at 11 and Emile Zola at 16. They still have a very good ranking in the present "Top 100".

As a side remark, the first "images" ever made available by Project Gutenberg were French Cave Paintings, posted in April 1995, with an XHTML version posted in November 2000. This eBook contains four photos of paleolithic paintings found in a grotto located in Ardèche, a region of south-eastern France. These photos, which are copyrighted, were made available to Project Gutenberg thanks to Jean Clottes, a French general curator for cultural heritage (conservateur général du patrimoine), for everyone to enjoy them.

Multilingualism is now one of the priorities of Project Gutenberg, like internationalization. In early 2004, Michael Hart went off to Europe, with stops in Paris, Brussels and Belgrade. He gave a lecture on February 12, 2004 at UNESCO (United Nations Educational, Scientific and Cultural Organization) headquarters in Paris. He chaired a discussion at the French National Assembly on February 13. The following week, he addressed the European Parliament, in Brussels. He also met with the team of Project Rastko, in Belgrade, to support the creation of Distributed Proofreaders Europe (launched in January 2004) and Project Gutenberg Europe (conceived at the same time, and launched in June 2005).

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 eBooks for Project Gutenberg Europe. Since its very beginnings, DP Europe has been 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 the European languages. When it gets up to speed, DP Europe will provide eBooks for several national and/or linguistic digital libraries, for example Projet Gutenberg France for France. 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 had the boldness to launch such a difficult and exciting project for Europe, and catalysed 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 eBook. In June 2005 Project Gutenberg Europe was launched with these first 100 eBooks. PG Europe operates under "life

+50" copyright laws. On August 3, 2005, 137 books were complete (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). DP Europe supports Unicode to be able to proofread eBooks in numerous languages. Unicode is an encoding system created in 1991 that gives a unique number for every character in any language. From the Past to the Future

10 books online in August 1989; 100 books in January 1994; 1,000 books in August 1997; 2,000 books in May 1999; 3,000 books in December 2000; 4,000 books in October 2001; 5,000 books in April 2002; 10,000 books in October 2003; 15,000 books in January 2005; and 1 million books planned for 2015.

But Project Gutenberg's results are not only measured in numbers, which can't compete yet with the number of print books in the public domain. The results also include the major influence that the project has had. As the oldest producer of free eBooks on the internet, Project Gutenberg has inspired many other digital libraries, for example Projekt Gutenberg-DE for classic German literature and Projekt Runeberg for classic Nordic (Scandinavian) literature, to name only two.

Project Gutenberg keeps its administrative and financial structure to the bare minimum. Its motto fits into three words: "Less is more". The minimal rules give much space to volunteers and to new ideas. The goal is to ensure its independence from loans and other funding and from ephemeral cultural priorities, to avoid pressure from politicians or economic interests. The aim is also to ensure respect for the volunteers, who can be confident their work will be used not just for decades but for centuries. Volunteers can network through mailing lists and weekly or monthly newsletters. Donations are used to buy equipment and supplies, mostly computers and scanners. Founded in 2000, the PGLAF (Project Gutenberg Literary Archive Foundation) has only three part-time employees.

More generally, Michael should be given more credit as the real inventor of the eBook. If we consider the eBook in its etymological sense, that is to say a book that has been digitized to be distributed as an electronic file, it is now 34 years old and was born with Project Gutenberg in July 1971. This is a much more comforting paternity than the various commercial launchings in proprietary formats that peppered the early 2000s. There is no reason for the term "eBook" to be the monopoly of Amazon, Barnes & Noble, Gemstar and others. The non-commercial eBook is a full eBook, and not a "poor" version, just as non-commercial ePublishing is a fully-fledged way of publishing, and as valuable as commercial ePublishing. Project Gutenberg eTexts are now called eBooks, to use the recent terminology in the field.

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

What about documents other than text?

In September 2003, Project Gutenberg launched Project Gutenberg Audio eBooks. As of 2005, there are 391 computer-generated audio books and a few human-read audio books. The number of human-read eBooks should greatly increase over the next few years. As for computer-generated eBooks, it seems they won't be stored in a specific section any more, but "converted" when requested from the existing electronic files in the main collections. Voice-activated requests will be possible, as a useful tool for visually impaired readers.

Launched at the same time, The Sheet Music Subproject is dedicated to digitized music sheet. It also contains a few music recordings. Some still pictures and moving pictures are also available. These new collections should take off in the future.

But digitizing books remains the priority, and there is a big demand, as confirmed by the tens of thousands of eBooks that are downloaded every day. For example, on July 31, 2005, there were 37,532 downloads for the day, 243,808 downloads for the week (July 24-31), and 1,154,765 downloads for the month. This only for transfers from ibiblio.org (University of North Carolina at Chapel Hill), the main eBook distribution site (which also hosts the website). The Internet Archive is the backup distribution site and provides unlimited disk space for storage and processing. Project Gutenberg has 44 mirror sites in many countries and is looking for new ones. It also encourages the use of P2P for sharing its eBooks. The "Top 100" lists the top 100 eBooks and the top 100 authors for the previous day, the last 7 days and the last 30 days.

Project Gutenberg eBooks can also help bridge the "digital divide." They can be read on a computer

or a secondhand PDA costing just a few dollars. Solar-powered PDAs offer a good solution in remote regions and developing countries.

eBooks are also copied on CDs and DVDs. Blank CDs and DVDs cost next to nothing, as does their burning on a CD or DVD writer. Project Gutenberg sends a free CD or DVD to anyone who asks for it, and people are encouraged to make copies for a friend, a library or a school. Released in August 2003, the "Best of Gutenberg" CD contains over 600 eBooks. Released in December 2003, the first Project Gutenberg DVD contains 9,400 eBooks. A new DVD is in preparation. The current prototype contains nearly 26,000 eBooks (with some titles in different versions and formats), and is about 3/4 full.

By the time the collections hit one million eBooks in 2015 or before, it is hoped machine translation software will be able to convert them from one to another of 100 languages. In ten years from now, it is possible that machine translation will be judged 99% satisfactory (research is very active on that front, but there is still a lot to do), allowing for the reading of literary classics in a choice of many languages. In 2004, Project Gutenberg was in touch with a European project studying how to combine translation software and human translators, somewhat as OCR software is now combined with the work of proofreaders.

34 years after the beginnings of Project Gutenberg, Michael Hart describes himself as a workaholic who devotes his entire life to his project, because he thinks eBooks will become the "killer ap(plication)" of the computer revolution. He considers himself a pragmatic and farsighted altruist. For years he was regarded as a nut but now he is respected. He wants to change the world through freely-available eBooks that can be used and copied endlessly. Reading and culture for everyone at minimal cost. Project Gutenberg's mission can be stated in eight words: "To encourage the creation and distribution of eBooks," by everybody, and by every possible means. While implementing new ideas, new methods and new software.

Let us give the last word to Michael, whom I asked in August 1998: "What is your best experience with the internet?" His answer was: "The notes I get that tell me people appreciate that I have spent my life putting books, etc., on the internet. Some are quite touching, and can make my whole day." Seven years later, he confirms that his answer would still be the same.

7. CHRONOLOGY [UPDATED IN 2006]

1971 (July): Michael Hart keyed in The United States Declaration of Independence (eBook # 1) and informed the first 100 internet users. Project Gutenberg was born.

1972: He keyed in The United States Bill of Rights (eBook # 2).

1973: He keyed in The United States Constitution (eBook # 5).

1974-1988: He keyed in parts of the Bible and several works by Shakespeare.

1989 (August): The King James Bible (eBook # 10).

1991 (January): Alice's Adventures in Wonderland (eBook # 11).

1991 (June): Peter Pan (eBook # 16).

1991: Digitization of one book per month.

1992: Digitization of two books per month.

1993: Digitization of four books per month.

1993 (December): Creation of three main sections: Light Literature, Heavy Literature and Reference Literature.

1994: Digitization of eight books per month.

1994 (January): The Complete Works of William Shakespeare (eBook # 100).

1995: Digitization of 16 books per month.

1996-1997: Digitization of 32 books per month.

1997 (August): La Divina Commedia di Dante, in Italian (eBook # 1000).

1997: Launching of the Project Gutenberg Consortia Center.

1998-2000: Digitization of 36 books per month.

1999 (May): Don Quijote, by Cervantes, in Spanish (eBook # 2000).

2000: Creation of the Project Gutenberg Literary Archive Foundation.

2000 (October): Charles Franks conceived Distributed Proofreaders to assist Project Gutenberg.

2000 (December): A l'ombre des jeunes filles en fleurs, 3rd volume, by Proust, in French (eBook # 3000).

2001 (August): Creation of Project Gutenberg of Australia.

2001 (October): The French Immortals Series, in English (eBook # 4000).

2001: Digitization of 103 books per month.

2001: Distributed Proofreaders became the main source of Project Gutenberg eBooks.

2002: Distributed Proofreaders became an official Project Gutenberg site.

2002 (April): The Notebooks of Leonardo da Vinci, in English (eBook # 5000).

2003 (August): "Best of Gutenberg" CD with 600 eBooks.

2002: Digitization of 203 books per month.

2003 (September): Launching of Project Gutenberg Audio eBooks.

2003 (October): The number of eBooks doubled in 18 months, going from 5,000 to 10,000.

2003 (October): The Magna Carta (eBook # 10000).

2003 (December): First DVD, with 9,400 eBooks.

2003: Project Gutenberg Consortia Center became an official Project Gutenberg site.

2003: Digitization of 355 books per month.

2004 (January): Launching of Project Gutenberg Europe by Project Rastko.

2004 (January): Launching of Distributed Proofreaders Europe by Project Rastko.

2004 (February): Michael Hart went off to Europe (Paris, Brussels, Belgrade).

2004 (February): Michael Hart's presentation at UNESCO headquarters, in Paris.

2004 (February): Michael Hart's visit to the European Parliament, in Brussels.

2004 (October): 5,000 eBooks processed by Distributed Proofreaders.

2004: Digitization of 336 books per month.

2005 (January): The Life of Reason, by George Santayana (eBook # 15000).

2005 (May): 7,000 eBooks processed by Distributed Proofreaders.

2005 (May): First 100 eBooks processed by Distributed Proofreaders Europe.

2005 (June): 16,000 eBooks in Project Gutenberg.

2005 (June): Project Gutenberg Europe has 100 eBooks.

2005 (July): First steps of Project Gutenberg of Canada.

2005 (October): 5th anniversary of Distributed Proofreaders.

2005: Digitization of 248 books per month.

2006 (January): Launching of Project Gutenberg PrePrints.

2006 (February): 8,000 eBooks processed by Distributed Proofreaders.

2006 (May): Creation of the Distributed Proofreaders Foundation.

2006 (July): 35th anniversary of Project Gutenberg.

2006 (July): New DVD, with 17,000 eBooks.

2006 (November): Launching of the Project Gutenberg News website.

2006 (December): 20,000 eBooks in Project Gutenberg.

2006 (December): 400 eBooks processed by Distributed Proofreaders Europe.

2006: Digitization of 360 books per month.

2010 (estimation): Automatic conversion in numerous formats.

2015 (estimation): 1,000,000 eBooks in Project Gutenberg.

2015 (estimation): Machine translation in 100 languages.

9. LINKS

Project Gutenberg: https://www.gutenberg.org/

Project Gutenberg's FAQ: https://www.gutenberg.org/faq/

Project Gutenberg Europe: http://pge.rastko.net/

Project Gutenberg of Australia: https://gutenberg.org.au/

Distributed Proofreaders: https://www.pgdp.net/

Distributed Proofreaders's FAQ Central: https://www.pgdp.net/c/faq/faq_central.php

Distributed Proofreaders Europe: http://dp.rastko.net/

Project Gutenberg - Online Book Catalog: https://www.gutenberg.org/catalog/

Project Gutenberg - Advanced Search: https://www.gutenberg.org/catalog/world/search

Project Gutenberg - Top 100: https://www.gutenberg.org/browse/scores/top

Project Gutenberg - By Language: French: https://www.gutenberg.org/browse/languages/fr

Project Gutenberg Audio eBooks: https://www.gutenberg.org/audio/

Project Gutenberg - The Sheet Music Subproject: https://www.gutenberg.org/music/

Project Gutenberg - The CD and DVD Project: https://www.gutenberg.org/cdproject/

10. SHORT VERSION [DATED 2004]

MICHAEL HART: CHANGING THE WORLD THROUGH EBOOKS

[English version published by Project Gutenberg, 21 June 2004. Original version published in French by Edition-Actu, 15 February 2004.]

When Michael Hart was a student at the University of Illinois (USA), in July 1971, he set up Project Gutenberg with the goal of making available for free, and electronically, the largest possible number of books whose copyright had expired.

This ground-breaking project became both the first Internet information site and the world's first digitized library. Michael himself typed in the first hundred books. When the Internet became widely-used, in the mid-1990s, the project got a boost and an international dimension. Michael still typed and scanned in books, but now coordinated the work of dozens and then hundreds of volunteers in 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). Project Gutenberg had 5,000 books online in April 2002 and topped 10,000 in October 2003, when it had a team of 1,000 volunteers around the world making 350 new books available every month. These 10,000 books are also available on DVD for US$1 each.

Michael hopes to have a million available by 2015.

The books are digitized in "text" format, with caps for terms in italic, bold or underlined, so they can be read easily by any machine, operating system or software. Digitization is done by scanning. The book is then proofread twice by two different people, who make any corrections necessary. When the original is in poor condition, as with very old books, it is typed in manually, word by word.

Digitization in text format means a book can be copied, indexed, searched, analyzed and compared with other books. It also makes a smaller and more easily sendable computer file, unlike with scanning each page, which produces a bulky "photo" file.

Hart describes himself as a workaholic who is devoting his entire life to the project, which he sees as the start of a new Industrial Revolution. He considers himself as a pragmatic and farsighted altruist. For years he was regarded as a nut but now he is respected. He wants to change the world through freely-available e-books that can be used and copied endlessly. Reading and culture for everyone at minimal cost, on a computer or a secondhand PDA costing just a few dollars, or even on a solar-powered PDA, which are starting to appear.

In early 2004, after a stay on the US west coast, in San Francisco and Berkeley, Hart went off to Europe, first Brussels and then Paris. He gave his first lecture in France on 12 February at UNESCO headquarters in Paris, organised with APRIL (Association pour la promotion et la recherche en informatique libre / Association for Promotion and Research in Free Computing) and AFUL (Association francophone des utilisateurs de Linux et des logiciels libres / French-speaking Linux and Free Software Users' Association). He chaired a discussion at the French National Assembly on 13 February at the invitation of the discussion group "Produire et gérer les savoirs" (Producing and Managing Knowledge), a branch of the "Les temps nouveaux" (New Times) group.

What about books in French? The first digitized books were mostly in English but now there are works in 25 different languages. Of the 11,340 e-books available as of 13 February 2004, 181 were in French. The launch of Project Gutenberg Europe in the next few weeks should see the number grow considerably, and so much the better.

There is much work to be done putting all the classics of French culture online freely available to all in a easy and practical format. A total of 1,117 books are currently accessible in text format on Gallica (Bibliothèque nationale de France / French National Library), 288 on ABU (Association des bibliophiles universels / The Universal Association of Booklovers), 195 in html and/or rtf format on Athena, and several dozen more on other websites. Some digital libraries specialize in shorter material. These include the Bibliothèque électronique de Lisieux (Lisieux Electronic Library), which digitizes mostly news and articles, or Miscellanées, which calls itself a "miscellaneous" library.

PROJECT GUTENBERG: QUESTIONS AND ANSWERS

[Original version published in French by Edition-Actu, 1st March 2004.]

Since my 15 February article about Michael Hart and Project Gutenberg, which mentioned the forthcoming launch of Project Gutenberg Europe (Hart recently spoke about it to the European Parliament), I've had a lot of questions from readers. Here are some answers:

Remember Project Gutenberg is becoming international. Its main office is in the United States, but Project Gutenberg Australia and Projekt Gutenberg-DE (Germany) have been going for a long time. Project Gutenberg Europe will be European, with a staff in Belgrade and links between the different projects. I think it's interesting to build a French-language online library working with other groups. It's preparing for the future, when machine translation will be 99% satisfactory (things are progressing well on that front, though there's a lot still to do). In about 10 years, everyone will be able to call up literary classics in a choice of about 100 languages. Let's work together instead of separately, since for once it's possible.

Let's also remember that everyone working with Project Gutenberg is a volunteer, including founder Michael Hart. The goal is to ensure its future independence of loans and other funding and of fleeting political and cultural priorities, to avoid any pressure from politicians or economic interests. The aim is also to ensure respect for the volunteers, who can be confident their work will be used for many years, even generations. Donations are used only to buy equipment and supplies, mostly computers and scanners.

And then let's remember that all the books scanned in are proofread twice, by two different people, to make sure they are 99.9% accurate. Software on the website (which is still being tested) allows users to convert books in ASCII, ISO-8859, Unicode and Big-5, for example, into other formats. Conversion

will eventually be possible into still more formats, including Braille and voice. So there's no point arguing about which format is best. Text format can either be used as is or to create others. Text-format books can also be easily used by those who want to offer them in more sophisticated formats, without any restriction except for respect for copyright laws in the country involved and the availability of new free versions produced.

Some readers have asked about how volunteer proofreaders work. You go to the Distributed Proofreaders Europe website that has just been put up (and is still being tested) by Project Rastko (Belgrade) to handle the shared proofreading done by Project Gutenberg Europe. Sign up and you'll then see detailed instructions (which are still being translated in several languages). For example, passages in bold, italic or underlined, like footnotes, are always treated the same way, to standardize presentation of all the e-books. A discussion forum allows you to ask questions or seek help at any time.

Each time you go to the website, you choose the book you want. Pages of the book appear side by side in two forms – one the scanned image and the other the text produced by OCR (optical character recognition) software. You compare the two and make corrections. OCR is usually 99% accurate, which makes for about 10 corrections a page. You save each page you do and can then either stop work or do another. All the books are proofread twice (the second time only by experienced proofreaders) before the final version is ready for the public (after which any further errors noted by readers are systematically corrected).

You don't have any quota to fulfill, but it's recommended you do a page a day if possible. It doesn't seem much but with hundreds of volunteers it really adds up. In 2003, on the original site of Distributed Proofreaders, about 250-300 people were working each day, producing a daily total of 2,500-3,000 pages, the equivalent of two pages a minute.

Volunteers can also work independently, by digitizing a whole book in any word-processing programme or else scan it in 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.

[These two articles appeared in French ("Michael Hart, ou la volonté de changer le monde par le biais de l'ebook" & "Project Gutenberg: quelques réponses à vos questions") in Edition Actu nos. 90 and 91, of 15 February and 1 March 2004. Edition Actu is the electronic newsletter of CyLibris (distributed free every fortnight) which aims to look at publishing from a different angle. CyLibris, founded in Paris in August 1996 and a pioneer of online publishing, was the first French publisher to use the Internet and digitization to bring out literary works.]

Copyright © 2005 Marie Lebert

End of Project Gutenberg's Project Gutenberg (1971-2005), by Marie Lebert

*** END OF THE PROJECT GUTENBERG EBOOK PROJECT GUTENBERG (1971-2005) ***

START: FULL LICENSE
THE FULL PROJECT GUTENBERG LICENSE
PLEASE READ THIS BEFORE YOU DISTRIBUTE OR USE THIS WORK

To protect the Project Gutenberg™ mission of promoting the free distribution of electronic works, by using or distributing this work (or any other work associated in any way with the phrase "Project Gutenberg"), you agree to comply with all the terms of the Full Project Gutenberg™ License available with this file or online at www.gutenberg.org/license.

Section 1. General Terms of Use and Redistributing Project Gutenberg™ electronic works

1.A. By reading or using any part of this Project Gutenberg™ electronic work, you indicate that you have read, understand, agree to and accept all the terms of this license and intellectual property (trademark/copyright) agreement. If you do not agree to abide by all the terms of this agreement, you must cease using and return or destroy all copies of Project Gutenberg™ electronic works in your possession. If you paid a fee for obtaining a copy of or access to a Project Gutenberg™ electronic work and you do not agree to be bound by the terms of this agreement, you may obtain a refund from the person or entity to whom you paid the fee as set forth in paragraph 1.E.8.

1.B. "Project Gutenberg" is a registered trademark. It may only be used on or associated in any way with an electronic work by people who agree to be bound by the terms of this agreement. There are a few things that you can do with most Project Gutenberg™ electronic works even without complying with the full terms of this agreement. See paragraph 1.C below. There are a lot of things you can do with Project Gutenberg™ electronic works if you follow the terms of this agreement and help preserve free future access to Project Gutenberg™ electronic works. See paragraph 1.E below.

1.C. The Project Gutenberg Literary Archive Foundation ("the Foundation" or PGLAF), owns a compilation copyright in the collection of Project Gutenberg™ electronic works. Nearly all the individual works in the collection are in the public domain in the United States. If an individual work is unprotected by copyright law in the United States and you are located in the United States, we do not claim a right to prevent you from copying, distributing, performing, displaying or creating derivative works based on the work as long as all references to Project Gutenberg are removed. Of course, we hope that you will support the Project Gutenberg™ mission of promoting free access to electronic works by freely sharing Project Gutenberg™ works in compliance with the terms of this agreement for keeping the Project Gutenberg™ name associated with the work. You can easily comply with the terms of this agreement by keeping this work in the same format with its attached full Project Gutenberg™ License when you share it without charge with others.

This particular work is one of the few individual works protected by copyright law in the United States and most of the remainder of the world, included in the Project Gutenberg collection with the permission of the copyright holder. Information on the copyright owner for this particular work and the terms of use imposed by the copyright holder on this work are set forth at the beginning of this work.

1.D. The copyright laws of the place where you are located also govern what you can do with this work. Copyright laws in most countries are in a constant state of change. If you are outside the United States, check the laws of your country in addition to the terms of this agreement before downloading, copying, displaying, performing, distributing or creating derivative works based on this work or any other Project Gutenberg™ work. The Foundation makes no representations concerning the copyright status of any work in any country other than the United States.

1.E. Unless you have removed all references to Project Gutenberg:

1.E.1. The following sentence, with active links to, or other immediate access to, the full Project Gutenberg™ License must appear prominently whenever any copy of a Project Gutenberg™ work (any work on which the phrase "Project Gutenberg" appears, or with which the phrase "Project Gutenberg" is associated) is accessed, displayed, performed, viewed, copied or distributed:

This eBook is for the use of anyone anywhere in the United States and most other parts of the world at no cost and with almost no restrictions whatsoever. You may copy it, give it away or re-use it under the terms of the Project Gutenberg License included with this eBook or online at www.gutenberg.org. If you are not located in the United States, you will have to check the laws of the country where you are located before using this eBook.

1.E.2. If an individual Project Gutenberg™ electronic work is derived from texts not protected by U.S. copyright law (does not contain a notice indicating that it is posted with permission of the copyright holder), the work can be copied and distributed to anyone in the United States without paying any fees or charges. If you are redistributing or providing access to a work with the phrase "Project Gutenberg" associated with or appearing on the work, you must comply either with the requirements of paragraphs 1.E.1 through 1.E.7 or obtain permission for the use of the work and the Project Gutenberg™ trademark as set forth in paragraphs 1.E.8 or 1.E.9.

1.E.3. If an individual Project Gutenberg™ electronic work is posted with the permission of the copyright holder, your use and distribution must comply with both paragraphs 1.E.1 through 1.E.7 and any additional terms imposed by the copyright holder. Additional terms will be linked to the Project Gutenberg™ License for all works posted with the permission of the copyright holder found at the beginning of this work.

1.E.4. Do not unlink or detach or remove the full Project Gutenberg™ License terms from this work,

or any files containing a part of this work or any other work associated with Project Gutenberg™.

1.E.5. Do not copy, display, perform, distribute or redistribute this electronic work, or any part of this electronic work, without prominently displaying the sentence set forth in paragraph 1.E.1 with active links or immediate access to the full terms of the Project Gutenberg™ License.

1.E.6. You may convert to and distribute this work in any binary, compressed, marked up, nonproprietary or proprietary form, including any word processing or hypertext form. However, if you provide access to or distribute copies of a Project Gutenberg™ work in a format other than "Plain Vanilla ASCII" or other format used in the official version posted on the official Project Gutenberg™ website (www.gutenberg.org), you must, at no additional cost, fee or expense to the user, provide a copy, a means of exporting a copy, or a means of obtaining a copy upon request, of the work in its original "Plain Vanilla ASCII" or other form. Any alternate format must include the full Project Gutenberg™ License as specified in paragraph 1.E.1.

1.E.7. Do not charge a fee for access to, viewing, displaying, performing, copying or distributing any Project Gutenberg™ works unless you comply with paragraph 1.E.8 or 1.E.9.

1.E.8. You may charge a reasonable fee for copies of or providing access to or distributing Project Gutenberg™ electronic works provided that:

• You pay a royalty fee of 20% of the gross profits you derive from the use of Project Gutenberg™ works calculated using the method you already use to calculate your applicable taxes. The fee is owed to the owner of the Project Gutenberg™ trademark, but he has agreed to donate royalties under this paragraph to the Project Gutenberg Literary Archive Foundation. Royalty payments must be paid within 60 days following each date on which you prepare (or are legally required to prepare) your periodic tax returns. Royalty payments should be clearly marked as such and sent to the Project Gutenberg Literary Archive Foundation at the address specified in Section 4, "Information about donations to the Project Gutenberg Literary Archive Foundation."

• You provide a full refund of any money paid by a user who notifies you in writing (or by e-mail) within 30 days of receipt that s/he does not agree to the terms of the full Project Gutenberg™ License. You must require such a user to return or destroy all copies of the works possessed in a physical medium and discontinue all use of and all access to other copies of Project Gutenberg™ works.

• You provide, in accordance with paragraph 1.F.3, a full refund of any money paid for a work or a replacement copy, if a defect in the electronic work is discovered and reported to you within 90 days of receipt of the work.

• You comply with all other terms of this agreement for free distribution of Project Gutenberg™ works.

1.E.9. If you wish to charge a fee or distribute a Project Gutenberg™ electronic work or group of works on different terms than are set forth in this agreement, you must obtain permission in writing from the Project Gutenberg Literary Archive Foundation, the manager of the Project Gutenberg™ trademark. Contact the Foundation as set forth in Section 3 below.

1.F.

1.F.1. Project Gutenberg volunteers and employees expend considerable effort to identify, do copyright research on, transcribe and proofread works not protected by U.S. copyright law in creating the Project Gutenberg™ collection. Despite these efforts, Project Gutenberg™ electronic works, and the medium on which they may be stored, may contain "Defects," such as, but not limited to, incomplete, inaccurate or corrupt data, transcription errors, a copyright or other intellectual property infringement, a defective or damaged disk or other medium, a computer virus, or computer codes that damage or cannot be read by your equipment.

1.F.2. LIMITED WARRANTY, DISCLAIMER OF DAMAGES - Except for the "Right of Replacement or Refund" described in paragraph 1.F.3, the Project Gutenberg Literary Archive Foundation, the owner of the Project Gutenberg™ trademark, and any other party distributing a Project Gutenberg™ electronic work under this agreement, disclaim all liability to you for damages, costs and expenses, including legal fees. YOU AGREE THAT YOU HAVE NO REMEDIES FOR NEGLIGENCE, STRICT LIABILITY, BREACH OF WARRANTY OR BREACH OF CONTRACT EXCEPT THOSE PROVIDED IN PARAGRAPH 1.F.3. YOU AGREE THAT THE FOUNDATION, THE TRADEMARK OWNER, AND ANY DISTRIBUTOR UNDER THIS AGREEMENT WILL NOT BE LIABLE TO YOU FOR ACTUAL, DIRECT, INDIRECT, CONSEQUENTIAL, PUNITIVE OR INCIDENTAL DAMAGES EVEN IF YOU GIVE NOTICE OF THE POSSIBILITY OF SUCH DAMAGE.

1.F.3. LIMITED RIGHT OF REPLACEMENT OR REFUND - If you discover a defect in this electronic work within 90 days of receiving it, you can receive a refund of the money (if any) you paid for it by sending a written explanation to the person you received the work from. If you received the work on a physical medium, you must return the medium with your written explanation. The person or entity that provided you with the defective work may elect to provide a replacement copy in lieu of a refund. If you received the work electronically, the person or entity providing it to you may choose to give you a second opportunity to receive the work electronically in lieu of a refund. If the second copy is also defective, you may demand a refund in writing without further opportunities to

fix the problem.

1.F.4. Except for the limited right of replacement or refund set forth in paragraph 1.F.3, this work is provided to you 'AS-IS', WITH NO OTHER WARRANTIES OF ANY KIND, EXPRESS OR IMPLIED, INCLUDING BUT NOT LIMITED TO WARRANTIES OF MERCHANTABILITY OR FITNESS FOR ANY PURPOSE.

1.F.5. Some states do not allow disclaimers of certain implied warranties or the exclusion or limitation of certain types of damages. If any disclaimer or limitation set forth in this agreement violates the law of the state applicable to this agreement, the agreement shall be interpreted to make the maximum disclaimer or limitation permitted by the applicable state law. The invalidity or unenforceability of any provision of this agreement shall not void the remaining provisions.

1.F.6. INDEMNITY - You agree to indemnify and hold the Foundation, the trademark owner, any agent or employee of the Foundation, anyone providing copies of Project Gutenberg™ electronic works in accordance with this agreement, and any volunteers associated with the production, promotion and distribution of Project Gutenberg™ electronic works, harmless from all liability, costs and expenses, including legal fees, that arise directly or indirectly from any of the following which you do or cause to occur: (a) distribution of this or any Project Gutenberg™ work, (b) alteration, modification, or additions or deletions to any Project Gutenberg™ work, and (c) any Defect you cause.

Section 2. Information about the Mission of Project Gutenberg™

Project Gutenberg™ is synonymous with the free distribution of electronic works in formats readable by the widest variety of computers including obsolete, old, middle-aged and new computers. It exists because of the efforts of hundreds of volunteers and donations from people in all walks of life.

Volunteers and financial support to provide volunteers with the assistance they need are critical to reaching Project Gutenberg™'s goals and ensuring that the Project Gutenberg™ collection will remain freely available for generations to come. In 2001, the Project Gutenberg Literary Archive Foundation was created to provide a secure and permanent future for Project Gutenberg™ and future generations. To learn more about the Project Gutenberg Literary Archive Foundation and how your efforts and donations can help, see Sections 3 and 4 and the Foundation information page at www.gutenberg.org.

Section 3. Information about the Project Gutenberg Literary Archive Foundation

The Project Gutenberg Literary Archive Foundation is a non-profit 501(c)(3) educational corporation organized under the laws of the state of Mississippi and granted tax exempt status by the Internal Revenue Service. The Foundation's EIN or federal tax identification number is 64-6221541. Contributions to the Project Gutenberg Literary Archive Foundation are tax deductible to the full extent permitted by U.S. federal laws and your state's laws.

The Foundation's business office is located at 809 North 1500 West, Salt Lake City, UT 84116, (801) 596-1887. Email contact links and up to date contact information can be found at the Foundation's website and official page at www.gutenberg.org/contact

Section 4. Information about Donations to the Project Gutenberg Literary Archive Foundation

Project Gutenberg™ depends upon and cannot survive without widespread public support and donations to carry out its mission of increasing the number of public domain and licensed works that can be freely distributed in machine-readable form accessible by the widest array of equipment including outdated equipment. Many small donations ($1 to $5,000) are particularly important to maintaining tax exempt status with the IRS.

The Foundation is committed to complying with the laws regulating charities and charitable donations in all 50 states of the United States. Compliance requirements are not uniform and it takes a considerable effort, much paperwork and many fees to meet and keep up with these requirements. We do not solicit donations in locations where we have not received written confirmation of compliance. To SEND DONATIONS or determine the status of compliance for any particular state visit www.gutenberg.org/donate.

While we cannot and do not solicit contributions from states where we have not met the solicitation requirements, we know of no prohibition against accepting unsolicited donations from donors in such states who approach us with offers to donate.

International donations are gratefully accepted, but we cannot make any statements concerning tax treatment of donations received from outside the United States. U.S. laws alone swamp our small staff.

Please check the Project Gutenberg web pages for current donation methods and addresses. Donations are accepted in a number of other ways including checks, online payments and credit

card donations. To donate, please visit: www.gutenberg.org/donate

Section 5. General Information About Project Gutenberg™ electronic works

Professor Michael S. Hart was the originator of the Project Gutenberg™ concept of a library of electronic works that could be freely shared with anyone. For forty years, he produced and distributed Project Gutenberg™ eBooks with only a loose network of volunteer support.

Project Gutenberg™ eBooks are often created from several printed editions, all of which are confirmed as not protected by copyright in the U.S. unless a copyright notice is included. Thus, we do not necessarily keep eBooks in compliance with any particular paper edition.

Most people start at our website which has the main PG search facility: www.gutenberg.org.

This website includes information about Project Gutenberg™, including how to make donations to the Project Gutenberg Literary Archive Foundation, how to help produce our new eBooks, and how to subscribe to our email newsletter to hear about new eBooks.