

# The Project Gutenberg eBook of Le Projet Gutenberg (1971-2008)

This is a \*copyrighted\* Project Gutenberg eBook, details below.

Title: Le Projet Gutenberg (1971-2008)

Author: Marie Lebert

Release date: October 26, 2008 [eBook #27046]

Most recently updated: January 4, 2021

Language: French

\*\*\* START OF THE PROJECT GUTENBERG EBOOK LE PROJET GUTENBERG (1971-2008) \*\*\*

Produced by Al Haines

## **LE PROJET GUTENBERG (1971-2008)**

**MARIE LEBERT**

NEF, University of Toronto & Project Gutenberg, 2008

Copyright © 2008 Marie Lebert

Cet article est disponible en anglais: [Project Gutenberg \(1971-2008\)](#).

### **TABLE**

1. Sommaire
2. Un pari depuis 1971
3. La méthode adoptée
4. La correction partagée
5. Des collections multilingues
6. Domaine public versus copyright
7. Du passé vers l'avenir
8. Chronologie
9. Statistiques

## 1. SOMMAIRE

Août 1997: 1.000 livres; avril 2002: 5.000 livres; octobre 2003: 10.000 livres; janvier 2005: 15.000 livres, décembre 2006: 20.000 livres, avril 2008: 25.000 livres.

En juillet 1971, Michael Hart crée le Projet Gutenberg pour diffuser gratuitement sous forme électronique les oeuvres littéraires du domaine public. Un projet longtemps considéré par ses détracteurs comme impossible à grande échelle. Site pionnier à tous égards, le Projet Gutenberg est à la fois le premier site d'information sur un réseau encore embryonnaire et la première bibliothèque numérique. Les livres sont numérisés en mode texte (TXT) en utilisant le code ASCII (American standard code for information interchange). Lorsque l'utilisation du web se généralise, au milieu des années 1990, le projet trouve un second souffle et un rayonnement international. Les collections atteignent 1.000 livres en août 1997, 2.000 livres en mai 1999, 3.000 livres en décembre 2000, 4.000 livres en octobre 2001, 5.000 livres en avril 2002, 10.000 livres en octobre 2003, 15.000 livres en janvier 2005, 20.000 livres en décembre 2006 et 25.000 livres en avril 2008, dans 55 langues, avec 340 nouveaux livres par mois, 340 sites miroirs dans de nombreux pays, plusieurs dizaines de milliers de téléchargements par jour et des milliers de volontaires de par le monde. En octobre 2001, la principale source des livres devient Distributed Proofreaders, mis sur pied par Charles Franks pour gérer la correction partagée entre volontaires. D'abord essentiellement anglophones, le Projet Gutenberg s'internationalise et les collections deviennent multilingues. Le Projet Gutenberg Europe est lancé en janvier 2004 par le Projet Rastko, basé à Belgrade, en Serbie. Suivent le Projet Gutenberg Canada et bien d'autres.

## 2. UN PARI DEPUIS 1971

= En quelques mots

Si le livre imprimé a cinq siècles et demi, le livre numérique a tout juste 37 ans. Il est né avec le Projet Gutenberg, créé en juillet 1971 par Michael Hart pour diffuser gratuitement sous forme électronique les oeuvres littéraires du domaine public. Site pionnier à tous égards, le Projet Gutenberg est à la fois le premier site d'information sur un réseau encore embryonnaire et la première bibliothèque numérique. Longtemps considéré par ses détracteurs comme totalement irréaliste, le Projet Gutenberg compte 25.00 titres en avril 2008, avec des dizaines de milliers de téléchargements quotidiens. A ce jour, personne n'a fait mieux pour mettre les classiques de la littérature mondiale à la disposition de tous, ni pour créer à moindres frais un immense réseau de volontaires de par le monde, sans gâchis de compétences ni d'énergie.

Les vingt premières années, Michael Hart numérise lui-même les cent premiers livres, avec l'aide occasionnelle de telle ou telle personne. Lorsque l'utilisation du web se généralise au milieu des années 1990, le projet trouve un second souffle et un rayonnement international. Tout en continuant de numériser des livres, Michael Hart coordonne désormais le travail de dizaines puis de centaines de volontaires de par le monde. Les collections atteignent 1.000 livres en août 1997, 2.000 livres en mai 1999, 3.000 livres en décembre 2000 et 4.000 livres en octobre 2001.

Trente ans après ses débuts, le Projet Gutenberg fonctionne à plein régime. La barre des 5.000 livres est franchie en avril 2002, celle des 10.000 livres en octobre 2003, celle des 15.000 livres en janvier 2005, celle des 20.000 livres en décembre 2006 et celle des 25.000 livres en avril 2008. Avec 340 nouveaux livres par mois, 40 sites miroirs dans de nombreux pays, plusieurs dizaines de milliers de téléchargements par jour et des dizaines de milliers de volontaires toutes équipes confondues.

Qu'ils aient été numérisés il y a trente ans ou qu'ils soient numérisés maintenant, tous les livres sont numérisés en mode texte, en utilisant l'ASCII (American standard code for information interchange) original sur sept bits, avec des règles précises pour le formatage. Grâce à quoi les textes peuvent être lus sans problème quels que soient la machine, la plateforme et le logiciel utilisés, y compris sur un PDA ou sur une tablette de lecture. Libre ensuite à chacun de convertir les livres dans d'autres formats, après avoir vérifié que les oeuvres sont également du domaine public dans le pays concerné.

En janvier 2004, le Projet Gutenberg essaime outre-Atlantique avec la création du Projet Gutenberg Europe. A la mission originelle s'ajoute le rôle de passerelle entre les langues et les cultures, avec de nombreuses sections nationales et linguistiques. Tout en conservant la même ligne de conduite, à savoir la lecture pour tous à moindres frais, par le biais du texte électronique gratuit, indéfiniment utilisable et reproductible. Et, dans un deuxième temps, la numérisation de l'image et du son, dans le

même esprit.

= Gestation puis persévérance

Revenons aux tous débuts du projet. Alors étudiant à l'Université d'Illinois (Etats-Unis), Michael Hart se voit attribuer 100 millions de dollars de "temps machine" par le laboratoire informatique (Materials Research Lab) de son université. Le 4 juillet 1971, jour de la fête nationale, il saisit The United States Declaration of Independence (Déclaration de l'indépendance des Etats-Unis, signée le 4 juillet 1776) sur le clavier de son ordinateur. En caractères majuscules, puisque les caractères minuscules n'existent pas encore. Le texte électronique représente 5 Ko (kilo-octets). Mais l'envoi d'un fichier de 5 Ko à la centaine de personnes que représente le réseau de l'époque aurait fait implorer celui-ci, la bande passante étant infime. Michael Hart diffuse donc un message indiquant où le texte est stocké - sans lien hypertexte toutefois, puisque le web ne voit le jour que vingt ans après - suite à quoi le fichier est téléchargé par six personnes. Le Projet Gutenberg est né.

Dans la foulée, Michael Hart décide de consacrer ce crédit-temps de 100 millions de dollars à la recherche des oeuvres du domaine public disponibles en bibliothèque et à la numérisation de celles-ci. Il décide aussi de stocker les textes électroniques de la manière la plus simple possible, au format ASCII, pour que ces textes puissent être lus sans problème quels que soient la machine, la plateforme et le logiciel utilisés. Au lieu d'un ensemble de pages reliées, le livre devient un texte électronique que l'on peut dérouler en continu, avec des lettres capitales pour les termes en italique, en gras et soulignés de la version imprimée.

Peu après, il définit la mission du Projet Gutenberg: mettre à la disposition de tous, par voie électronique, le plus grand nombre possible d'oeuvres du domaine public. "Nous considérons le texte électronique comme un nouveau médium, sans véritable relation avec le papier", explique-t-il beaucoup plus tard, en août 1998. "Le seul point commun est que nous diffusons les mêmes oeuvres, mais je ne vois pas comment le papier peut concurrencer le texte électronique une fois que les gens y sont habitués, particulièrement dans les écoles."

Après avoir saisi The United States Declaration of Independence en 1971, Michael Hart poursuit ses efforts en 1972 en saisissant un texte plus long, The United States Bill of Rights (Déclaration des droits américaine). Cette Déclaration des droits comprend les dix premiers amendements ajoutés en 1789 à la Constitution des Etats-Unis (qui date elle-même de 1787), et définissant les droits individuels des citoyens et les pouvoirs respectifs du gouvernement fédéral et des Etats. En 1973, Michael Hart saisit The United States Constitution (Constitution des Etats-Unis) dans son entier.

D'année en année, la capacité de la disquette augmente régulièrement - le disque dur n'existe pas encore - si bien qu'il est possible d'envisager des fichiers de plus en plus volumineux. Michael Hart entreprend la numérisation de la Bible, composée elle-même de plusieurs livres, qui peuvent être traités séparément et occuper chacun un fichier différent. Il débute aussi la saisie des oeuvres complètes de Shakespeare, une pièce après l'autre, avec un fichier pour chaque pièce. Cette édition n'est d'ailleurs jamais mise en ligne, du fait d'une loi plus contraignante sur le copyright entrée en vigueur dans l'intervalle, et qui vise non pas le texte de Shakespeare, tombé depuis longtemps dans le domaine public, mais les commentaires et notes de cette édition. D'autres éditions annotées appartenant au domaine public sont mises en ligne quelques années après.

Parallèlement, l'internet, qui était encore embryonnaire en 1971, débute véritablement en 1974, suite à la création du protocole TCP/IP (transmission control protocol / internet protocol). En 1983, le réseau est en plein essor.

= De dix à mille livres

En août 1989, le Projet Gutenberg met en ligne son dixième texte, The King James Bible, publiée pour la première fois en 1611 et dont la version standard date de 1769. En 1990, les internautes sont au nombre de 250.000, et le standard en vigueur est la disquette de 360 Ko (kilo-octets). En janvier 1991, Michael Hart saisit Alice's Adventures in Wonderland (Alice au pays des merveilles) de Lewis Carroll (paru en 1865). En juillet de la même année, il saisit Peter Pan de James M. Barrie (paru en 1904). Ces deux classiques de la littérature enfantine tiennent chacun sur une disquette standard.

Arrive ensuite le web, opérationnel en 1991. Le premier navigateur, Mosaic, apparaît en novembre 1993. Lorsque l'utilisation du web se généralise, il devient plus facile de faire circuler les textes électroniques et de recruter des volontaires. Le Projet Gutenberg rode sa méthode de travail, avec la numérisation d'un texte par mois en 1991, deux textes par mois en 1992, quatre textes par mois en 1993 et huit textes par mois en 1994. En janvier 1994, le Projet Gutenberg fête son centième livre avec la mise en ligne de The Complete Works of William Shakespeare (Les oeuvres complètes de William Shakespeare). Shakespeare aurait écrit l'essentiel de son oeuvre entre 1590 et 1613. La production

continue ensuite d'augmenter, avec une moyenne de 8 textes par mois en 1994, 16 textes par mois en 1995 et 32 textes par mois en 1996.

Comme on le voit, entre 1991 et 1996, la production double chaque année. Tout en continuant de numériser des livres, Michael Hart coordonne désormais le travail de dizaines de volontaires. Depuis la fin 1993, le Projet Gutenberg s'articule en trois grands secteurs: a) Light Literature (littérature de divertissement), qui inclut par exemple Alice's Adventures in Wonderland, Peter Pan ou Aesop's Fables (Les Fables d'Esopé); b) Heavy Literature (littérature "sérieuse"), qui inclut par exemple La Bible, les oeuvres de Shakespeare ou Moby Dick; c) Reference Literature (littérature de référence), composée d'encyclopédies et de dictionnaires, par exemple le Roget's Thesaurus. Cette présentation en trois secteurs est abandonnée par la suite pour laisser place à un classement par rubriques plus détaillé.

Le Projet Gutenberg se veut universel, aussi bien pour les oeuvres choisies que pour le public visé, le but étant de mettre la littérature à la disposition de tous, en dépassant largement le public habituel des étudiants et des enseignants. Le secteur consacré à la littérature de divertissement est destiné à amener devant l'écran un public très divers, par exemple des enfants et leurs grands-parents recherchant le texte électronique de Peter Pan après avoir vu le film Hook, ou bien la version électronique d'Alice au pays des merveilles après avoir regardé l'adaptation filmée à la télévision, ou encore l'origine d'une citation littéraire après avoir vu un épisode de Star Trek. Pratiquement tous les épisodes de Star Trek citent des livres ayant leur correspondant numérique dans les collections du Projet Gutenberg.

L'objectif est donc que tous les publics, qu'ils soient familiers ou non avec le livre imprimé, puissent facilement retrouver des textes entendus dans des conversations, des films, des musiques, ou alors lus dans d'autres livres, journaux et magazines. Les fichiers électroniques prennent peu de place grâce à l'utilisation du format ASCII. On peut facilement les télécharger par le biais de la ligne téléphonique. La recherche textuelle est tout aussi simple. Il suffit d'utiliser la fonction "recherche" présente dans n'importe quel logiciel.

En 1997, la production est toujours de 32 titres par mois. En juin 1997, le Projet Gutenberg met en ligne The Merry Adventures of Robin Hood (Les aventures de Robin des Bois) de Howard Pyle (paru en 1883). En août 1997, il met en ligne son millièmè texte électronique, La Divina Commedia di Dante (La Divine Comédie de Dante, parue en 1321), dans sa langue d'origine, en italien.

En août 1998, Michael Hart écrit: "Mon projet est de mettre 10.000 textes électroniques sur l'internet. (Ce sera chose faite en octobre 2003, ndlr.) Si je pouvais avoir des subventions importantes, j'aimerais aller jusqu'à un million et étendre aussi le nombre de nos usagers potentiels de 1,x% à 10% de la population mondiale, ce qui représenterait la diffusion de 1.000 fois un milliard de textes électroniques au lieu d'un milliard seulement."

= De mille à dix mille livres

Entre 1998 et 2000, la moyenne est constante, avec 36 textes par mois. En mai 1999, les collections comptent 2.000 livres. Le 2.000e texte est Don Quijote (Don Quichotte) de Cervantès (paru en 1605), dans sa langue d'origine, en espagnol.

Disponible en décembre 2000, le 3.000e titre est le troisième volume de A l'ombre des jeunes filles en fleurs de Marcel Proust (paru en 1919), dans sa langue d'origine, en français. La moyenne passe à 104 livres par mois en 2001.

Mis en ligne en octobre 2001, le 4.000e texte est The French Immortals Series (La série des Immortels français), dans sa traduction anglaise. Publié en 1905 par la Maison Mazarin (Paris), ce livre rassemble plusieurs fictions d'écrivains couronnés par l'Académie française, comme Emile Souvestre, Pierre Loti, Hector Malot, Charles de Bernard, Alphonse Daudet, etc.

Disponible en avril 2002, le 5.000e texte est The Notebooks of Leonardo da Vinci (Les Carnets de Léonard de Vinci), qui datent du début du 16e siècle. Un texte qui, en 2008, se trouve toujours dans le Top 100 des livres téléchargés.

En 1988, Michael Hart choisit de numériser Alice's Adventures in Wonderland et Peter Pan parce que, dans l'un et l'autre cas, leur version numérisée tient sur la disquette standard de l'époque de 360 Ko (kilo-octets). Quinze ans plus tard, en 2002, on dispose de disquettes de 1,44 Mo (méga-octets) et on peut aisément compresser les fichiers en les zippant. Un fichier standard peut désormais comporter trois millions de caractères, plus qu'il n'en faut pour un livre de taille moyenne. Un roman de 300 pages numérisé au format ASCII représente un méga-octet. Un livre volumineux représente deux fichiers ASCII, téléchargeables tels quels ou en version zippée.

Cinquante heures environ sont nécessaires pour sélectionner un livre de taille moyenne, vérifier qu'il

est bien du domaine public, le scanner, le corriger, le formater et le mettre en page.

Quelques numéros de livres sont réservés pour l'avenir, par exemple le numéro 1984 (eBook #1984) pour le roman éponyme de George Orwell, publié en 1949, et qui est donc loin d'être tombé dans le domaine public.

En 2002, les collections s'accroissent de 203 titres par mois. Au printemps 2002, elles représentent le quart des oeuvres du domaine public en accès libre sur le web, recensées de manière pratiquement exhaustive par l'Internet Public Library (IPL). Un beau résultat dû au patient travail de milliers de volontaires actifs dans plusieurs pays.

1.000 livres en août 1997, 2.000 livres en mai 1999, 3.000 livres en décembre 2000, 4.000 livres en octobre 2001, 5.000 livres en avril 2002, 10.000 livres en octobre 2003. Le 10.000e livre est The Magna Carta, qui fut le premier texte constitutionnel anglais, signé en 1215.

Entre avril 2002 et octobre 2003, les collections doublent, passant de 5.000 à 10.000 livres en dix-huit mois. La moyenne mensuelle est de 348 livres numérisés en 2003.

Dix mille livres. Un chiffre impressionnant quand on pense à ce que cela représente de pages scannées, relues et corrigées. Cette croissance rapide est due à l'activité de Distributed Proofreaders (DP), un site conçu en 2000 par Charles Franks pour permettre la correction partagée. Les volontaires choisissent un livre en cours de traitement pour relire et corriger une page donnée. Chacun travaille à son propre rythme. A titre indicatif, le site conseille de relire une page par jour. C'est peu de temps sur une journée, et c'est beaucoup pour le projet.

En août 2003, un CD Best of Gutenberg est disponible avec une sélection de 600 livres. En décembre 2003, date à laquelle le Projet Gutenberg franchit la barre des 10.000 livres, la quasi-totalité des livres (9.400 livres) est gravée sur un DVD. CD et DVD sont envoyés gratuitement à qui en fait la demande. Libre ensuite à chacun de faire autant de copies que possible et de les distribuer autour de soi.

= De dix mille à vingt mille livres

En décembre 2003, les collections approchent les 11.000 livres. Plusieurs formats sont désormais présents, à commencer par les formats HTML, XML et RTF, le format principal (et obligatoire) restant l'ASCII. Le tout représente 46.000 fichiers, soit une capacité totale de 110 gigaoctets. Le 13 février 2004, date de la conférence de Michael Hart au siège de l'UNESCO à Paris, les collections comprennent très exactement 11.340 livres dans 25 langues différentes. En mai 2004, les 12.581 livres disponibles représentent 100.000 fichiers dans vingt formats différents, soit une capacité totale de 135 gigaoctets, destinée à doubler chaque année avec l'ajout de plus de 300 livres par mois (338 livres en 2004).

Parallèlement, le Project Gutenberg Consortia Center (PGCC), qui avait été lancé en 1997 pour rassembler des collections de livres numériques avec point d'accès unique, est officiellement affilié au Projet Gutenberg en 2003.

Par ailleurs, à l'instigation du Projet Rastko, basé à Belgrade (Serbie), les activités de Distributed Proofreaders Europe débutent en décembre 2003, et celle du Projet Gutenberg Europe en janvier 2004, avec cent livres disponibles en avril 2005. Les livres sont en plusieurs langues pour refléter la diversité linguistique prévalant en Europe, avec cent langues prévues sur le long terme.

En janvier 2005, le Projet Gutenberg fête ses 15.000 livres, avec la mise en ligne de The Life of Reason de George Santayana (paru en 1906).

En juin 2005, le nombre de livres s'élève à 16.000. Si 25 langues seulement étaient présentes en février 2004, 42 langues sont représentées en juillet 2005, dont l'iroquois, le sanscrit et les langues mayas. A la date du 27 juillet 2005, sur un total de 16.800 livres, sept langues disposent de plus de cinquante titres: l'anglais (14.548 livres), le français (577 livres), l'allemand (349 livres), le finnois (218 livres), le hollandais (130 livres), l'espagnol (103 livres) et le chinois (69 livres).

En décembre 2006, on compte 50 langues. A la date du 16 décembre 2006, Les langues comprenant plus de 50 titres sont l'anglais (17.377 livres), le français (966 titres), l'allemand (412 titres), le finnois (344 titres), le hollandais (244 titres), l'espagnol (140 titres), l'italien (102 titres), le chinois (69 titres), le portugais (68 titres) et le tagalogue (51 titres).

Lancé en août 2001, le Project Gutenberg of Australia fête ses 500 livres en juillet 2005, tandis que le Project Gutenberg Canada est en gestation, tout comme un Projet Gutenberg au Portugal et aux Philippines.

En décembre 2006, le Projet Gutenberg franchit la barre des 20.000 livres. Le 20.000e livre est un livre audio, *Twenty Thousand Leagues Under the Sea*, version anglaise de *Vingt mille lieues sous les mers* de Jules Verne (publié en 1869). Sur ces 20.000 livres, 10.000 ont été produits par Distributed Proofreaders depuis octobre 2000. La moyenne est de 345 nouveaux livres par mois en 2006. S'il a fallu 32 ans, de juillet 1971 à octobre 2003, pour numériser les 10.000 premiers livres, il n'aura fallu que trois ans et deux mois, d'octobre 2003 à décembre 2006, pour numériser les 10.000 livres suivants.

A la même date, le Project Gutenberg of Australia approche les 1.500 livres (c'est chose faite en avril 2007) et le Projet Gutenberg Europe compte 400 livres.

La section Project Gutenberg PrePrints débute en janvier 2006 pour accueillir de nouveaux documents suffisamment intéressants pour être mis en ligne, mais ne pouvant être intégrés aux collections existantes sans traitement ultérieur par des volontaires, pour diverses raisons: collections incomplètes, qualité insuffisante, conversion souhaitée dans un autre format, etc. Cette section comprend 379 titres en décembre 2006.

= De vingt mille à vingt-cinq mille livres

Le site Project Gutenberg News débute en novembre 2006 à l'instigation de Mike Cook, en tant que complément de la lettre d'information hebdomadaire et mensuelle existant déjà depuis nombre d'années. Le site offre par exemple les statistiques de production hebdomadaires, mensuelles et annuelles depuis 2001. La production hebdomadaire est de 24 livres en 2001, 47 livres en 2002, 79 livres en 2003, 78 livres en 2004, 58 livres en 2005, 80 livres en 2006 et 78 livres en 2007. La production mensuelle est de 104 livres en 2001, 203 livres en 2002, 348 livres en 2003, 338 livres en 2004, 252 livres en 2005, 345 livres en 2006 et 338 livres en 2007. La production annuelle est de 1.244 livres en 2001, 2.432 livres en 2002, 4.176 livres en 2003, 4.058 livres en 2004, 3.019 livres en 2005, 4.141 livres en 2006 et 4.049 livres en 2007.

Le Project Gutenberg Canada (PGC) voit le jour le 1er juillet 2007, le jour de la fête nationale, à l'instigation de Michael Shepard et David Jones. Distributed Proofreaders Canada (DPC) suit et la production débute en décembre 2007. Les cent premiers livres sont disponibles en mars 2008, avec des livres en anglais, en français et en italien.

Le total pour les différents sites (USA, Australie, Europe) est de 26.161 titres en 2007.

Le Project Gutenberg Consortia Center (PGCC) - dont le but est de rassembler des collections de livres déjà numérisés - comptabilise 75.000 livres en décembre 2007.

Le Projet Gutenberg envoie 15 millions de livres par voie postale en 2007, sous forme de CD et DVD. Daté de juillet 2006, le dernier DVD comprend 17.000 livres. Les images de ces CD et DVD peuvent aussi être téléchargées sur le site de BitTorrent pour gravage individuel.

Distributed Proofreaders (DP), lancé en octobre 2000, comptabilise 52.000 volontaires en janvier 2008, avec un nombre total de 11.934 livres traités en sept ans et trois mois. Distributed Proofreaders Europe (DP Europe), lancé en décembre 2003, comptabilise 1.500 volontaires. Distributed Proofreaders Canada (DPC), lancé en décembre 2007, comptabilise 250 volontaires en janvier 2008.

Le Projet Gutenberg franchit la barre des 25.000 livres en avril 2008. Le 25.000e livre est *English Book Collectors*, de William Younger Fletcher (publié en 1902). A la date du 21 avril 2008, sur un nombre total de 25.004 livres, les langues les plus représentées sont l'anglais (21.475 livres), le français (1.168 livres), l'allemand (530 livres), le finnois (433 livres), le hollandais (326 livres), le portugais (217 livres), le chinois (196 livres), l'espagnol (180 livres), l'italien (128 livres), le latin (55 livres) et le tagalogue (54 livres). Viennent ensuite l'espéranto (45 livres), le suédois (40 livres), le danois (20 livres), le catalan (19 livres), l'irlandais (10 livres), le norvégien (10 livres), le russe (7 livres), l'islandais (7 livres), le hongrois (7 livres), l'anglais ancien (6 livres), le grec (6 livres) et le bulgare (6 livres).

### **3. LA METHODE ADOPTEE**

Qu'ils aient été numérisés il y a des années ou qu'ils soient numérisés maintenant, tous les livres sont numérisés en mode texte, en utilisant l'ASCII original. Présent dès les débuts de l'informatique et dénommé Plain Vanilla ASCII, cet ASCII sur sept bits traite 128 caractères, dont 97 caractères imprimables correspondant aux touches du clavier anglais ou américain (A-Z, a-z, chiffres, ponctuation et quelques symboles). Dans le cas de langues autres que l'anglais, on utilise des extensions de l'ASCII (appelées ISO-8859 ou ISO-Latin) prenant en compte les caractères accentués. Mais, même dans ce

cas, le Projet Gutenberg propose systématiquement en complément une version ASCII sur sept bits sans accents. Sauf, bien entendu, dans le cas de langues non traduisibles en ASCII, comme le chinois, qui est encodé au format Big-5.

Dénoté à juste titre le plus petit dénominateur commun, l'ASCII sur sept bits est le seul format compatible avec 99% des machines et des logiciels, et pouvant être converti dans d'autres formats. Il sera toujours utilisé quand d'autres formats auront disparu, à commencer par les formats éphémères liés aux tablettes de lecture commercialisées depuis 1999 et déjà disparues du marché. Il est l'assurance que les collections ne deviendront jamais obsolètes, et survivront aux changements technologiques des prochaines décennies ou même des prochains siècles. Il n'existe pas d'autre standard aussi largement utilisé pour le moment, y compris l'Unicode, système d'encodage universel créé en 1991.

Le Projet Gutenberg propose toutefois certains livres dans d'autres formats, notamment dans les trois formats répandus que sont les formats HTML, XML et RTF. Des fichiers Unicode sont également présents. De plus, tout format proposé par tel ou tel volontaire - PDF, LIT, TeX et beaucoup d'autres - est généralement accepté, dans la mesure où un fichier ASCII est également présent.

Pour une conversion à grande échelle dans un format donné, le relais est passé à d'autres organismes. Par exemple Blackmask Online, qui puise dans les collections du Projet Gutenberg pour proposer des milliers de livres gratuits dans huit formats différents, tous issus du format Open eBook (OeB). Ou encore Manybooks.net, qui convertit les collections du Projet Gutenberg dans des formats lisibles sur PDA. Ou encore GutenMark, un outil permettant de reformater les livres aux formats HTML et LaTeX pour une lecture plus attractive ou bien de les reformater au format PDF pour une impression à la demande. Ou encore MobileBooks, qui propose 5.000 livres en Java pour lecture sur l'écran d'un téléphone portable. Ou encore Wattpad, un service gratuit permettant un accès immédiat aux livres du Projet Gutenberg sur son téléphone portable.

En quoi consiste exactement le travail des volontaires, une fois reçue la confirmation que le livre est bien du domaine public? Il consiste à scanner le livre page après page, ce qui donne des fichiers numérisés en mode image, puis à utiliser un logiciel OCR (optical character recognition), qui permet de convertir chaque fichier image en un fichier texte. Il consiste ensuite à relire le contenu du fichier texte au regard de l'original (image scannée ou livre imprimé) en corrigeant les erreurs, à savoir dix erreurs par page en moyenne quand le logiciel OCR est de qualité.

Le livre est relu et corrigé à deux reprises par deux personnes différentes. Les livres anciens sont parfois saisis ligne après ligne si le texte original manque de clarté. Certains volontaires préfèrent taper eux-mêmes des textes courts ou des oeuvres qu'ils aiment particulièrement. Mais les livres sont le plus souvent scannés et OCRisés, puis relus et corrigés.

Contrairement à la numérisation en mode image, qui s'arrête à l'étape du scanner, la numérisation en mode texte permet la copie du texte, l'indexation, la recherche plein texte, l'analyse textuelle, une étude comparative entre plusieurs textes, etc. On peut aussi lancer une recherche à partir de la fonction "chercher" proposée par n'importe quel programme, sans logiciel de recherche intermédiaire.

Les avantages de la numérisation en mode texte sont multiples. Les fichiers prennent peu de place et circulent d'autant plus facilement. Contrairement à d'autres formats, le téléchargement d'un livre au format texte ne requiert pas de bande passante large. Le fichier texte peut être copié à l'infini, et constituer la base de centaines de nouvelles versions numériques et imprimées, pour un coût pratiquement nul. A tout moment, on peut corriger les erreurs typographiques qui auraient pu subsister. Les lecteurs peuvent changer à volonté la taille et la police des caractères, ainsi que les marges ou le nombre de lignes par page. Le lecteur malvoyant peut grossir la taille des polices et le lecteur aveugle utiliser un logiciel de reconnaissance vocale. Tout ceci est nettement plus difficile, sinon impossible, avec nombre d'autres formats.

Si la correction par deux personnes différentes permet de mettre en ligne un texte fiable à 99,9%, le but n'est pas pour autant de créer des éditions faisant autorité, ou d'épiloguer sans fin avec un lecteur pointilleux sur le bien-fondé ou non d'un signe de ponctuation tel que deux points à la place d'un point virgule entre deux propositions.

Le Projet Gutenberg insiste régulièrement sur la nécessité de la relecture, qu'il juge essentielle. Utiliser directement des livres scannés puis convertis au format texte par un logiciel OCR, sans relecture, donne un résultat de bien moindre qualité, avec une fiabilité de 99% dans le meilleur des cas. L'étape de la relecture avec correction permet d'atteindre une fiabilité de 99,95%, un pourcentage élevé qui est aussi le standard de la Library of Congress.

Le Projet Gutenberg s'inscrit donc dans une perspective assez différente de la bibliothèque de

l'Internet Archive (qui héberge également les collections du Projet Gutenberg, en tant que deuxième site de distribution et site de sauvegarde). Dans le cas de l'Internet Archive, les livres sont scannés puis OCRisés, mais ils ne sont pas relus par des correcteurs s'attachant à traquer les erreurs. Plus rapide et moins fiable quant au résultat, la numérisation des livres sans relecture est aussi la méthode adoptée par Google, Microsoft et bien d'autres pour leurs propres bibliothèques numériques.

Le Projet Gutenberg dispose d'un moteur de recherche pour l'ensemble de ses collections, grâce à un partenariat avec Google, avec mise à jour mensuelle. Tout comme une recherche sur les métadonnées (auteur, titre, descriptif, mots-clés) grâce à un partenariat avec Yahoo!, avec mise à jour hebdomadaire. Pour la recherche avancée (Advanced Search), la recherche multicritères (auteur, titre, sujet, langue, catégorie, classification, format, numéro) inclut désormais un critère supplémentaire de recherche plein texte (Full Text), à titre expérimental.

Disponible sur le site du Projet Gutenberg, le File Recode Service permet de convertir les fichiers d'un système d'encodage (ASCII, ISO-8859, Unicode, etc.) à un autre. A l'avenir, un logiciel de conversion beaucoup plus puissant devrait permettre la conversion automatique dans bien d'autres formats (XML, HTML, PDF, TeX, RTF, BRF, etc.). Il sera également possible de choisir d'emblée la taille et la police des caractères, ainsi que le fonds d'écran. Une autre conversion très attendue est la conversion d'une langue à une autre par le biais d'un logiciel de traduction automatique. Une telle conversion pourrait être possible dans quelques années, quand ce type de logiciel aura gagné en qualité.

#### **4. LA CORRECTION PARTAGÉE**

La croissance rapide des collections depuis 2001 est due à l'activité de Distributed Proofreaders, site lancé en octobre 2000 par Charles Franks pour gérer la correction partagée entre les volontaires. A l'origine, il s'agit seulement d'intensifier la production de livres du Projet Gutenberg. Mais le succès est tel que le site devient la principale source des collections. En 2002, Distributed Proofreaders est officiellement affilié au Projet Gutenberg. En mai 2006, Distributed Proofreaders devient une entité indépendante tout en poursuivant une collaboration étroite avec le Projet Gutenberg.

Les volontaires n'ont aucun quota à respecter. A titre indicatif, il est suggéré de relire une page par jour, si possible. Cela semble peu, mais une page multipliée par des centaines de volontaires représente un chiffre considérable. La progression est rapide. En 2003, une moyenne de 250 à 300 relecteurs quotidiens permet de produire entre 2.500 et 3.000 pages par jour, ce qui représente deux pages par minute. En 2004, la moyenne est de 300 à 400 relecteurs quotidiens produisant entre 4.000 et 7.000 pages par jour, à savoir quatre pages par minute. Distributed Proofreaders comptabilise un total de 3.000 livres en février 2004, 5.000 livres en octobre 2004, 7.000 livres en mai 2005, 8.000 livres en février 2006 et 10.000 livres en mars 2007, avec une production de cinq livres par jour et 52.000 volontaires en décembre 2007.

Le 3 août 2005, le site recense 7.639 livres versés dans le Projet Gutenberg, 1.250 livres prêts à y être versés (en cours de dernière relecture et assemblage) et 831 livres en cours de correction. Le 1er mai 2008, le site recense 13.039 livres versés dans le Projet Gutenberg, 1.840 livres prêts à y être versés (en cours de dernière relecture et assemblage) et 1.000 livres en cours de correction.

Le site a pour but de permettre à plusieurs correcteurs de travailler simultanément au même livre, sur des pages différentes. Le volontaire commence par s'inscrire. Il reçoit des directives détaillées. Ces directives concernent par exemple les parties en gras, en italique et soulignées, ou les notes, qui sont toutes traitées de la même manière. Un forum permet de poser des questions et de demander de l'aide si nécessaire. Quand le volontaire se connecte au site, il sélectionne le livre de son choix à partir d'une liste donnée. Une page du livre choisi apparaît simultanément en deux versions: d'une part l'image scannée, d'autre part le texte issu de cette image, produit par un logiciel OCR. Le relecteur compare les deux versions et corrige les différences. Un logiciel OCR étant fiable à 99%, ceci représente une moyenne de dix erreurs à corriger par page. La page est ensuite sauvegardée. Le relecteur peut soit cesser le travail, soit opter pour la correction d'une autre page. Tous les livres sont relus et corrigés deux fois de suite et, la deuxième fois, uniquement par des correcteurs expérimentés. Les pages corrigées sont ensuite formatées selon des règles précises et assemblées par d'autres volontaires pour obtenir un livre numérique. Durant tout le processus, un livre donné est suivi par un responsable (project manager) qui s'assure du bon déroulement des opérations. Après la mise en forme suit la mise en ligne, avec indexation (titre, sous-titre, numéro de l'ebook et format) puis catalogage (dates de naissance et de décès de l'auteur, classification de la Library of Congress, etc.).

Les volontaires peuvent aussi travailler de manière indépendante, en s'adressant directement au Projet Gutenberg. Ils peuvent saisir leur livre préféré de bout en bout sur le traitement de texte de leur

choix. Ils peuvent aussi scanner eux-mêmes un livre, le convertir en texte par le biais d'un logiciel OCR et faire les corrections nécessaires en comparant le résultat à l'original. Dans les deux cas, une deuxième relecture est faite par une autre personne. Toute participation est bienvenue, quelle que soit la méthode adoptée. Il est tout à fait possible d'envoyer des fichiers dans d'autres formats en complément du fichier ASCII.

Aussi bien pour Distributed Proofreaders (DP) que pour Distributed Proofreaders Europe (DP Europe), de nouveaux volontaires sont bienvenus, y compris pour les livres en français. La tâche est immense. Comme indiqué sur les deux sites, "DP ne s'attend pas à un engagement inconditionnel de votre part. Corrigez des textes aussi souvent que vous le voulez, et le nombre de pages que vous voulez. Nous encourageons les gens à corriger une page par jour, mais vous êtes tout à fait libre de faire ce qui vous plaît. Nous espérons que vous vous joindrez à notre mission de préserver 'la littérature mondiale dans un format gratuit et disponible pour tous'."

## 5. DES COLLECTIONS MULTILINGUES

Qu'en est-il exactement des langues? Le Projet Gutenberg est d'abord essentiellement anglophone, puisqu'il est basé aux Etats-Unis et qu'il sert en priorité la communauté anglophone nationale et internationale. Sa vocation multilingue débute en 1997.

En octobre 1997, Michael Hart annonce son intention de produire des livres dans d'autres langues que l'anglais. Début 1998, le catalogue comprend quelques oeuvres en allemand, en espagnol, en français (dix titres), en italien et en latin. En juillet 1999, Michael Hart écrit: "J'introduis une nouvelle langue par mois maintenant, et je vais poursuivre cette politique aussi longtemps que possible."

Si 25 langues seulement sont représentées début 2004, 42 langues sont représentées en juillet 2005, dont l'iroquois, le sanscrit et les langues mayas. Outre l'anglais, six langues disposent de plus de cinquante livres: le français, l'allemand, le finnois, le hollandais, l'espagnol et le chinois. 50 langues sont représentées en décembre 2006. Les langues comprenant plus de 50 livres incluent désormais l'italien, le portugais et le tagalogue. 55 langues sont représentées en avril 2008. Le latin s'ajoute aux langues ayant plus de 50 livres. L'espéranto est en bonne place avec 45 livres, tout comme le suédois avant 40 livres.

Le français se trouve être la deuxième langue du Projet Gutenberg, après l'anglais. Le 13 février 2004, on compte seulement 181 livres en français (sur un total de 11.340 livres). Le 16 mai 2005, on compte 547 livres en français (sur un total de 15.505 livres). Soit trois fois plus en quinze mois. Le 27 juillet 2005, on compte 577 livres en français (sur un total de 16.800 livres). Le 16 décembre 2006, on compte 966 livres en français (sur un total de 19.996 livres), soit un peu moins du double en dix-huit mois. Le 21 avril 2008, on compte 1.168 livres en français (sur un total de 25.004 livres). Le mouvement devrait sensiblement s'accélérer dans quelques années, quand le Projet Gutenberg Europe tournera à plein régime.

Les premiers titres disponibles dans la langue de Molière sont six romans de Stendhal et deux romans de Jules Verne, tous mis en ligne au début de 1997. Les six romans de Stendhal sont L'Abbésse de Castro, Les Cenci, La Chartreuse de Parme, La Duchesse de Palliano, Le Rouge et le Noir et Vittoria Accoramboni, et les deux romans de Jules Verne De la terre à la lune et Le tour du monde en quatre-vingts jours. A la même date, si aucun titre de Stendhal n'est disponible en anglais, trois romans de Jules Verne le sont: 20,000 Leagues Under the Seas (Vingt mille lieues sous les mers, mis en ligne en septembre 1994), Around the World in 80 Days (Le tour du monde en quatre-vingts jours, mis en ligne en janvier 1994) et From the Earth to the Moon (De la terre à la lune, mis en ligne en septembre 1993). Stendhal et Jules Verne sont suivis par Edmond Rostand, avec Cyrano de Bergerac, mis en ligne en mars 1998.

A la fin de 1999, le Top 20, à savoir la liste des vingt auteurs les plus lus, mentionne Jules Verne à la onzième place, et Emile Zola à la seizième place. Ils sont toujours en bonne position dans le Top 100 actuel.

A titre anecdotique, le premier document illustré disponible toutes langues confondues est French Cave Paintings (Peintures des cavernes en France), mis en ligne dès avril 1995, avec une version XHTML ajoutée en novembre 2000. Il s'agit de quatre photos de peintures paléolithiques retrouvées dans une grotte de l'Ardèche (département de la région Rhône-Alpes). Ces photos sous droits ont été mises à la disposition du Projet Gutenberg par Jean Clottes, conservateur général du patrimoine, pour être largement diffusées.

En 2004, le multilinguisme devient l'une des priorités du Projet Gutenberg, tout comme l'internationalisation. Michael Hart prend son bâton de pèlerin vers l'Europe, avec des étapes à

Bruxelles, Paris et Belgrade. Le 12 février 2004, il donne une conférence au siège de l'UNESCO (Organisation des Nations Unies pour l'éducation, la science et la culture) à Paris. Le lendemain, toujours à Paris, il anime un débat à l'Assemblée nationale. La semaine suivante, il s'adresse au Parlement européen à Bruxelles. Puis il rend visite à l'équipe du Projet Rastko à Belgrade, pour soutenir la création du Projet Gutenberg Europe et de Distributed Proofreaders Europe.

Le lancement de Distributed Proofreaders Europe par le Projet Rastko en décembre 2003 représente une étape importante. Le site est calqué sur le site original de Distributed Proofreaders, pour gérer la relecture partagée du Projet Gutenberg Europe. Dès ses débuts, il s'agit d'un site multilingue, qui prend en compte les principales langues nationales. Grâce à des traducteurs volontaires, le site est disponible en douze langues en avril 2004 et vingt-deux langues en mai 2008. L'objectif à moyen terme est soixante langues - et donc soixante équipes linguistiques - avec prise en compte de toutes les langues européennes.

Quand il aura atteint sa vitesse de croisière, Distributed Proofreaders Europe devrait alimenter plusieurs bibliothèques numériques nationales et/ou linguistiques, par exemple le Projet Gutenberg France pour la France. Le but étant que chaque pays ou région ait son propre accès réseau autorisé (respectant la législation en vigueur dans le pays donné), qui sera un accès local au sein d'un réseau continental (dans le cas de la France, le réseau européen) et d'un réseau global (à l'échelle de la planète).

Quelques mots maintenant sur le Projet Rastko, qui s'est porté volontaire pour un pari aussi fou, catalysant du même coup les bonnes volontés européennes à l'est comme à l'ouest. Fondé en 1997, le Projet Rastko est une initiative non gouvernementale à vocation culturelle et pédagogique. L'un de ses objectifs est la mise en ligne de la culture serbe. Il fait partie de la Balkans Cultural Network Initiative, un réseau culturel régional couvrant la péninsule des Balkans, située au sud-est de l'Europe.

En mai 2005, Distributed Proofreaders Europe compte cent livres numérisés par ses soins, avec mise en ligne de ces cent livres le mois suivant sur le site du Projet Gutenberg Europe. En décembre 2006, DP Europe comptabilise 400 livres numérisés. La règle utilisée pour définir le domaine public est l'équation "décès de l'auteur + 50 ans", qui correspond à la législation en vigueur en Serbie. DP Europe utilise l'Unicode pour pouvoir traiter des livres dans un grand nombre de langues. Créé en 1991 et largement répandu à partir de 1998, l'Unicode est un système d'encodage qui attribue un code unique à chaque caractère pour être en mesure de traiter toutes les langues, contrairement à l'ASCII qui ne peut traiter que l'anglais et quelques langues européennes.

Le 3 août 2005, le site recense 135 livres versés dans le Projet Gutenberg Europe, 418 livres prêts à y être versés (en cours de dernière relecture et assemblage) et 125 livres en cours de correction. Le 10 mai 2008, le site recense 496 livres versés dans le Projet Gutenberg Europe, 653 livres prêts à y être versés (en cours de dernière relecture et assemblage) et 91 livres en cours de correction.

## **6. DOMAINE PUBLIC VERSUS COPYRIGHT**

Chose inquiétante à l'heure d'une société dite de l'information, le domaine public se réduit comme peau de chagrin. A une époque qui n'est pas si lointaine, 50% des oeuvres appartenaient au domaine public, et pouvaient donc être librement utilisées par tous. Mais le domaine public se dégrade régulièrement au fil des siècles, particulièrement au 20e siècle. D'ici 2100, 99% des oeuvres devraient être régies par le droit d'auteur, avec un maigre 1% laissé au domaine public. Un problème épineux pour tous ceux qui gèrent des bibliothèques numériques. La tâche du Projet Gutenberg ne se trouve donc guère facilitée par les coups de boutoir portés au domaine public. Michael Hart se penche sur la question depuis quelque trente ans, avec l'aide d'un groupe d'avocats spécialisés dans le droit d'auteur.

Dans la section Copyright HowTo, le Projet Gutenberg détaille les calculs à faire pour déterminer si un titre publié aux Etats-Unis appartient ou non au domaine public. Les oeuvres publiées avant 1923 sont soumises au droit d'auteur pendant 75 ans à partir de leur date de publication (elles sont donc maintenant du domaine public). Les oeuvres publiées entre 1923 et 1977 sont soumises au droit d'auteur pendant 95 ans à partir de leur date de publication (rien ne tombera dans le domaine public avant 2019). Une oeuvre publiée en 1998 et les années suivantes est soumise au droit d'auteur pendant 70 ans à partir de la date du décès de l'auteur s'il s'agit d'un auteur personnel (rien dans le domaine public avant 2049), ou alors pendant 95 ans à partir de la date de publication - ou 120 ans à partir de la date de création - s'il s'agit d'un auteur collectif (rien dans le domaine public avant 2074). Tout ceci dans les grandes lignes. D'autres règles viennent s'ajouter à ces règles de base, et la loi sur le copyright est retouchée onze fois au cours des quarante dernières années.

Nettement plus contraignante que la précédente, la législation actuelle est entérinée par le Congrès

le 27 octobre 1998 pour contrer le formidable véhicule de diffusion qu'est l'internet. "Le copyright a été augmenté de 20 ans, explique Michael Hart en juillet 1999. Auparavant on devait attendre 75 ans, on est maintenant passé à 95 ans. Bien avant, le copyright durait 28 ans (plus une extension de 28 ans si on la demandait avant l'expiration du délai) et, avant cela, le copyright durait 14 ans (plus une extension de 14 ans si on la demandait avant l'expiration du délai). Comme on le voit, on assiste à une dégradation régulière et constante du domaine public."

Au fil des siècles, chaque avancée technique est accompagnée d'un durcissement du copyright, qui semble être la réponse des éditeurs à un accès plus facile au savoir, et la peur afférente de perdre des royalties. Ce fut le cas pour l'apparition des nouvelles presses à vapeur, puis l'apparition des nouvelles presses électriques, puis l'apparition de la photocopieuse lancée par Xerox, et enfin l'apparition des technologies numériques permettant la diffusion de centaines de milliers d'oeuvres désormais disponibles sur CD et DVD et sur le web, gratuitement ou à un prix très bas.

Pour ne prendre qu'un exemple, le classique mondial *Autant en emporte le vent* (*Gone With the Wind*), publié en 1939, aurait dû tomber dans le domaine public au bout de 56 ans, en 1995, conformément à la législation de l'époque, libérant ainsi les droits pour les adaptations en tous genres. Suite aux législations de 1976 et 1998, ce classique ne devrait désormais tomber dans le domaine public qu'en 2035.

La législation de 1998 porte un coup très rude aux bibliothèques numériques, en plein essor avec le développement du web, et scandalisent ceux qui les gèrent, à commencer par Michael Hart, fondateur du *Projet Gutenberg* en 1971, et John Mark Ockerbloom, fondateur de *l'Online Books Page* en 1993. Nombre de titres doivent être retirés des collections.

Mais comment faire le poids vis-à-vis des majors de l'édition? Michael Hart raconte en juillet 1999: "J'ai été le principal opposant aux extensions du copyright, mais Hollywood et les grands éditeurs ont fait en sorte que le Congrès ne mentionne pas mon action en public. Les débats actuels sont totalement irréalistes. Ils sont menés par 'l'aristocratie terrienne de l'âge de l'information' et servent uniquement ses intérêts. Un âge de l'information? Et pour qui?"

John Mark Ockerbloom ajoute en août 1999: "Il est important que les internautes comprennent que le copyright est un contrat social conçu pour le bien public - incluant à la fois les auteurs et les lecteurs. Ceci signifie que les auteurs devraient avoir le droit d'utiliser de manière exclusive et pour un temps limité les oeuvres qu'ils ont créées, comme ceci est spécifié dans la loi actuelle sur le copyright. Mais ceci signifie également que leurs lecteurs ont le droit de copier et de réutiliser ce travail autant qu'ils le veulent à l'expiration de ce copyright. Aux Etats-Unis, on voit maintenant diverses tentatives visant à retirer ces droits aux lecteurs, en limitant les règles relatives à l'utilisation de ces oeuvres, en prolongeant la durée du copyright (y compris avec certaines propositions visant à le rendre permanent) et en étendant la propriété intellectuelle à des travaux distincts des oeuvres de création (comme on en trouve dans les propositions de copyright pour les bases de données). Il existe même des propositions visant à entièrement remplacer la loi sur le copyright par une loi instituant un contrat beaucoup plus lourd."

Les instances politiques ne cessent de parler d'âge de l'information alors que, en parallèle, elles durcissent la réglementation relative à la mise à disposition de cette information. La contradiction est flagrante. Le copyright est passé d'une durée de 30 ans en moyenne en 1909 à une durée de 95 ans en moyenne en 1998, explique Michael Hart sur son blog. En 89 ans, de 1909 à 1998, le copyright a subi une extension de 65 ans qui affecte les trois quarts de la production du 20e siècle. Seul un livre publié avant 1923 peut être considéré avec certitude comme du domaine public. Un durcissement similaire touche les pays de l'Union européenne. La règle générale est désormais un copyright de 70 ans après le décès de l'auteur, alors qu'il était auparavant de 50 ans. Ceci suite aux pressions exercées par les éditeurs de contenu, sous le prétexte d'harmoniser les lois nationales relatives au copyright pour répondre à la mondialisation du marché.

Pour éviter les poursuites judiciaires, le *Projet Gutenberg* se lance régulièrement dans des recherches pouvant durer des années pour déterminer si tel ou tel livre est du domaine public ou non. A ce jour, il peut s'enorgueillir d'un résultat exact à 99,9% pour l'ensemble de ses collections, ce pourcentage signifiant que 20 livres seulement sur les 20.000 livres que comptent les collections (en décembre 2006) pourraient poser problème. Sur le site du *Project Gutenberg Consortia Center* (PGCC), Michael Hart raconte que la version originale en tchèque de *Universal Robots* de Rossum lui a demandé des années de recherches dans le monde entier pour avoir l'assurance que cette version était du domaine public aux Etats-Unis. Quant au discours *I Have a Dream* de Martin Luther King, tombé pendant un temps court dans le domaine public, il a dû être retiré des collections suite à une législation plus contraignante sur le copyright.

Une lueur d'espoir existe toutefois pour les livres parus après 1923. D'après Greg Newby, directeur

de la PFLAF (Project Gutenberg Literary Archive Foundation), un million de livres publiés aux Etats-Unis entre 1923 et 1963 appartiendrait en fait au domaine public, puisque seuls 10% des copyrights sont effectivement renouvelés - avec renouvellement demandé entre 1950 et 1993. Les livres dont le copyright n'a pas été renouvelé peuvent donc légalement intégrer les collections du Projet Gutenberg. Pourquoi la période 1923-1963? Parce que les livres parus avant le 1er janvier 1923 sont du domaine public et que les livres parus à compter du 1er janvier 1964 ont vu leur copyright automatiquement renouvelé suite à l'adoption du 1976 Copyright Act.

Pour un titre donné, comment savoir si le copyright a été renouvelé ou non? Pour les livres aux copyrights renouvelés en 1978 et après, on dispose de la base de données en ligne du US Copyright Office. Pour les livres aux copyrights renouvelés entre 1950 et 1977, on ne disposait que des publications imprimées bisannuelles (deux fois par an) du même Copyright Office. En 2004, ces listes sont numérisées par Distributed Proofreaders et mises en ligne sur le site du Projet Gutenberg. Si un livre publié entre 1923 et 1963 ne figure sur aucune de ces listes, cela signifie que son copyright n'a pas été renouvelé, qu'il est tombé dans le domaine public et qu'on peut donc le traiter. En avril 2007, l'Université de Stanford (Californie) convertit les listes numérisées du Projet Gutenberg en base de données - la Copyright Renewal Database - avec recherche possible par titre, auteur, date du copyright et date de renouvellement du copyright.

## **7. DU PASSE VERS L'AVENIR**

Le pari fait par Michael Hart en 1971 est donc réussi. Le Projet Gutenberg compte 10 livres en août 1989, 100 livres en janvier 1994, 1.000 livres en août 1997, 2.000 livres en mai 1999, 3.000 livres en décembre 2000, 4.000 livres en octobre 2001, 5.000 livres en avril 2002, 10.000 livres en octobre 2003, 15.000 livres en janvier 2005, 20.000 livres en décembre 2006 et 25.000 livres en avril 2008.

Mais les résultats du Projet Gutenberg ne se mesurent pas seulement à ces chiffres, qui restent assez modestes par rapport à la production imprimée. Les résultats se mesurent également à l'influence du projet, qui est considérable. Premier site d'information sur l'internet et première bibliothèque numérique, le Projet Gutenberg a inspiré bien d'autres bibliothèques numériques depuis, par exemple le Projekt Runeberg pour la littérature scandinave ou le Projekt Gutenberg-DE pour la littérature allemande, pour n'en citer que deux. Fondé en décembre 1992 par Lysator, un club informatique d'étudiants, en collaboration avec la Linköping University Library (Suède), le Projekt Runeberg regroupe 200 oeuvres appartenant à la littérature nordique. Créé en 1994, le Projekt Gutenberg-DE - désormais hébergé sur le site de l'hebdomadaire Der Spiegel - comprend 200 titres de littérature allemande et de littérature étrangère en allemand.

La structure administrative et financière du Projet Gutenberg se limite au strict minimum, avec une devise qui tient en trois mots: «Less is more.» Michael Hart insiste régulièrement sur la nécessité d'un cadre aussi souple que possible laissant toute initiative aux volontaires, et la porte grande ouverte aux idées nouvelles. Le but est d'assurer la pérennité du projet indépendamment des crédits, des coupures de crédits et des priorités politiques et culturelles du moment. Pas de pression possible donc par le pouvoir et par l'argent. Et respect à l'égard des volontaires, qui sont assurés de voir leur travail utilisé pendant de nombreuses années, si ce n'est pour plusieurs générations, d'où l'intérêt d'un format numérique qui soit toujours valable dans quelques siècles. Le suivi régulier du projet est assuré grâce à une lettre d'information hebdomadaire et mensuelle, des forums de discussion, des wikis et des blogs.

Les dons servent à financer des ordinateurs et des scanners, et à envoyer des CD et DVD gratuits à ceux qui en font la demande. Suite au CD Best of Gutenberg disponible en août 2003 avec une sélection de 600 titres et à un premier DVD disponible en décembre 2003 avec 9.400 titres (sur les 10.000 titres que comprennent les collections), un deuxième DVD est disponible en juillet 2006 avec 17.000 titres (sur les 19.000 titres que comprennent désormais les collections). En 2005, CD et DVD sont disponibles sous forme d'images ISO sur le site de BitTorrent, ces images pouvant être téléchargées pour graver des CD et DVD sur place à titre personnel. En 2007, le Projet Gutenberg envoie 15 millions de livres par voie postale sous forme de CD et DVD.

Chose souvent passée sous silence, Michael Hart est le véritable inventeur de l'ebook. Si on considère l'ebook dans son sens étymologique, à savoir un livre numérisé pour diffusion sous forme de fichier électronique, celui-ci aurait 37 ans et serait né avec le Projet Gutenberg en juillet 1971. Une paternité beaucoup plus réconfortante que les divers lancements commerciaux dans un format propriétaire ayant émaillé le début des années 2000. Il n'y a aucune raison pour que la dénomination "ebook" ne désigne que l'ebook commercial et soit réservée aux Amazon, Barnes & Noble, et autres. L'ebook non commercial est un ebook à part entière, et non un parent pauvre, tout comme l'édition électronique non commerciale est une forme d'édition à part entière, et tout aussi valable que l'édition commerciale. En 2003, les etexts du Projet Gutenberg deviennent des ebooks, pour coller à la

terminologie ambiante.

En juillet 1971, l'envoi d'un fichier de 5 Ko (kilo-octets) à cent personnes aurait fait sauter l'embryon de réseau disponible à l'époque. En novembre 2002, le Projet Gutenberg peut mettre en ligne les 75 fichiers du Human Genome Project (Le séquençage du génome humain), chaque fichier se chiffrant en dizaines sinon en centaines de méga-octets. Ceci peu de temps après sa parution initiale en février 2001, puisqu'il appartient d'emblée au domaine public. En 2004, la capacité de stockage des disques durs est telle qu'il serait possible de faire tenir l'intégralité de la Library of Congress sur un support de stockage coûtant 140 dollars US. Et quelques années seulement nous sépareront d'une clé USB permettant de stocker l'intégralité du patrimoine écrit de l'humanité.

Qu'en est-il des documents autres que l'écrit? En septembre 2003, le Projet Gutenberg se lance dans la diffusion de livres audio. En décembre 2006, on compte 367 livres lus par une synthèse vocale (Audio Book, computer-generated) et 132 livres lus par l'être humain (Audio Book, human-read). Le nombre de ces derniers devrait fortement augmenter dans un proche avenir. Ils sont au nombre de 412 en mai 2008. Par contre, les livres lus par une synthèse vocale ne sont plus être stockés dans une section spécifique, mais réalisés à la demande à partir des fichiers électroniques existant dans les collections générales. Les lecteurs aveugles ou malvoyants pourront utiliser une commande vocale pour demander le texte de tel ou tel livre.

Lancée à la même époque, la section The Sheet Music Subproject est consacrée aux partitions musicales numérisées (Music, Sheet). Elle est complétée par une section d'enregistrements musicaux (Music, recorded). Des sections sont également disponibles pour les images fixes (Pictures, still) et animées (Pictures, moving). Ces nouvelles collections devraient être développées dans les prochaines années.

Mais la numérisation des livres reste prioritaire. Et la demande est énorme. En témoigne le nombre de téléchargements, qui se comptent désormais en dizaines de milliers par jour. A la date du 31 juillet 2005, on compte 37.532 fichiers téléchargés dans la journée, 243.808 fichiers téléchargés dans la semaine et 1.154.765 fichiers téléchargés dans le mois. A la date du 6 mai 2007, on compte 89.841 fichiers téléchargés dans la journée, 697.818 fichiers téléchargés dans la semaine et 2.995.436 fichiers téléchargés dans le mois. Courant mai, ce nombre atteint les 3 millions. A la date du 8 mai 2008, on compte 115.138 fichiers téléchargés dans la journée, 714.323 fichiers téléchargés dans la semaine et 3.055.327 fichiers téléchargés dans le mois. Ceci uniquement pour le principal site de téléchargement, [ibiblio.org](http://ibiblio.org) (basé à l'Université de Caroline du Nord, Chapel Hill, Etats-Unis), qui héberge aussi le site du Projet Gutenberg. Le deuxième site de téléchargement est l'Internet Archive, qui est aussi le site de sauvegarde et qui met à la disposition du Projet Gutenberg une capacité de stockage illimitée.

Le Projet Gutenberg dispose de 40 sites miroirs répartis dans de nombreux pays, et il en cherche d'autres. La circulation des fichiers se fait aussi en mode P2P (peer-to-peer), qui permet d'échanger des fichiers directement d'un utilisateur à l'autre.

Un Top 100 recense les cent titres et les cent auteurs les plus téléchargés dans la journée, dans la semaine et dans le mois.

Les livres du Projet Gutenberg peuvent aider à combler la fracture numérique. Ils sont aisément téléchargeables sur PDA. Un ordinateur ou un PDA d'occasion ne coûte que quelques dollars ou quelques dizaines de dollars, en fonction du modèle. Certains PDA fonctionnent à l'énergie solaire, permettant la lecture dans les régions reculées et les pays en développement.

Plus tard, il sera peut-être possible d'envisager une traduction simultanée dans une centaine de langues. En utilisant la traduction automatique qui, d'ici là, pourrait avoir atteint un taux de fiabilité de l'ordre de 99%, un pourcentage dont on est encore loin. En 2004, le Projet Gutenberg était en lien avec un projet européen envisageant un logiciel de traduction automatique relayé par des traducteurs (non pas des machines, mais des êtres humains), sur un modèle comparable à la technologie OCR relayée par des correcteurs (non pas des logiciels, mais des êtres humains) pour offrir un contenu de grande qualité.

37 ans après les débuts du Projet Gutenberg, Michael Hart se définit toujours comme un fou de travail dédiant toute sa vie à son projet, qu'il voit comme étant à l'origine d'une révolution néo-industrielle. Il se définit aussi comme altruiste, pragmatique et visionnaire. Après avoir été traité de toqué pendant de nombreuses années, il force maintenant le respect. Au fil des ans, la mission du Projet Gutenberg reste la même, à savoir changer le monde par le biais de l'ebook gratuit indéfiniment utilisable et reproductible. L'objectif reste lui aussi le même, à savoir la lecture et la culture pour tous à moindres frais. Quant à la mission, elle se résume en quelques mots: "encourager la création et la distribution d'ebooks", par autant de personnes que possible, et par tous les moyens. Tout en prenant les virages nécessaires pour intégrer de nouvelles idées, de nouvelles méthodes et de nouveaux

supports.

D'après Michael Hart, le patrimoine écrit de l'humanité représenterait 25 millions de livres appartenant au domaine public, qui pourraient être collectés auprès des grandes bibliothèques nationales et régionales, à raison d'un exemplaire par livre, sans tenir compte des nombreuses éditions annotées et commentées. Si Gutenberg a permis à chacun d'avoir des livres grâce à l'invention de l'imprimerie, le Projet Gutenberg permet à chacun d'avoir une bibliothèque de livres grâce au stockage de ceux-ci sur un support numérique tenant dans un sac sinon dans une poche. En avril 2008, le Projet Gutenberg permet d'ores et déjà à chacun d'avoir une bibliothèque numérique de 25.000 livres au contenu parfaitement fiable.

Laissons le mot de la fin à Michael Hart, à qui je demandais en août 1998 quel était son meilleur souvenir. A l'époque, il répondait: "Le courrier que je reçois me montre combien les gens apprécient que j'aie passé ma vie à mettre des livres sur l'internet. Certaines lettres sont vraiment émouvantes, et elles me rendent heureux pour toute la journée." Dix ans après, il confirme que sa réponse serait toujours la même.

## 8. CHRONOLOGIE

[\*1971/07 = année/mois]

1971/07: Saisie par Michael Hart de The United States Declaration of Independence (ebook #1) et diffusion d'un message auprès des cent premiers usagers du réseau. Le Projet Gutenberg est né.

1972: Saisie de The United States Bill of Rights (eBook #2).

1973: Saisie de The United States Constitution (eBook #5).

1974-88: Saisie de la Bible et de plusieurs pièces de Shakespeare.

1989/08: The King James Bible (eBook #10).

1991/01: Alice's Adventures in Wonderland (eBook #11).

1991/06: Peter Pan (eBook #16).

1991: Numérisation d'un livre par mois.

1992: Numérisation de deux livres par mois.

1993: Numérisation de quatre livres par mois.

1993/12: Constitution de trois grands secteurs: Light Literature, Heavy Literature, Reference Literature.

1994: Numérisation de huit livres par mois.

1994/01: The Complete Works of William Shakespeare (eBook #100).

1995: Numérisation de 16 livres par mois.

1996-97: Numérisation de 32 livres par mois.

1997/08: La Divina Commedia di Dante, en italien (eBook #1000).

1997: Lancement du Project Gutenberg Consortia Center (PGCC).

1998-2000: Numérisation de 36 livres par mois.

1999/05: Don Quijote, de Cervantès, en espagnol (eBook #2000).

2000: Création de la Project Gutenberg Literary Archive Foundation (PGLAF).

2000/10: Lancement de Distributed Proofreaders par Charles Franks pour permettre la correction partagée.

2000/12: A l'ombre des jeunes filles en fleurs (vol. 3), de Proust, en français (eBook #3000).

2001/08: Création du Project Gutenberg of Australia.

2001/10: The French Immortals Series, en anglais (eBook #4000).

2001: Numérisation de 104 livres par mois.

2002: Affiliation officielle de Distributed Proofreaders au Projet Gutenberg.

2002/04: The Notebooks of Leonardo da Vinci, en anglais (eBook #5000).

2002: Numérisation de 203 livres par mois.

2003/08: Edition d'un CD «Best of Gutenberg» contenant 600 livres.

2003/09: Lancement de la section Project Gutenberg Audio eBooks.

2003/10: Les collections doublent en dix-huit mois, passant de 5.000 à 10.000 livres.

2003/10: The Magna Carta (eBook # 10000).

2003/12: Edition du premier DVD, qui contient 9.400 livres.

2003: Numérisation de 348 livres par mois.

2003: Affiliation officielle du Project Gutenberg Consortia Center (PGCC) au Projet Gutenberg.

2003/12: Lancement de Distributed Proofreaders Europe (DP-EU) par le Projet Rastko.

2004/01: Lancement du Projet Gutenberg Europe par le Projet Rastko.

2004/02: Voyage de Michael Hart en Europe (Paris, Bruxelles, Belgrade).

2004/02: Conférence de Michael Hart au siège de l'UNESCO, à Paris.

2004/02: Visite de Michael Hart au Parlement européen, à Bruxelles.

2004/10: 5.000 livres produits par Distributed Proofreaders.

2004: Numérisation de 338 livres par mois.

2005/01: The Life of Reason, par George Santayana (eBook #15000).

2005/05: 7.000 livres produits par Distributed Proofreaders.

2005/05: 100 premiers livres produits par Distributed Proofreaders Europe.

2005/06: Le Projet Gutenberg compte 16.000 livres.

2005/06: Le Projet Gutenberg Europe compte 100 livres.

2005/07: Le Project Gutenberg of Australia compte 500 livres.

2005/10: 5e anniversaire de Distributed Proofreaders.

2005: Numérisation de 252 livres par mois.

2006/01: Lancement de la section Project Gutenberg PrePrints.

2006/02: 8.000 livres produits par Distributed Proofreaders.

2006/05: Création de la Distributed Proofreaders Foundation.

2006/07: 35e anniversaire du Projet Gutenberg.

2006/07: Edition d'un nouveau DVD, qui contient 17.000 livres.

2006/11: Lancement du site Project Gutenberg News.

2006/12: Le Projet Gutenberg compte 20.000 livres.

2006/12: Twenty Thousand Leagues Under the Sea, par Jules Verne, audiobook en anglais (eBook #20000).

2006/12: 400 livres produits par Distributed Proofreaders Europe.

2006: Numérisation de 345 livres par mois.

2007/03: 10.000 livres produits par Distributed Proofreaders.

2007/04: Le Project Gutenberg of Australia compte 1.500 livres.

2007/07: Création de Project Gutenberg Canada (PGC).

2007/12: Lancement de Distributed Proofreaders Canada (DPC).

2007: Numérisation de 338 livres par mois.

2008/03: Le Project Gutenberg of Canada compte 100 livres.

2008/04: Le Projet Gutenberg compte 25.000 livres.

2008/04: English Book Collectors, par William Younger Fletcher (eBook #25000).

2008/05: Le Projet Gutenberg Europe compte 500 livres.

## 9. STATISTIQUES

\*Ces statistiques concernent uniquement le site original du Projet Gutenberg. Des statistiques relatives aux autres sites du Projet Gutenberg (Australie, Europe, Canada) sont disponibles dans Project Gutenberg News.

= Quelques chiffres

1.000 livres en août 1997.

2.000 livres en mai 1999.

3.000 livres en décembre 2000.

4.000 livres en octobre 2001.

5.000 livres en avril 2002.

10.000 livres en octobre 2003.

15.000 livres en janvier 2005.

20.000 livres en décembre 2006.

25.000 livres en avril 2008.

= Nouveaux livres: moyennes annuelles

2001: 1.244 livres par an.

2002: 2.432 livres par an.

2003: 4.176 livres par an.

2004: 4.058 livres par an.

2005: 3.019 livres par an.

2006: 4.141 livres par an.

2007: 4.049 livres par an.

= Nouveaux livres: moyennes mensuelles

2001: 104 livres par mois.

2002: 203 livres par mois.

2003: 348 livres par mois.

2004: 338 livres par mois.

2005: 252 livres par mois.

2006: 345 livres par mois.

2007: 338 livres par mois.

= Nouveaux livres: moyennes hebdomadaires

2001: 24 livres par semaine.

2002: 47 livres par semaine.

2003: 79 livres par semaine.

2004: 78 livres par semaine.

2005: 58 livres par semaine.

2006: 80 livres par semaine.

2007: 78 livres par semaine.

= Quelques eBooks

eBook #1: The United States Declaration of Independence (1776) [en ligne en juillet 1971].

eBook #2: The United States Bill of Rights (1789) [en ligne en 1972].

eBook #5: The United States Constitution (1787) [en ligne en 1973].

eBook #10: The King James Bible (1769) [en ligne en août 1989].

eBook #11: Alice's Adventures in Wonderland, de Lewis Carroll (1865) [en ligne en janvier 1991].

eBook #16: Peter Pan, de James Barrie (1904) [en ligne en juin 1991].

eBook #100: The Complete Works of William Shakespeare (1590-1613) [en ligne en janvier 1994].

eBook #1000: La Divina Commedia di Dante (1321, en italien) [en ligne en août 1997].

eBook #2000: Don Quichote, de Cervantès (1605, en espagnol) [en ligne en mai 1999].

eBook #3000: A l'ombre des jeunes filles en fleurs, vol. 3, de Marcel Proust (1919) [en ligne en décembre 2000].

eBook #4000: The French Immortals Series (1905) [en ligne en octobre 2001].

eBook #5000: The Notebooks of Leonardo da Vinci (début 16e s.) [en ligne en avril 2002].

eBook #10000: The Magna Carta (début 13e s.) [en ligne en octobre 2003].

eBook #15000: The Life of Reason, de George Santayana (1906) [en ligne en janvier 2005].

eBook #20000: Twenty Thousand Leagues Under the Sea, de Jules Verne (1869), livre audio [en ligne en décembre 2006].

eBook #25000: English Book Collectors, de William Younger Fletcher (1902) [en ligne en avril 2008].

= Nombre de langues avec 50+ livres

Janvier 2004: 25 langues.

Juillet 2005: 42 langues.

Décembre 2006: 50 langues.

Avril 2008: 55 langues.

= Langues majeures

Juillet 2005: anglais, français, allemand, finnois, hollandais, espagnol, chinois. [Sur un total de 16.800 livres le 27 juillet 2005, 14.548 livres sont en anglais, 577 livres en français, 349 livres en allemand, 218 livres en finnois, 130 livres en hollandais, 103 livres en espagnol et 69 livres en chinois.]

Décembre 2006: anglais, français, allemand, finnois, hollandais, espagnol, italien, chinois, portugais, tagalogue. [Sur un total de 19.996 livres le 16 décembre 2006, 17.377 livres sont en anglais, 966 livres en français, 412 livres en allemand, 344 livres en finnois, 244 livres en hollandais, 140 livres en espagnol, 102 livres en italien, 69 livres en chinois, 68 livres en portugais et 51 livres en tagalogue.]

Avril 2008: anglais, français, allemand, finnois, hollandais, portugais, chinois, espagnol, italien, latin, tagalogue. [Sur un total de 25,004 livres le 21 avril 2008, 21.475 livres en anglais, 1.168 livres en français, 530 livres en allemand, 433 livres en finnois, 326 livres en hollandais, 217 livres en portugais, 196 livres en chinois, 180 livres en espagnol, 128 livres en italien, 55 livres en latin et 54 livres en tagalogue.]

= Téléchargements sur [ibiblio.org](http://ibiblio.org)

\*[ibiblio.org](http://ibiblio.org) est le principal site de téléchargement. Les téléchargements des 40 sites miroirs ne sont pas comptabilisés ici.

31 juillet 2005: 37.532 fichiers téléchargés dans la journée, 243.808 fichiers téléchargés dans la semaine, 1.154.765 fichiers téléchargés dans le mois.

6 mai 2007: 89.841 fichiers téléchargés dans la journée, 697.818 fichiers téléchargés dans la semaine, 2.995.436 fichiers téléchargés dans le mois.

8 mai 2008: 115.138 fichiers téléchargés dans la journée, 714.323 fichiers téléchargés dans la semaine, 3.055.327 fichiers téléchargés dans le mois.

## 10. LIENS

Distributed Proofreaders (DP): <https://www.pgdp.net/>

Distributed Proofreaders Canada (DPC): <http://www.pgdpcanada.net/>

Distributed Proofreaders Europe (DP Europe): <http://dp.rastko.net/>

Hart, Michael (blog): <http://hart.pglaf.org/>

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

Project Gutenberg / Catalog: <https://www.gutenberg.org/catalog/>

Project Gutenberg / File Recode Service: <https://www.gutenberg.org/catalog/world/recode.php>

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

Project Gutenberg Canada (PGC): <http://www.gutenberg.ca/>

Project Gutenberg Consortia Center (PGCC): <http://www.gutenberg.us/>

Project Gutenberg Europe (PGE): <http://pge.rastko.net/>

Project Gutenberg Literary Archive Foundation (PGLAF): <https://www.pglaf.org/>

Project Gutenberg News (PG News): <http://www.pg-news.org/>

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

Project Gutenberg of the Philippines (PGPH): <http://www.gutenberg.ph/>

Project Gutenberg PrePrints: <http://preprints.readingroo.ms/>

Projekt Gutenberg-DE: <http://gutenberg.spiegel.de/>

Projekt Runeberg: <http://runeberg.org/>

Copyright © 2008 Marie Lebert

End of Project Gutenberg's Le Projet Gutenberg (1971-2008), by Marie Lebert

\*\*\* END OF THE PROJECT GUTENBERG EBOOK LE PROJET GUTENBERG (1971-2008) \*\*\*

Updated editions will replace the previous one—the old editions will be renamed.

Creating the works from print editions not protected by U.S. copyright law means that no one owns a United States copyright in these works, so the Foundation (and you!) can copy and distribute it in the United States without permission and without paying copyright royalties. Special rules, set forth in the General Terms of Use part of this license, apply to copying and distributing Project Gutenberg™ electronic works to protect the PROJECT GUTENBERG™ concept and trademark. Project Gutenberg is a registered trademark, and may not be used if you charge for an eBook, except by following the terms of the trademark license, including paying royalties for use of the Project Gutenberg trademark. If you do not charge anything for copies of this eBook, complying with the trademark license is very easy. You may use this eBook for nearly any purpose such as creation of derivative works, reports, performances and research. Project Gutenberg eBooks may be modified and printed and given away—you may do practically ANYTHING in the United States with eBooks not protected by U.S. copyright law. Redistribution is subject to the trademark license, especially commercial redistribution.

**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](http://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](http://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](http://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](http://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](http://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](http://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](http://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](http://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.