

Technologies et livre pour tous

Marie Lebert

NEF, Université de Toronto, 2008

Copyright © 2008 Marie Lebert

Du Projet Gutenberg en 1971 à l'Encyclopedia of Life en 2007, 38 étapes avec résumé et description plus détaillée. Ce dossier est disponible aussi en anglais, avec un texte différent. Les deux versions sont disponibles sur le NEF
<<http://www.etudes-francaises.net/dossiers/technologies.htm>>.

Marie Lebert, chercheuse et journaliste, s'intéresse aux technologies dans le monde du livre, des autres médias et des langues. Elle est l'auteure des ouvrages Les mutations du livre (2007) et Le Livre 010101 (2003). Tous ses livres et dossiers sont publiés par le NEF (Net des études françaises), Université de Toronto, et sont librement disponibles en ligne sur le NEF
<<http://www.etudes-francaises.net>>.

Ce dossier doit beaucoup à toutes les personnes ayant accepté de répondre à mes questions, dont certaines pendant plusieurs années. La plupart des entretiens sont disponibles sur le NEF <www.etudes-francaises.net/entretiens/index.htm>.

Table

Introduction

1968: ASCII
1971: Projet Gutenberg
1974: Internet
1977: UNIMARC
1984: Copyleft
1990: Web
1991: Unicode
1993: Online Books Page
1993: Format PDF
1994: Premier site de bibliothèque
1994: Des titres gratuits / payants
1995: Amazon.com
1995: Presse en ligne
1996: Palm Pilot
1996: Internet Archive
1996: Enseignement et internet
1997: Publication numérique
1997: Dictionnaires en ligne
1997: Convergence multimédia
1998: Gallica
1998: Bibliothécaires numériques
1998: Web multilingue
1999: Format Open eBook
1999: Auteurs numériques
2000: yourDictionary.com
2000: Bible de Gutenberg en ligne
2000: Distributed Proofreaders
2000: Public Library of Science
2001: Wikipedia
2001: Creative Commons
2002: Cours du MIT gratuits
2004: Projet Gutenberg Europe
2004: Google Books
2005: Open Content Alliance
2006: WorldCat gratuit
2007: Citizendium
2007: Encyclopedia of Life

Introduction

Michael Hart fonde le Projet Gutenberg en 1971. Il écrit en 1998: "Nous considérons le texte électronique comme un nouveau médium, sans véritable relation avec le papier. 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 établissements d'enseignement." (entretien du NEF)

Tim Berners-Lee invente le web en 1990. Il écrit en 1998: "Le rêve derrière le web est un espace d'information commun dans lequel nous communiquons en partageant l'information. Son universalité est essentielle, à savoir le fait qu'un lien hypertexte puisse pointer sur quoi que ce soit, quelque chose de personnel, de local ou de global, aussi bien une ébauche qu'une réalisation très sophistiquée. Deuxième partie de ce rêve, le web deviendrait d'une utilisation tellement courante qu'il serait un miroir réaliste (sinon la principale incarnation) de la manière dont nous travaillons, jouons et nouons des relations sociales. Une fois que ces interactions seraient en ligne, nous pourrions utiliser nos ordinateurs pour nous aider à les analyser, donner un sens à ce que nous faisons, et voir comment chacun trouve sa place et comment nous pouvons mieux travailler ensemble." (The World Wide Web: A very short personal history)

here was a second part of the dream, too, dependent on the web being so generally used that it became a realistic mirror (or in fact the primary embodiment) of the ways in which we work and play and socialize. That was that once the state of our interactions was on line, we could then use computers to help us analyse it, make sense of what we are doing, where we individually fit in, and how we can better work together."

John Mark Ockerbloom crée l'Online Books Page en 1993. Il écrit en 1998: "Je me suis passionné pour l'énorme potentiel qu'a l'internet de rendre la littérature accessible au plus grand nombre. (...) Je suis très intéressé par le développement de l'internet en tant que médium de communication de masse dans les prochaines années. J'aimerais aussi rester impliqué dans la mise à disposition gratuite de livres sur l'internet, que ceci fasse partie intégrante de mon activité professionnelle, ou que ceci soit une activité bénévole menée sur mon temps libre." (entretien du NEF)

Voici le voyage virtuel que nous allons suivre:

1968: Le code ASCII est le premier système d'encodage informatique.

1971: Le Projet Gutenberg est la première bibliothèque numérique.

1974: L'internet prend son envol.

1977: L'UNIMARC est créé en tant que format bibliographique commun.

1984: Le copyleft est une adaptation du copyright pour les logiciels.

1990: Le web fait ses débuts.

1991: L'Unicode est un système d'encodage universel pour toutes les langues.

1993: L'Online Books Page est le premier répertoire de livres gratuits.

1993: Le format PDF est lancé par Adobe.

1994: Le premier site de bibliothèque est mis en ligne.

1994: Les éditeurs utilisent le web comme outil de marketing.
1995: Amazon.com est la première grande librairie en ligne.
1995: La grande presse se met en ligne.
1996: Le Palm Pilot est le premier assistant personnel.
1996: L'Internet Archive est créée pour archiver le web.
1996: De nouvelles méthodes d'enseignement sont à l'étude.
1997: L'édition en ligne commence à se généraliser.
1997: Des dictionnaires sont mis en ligne gratuitement.
1997: La convergence multimédia est le sujet d'un colloque.
1998: Gallica est le secteur numérique de la Bibliothèque nationale de France.
1998: Le web devient multilingue.
1998: Les bibliothécaires numériques font carrière.
1999: L'Open eBook (OeB) sera le format standard des livres numériques.
1999: Certains auteurs se mettent au numérique.
2000: yourDictionary.com est créé en tant que portail pour les langues.
2000: La Bible de Gutenberg est mise en ligne.
2000: Distributed Proofreaders oeuvre pour le Projet Gutenberg.
2000: La Public Library of Science (PLOS) lancera des revues scientifiques gratuites.
2001: Wikipedia est la première grande encyclopédie collaborative.
2001: La licence Creative Commons renoue le droit d'auteur sur le web.
2003: Le MIT OpenCourseWare met les cours du MIT à la disposition de tous.
2004: Le Projet Gutenberg Europe débute une bibliothèque multilingue.
2004: Google Print laisse la place à Google Books.
2005: L'Open Content Alliance (OCA) lance une bibliothèque universelle publique.
2006: Le catalogue collectif WorldCat devient gratuit.
2007: Citizendium promeut une encyclopédie collaborative "fiable".
2007: L'Encyclopedia of Life répertoriera toutes les espèces végétales et animales.

[La plupart des citations sont des extraits des Entretiens du NEF.]

1968: ASCII

[Résumé]

Publié par l'ANSI (American National Standards Institute) en 1968, aux débuts de l'informatique, avec actualisation en 1977 et en 1986, le code ASCII (American standard code for information interchange) est un code standard de 128 caractères traduits en langage binaire sur sept bits (A est traduit par "1000001", B est traduit par "1000010", etc.). Les 128 caractères comprennent 33 caractères de contrôle (qui ne représentent donc pas de symbole écrit) et 95 caractères imprimables: les 26 lettres sans accent en majuscules (A-Z) et minuscules (a-z), les chiffres, les signes de ponctuation et quelques symboles, à savoir les touches du clavier anglais ou américain. L'ASCII permet donc uniquement la lecture de l'anglais et du latin. Par la suite, des variantes de l'ASCII (nomme ISO-8859 ou ISO-Latin) prennent en compte les caractères accentués de quelques langues européennes. La variante pour le français est définie par la norme ISO 8859-1 (Latin-1).

[En détail]

Le premier système d'encodage informatique est l'ASCII (American standard code for information interchange). Publié par l'American National Standards Institute (ANSI) en 1968, avec actualisation en 1977 et 1986, l'ASCII est un code standard de 128 caractères traduits en langage binaire sur sept bits (A est traduit par "1000001", B est traduit par "1000010", etc.). Les 128 caractères comprennent 33 caractères de contrôle (qui ne représentent donc pas de symbole écrit) et 95 caractères imprimables: les 26 lettres sans accent en majuscules (A-Z) et minuscules (a-z), les chiffres, les signes de ponctuation et quelques symboles, le tout correspondant aux touches du clavier anglais ou américain.

L'ASCII permet uniquement la lecture de l'anglais et du latin. Il ne permet pas de prendre en compte les lettres accentuées présentes dans bon nombre de langues européennes, et à plus forte raison les systèmes non alphabétiques (chinois, japonais, coréen, etc.). Ceci ne pose pas de problème majeur les premières années, tant que l'échange de fichiers électroniques se limite essentiellement à l'Amérique du Nord. Des variantes de l'ASCII (norme ISO-8859 ou ISO-Latin) prennent en compte les caractères accentués de quelques langues européennes. La variante pour le français est définie par la norme ISO 8859-1 (Latin-1). Mais le passage de l'ASCII original à ses différentes extensions devient vite un véritable casse-tête, y compris au sein de l'Union européenne, les problèmes étant entre autres la multiplication des variantes, la corruption des données dans les échanges informatiques ou encore l'incompatibilité des systèmes, les pages ne pouvant être affichées que dans une seule langue à la fois.

Avec le développement du web, l'échange des données s'internationalise de plus en plus. On ne peut plus se limiter à l'utilisation de l'anglais et de quelques langues européennes, traduites par un système d'encodage datant des années 1960. Publié pour la première fois en janvier 1991, l'Unicode est un système d'encodage universel sur 16 bits spécifiant un nombre unique pour chaque caractère. Ce nombre est lisible quels que soient la plateforme, le logiciel et la langue utilisés. L'Unicode peut traiter 65.000 caractères uniques et prendre en compte tous les systèmes d'écriture de la planète.

L'ASCII garde toutefois une place prépondérante. Dénommé 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 à quelques tablettes de lecture commercialisées depuis 1999 et déjà disparues du marché. Il est l'assurance que les documents survivront aux changements technologiques des prochaines décennies ou même des prochains siècles. Il n'existe pas d'autre standard aussi largement utilisé, y compris l'Unicode, ce jusqu'en 2008, date à laquelle les deux systèmes d'encodage sont également représentés sur le web.

1971: Projet Gutenberg

[Résumé]

Fondé par Michael Hart en juillet 1971 alors qu'il était étudiant à l'Université d'Illinois (Etats-Unis), le Projet Gutenberg a pour but de diffuser gratuitement par voie électronique le plus grand nombre possible d'oeuvres du domaine public. Il est le premier site d'information sur un internet encore embryonnaire, qui débute véritablement en 1974 et prend son essor en 1983. Vient ensuite le web (sous-ensemble de l'internet), opérationnel en 1991, puis le premier navigateur, qui apparaît en novembre 1993. Lorsque l'utilisation du web se généralise, le Projet Gutenberg trouve un second souffle et un rayonnement international. Au fil des ans, des centaines d'oeuvres sont patiemment numérisées en mode texte par des milliers de volontaires. D'abord essentiellement anglophones, les collections deviennent peu à peu multilingues. Le Projet Gutenberg Europe débute en janvier 2004. Le Projet Gutenberg franchit la barre des 20.000 titres en décembre 2006 et celle des 25.000 titres en avril 2008.

[En détail]

Si le livre imprimé a cinq siècles et demi, le livre numérique n'a pas encore 40 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 30.00 titres en novembre 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 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.

1974: Internet

[Résumé]

Apparu en 1974, l'internet se développe rapidement à partir de 1983. L'internet est un vaste réseau informatique opérant à l'échelle mondiale et reliant une multitude de sous-réseaux au moyen d'un même protocole (IP: internet protocol), ce protocole permettant à des ordinateurs différents de communiquer entre eux. Le réseau internet regroupe les réseaux publics, réseaux privés, réseaux commerciaux, réseaux d'enseignement, réseaux de services, etc., opérant à l'échelle planétaire pour offrir d'énormes ressources en information, communication et diffusion. Vinton Cerf est souvent appelé le père de l'internet parce qu'il crée en 1974 (avec Bob Kahn) le protocole TCP/IP (transmission control protocol / internet protocol), à la base de tout échange de données. En 1992, Vinton Cerf fonde l'Internet Society (ISOC), un organisme international visant à promouvoir le développement de l'internet. Quelque trente ans après les débuts de l'internet, "ses trois pouvoirs - l'ubiquité, la variété et l'interactivité - rendent son potentiel d'usages quasi infini" (Le Monde, 19 août 2005).

[En détail]

Apparu en 1974, l'internet est d'abord un phénomène expérimental enthousiasmant quelques branchés. A partir de 1983, il relie les centres de recherche et les universités. Suite à l'apparition du web en 1990 et du premier navigateur en 1993, il envahit notre vie quotidienne. Les signes cabalistiques des adresses web fleurissent sur les livres, les journaux, les affiches et les publicités. La presse s'enflamme pour ce nouveau médium. La majuscule d'origine d'Internet s'estompe. Internet devient l'internet, avec un "i" minuscule. De nom propre il devient nom commun, au même titre que l'ordinateur, le téléphone, le fax et le minitel. La même remarque vaut pour le World Wide Web, qui devient tout simplement le web.

Mais comment définir l'internet autrement que par ses composantes techniques? Sur le site de l'Internet Society (ISOC), organisme international coordonnant le développement du réseau, A Brief History of the Internet propose une triple définition. L'internet est: (a) un instrument de diffusion internationale, (b) un mécanisme de diffusion de l'information, (c) un moyen de collaboration et d'interaction entre les individus et les ordinateurs, indépendamment de leur situation géographique.

Selon ce document, bien plus que toute autre invention (télégraphe, téléphone, radio, ordinateur), l'internet révolutionne de fond en comble le monde des communications. Il représente l'un des exemples les plus réussis d'interaction entre un investissement soutenu dans la recherche et le développement d'une infrastructure de l'information, dans le cadre d'un réel partenariat entre les gouvernements, les entreprises et les universités.

Sur le site du World Wide Web Consortium (W3C), organisme international de normalisation du web, Bruce Sterling décrit le développement spectaculaire de l'internet dans Short History of the Internet. L'internet se développe plus vite que les téléphones cellulaires et les

télécopieurs. En 1996, sa croissance est de 20% par mois. Le nombre de machines ayant une connexion directe TCP/IP (transmission control protocol / internet protocol) a doublé depuis 1988. D'abord présent dans l'armée et dans les instituts de recherche, l'internet déferle dans les écoles, les universités et les bibliothèques, et il est également pris d'assaut par le secteur commercial.

Bruce Sterling s'intéresse aux raisons pour lesquelles on se connecte à l'internet. Une raison majeure lui semble être la liberté. L'internet est un exemple d'"anarchie réelle, moderne et fonctionnelle". Il n'y a pas de société régissant l'internet. Il n'y a pas non plus de censeurs officiels, de patrons, de comités de direction ou d'actionnaires. Toute personne peut parler d'égal à égal avec une autre, du moment qu'elle se conforme aux protocoles TCP/IP, des protocoles qui ne sont pas sociaux ni politiques mais strictement techniques. Malgré tous les efforts des "dinosaures" politiques et commerciaux, il est difficile à quelque organisme que ce soit de mettre la main sur l'internet. C'est ce qui fait sa force.

On y voit aussi une réelle solidarité. Christiane Jadelot, ingénieur d'études à l'INaLF-Nancy (INaLF: Institut national de la langue française), relate en juin 1998: "J'ai commencé à utiliser vraiment l'internet en 1994, je crois, avec un logiciel qui s'appelait Mosaic. J'ai alors découvert un outil précieux pour progresser dans ses connaissances en informatique et linguistique, littérature... Tous les domaines sont couverts. Il y a le pire et le meilleur, mais en consommateur averti, il faut faire le tri de ce que l'on trouve. J'ai surtout apprécié les logiciels de courrier, de transfert de fichiers, de connexion à distance. J'avais à cette époque des problèmes avec un logiciel qui s'appelait Paradox et des polices de caractères inadaptées à ce que je voulais faire. J'ai tenté ma chance et posé la question dans un groupe de News approprié. J'ai reçu des réponses du monde entier, comme si chacun était soucieux de trouver une solution à mon problème!"

En janvier 1998, lors d'un entretien avec Annick Rivoire, journaliste du quotidien Libération, Pierre Lévy, philosophe, explique que l'internet ouvre la voie à une forme d'intelligence collective: "Les réseaux permettent de mettre en commun nos mémoires, nos compétences, nos imaginations, nos projets, nos idées, et de faire en sorte que toutes les différences, les singularités se relancent les unes les autres, entrent en complémentarité, en synergie."

D'après Timothy Leary, philosophe adepte du cyberespace dès ses débuts, le 21^e siècle verrait l'émergence d'un nouvel humanisme, dont les idées-force seraient la contestation de l'autorité, la liberté de pensée et la créativité personnelle, le tout soutenu et encouragé par la vulgarisation de l'ordinateur et des technologies de la communication. Dans son livre Chaos et cyberculture (éditions du Léopard, 1998), il écrit: "Jamais l'individu n'a eu à sa portée un tel pouvoir. Mais, à l'âge de l'information, il faut saisir les signaux. Populariser signifie 'rendre accessible au peuple'. Aujourd'hui, le rôle du philosophe est de personnaliser, de populariser et d'humaniser les concepts informatiques, de façon à ce que personne ne se sente exclu."

Outre ce changement radical dans la relation information-utilisateur, on assiste à une transformation radicale de la nature même de l'information. L'information contenue dans les livres reste la même, au moins pendant une période donnée, alors que l'internet privilégie l'information la plus récente qui, elle, est en constante mutation. Vinton Cerf co-invente avec Bob Kahn en 1974 le protocole TCP/IP, à la base de tout échange de données sur le réseau.

Sur le site de l'Internet Society (ISOC), qu'il fonde en 1992 pour promouvoir le développement de l'internet, il explique: "Le réseau fait deux choses (...): comme les livres, il permet d'accumuler de la connaissance. Mais, surtout, il la présente sous une forme qui la met en relation avec d'autres informations. Alors que, dans un livre, l'information est maintenue isolée."

1977: UNIMARC

[Résumé]

Créé en 1977 par l'IFLA (International Federation of Library Associations), le format UNIMARC (universal machine readable cataloging) est un format universel permettant le stockage et l'échange de notices bibliographiques au moyen d'une codification des différentes parties de la notice (auteur, titre, éditeur, etc.) pour traitement informatique. Ce format permet de favoriser les échanges de données entre les différents formats MARC existants, qui correspondent chacun à une pratique nationale de catalogage (INTERMARC en France, UKMARC au Royaume-Uni, USMARC aux Etats-Unis, CAN/MARC au Canada, etc.). Les notices dans le format MARC d'origine sont d'abord converties au format UNIMARC avant d'être converties à nouveau dans le format MARC de destination. UNIMARC peut également être utilisé comme "modèle" pour le développement de nouveaux formats MARC.

[En détail]

A la fin des années 1990, l'avenir des catalogues en réseau tient à l'harmonisation du format MARC (machine readable cataloging) par le biais de l'UNIMARC (universal machine readable cataloging). Créé en 1977 par l'IFLA (International Federation of Library Associations), le format UNIMARC est un format universel permettant le stockage et l'échange de notices bibliographiques au moyen d'une codification des différentes parties de la notice (auteur, titre, éditeur, etc.) pour traitement informatique. Ce format favorise les échanges de données entre la vingtaine de formats MARC existants, qui correspondent chacun à une pratique nationale de catalogage (INTERMARC en France, UKMARC au Royaume-Uni, USMARC aux Etats-Unis, CAN/MARC au Canada, etc.). Les notices dans le format MARC d'origine sont d'abord converties au format UNIMARC avant d'être converties à nouveau dans le format MARC de destination. UNIMARC peut aussi être utilisé comme standard pour le développement de nouveaux formats MARC.

Dans le monde anglophone, la British Library (qui utilise UKMARC), la Library of Congress (qui utilise USMARC) et la Bibliothèque nationale du Canada (qui utilise CAN/MARC) décident d'harmoniser leurs formats MARC nationaux. Un programme de trois ans (de décembre 1995 à décembre 1998) permet de mettre au point un format MARC commun aux trois bibliothèques.

Parallèlement, en 1996, dans le cadre de son Programme des bibliothèques, la Commission européenne promeut l'utilisation du format UNIMARC comme format commun d'échange entre tous les formats MARC utilisés dans les pays de l'Union européenne. Le groupe de travail correspondant étudie aussi les problèmes posés par les différentes polices de caractères, ainsi que la manière d'harmoniser le format bibliographique, tout comme le format du document lui-même pour les documents disponibles en ligne.

1984: Copyleft

[Résumé]

L'idée du copyleft est lancée en 1984 par Richard Stallman, ingénieur en informatique et défenseur inlassable du mouvement «open source» au sein de la Free Software Foundation (FSF). Conçu à l'origine pour les logiciels, le copyleft est formalisé par la GPL (general public license) et étendu par la suite à toute oeuvre de création, le copyleft contient la déclaration normale du copyright affirmant la propriété et l'identification de l'auteur. Son originalité est de donner à l'utilisateur le droit de librement redistribuer le document et de le modifier. L'utilisateur ne peut toutefois revendiquer ni la paternité du travail original ni celle des changements effectués par d'autres. De plus, tous les travaux dérivés sont eux-mêmes soumis au copyleft. Chez les auteurs et les créateurs numériques, les adeptes du copyleft sont de plus en plus nombreux, afin de respecter la vocation première du web, réseau de communication et de diffusion à l'échelon mondial.

1990: Web

[Résumé]

Nom usuel du World Wide Web, le web (avec ou sans majuscule) est conçu en 1989-90 par Tim Berners-Lee, alors chercheur au CERN (Organisation européenne pour la recherche nucléaire) à Genève. En 1989, il met en réseau des documents utilisant l'hypertexte. En 1990, il met au point le premier serveur HTTP (hypertext transfert protocol) et le premier navigateur web. En 1991, le web est opérationnel et change radicalement l'utilisation de l'internet (qui existe depuis 1974). Selon les termes mêmes de son inventeur, le web est "un espace commun d'information dans lequel nous communiquons en partageant cette information". En novembre 1993, le web prend son essor grâce à Mosaic, premier navigateur à destination du grand public. En octobre 1994 est fondé le consortium W3C (World Wide Web Consortium), un consortium international chargé de développer les normes et protocoles nécessaires au bon fonctionnement du web. Le W3C est présidé par Tim Berners-Lee. Quinze ans après la création du web, le magazine Wired constate dans son numéro d'août 2005 que "moins de la moitié du web est commercial, le reste fonctionne avec la passion".

[En détail]

Tim Berners-Lee invente le web en 1989-90. Dans *The World Wide Web: A very short personal history*, il écrit en avril 1998: "Le rêve derrière le web est un espace d'information commun dans lequel nous communiquons en partageant l'information. Son universalité est essentielle, à savoir le fait qu'un lien hypertexte puisse pointer sur quoi que ce soit, quelque chose de personnel, de local ou de global, aussi bien une ébauche qu'une réalisation très sophistiquée."

Le web est d'abord conçu comme un grand livre composé de pages reliées entre elles par des liens hypertextes et reproduisant les modèles connus de l'édition papier. Certains appellent cette étape le web 1.0.

Hébergé sur un serveur web et défini par une adresse web, appelée aussi URL (uniform service locator), un site web est constitué d'un ensemble de pages web reliées par des liens hypertextes (reliant entre eux des textes et des images) et hypermédias (reliant des textes et des images à des graphiques, des images animées, des bandes sonores et des vidéos). Ces hyperliens sont utilisés au sein d'une même page web, au sein du même site web (pour relier les pages les unes aux autres) et vers d'autres sites web.

Le web permet au livre de se convertir. On voit apparaître les textes électroniques, les bibliothèques numériques, les librairies en ligne, les éditeurs électroniques, les encyclopédies en ligne, les oeuvres hypermédias, les logiciels de lecture et les appareils de lecture dédiés. Le web devient une vaste encyclopédie.

Au début des années 2000, des milliers d'oeuvres du domaine public sont en accès libre. Les libraires et les éditeurs ont pour la plupart un site web. Certains naissent directement sur le web, avec la totalité de leurs transactions s'effectuant via l'internet. De plus en plus de livres

et revues ne sont disponibles qu'en version numérique, pour éviter les coûts d'une publication imprimée. L'internet devient un outil indispensable pour se documenter, avoir accès aux documents et élargir ses connaissances. Le web est non seulement une gigantesque encyclopédie mais aussi une énorme bibliothèque, une immense librairie et un médium des plus complets. De statique dans les livres imprimés, l'information devient fluide, avec possibilité d'actualisation constante.

Dans *The World Wide Web: A very short personal history*, document écrit en avril 1998, Tim Berners-Lee ajoute: "Deuxième partie de ce rêve, le web deviendrait d'une utilisation tellement courante qu'il serait un miroir réaliste (sinon la principale incarnation) de la manière dont nous travaillons, jouons et nouons des relations sociales. Une fois que ces interactions seraient en ligne, nous pourrions utiliser nos ordinateurs pour nous aider à les analyser, donner un sens à ce que nous faisons, et voir comment chacun trouve sa place et comment nous pouvons mieux travailler ensemble."

En effet. Apparue en 2004, le web 2.0 serait un web de 2e génération, caractérisé par les notions de communauté et de partage, avec une flopée de sites dont le contenu est alimenté par les utilisateurs, par exemple les blogs, les wikis et les sites "sociaux". L'expression "web 2.0" émane d'ailleurs d'un éditeur puisqu'elle est utilisée pour la première fois en 2004 par Tim O'Reilly, fondateur O'Reilly Media, en tant que titre pour une série de conférences. Certains parlent de World Live Web au lieu de World Wide Web, le nom d'origine du web. Le web vise non seulement à utiliser l'information, mais aussi à collaborer en ligne, par exemple en tenant un blog personnel ou collectif, ou encore en participant aux encyclopédies Citizendium ou Wikipedia, cette dernière étant devenue l'un des dix sites les plus visités du web. Certains sites communautaires sont "incontournables", par exemple le site de photos Flickr, le site de vidéos YouTube ou les réseaux sociaux Facebook et MySpace.

Le web 3.0 serait le web du futur, un web de 3e génération qui prendrait logiquement le relais du web 2.0. Il s'agirait d'un web capable d'apporter une réponse complète à une requête exprimée en langage courant. D'après la société Radar Networks, ce web serait "doté d'une forme d'intelligence artificielle globale et collective", avec des données qui seraient rassemblées sur les nombreux sites sociaux existant sur le web, tout comme les sites sur lesquels les utilisateurs donnent leur avis. Ces données pourraient ensuite être traitées automatiquement après avoir été structurées sur la base du langage descriptif RDF (resource description framework) développé par le W3C (World Wide Web Consortium), l'organisme international chargé du développement du web. Cette définition du web 3.0 est d'ailleurs loin de faire l'unanimité.

Terminons par quelques chiffres, ceux de Netcraft, une société de services internet qui fait un décompte des sites web au fil des ans. Netcraft recense un million de sites en avril 1997, 10 millions de sites en février 2000, 20 millions de sites en septembre 2000, 30 millions de sites en juillet 2001, 40 millions de sites en avril 2003, 50 millions de sites en mai 2004, 60 millions de sites en mars 2005, 70 millions de sites en août 2005, 80 millions de sites en avril 2006, 90 millions de sites en août 2006 et enfin 100 millions de sites le 1er novembre 2006. La forte croissance de l'année 2006 serait due à l'explosion des sites de petites entreprises et des blogs.

1991: Unicode

[Résumé]

Publié pour la première fois en janvier 1991, l'Unicode est un système d'encodage informatique sur 16 bits spécifiant un nombre unique pour chaque caractère. Ce nombre est lisible quels que soient la plateforme, le logiciel et la langue utilisés. L'Unicode peut traiter 65.000 caractères uniques et prendre en compte tous les systèmes d'écriture de la planète. A la grande satisfaction des linguistes, l'Unicode remplace progressivement l'ASCII (American standard code for information interchange), un système d'encodage sur sept bits datant de 1968 et ne pouvant traiter que l'anglais et quelques langues européennes. L'utilisation de l'Unicode commence à se généraliser en 1998. Les versions récentes du système d'exploitation Windows de Microsoft (Windows NT, Windows 2000, Windows XP, Windows Vista) utilisent l'Unicode pour les fichiers texte, alors que les versions précédentes utilisaient l'ASCII. L'Unicode dispose de plusieurs variantes en fonction des besoins, par exemple UTF-8, UTF-16 et UTF-32 (UTF: Unicode transformation format).

[En détail]

Avec l'apparition du web en 1990, l'échange des données s'internationalise de plus en plus. On ne peut plus se limiter à l'utilisation de l'anglais et de quelques langues européennes, transcrites en ASCII (American standard code for information interchange), un système d'encodage sur sept bits datant de 1968.

Publié pour la première fois en janvier 1991, l'Unicode est un système d'encodage universel sur 16 bits spécifiant un nombre unique pour chaque caractère. Ce nombre est lisible quels que soient la plateforme, le logiciel et la langue utilisés. L'Unicode peut traiter 65.000 caractères uniques et prendre en compte tous les systèmes d'écriture de la planète. A la grande satisfaction des linguistes, il remplace progressivement l'ASCII. L'Unicode dispose de plusieurs variantes en fonction des besoins, par exemple UTF-8, UTF-16 et UTF-32 (UTF: Unicode transformation format). Il devient une composante des spécifications du W3C (World Wide Web Consortium), l'organisme international chargé du développement du web.

L'utilisation de l'Unicode se généralise en 1998, par exemple pour les fichiers texte sous plateforme Windows (Windows NT, Windows 2000, Windows XP, Windows Vista), qui étaient jusque-là en ASCII. Mais l'Unicode ne peut résoudre tous les problèmes, comme le souligne en juin 2000 Luc Dall'Armellina, co-auteur et webmestre d'oVosite, un espace d'écriture multimédia: "Les systèmes d'exploitation se dotent peu à peu des kits de langues et bientôt peut-être de polices de caractères Unicode à même de représenter toutes les langues du monde; reste que chaque application, du traitement de texte au navigateur web, emboîte ce pas. Les difficultés sont immenses: notre clavier avec ses ± 250 touches avoue ses manques dès lors qu'il faille saisir des Katakana ou Hiragana japonais, piè encore avec la langue chinoise. La grande variété des systèmes d'écritures de par le monde et le nombre de leurs signes font barrage. Mais les écueils culturels ne sont pas moins importants, liés aux codes et modalités de représentation propres à chaque culture ou ethnie."

Que préconise Olivier Gainon, créateur de CyLibris et pionnier de l'édition littéraire en ligne?
"Première étape: le respect des particularismes au niveau technique, explique-t-il en décembre 2000. Il faut que le réseau respecte les lettres accentuées, les lettres spécifiques, etc. Je crois très important que les futurs protocoles permettent une transmission parfaite de ces aspects - ce qui n'est pas forcément simple (dans les futures évolutions de l'HTML, ou des protocoles IP, etc.). Donc, il faut que chacun puisse se sentir à l'aise avec l'internet et que ce ne soit pas simplement réservé à des (plus ou moins) anglophones. Il est anormal aujourd'hui que la transmission d'accents puisse poser problème dans les courriers électroniques. La première démarche me semble donc une démarche technique. Si on arrive à faire cela, le reste en découle: la représentation des langues se fera en fonction du nombre de connectés, et il faudra envisager à terme des moteurs de recherche multilingues."

1993: Online Books Page

[Résumé]

L'Online Books Page est créée en janvier 1993 par John Mark Ockerbloom pour répertorier les textes électroniques de langue anglaise en accès libre sur le web. A cette date, John Mark Ockerbloom est doctorant à l'Université Carnegie Mellon (à Pittsburgh, dans l'Etat de Pennsylvanie). En 1999, il rejoint l'Université de Pennsylvanie pour travailler à la R&D (recherche et développement) de la bibliothèque numérique. A la même époque, il y transfère l'Online Books Page, tout en gardant la même présentation, très sobre, et il poursuit son travail d'inventaire dans le même esprit. En 2003, ce répertoire fête ses dix ans et recense 20.000 textes électroniques, dont 4.000 textes publiés par des femmes. En 2006, il recense 25.000 textes. En 2008, il recense 30.000 textes, dont 7.000 titres du Projet Gutenberg.

[En détail]

Si certains se donnent pour tâche de numériser des oeuvres, comme le Projet Gutenberg, d'autres décident de répertorier celles qui sont en accès libre sur le web, en offrant au lecteur un point d'accès commun. C'est le cas de John Mark Ockerbloom, doctorant à l'Université Carnegie Mellon (à Pittsburgh, dans l'Etat de Pennsylvanie), qui crée l'Online Books Page pour recenser les oeuvres anglophones.

Cinq ans plus tard, en septembre 1998, John Mark relate: "J'étais webmestre ici pour la section informatique de la CMU (Carnegie Mellon University), et j'ai débuté notre site local en 1993. Il comprenait des pages avec des liens vers des ressources disponibles localement, et à l'origine l'Online Books Page était l'une de ces pages, avec des liens vers des livres mis en ligne par des collègues de notre département (par exemple Robert Stockton, qui a fait des versions web de certains textes du Projet Gutenberg). Ensuite les gens ont commencé à demander des liens vers des livres disponibles sur d'autres sites. J'ai remarqué que de nombreux sites (et pas seulement le Projet Gutenberg ou Wiretap) proposaient des livres en ligne, et qu'il serait utile d'en avoir une liste complète qui permette de télécharger ou de lire des livres où qu'ils soient sur l'internet. C'est ainsi que mon index a débuté. J'ai quitté mes fonctions de webmestre en 1996, mais j'ai gardé la gestion de l'Online Books Page, parce qu'entre temps je m'étais passionné pour l'énorme potentiel qu'a l'internet de rendre la littérature accessible au plus grand nombre. Maintenant il y a tant de livres mis en ligne que j'ai du mal à rester à jour. Je pense pourtant poursuivre cette activité d'une manière ou d'une autre. Je suis très intéressé par le développement de l'internet en tant que médium de communication de masse dans les prochaines années. J'aimerais aussi rester impliqué dans la mise à disposition gratuite de livres sur l'internet, que ceci fasse partie intégrante de mon activité professionnelle, ou que ceci soit une activité bénévole menée sur mon temps libre."

Fin 1998, John Mark Ockerbloom obtient son doctorat en informatique. En 1999, il rejoint l'Université de Pennsylvanie, où il travaille à la R&D (recherche et développement) de la bibliothèque numérique de l'université. A la même époque, il y transfère l'Online Books Page, tout en gardant la même présentation, très sobre, et il poursuit son travail d'inventaire dans

le même esprit. Ce répertoire recense 20.000 livres en septembre 2003, 25.000 livres en décembre 2006 et 30.000 livres en décembre 2007.

1993: Format PDF

[Résumé]

Le format PDF (portable document format) est lancé en juin 1993 par la société Adobe, en même temps que le logiciel Acrobat Reader. Défini par l'extension de fichier ".pdf", ce format conserve la présentation, les polices, les couleurs et les images du document source, quelle que soit la plateforme utilisée (Macintosh, Windows, Unix, Linux, etc.) pour le créer et pour le lire. Lisible à l'aide de l'Acrobat Reader - logiciel de lecture téléchargeable gratuitement - ce format devient au fil des ans la norme internationale de diffusion des documents dont la présentation originale doit être conservée. Tout document peut être converti en PDF à l'aide du logiciel Adobe Acrobat, disponible dans de nombreuses langues et pour de nombreuses plateformes. Dix ans plus tard, 10% des documents disponibles sur l'internet sont au format PDF, et ce format est également le format de livre numérique le plus répandu.

[En détail]

Lancé en juin 1993 par la société Adobe et diffusé gratuitement, le premier logiciel de lecture du marché est l'Acrobat Reader, qui permet de lire des documents au format PDF (portable document format). Ce format permet de figer les documents numériques dans une présentation donnée, pour conserver les polices, les couleurs et les images du document source, quelle que soit la plateforme utilisée pour le créer et pour le lire. Vendu en parallèle, le logiciel Adobe Acrobat permet de convertir n'importe quel document au format PDF.

Au fil des ans, le format PDF devient la norme internationale de diffusion des documents électroniques, pour impression ou pour transfert d'une plateforme à l'autre. Des millions de documents PDF sont présents sur le web pour lecture ou téléchargement, ou bien transitent par courriel. L'Acrobat Reader pour ordinateur est progressivement disponible dans plusieurs langues et pour diverses plateformes (Windows, Mac, Unix, Linux). En 2001, Adobe lance également un Acrobat Reader pour assistant personnel (PDA), utilisable sur le Palm Pilot (en mai 2001) puis sur le Pocket PC (en décembre 2001).

Face à la concurrence représentée par le Microsoft Reader (lancé en avril 2000), Adobe annonce en août 2000 l'acquisition de la société Glassbook, spécialisée dans les logiciels de distribution de livres numériques pour éditeurs, libraires, distributeurs et bibliothèques. Adobe passe aussi un partenariat avec Amazon.com et Barnes & Noble.com afin de proposer des titres lisibles sur l'Acrobat Reader et le Glassbook Reader.

En janvier 2001, Adobe lance deux nouveaux logiciels. Le premier logiciel, gratuit, est l'Acrobat eBook Reader. Il permet de lire les fichiers PDF de livres numériques sous droits, avec gestion des droits par l'Adobe Content Server. Il permet aussi d'ajouter des notes et des signets, de choisir l'orientation de lecture des livres (paysage ou portrait), ou encore de visualiser leur couverture dans une bibliothèque personnelle. Il utilise la technique d'affichage CoolType et comporte un dictionnaire intégré. Le deuxième logiciel, payant, est l'Adobe Content Server, destiné aux éditeurs et distributeurs. Il s'agit d'un logiciel serveur de contenu assurant le conditionnement, la protection, la distribution et la vente sécurisée de

livres numériques au format PDF. Ce système de gestion des droits numériques (DRM - digital rights management) permet de contrôler l'accès aux livres numériques sous droits, et donc de gérer les droits d'un livre selon les consignes données par le gestionnaire des droits, par exemple en autorisant ou non l'impression ou le prêt. En avril 2001, Adobe conclut un partenariat avec la grande librairie en ligne Amazon.com, qui met en vente 2.000 livres numériques lisibles sur l'Acrobat eBook Reader: titres de grands éditeurs, guides de voyages, livres pour enfants, etc.

En dix ans, entre 1993 et 2003, l'Acrobat Reader aurait été téléchargé 500 millions de fois. Ce logiciel gratuit est désormais disponible dans de nombreuses langues et pour de nombreuses plateformes (Windows, Mac, Linux, Unix, Palm OS, Pocket PC, Symbian OS, etc.). En mai 2003, l'Acrobat Reader (5e version) fusionne avec l'Acrobat eBook Reader (2e version) pour devenir l'Adobe Reader (débutant à la version 6), qui permet de lire aussi bien les fichiers PDF standard que les fichiers PDF sécurisés.

Fin 2003, Adobe ouvre sa librairie en ligne, dénommée Digital Media Store, avec les titres au format PDF de grands éditeurs (HarperCollins Publishers, Random House, Simon & Schuster, etc.) ainsi que des versions électroniques de journaux et magazines (New York Times, Popular Science, etc.). Adobe lance aussi Adobe eBooks Central, un service permettant de lire, publier, vendre et prêter des livres numériques, et l'Adobe eBook Library, qui se veut un prototype de bibliothèque de livres numériques. En novembre 2004, l'Adobe Content Server est remplacé par l'Adobe LiveCycle Policy Server. Les versions récentes d'Adobe Acrobat permettent de créer des PDF compatibles avec les formats Open eBook (OeB) et ePub (ePub ayant succédé à OeB), devenus eux aussi des standards du livre numérique.

1994: Premier site de bibliothèque

[Résumé]

La première bibliothèque présente sur le web est la Bibliothèque municipale d'Helsinki (Finlande), qui inaugure son site en février 1994. Puis nombre de bibliothèques créent un site web pour y mettre des informations pratiques, leur catalogue en ligne et une sélection de sites web, et développent une bibliothèque numérique à côté de leurs collections traditionnelles. Qu'elles soient des bibliothèques de textes, des bibliothèques d'images (fixes ou animées) ou des bibliothèques sonores, ou qu'elles associent les trois supports, ces bibliothèques numériques permettent à un large public d'avoir accès à des documents jusque-là difficiles à consulter parce qu'appartenant à des fonds anciens, des fonds locaux et régionaux, ou des fonds spécialisés.

[En détail]

La première bibliothèque présente sur le web est la Bibliothèque municipale d'Helsinki (Finlande), qui inaugure son site en février 1994. Puis nombre de bibliothèques créent un site web pour y mettre des informations pratiques, leur catalogue en ligne et une sélection de sites web, et développent aussi une bibliothèque numérique à côté de leurs collections traditionnelles. Qu'elles soient des bibliothèques de textes, des bibliothèques d'images (fixes ou animées) ou des bibliothèques sonores, ou qu'elles associent les trois supports, ces bibliothèques numériques permettent à un large public d'avoir accès à des documents jusque-là difficiles à consulter parce qu'appartenant à des fonds anciens, des fonds locaux et régionaux, ou des fonds spécialisés.

Ces fonds sont difficilement accessibles pour des raisons diverses: souci de conservation des documents rares et fragiles, heures d'ouverture réduites, nombreux formulaires à remplir, délais significatifs de communication, pénurie de personnel, qui sont autant de barrières à franchir et qui demandent parfois au chercheur une patience à toute épreuve et une détermination hors du commun pour arriver jusqu'au document.

Grâce à sa "cyberbibliothèque", la bibliothèque peut enfin rendre comptables deux objectifs qui jusque-là ne l'étaient guère, à savoir la conservation des documents et la communication de ceux-ci. D'une part le document ne quitte son rayonnage qu'une seule fois pour être scanné, d'autre part le grand public y a facilement accès. Assis sur sa chaise ou dans son fauteuil, le lecteur peut consulter ces oeuvres en cliquant de l'une à l'autre, au gré de son humeur, de ses centres d'intérêt ou d'une curiosité passagère, de manière plus pratique qu'en se promenant dans les rayonnages, et surtout de manière beaucoup plus exhaustive puisque, dans les bibliothèques, une bonne partie des collections se trouve rangée dans des magasins et n'est donc pas accessible au lecteur qui chine. Si le lecteur veut ensuite consulter le document lui-même - dans certains cas, la consultation à l'écran ne peut remplacer le contact direct avec l'oeuvre - il pourra ensuite se lancer dans le parcours parfois laborieux qu'implique une consultation traditionnelle, mais ceci en connaissance de cause, après que le "feuilletage" à l'écran lui ait permis de sélectionner les documents en question et de ne demander que ceux qui l'intéressent vraiment.

Certaines "cyberbibliothèques" naissent directement sur le web. Créée en 1994 et hébergée sur le site de l'Université de Genève, Athena est l'oeuvre de Pierre Perroud, qui y consacre trente heures par semaine, en plus de son activité de professeur au Collège Voltaire. Pierre-Louis Chantre, journaliste, raconte dans L'Hebdo du 13 février 1997: "Il numérise des livres, met en page des textes que des correspondants inconnus lui envoient, crée des liens électroniques avec des livres disponibles ailleurs, tout en essayant de répondre le mieux possible aux centaines de lettres électroniques qu'il reçoit (mille personnes consultent Athena chaque jour). Un travail artisanal qu'il accomplit seul, sans grande rémunération. Malgré des demandes répétées, le Département de l'instruction publique de Genève ne lui paie que deux heures par semaine." (L'Hebdo, n° 7)

En 1997, le site bilingue français-anglais donne accès à 3.500 textes électroniques dans des domaines aussi variés que la philosophie, les sciences, la période classique, la littérature, l'histoire, l'économie, etc. En décembre 1998, les collections comprennent 8.000 textes. Un des objectifs d'Athena est de mettre en ligne des textes de langue française (dans la section: Textes français). Une section spécifique (Swiss Authors and Texts) regroupe les auteurs et textes suisses. On trouve aussi un répertoire mondial des ressources littéraires en ligne (Athena Literature Resources). Par ailleurs, Athena propose une table de minéralogie qui est l'oeuvre de Pierre Perroud et qui est consultée dans le monde entier.

Dans un article de la revue Informatique-Informations de février 1997, Pierre Perroud insiste sur la complémentarité du texte électronique et du livre imprimé. Selon lui, "les textes électroniques représentent un encouragement à la lecture et une participation conviviale à la diffusion de la culture", notamment pour l'étude et la recherche textuelle. Ces textes "sont un bon complément du livre imprimé - celui-ci restant irremplaçable lorsqu'il s'agit de lire". S'il est persuadé de l'utilité du texte électronique, le livre imprimé reste "un compagnon mystérieusement sacré vers lequel convergent de profonds symboles: on le serre dans la main, on le porte contre soi, on le regarde avec admiration; sa petitesse nous rassure autant que son contenu nous impressionne; sa fragilité renferme une densité qui nous fascine; comme l'homme il craint l'eau et le feu, mais il a le pouvoir de mettre la pensée de celui-là à l'abri du Temps."

Créée en mars 1995 par l'Université du Michigan (Etats-Unis) dans le cadre de la School of Information and Library Studies, l'Internet Public Library (IPL) se définit comme la première bibliothèque publique de l'internet sur l'internet, à savoir une bibliothèque sélectionnant, organisant et cataloguant les ressources disponibles sur l'internet, et n'existant elle-même que sur le réseau. Cette bibliothèque publique d'un genre nouveau devient vite une référence. L'IPL recense de manière pratiquement exhaustive les livres (Online Texts), les journaux (Newspapers) et les magazines (Magazines) disponibles sur le web. Les livres sont essentiellement des oeuvres du domaine public, avec 22.500 titres en 2006, dont le quart provient du Projet Gutenberg.

1994: Titres gratuits / payants

[Résumé]

La publication en ligne d'un livre à titre gratuit nuit-elle aux ventes de la version imprimée ou non? La National Academy Press (NAP) est la première à prendre un tel risque, dès 1994, avec un pari gagné. La même expérience est menée ensuite par la MIT Press (MIT: Massachusetts Institute of Technology). "A première vue, cela paraît illogique", écrit Beth Berselli, journaliste au Washington Post, dans un article repris par Le Courrier international de novembre 1997. "Un éditeur de Washington, la National Academy Press (NAP), qui a publié sur internet 700 titres de son catalogue actuel, permettant ainsi à tout un chacun de lire gratuitement ses livres, a vu ses ventes augmenter de 17% l'année suivante. Qui a dit que personne n'achèterait la vache si on pouvait avoir le lait gratuitement?"

[En détail]

La publication en ligne d'un livre à titre gratuit nuit-elle aux ventes de la version imprimée ou non? La National Academy Press (NAP) est la première à prendre un tel risque, dès 1994, avec un pari gagné. La même expérience est menée par la MIT Press (MIT: Massachusetts Institute of Technology) un an plus tard.

"A première vue, cela paraît illogique", écrit Beth Berselli, journaliste au Washington Post, dans un article repris par Le Courrier international de novembre 1997. "Un éditeur de Washington, la National Academy Press (NAP), qui a publié sur internet 700 titres de son catalogue actuel, permettant ainsi à tout un chacun de lire gratuitement ses livres, a vu ses ventes augmenter de 17% l'année suivante. Qui a dit que personne n'achèterait la vache si on pouvait avoir le lait gratuitement?"

Une politique atypique porte donc ses fruits. Editeur universitaire, la NAP (National Academy Press, qui devient ensuite la National Academies Press) publie environ 200 livres par an, essentiellement des livres scientifiques et techniques et des ouvrages médicaux. En 1994, l'éditeur choisit de mettre en accès libre sur le web le texte intégral de plusieurs centaines de livres, afin que les lecteurs puissent les "feuilleter" à l'écran, comme ils le feraient dans une librairie, avant de les acheter ensuite si utile. La NAP est le premier éditeur à se lancer dans un tel pari, une initiative saluée par les autres maisons d'édition, qui hésitent cependant à se lancer elles aussi dans l'aventure, et ce pour trois raisons: le coût excessif qu'entraîne la mise en ligne de milliers de pages, les problèmes liés au droit d'auteur, et enfin une concurrence qu'ils estiment nuisible à la vente.

Dans le cas de la NAP, l'éditeur est soutenu par les auteurs eux-mêmes qui, pour se faire mieux connaître, insistent pour que leurs livres soient mis en ligne sur le site. Le web est un nouvel outil de marketing face aux 50.000 ouvrages publiés chaque année aux Etats-Unis. Une réduction de 20% est accordée pour toute commande effectuée en ligne. La présence de ces livres sur le web entraîne aussi une augmentation des ventes par téléphone. En 1998, le site de la NAP propose le texte intégral d'un millier de titres.

La solution choisie par la NAP est également adoptée en 1995 par la MIT Press, qui voit rapidement ses ventes doubler pour les livres disponibles en version intégrale sur son site.

1995: Amazon.com

[Résumé]

En juillet 1995, Jeff Bezos fonde à Seattle (Etat de Washington, Etats-Unis) la librairie en ligne Amazon.com, futur géant du commerce électronique. Suite à une étude de marché démontrant que les livres sont les meilleurs "produits" à vendre sur l'internet, Amazon.com débute avec dix salariés et trois millions d'articles. Pionnier d'un nouveau modèle économique, Amazon.com devient vite un géant du commerce électronique. Cinq ans plus tard, en novembre 2000, la société compte 7.500 salariés, 28 millions d'articles, 23 millions de clients et quatre filiales au Royaume-Uni (filiale ouverte en octobre 1998), en Allemagne (filiale ouverte à la même date), en France (filiale ouverte en août 2000) et au Japon (filiale ouverte en octobre 2000). Une cinquième filiale est ouverte au Canada (en juin 2002), suivie d'une sixième filiale, Joyo, en Chine (en septembre 2004).

[En détail]

Fondé par Jeff Bezos, Amazon.com voit le jour en juillet 1995 à Seattle, dans l'Etat de Washington, sur la côte ouest des Etats-Unis. Quinze mois auparavant, au printemps 1994, Jeff Bezos fait une étude de marché pour décider du meilleur "produit de consommation" à vendre sur l'internet. Dans sa liste de vingt produits marchands, qui comprennent entre autres les vêtements et les instruments de jardinage, les cinq premiers du classement se trouvent être les livres, les CD, les vidéos, les logiciels et le matériel informatique.

"J'ai utilisé tout un ensemble de critères pour évaluer le potentiel de chaque produit", relate Jeff Bezos dans le kit de presse d'Amazon.com. "Le premier critère a été la taille des marchés existants. J'ai vu que la vente des livres représentait un marché mondial de 82 milliards de dollars US. Le deuxième critère a été la question du prix. Je voulais un produit bon marché. Mon raisonnement était le suivant: puisque c'était le premier achat que les gens allaient faire en ligne, il fallait que la somme à payer soit modique. Le troisième critère a été la variété dans le choix: il y avait trois millions de titres de livres alors qu'il n'y avait que 300.000 titres pour les CD, par exemple."

La vente de livres en ligne débute en juillet 1995, avec dix salariés et un catalogue dix fois supérieur à celui des plus grands hypemarchés. Le catalogue en ligne permet de rechercher les livres par titre, auteur, sujet ou rubrique. On y trouve aussi des CD, des DVD, des jeux informatiques, etc. Très attractif, le contenu éditorial du site change quotidiennement et forme un véritable magazine littéraire proposant des extraits de livres, des entretiens avec des auteurs et des conseils de lecture. Amazon.com devient le pionnier d'un nouveau modèle économique. Son évolution rapide est suivie de près par des analystes de tous bords.

En 1998, avec 1,5 million de clients dans 160 pays et une très bonne image de marque, Amazon.com est régulièrement cité comme un symbole de réussite dans le cybercommerce. Si la librairie en ligne est toujours déficitaire, sa cotation boursière est excellente, suite à une introduction à la Bourse de New York en mai 1997.

En novembre 2000, la société compte 7.500 salariés, 28 millions d'articles, 23 millions de clients et quatre filiales (Royaume-Uni, Allemagne, France, Japon), auxquelles s'ajoute en juin 2002 une cinquième filiale au Canada. La maison mère diversifie ses activités. Elle vend non seulement des livres, des vidéos, des CD et des logiciels, mais aussi des produits de santé, des jouets, des appareils électroniques, des ustensiles de cuisine, des outils de jardinage, etc. En novembre 2001, la vente des livres, disques et vidéos ne représente plus que 58% du chiffre d'affaires global. Admiré par certains, le modèle économique d'Amazon.com est contesté par d'autres, notamment en matière de gestion du personnel, avec des contrats de travail précaires et de bas salaires.

Tout comme la grande librairie en ligne britannique Internet Bookshop, Amazon.com offre une part des bénéfices à ses "associés" en ligne. Depuis le printemps 1997, tous les possesseurs d'un site web peuvent vendre des livres appartenant au catalogue de la librairie et toucher un pourcentage de 15% sur les ventes. Ces associés font une sélection dans les titres du catalogue et rédigent leurs propres résumés. Amazon.com reçoit les commandes par leur intermédiaire, expédie les livres et rédige les factures. Les associés reçoivent un rapport hebdomadaire d'activité. Au printemps 1998, la librairie en ligne compte plus de 30.000 sites affiliés.

La présence européenne d'Amazon.com débute en octobre 1998. Les deux premières filiales sont implantées en Allemagne et au Royaume-Uni. En août 2000, avec 1,8 million de clients en Grande-Bretagne, 1,2 million de clients en Allemagne et quelques centaines de milliers de clients en France, la librairie réalise 23% de ses ventes hors des Etats-Unis. A la même date, elle ouvre sa filiale française. Une filiale japonaise est ouverte en octobre 2000. En novembre 2000, Amazon ouvre un secteur eBooks, à savoir un secteur vendant des livres numériques. En 2001, les 29 millions de clients d'Amazon génèrent un chiffre d'affaires de 4 milliards de dollars US. En juin 2002, une cinquième filiale est ouverte au Canada. Au 3e trimestre 2003, la société devient bénéficiaire pour la première fois de son histoire. En octobre 2003, Amazon.com lance un service de recherche plein texte (Search Inside the Book) dans le texte intégral de 120.000 titres, un nombre promis à une croissance rapide. Une sixième filiale est ouverte en Chine sous le nom de Joyo en septembre 2004.

En 2004, le bénéfice net d'Amazon est de 588 millions de dollars US, dont 45% généré par ses filiales, avec un chiffre d'affaires de 6,9 milliards de dollars. Présent dans sept pays (Etats-Unis, Canada, Royaume-Uni, Allemagne, France, Japon, Chine) et devenu une référence mondiale du commerce en ligne, Amazon fête ses dix ans d'existence en juillet 2005, avec 9.000 salariés et 41 millions de clients actifs, attirés par des produits culturels, high-tech et autres aux prix attractifs et une livraison en 48 heures maximum dans les pays hébergeant une plateforme Amazon.

1995: Presse en ligne

[Résumé]

Au début des années 1990, les premières éditions électroniques de journaux sont disponibles par le biais de services commerciaux tels que America Online ou CompuServe. Suite à l'apparition du premier navigateur fin 1993 et à la croissance rapide du web qui s'ensuit, les organes de presse créent leurs propres sites. Mis en ligne en février 1995, le site web du mensuel Le Monde diplomatique est le premier site d'un périodique imprimé français. Monté dans le cadre d'un projet expérimental avec l'Institut national de l'audiovisuel (INA), ce site est inauguré lors du forum des images Imagina. Il donne accès à l'ensemble des articles depuis janvier 1994, par date, par sujet et par pays. Fin 1995, le quotidien Libération met en ligne son site web. Le site du quotidien Le Monde est lancé en 1996.

[En détail]

Au début des années 1990, les premières éditions électroniques de journaux sont disponibles par le biais de services commerciaux tels que America Online ou CompuServe. Suite à l'apparition du premier navigateur fin 1993 et à la croissance rapide du web qui s'ensuit, les organes de presse créent leurs propres sites.

Au Royaume-Uni, le Times et le Sunday Times font web commun sur un site dénommé Times Online, avec possibilité de créer une édition personnalisée. Aux Etats-Unis, la version en ligne du Wall Street Journal est payante, avec 100.000 abonnés en 1998. Celle du New York Times est disponible sur abonnement gratuit. Le Washington Post propose l'actualité quotidienne en ligne et de nombreux articles archivés, le tout avec images, sons et vidéos. Pathfinder (rebaptisé ensuite Time) est le site web du groupe Time-Warner, éditeur de Time Magazine, Sports Illustrated, Fortune, People, Southern Living, Money, Sunset, etc. On peut lire les articles "maison" et les rechercher par date ou par sujet. Lancé en 1992 en Californie, Wired, premier magazine imprimé entièrement consacré à la culture cyber, est bien évidemment présent sur le web.

Mis en ligne en février 1995, le site web du mensuel Le Monde diplomatique est le premier site d'un périodique imprimé français. Monté dans le cadre d'un projet expérimental avec l'Institut national de l'audiovisuel (INA), ce site est inauguré lors du forum des images Imagina. Il donne accès à l'ensemble des articles depuis janvier 1994, par date, par sujet et par pays. L'intégralité du mensuel en cours est consultable gratuitement pendant deux semaines suivant sa parution. Un forum de discussion permet au journal de discuter avec ses lecteurs. En juin 1998, Philippe Rivière, responsable du site, précise que, trois ans après sa mise en ligne, celui-ci a "bien grandi, autour des mêmes services de base: archives et annonce de sommaire". Grâce à l'internet, "le travail journalistique s'enrichit de sources faciles d'accès, aisément disponibles. Le travail éditorial est facilité par l'échange de courriers électroniques; par contre, une charge de travail supplémentaire due aux messages reçus commence à peser fortement."

Fin 1995, le quotidien Libération met en ligne son site web, peu après le lancement du Cahier

Multimédia, un cahier imprimé hebdomadaire inclus dans l'édition du jeudi. Le site propose la Une du quotidien, la rubrique Multimédia, qui regroupe les articles du Cahier Multimédia et les archives des cahiers précédents, le Cahier Livres complété par Chapitre Un (à savoir les premiers chapitres des nouveautés), et bien d'autres rubriques. La rubrique Multimédia est ensuite rebaptisée Numériques.

Le site du quotidien Le Monde est lancé en 1996. On y trouve des dossiers en ligne, la Une en version graphique à partir de 13 h, l'intégralité du journal avant 17 h, l'actualité en liaison avec l'AFP (Agence France-Presse), et des rubriques sur la Bourse, les livres, le multimédia et les sports. En 1998, le journal complet en ligne coûte 5 FF (0,76 euros) alors que l'édition papier coûte 7,50 FF (1,15 euros). S'ils concernent le multimédia, les articles du supplément imprimé hebdomadaire Télévision-Radio-Multimédia sont disponibles gratuitement en ligne dans la rubrique Multimédia, rebaptisée ensuite Nouvelles technologies.

L'Humanité est le premier quotidien français à proposer la version intégrale du journal en accès libre. Classés par rubriques, les articles sont disponibles entre 10 h et 11 h du matin, à l'exception de L'Humanité du samedi, disponible en ligne le lundi suivant. Tous les articles sont archivés sur le site.

Jacques Coubard, responsable du site web, explique en juillet 1998: "Le site de L'Humanité a été lancé en septembre 1996 à l'occasion de la Fête annuelle du journal. Nous y avons ajouté depuis un forum, un site consacré à la récente Coupe du monde de football (avec d'autres partenaires), et des données sur la Fête et sur le meeting d'athlétisme, parrainé par L'Humanité. Nous espérons pouvoir développer ce site à l'occasion du lancement d'une nouvelle formule du quotidien qui devrait intervenir à la fin de l'année ou au début de l'an prochain. Nous espérons également mettre sur le web L'Humanité Hebdo, dans les mêmes délais. Jusqu'à présent on ne peut pas dire que l'arrivée d'internet ait bouleversé la vie des journalistes, faute de moyens et de formation (ce qui va ensemble). Les rubriques sont peu à peu équipées avec des postes dédiés, mais une minorité de journalistes exploite ce gisement de données. Certains s'en servent pour transmettre leurs articles, leurs reportages. Il y a sans doute encore une "peur" culturelle à plonger dans l'univers du net. Normal, en face de l'inconnu. L'avenir devrait donc permettre par une formation (peu compliquée) de combler ce handicap. On peut rêver à un enrichissement par une sorte d'édition électronique, mais nous sommes sévèrement bridés par le manque de moyens financiers."

L'internet permet une information en profondeur qu'aucun organe de presse ne pouvait donner jusqu'ici: rapidité de propagation des informations, accès immédiat à de nombreux sites d'information, liens vers des articles et sources connexes, énormes capacités documentaires allant du général au spécialisé et réciproquement (cartes géographiques, notices biographiques, textes officiels, informations d'ordre politique, économique, social, culturel, etc.), grande variété d'illustrations (photos, graphiques, tableaux, vidéos, etc.), possibilité d'archivage avec moteur de recherche, etc.

1996: Palm Pilot

[Résumé]

Basée en Californie, la société Palm lance en mars 1996 le premier PDA (personal digital assistant), dénommé Palm Pilot, qui utilise le système d'exploitation Palm OS. 23 millions de Palm Pilot sont vendus entre 1996 et 2002. En mars 2001, on peut utiliser son Palm Pilot pour lire des livres numériques sur les logiciels de lecture Palm Reader et Mobipocket Reader. En 2002, malgré la concurrence, Palm est toujours le leader du marché (36,8% des machines vendues), suivi par Hewlett-Packard (13,5%), Sony (11%), Handspring (5,8%), Toshiba (3,7%) et Casio (3,3%). Les systèmes d'exploitation utilisés sont essentiellement le Palm OS (pour 55% des machines) et le Pocket PC de Microsoft (pour 25,7% des machines). La division PalmSource est plus précisément en charge des logiciels, notamment le logiciel de lecture Palm Reader, utilisable sur assistant personnel en mars 2001 puis sur ordinateur en juillet 2002. En 2003, 10.000 titres dans plusieurs langues sont lisibles sur le Palm Pilot et ses successeurs.

[En détail]

La société Palm lance le premier Palm Pilot en mars 1996 et vend 23 millions de machines entre 1996 et 2002. Son système d'exploitation est le Palm OS et son logiciel de lecture le Palm Reader. En mars 2001, les modèles Palm permettent aussi la lecture de livres numériques sur le Mobipocket Reader.

A la même date, la société Palm fait l'acquisition de Peanutpress.com, éditeur et distributeur de livres numériques pour PDA. Le Peanut Reader devient le Palm Reader, utilisable aussi bien sur le Palm Pilot que sur le Pocket PC (l'assistant personnel de Microsoft), et les 2.000 titres de Peanutpress.com sont transférés dans la librairie numérique de Palm (Palm Digital Media). Développé par PalmSource, une division de Palm, le Palm Reader permet de lire des livres numériques au format PDB (Palm database). D'abord utilisable uniquement sur les gammes Palm et Pocket PC, le Palm Reader est utilisable sur ordinateur en juillet 2002. A la même date, Palm Digital Media distribue 5.500 titres dans plusieurs langues. En 2003, le catalogue approche les 10.000 titres, téléchargeables à partir de la librairie Palm Digital Media, qui devient ensuite le Palm eBook Store.

Le Palm DRM System (DRM: digital rights management) est un serveur payant destiné aux éditeurs et distributeurs pour gérer le conditionnement, la protection, la distribution et la vente sécurisée de livres numériques au format PDB (Palm database), le format des fichiers lus par le Palm Reader. Ce système de DRM permet de contrôler l'accès aux livres numériques sous droits, et donc de gérer les droits d'un livre en fonction des consignes données par le gestionnaire des droits, par exemple en autorisant ou non l'impression ou le prêt.

Développé par PalmSource, une division de Palm, le Palm OS (OS: operating system) est le système d'exploitation du Palm Pilot, qui est d'abord un PDA unique avant de devenir une gamme de PDA. Le Palm OS équipe 55% des PDA vendus en 2002. L'autre "grand" système

d'exploitation est le Pocket PC de Microsoft, qui équipe pour sa part 25,7% des machines, y compris les assistants personnels du même nom.

Commercialisé par Microsoft en avril 2000 pour concurrencer le Palm Pilot, l'assistant personnel Pocket PC utilise d'abord un système d'exploitation spécifique, Windows CE, qui intègre le nouveau logiciel de lecture Microsoft Reader. En octobre 2001, Windows CE est remplacé par Pocket PC 2002, qui permet entre autres de lire des livres numériques sous droits. Ces livres sont protégés par un système de gestion des droits numériques dénommé Microsoft DAS Server (DAS: digital asset server). En 2002, le Pocket PC permet la lecture sur trois logiciels: le Microsoft Reader bien sûr, le Mobipocket Reader et le Palm Reader.

D'après Seybold Reports.com, en avril 2001, on compte 100.000 tablettes de lecture pour 17 millions d'assistants personnels (PDA). Deux ans plus tard, en juin 2003, plus aucune tablette n'est commercialisée. De nouveaux modèles apparaissent ensuite, mais on se demande s'ils peuvent vraiment réussir à s'imposer face à l'assistant personnel, qui offre aussi d'autres fonctionnalités. On se demande aussi s'il existe une clientèle spécifique pour les deux machines, la lecture sur assistant personnel étant destinée au grand public, et la lecture sur tablette électronique étant réservée aux gros consommateurs de documents que sont les lycéens, les étudiants, les professeurs, les chercheurs ou les juristes. Le débat reste toujours d'actualité dans les années qui suivent.

1996: Internet Archive

[Résumé]

Fondée en avril 1996 par Brewster Kahle à San Francisco, l'Internet Archive a pour but premier de constituer, stocker, préserver et gérer une "bibliothèque" de l'internet, en archivant régulièrement la totalité du web, afin d'offrir un outil de travail aux universitaires, chercheurs et historiens, et de préserver un historique de l'internet pour les générations futures. En octobre 2001, l'Internet Archive met ses archives en accès libre sur le web grâce à la Wayback Machine, qui permet à tout un chacun de consulter l'historique d'un site web, à savoir le contenu et la présentation d'un site web à différentes dates, théoriquement tous les deux mois, à partir de 1996. L'Internet Archive constitue aussi des collections numériques spécifiques, y compris pour les livres et les documents multimédias. Toutes ces collections sont en consultation libre sur le web.

[En détail]

Fondée en avril 1996 par Brewster Kahle à San Francisco, l'Internet Archive a pour but premier de constituer, stocker, préserver et gérer une "bibliothèque" de l'internet, en archivant régulièrement la totalité du web, afin d'offrir un outil de travail aux universitaires, chercheurs et historiens, et de préserver un historique de l'internet pour les générations futures. En octobre 2001, l'Internet Archive met ses archives en accès libre sur le web grâce à la Wayback Machine, qui permet à tout un chacun de consulter l'historique d'un site web, à savoir le contenu et la présentation d'un site web à différentes dates, théoriquement tous les deux mois, à partir de 1996. Toutes ces collections sont en consultation libre sur le web. Les archives du web représentent 300 téraoctets (To) de données en 2004, avec une croissance de 12 téraoctets par mois. Ces archives représentent 30 millions de pages web en 1996, 65 milliards de pages web (provenant de 50 millions de sites web) en décembre 2006 et 85 milliards de pages web en mai 2007.

Depuis la fin 1999, l'Internet Archive constitue aussi des collections numériques spécifiques, par exemple les sites relatifs au 11 septembre 2001, les sites relatifs aux élections de 2000 (présidentielles) et 2002 (élections du Congrès et des gouverneurs des Etats), les sites des pionniers du web, le Million Book Project (10.520 livres en avril 2005), des archives de films de la période 1903-1973, des archives de concerts "live" récents, des archives de logiciels, etc. Toutes ces collections sont en consultation libre sur le web, y compris la grande bibliothèque numérique (Text Archive) en cours de constitution. A cet effet, l'Internet Archive fonde en janvier 2005 l'Open Content Alliance (OCA), une initiative visant à créer un répertoire libre et multilingue de livres numérisés et de documents multimédia pour consultation et téléchargement sur n'importe quel moteur de recherche.

1996: Enseignement et internet

[Résumé]

Sur le site de l'Internet Society (ISOC), qu'il fonde en 1992 pour promouvoir le développement de l'internet, Vinton Cerf explique: "Le réseau fait deux choses (...): comme les livres, il permet d'accumuler de la connaissance. Mais, surtout, il la présente sous une forme qui la met en relation avec d'autres informations. Alors que, dans un livre, l'information est maintenue isolée." De plus, l'information contenue dans les livres reste la même, au moins pendant une période donnée, alors que l'internet privilégie l'information la plus récente qui, elle, est en constante mutation. Il s'ensuit un changement dans la manière d'enseigner. En septembre 1996, Dale Spender, professeur à l'Université de Queensland (Australie), tente d'analyser ce changement lors d'une communication de la 14e conférence mondiale de l'IFIP (International Federation of Information Processing).

[En détail]

Vinton Cerf co-invente en 1974 avec Bob Kahn le protocole TCP/IP, à la base de tout échange de données sur le réseau. Sur le site de l'Internet Society (ISOC), qu'il fonde en 1992 pour promouvoir le développement de l'internet, Vinton Cerf explique: "Le réseau fait deux choses (...): comme les livres, il permet d'accumuler de la connaissance. Mais, surtout, il la présente sous une forme qui la met en relation avec d'autres informations. Alors que, dans un livre, l'information est maintenue isolée."

De plus, l'information contenue dans les livres reste la même, au moins pendant une période donnée, alors que l'internet privilégie l'information la plus récente qui, elle, est en constante mutation. Il s'ensuit un changement dans la manière d'enseigner. Dès septembre 1996, dans Creativity and the Computer Education Industry, une communication de la 14e conférence mondiale de l'IFIP (International Federation of Information Processing), Dale Spender, professeur à l'Université de Queensland (Australie), tente d'analyser ce changement. Voici son argumentation résumée en deux paragraphes.

Depuis plus de cinq siècles, l'enseignement est essentiellement basé sur l'information procurée par les livres. Or les habitudes liées à l'imprimé ne peuvent être transférées dans l'univers numérique. L'enseignement en ligne offre des possibilités tellement nouvelles qu'il n'est guère possible d'effectuer les distinctions traditionnelles entre enseignant et enseigné. Le passage de la culture imprimée à la culture numérique exige donc d'entièrement repenser le processus d'acquisition du savoir, puisqu'on a maintenant l'opportunité sans précédent de pouvoir influencer sur le type d'enseignement qu'on souhaite recevoir.

Dans la culture imprimée, l'information contenue dans les livres reste la même pendant un certain temps, ce qui encourage à penser que l'information est stable. La nature même de l'imprimé est liée à la notion de vérité, stable elle aussi. Cette stabilité et l'ordre qu'elle engendre sont un des fondements de l'âge industriel et de l'ère des sciences et techniques. Les notions de vérité, de loi, d'objectivité et de preuve sont le fondement de nos croyances et de nos cultures. Mais le numérique change tout ceci. Soudain l'information en ligne

supplante l'information imprimée pour devenir la plus fiable et la plus utile, et l'utilisateur est prêt à la payer en conséquence. Cette transformation radicale de la nature même de l'information doit être au cœur du débat sur les nouvelles méthodes d'enseignement.

En juillet 1998, Patrick Rebollar, professeur de français et d'informatique dans des universités japonaises, analyse l'impact de l'internet sur sa vie professionnelle: "Mon travail de recherche est différent, mon travail d'enseignant est différent, mon image en tant qu'enseignant-chercheur de langue et de littérature est totalement liée à l'ordinateur, ce qui a ses bons et ses mauvais côtés (surtout vers le haut de la hiérarchie universitaire, plutôt constituée de gens âgés et technologiquement récalcitrants). J'ai cessé de m'intéresser à certains collègues proches géographiquement mais qui n'ont rien de commun avec mes idées, pour entrer en contact avec des personnes inconnues et réparties dans différents pays (et que je rencontre parfois, à Paris ou à Tokyo, selon les vacances ou les colloques des uns ou des autres). La différence est d'abord un gain de temps, pour tout, puis un changement de méthode de documentation, puis de méthode d'enseignement privilégiant l'acquisition des méthodes de recherche par mes étudiants, au détriment des contenus (mais cela dépend des cours). Progressivement, le paradigme réticulaire l'emporte sur le paradigme hiérarchique - et je sais que certains enseignants m'en veulent à mort d'enseigner ça, et de le dire d'une façon aussi crue. Cependant ils sont obligés de s'y mettre..."

Professeur au département d'études françaises de l'Université de Toronto, Russon Wooldridge relate en mai 2001: "Mes activités de recherche, autrefois menées dans une tour d'ivoire, se font maintenant presque uniquement par des collaborations locales ou à distance. (...) Tout mon enseignement exploite au maximum les ressources d'internet (le web et le courriel): les deux lieux communs d'un cours sont la salle de classe et le site du cours, sur lequel je mets tous les matériaux des cours. Je mets toutes les données de mes recherches des vingt dernières années sur le web (réédition de livres, articles, textes intégraux de dictionnaires anciens en bases de données interactives, de traités du 16e siècle, etc.). Je publie des actes de colloques, j'édite un journal, je collabore avec des collègues français, mettant en ligne à Toronto ce qu'ils ne peuvent pas publier en ligne chez eux."

1997: Publication numérique

[Résumé]

Pour la publication d'ouvrages et de périodiques éducatifs et scientifiques, dans lesquels l'information la plus récente est primordiale, la numérisation conduit à repenser complètement la signification même de publication, et à s'orienter vers une diffusion en ligne qui rend beaucoup plus facile les réactualisations régulières. On observe un tournant majeur en 1997. Certaines universités diffusent par exemple des manuels "sur mesure" composés de quelques chapitres sélectionnés dans une vaste base de données, un choix complété par des articles et par les commentaires des professeurs. Pour un séminaire, un très petit tirage peut être effectué à la demande pour un document rassemblant quelques textes et transmis par voie électronique à un imprimeur.

[En détail]

Depuis les années 1970, la chaîne traditionnelle de l'édition est soumise à de nombreux bouleversements. Le marché de l'imprimerie traditionnelle est d'abord ébranlé par l'apparition des machines de photocomposition. Les coûts d'impression continuent ensuite de baisser avec les procédés d'impression assistée par ordinateur, les photocopieurs, les photocopieurs couleur et le matériel d'impression numérique. L'impression est désormais souvent assurée à bas prix par des ateliers de PAO (publication assistée par ordinateur) et des entreprises d'arts graphiques. La numérisation accélère aussi le processus de rédaction, puisque le rédacteur, le concepteur artistique et le personnel chargé de la mise en page peuvent travailler simultanément sur le même ouvrage.

Pour la publication d'ouvrages et de périodiques éducatifs et scientifiques, dans lesquels l'information la plus récente est primordiale, la numérisation conduit à repenser complètement la signification même de publication, et à s'orienter vers une diffusion en ligne qui rend beaucoup plus facile les réactualisations régulières. Certaines universités diffusent par exemple des manuels "sur mesure" composés de quelques chapitres sélectionnés dans une vaste base de données, un choix complété par des articles et par les commentaires des professeurs. Pour un séminaire, un très petit tirage peut être effectué à la demande pour quelques textes transmis par voie électronique à un imprimeur.

L'interaction entre document imprimé et document électronique devient omniprésente, et elle s'accroîtra encore dans les prochaines années, à tel point qu'il deviendra probablement ridicule d'établir une distinction entre document électronique et document imprimé. Déjà, à l'heure actuelle, pratiquement tous les documents imprimés récents sont issus d'une version électronique sur traitement de texte, tableur ou base de données. De plus en plus de documents n'existent désormais qu'en version électronique, et de plus en plus de documents imprimés sont numérisés.

Outre sa facilité d'accès et son faible coût, le document électronique peut être régulièrement actualisé. Point n'est besoin d'attendre une nouvelle édition imprimée soumise aux contraintes commerciales et aux exigences de l'éditeur. L'édition électronique apparaît donc

comme une bonne solution pour résoudre les problèmes budgétaires des presses universitaires et des éditeurs axés sur la publication d'ouvrages de recherche.

1997: Dictionnaires en ligne

[Résumé]

La société de traduction Logos décide en 1997 de mettre le Logos Dictionary et tous ses outils professionnels en accès libre sur le web. Dans les années qui suivent, les dictionnaires en ligne sont soit en accès libre comme le Grand dictionnaire terminologique (GDT) de l'Office québécois de la langue française (OQLF), soit disponibles moyennant une inscription gratuite ou payante. Nombre de dictionnaires sont également disponibles sur CD-ROM (compact disc - read only memory) ou sur DVD (digital versatile disc), parallèlement à la version en ligne. En accès libre sur le site de l'ATILF (Analyse et traitement informatique de la langue française), Dictionnaires est une collection de dictionnaires informatisés anciens et récents, qui comprend notamment le TLFi (Trésor de la langue française informatisé, 1971-1994). Des dictionnaires gratuits écrits collectivement utilisent la forme du wiki, par exemple le dictionnaire multilingue Wiktionary, lancé en 2002 par la Wikipedia Foundation. Le portail yourDictionary.com répertorie 3.500 dictionnaires et grammaires dans 300 langues en avril 2007.

[En détail]

Fondé en 1979 à Modène (Italie) par Rodrigo Vergara, Logos est une société de traduction offrant des services dans 35 langues en 1997, avec 300 traducteurs travaillant sur place et un réseau mondial de 2.500 traducteurs travaillant en free-lance. Logos traduit en moyenne 200 textes par jour.

Initiative peu courante à l'époque, Logos décide de mettre tous ses outils professionnels en accès libre sur le web. Dans un entretien publié dans le quotidien Le Monde du 7 décembre 1997, Rodrigo Vergara relate: "Nous voulions que nos traducteurs aient tous accès aux mêmes outils de traduction. Nous les avons donc mis à leur disposition sur internet, et tant qu'à faire nous avons ouvert le site au public. Cela nous a rendus très populaires, nous a fait beaucoup de publicité. L'opération a drainé vers nous de nombreux clients, mais aussi nous a permis d'étoffer notre réseau de traducteurs grâce aux contacts établis à la suite de cette initiative."

Les outils de traduction disponibles sur le web comprennent un dictionnaire multilingue de 7,5 millions d'entrées (Logos Dictionary), une base de données de 553 glossaires (Linguistic Resources), des tables de conjugaison en 17 langues (Conjugation of Verbs), et enfin la Wordtheque, une base de données multilingue de 328 millions de termes issus de traductions de romans et documents techniques. La recherche dans la Wordtheque est possible par langue, mot, auteur ou titre. En 2007, la Wordtheque, devenue la Logos Library, comprend 710 millions de termes, Conjugation of Verbs, devenu l'Universal Conjugator, propose des tableaux de conjugaison dans 36 langues, et Linguistic Resources offre un point d'accès unique pour 1.215 glossaires.

Quand Logos met ses outils en accès libre en 1997, il ouvre la voie à d'autres initiatives.

Le premier dictionnaire de langue française en accès libre est le Dictionnaire universel francophone en ligne, qui répertorie 45.000 mots et 116.000 définitions tout en présentant "sur un pied d'égalité, le français dit "standard" et les mots et expressions en français tel qu'on le parle sur les cinq continents". Issu de la collaboration entre Hachette et l'AUPELF-UREF (devenu depuis l'AUF - Agence universitaire de la francophonie), il correspond à la partie "noms communs" du dictionnaire imprimé disponible chez Hachette. L'équivalent pour la langue anglaise est le site Merriam-Webster OnLine, qui donne librement accès au Collegiate Dictionary et au Collegiate Thesaurus.

En mars 2000, les 20 volumes de l'Oxford English Dictionary sont mis en ligne par l'Oxford University Press (OUP). La consultation du site est payante. Le dictionnaire bénéficie d'une mise à jour trimestrielle d'environ 1.000 entrées nouvelles ou révisées. Deux ans après cette première expérience, en mars 2002, l'Oxford University Press met en ligne l'Oxford Reference Online, une vaste encyclopédie conçue directement pour le web et consultable elle aussi sur abonnement payant. Avec 60.000 pages et un million d'entrées, elle représente l'équivalent d'une centaine d'ouvrages de référence.

En septembre 2000, l'Office québécois de la langue française (OQLF) lance le Grand dictionnaire terminologique (GDT), mis en ligne en septembre 2000. Le GDT comprend 3 millions de termes français et anglais du vocabulaire industriel, scientifique et commercial, dans 200 domaines d'activité. Il représente l'équivalent de 3.000 ouvrages de référence imprimés. Cette mise en ligne est le résultat d'un partenariat entre l'Office québécois de la langue française (OQLF), auteur du dictionnaire, et Semantix, société spécialisée dans les solutions logicielles linguistiques. Événement célébré par de nombreux linguistes, cette mise en ligne est un succès sans précédent pour un dictionnaire. Dès le premier mois, le GDT est consulté par 1,3 million de personnes, avec des pointes de 60.000 requêtes quotidiennes. La gestion de la base est ensuite assurée par Convera Canada. En février 2003, les requêtes sont au nombre de 3,5 millions par mois. Une nouvelle version du GDT est mise en ligne en mars 2003. Sa gestion est désormais assurée par l'OQLF lui-même, et non plus par une société prestataire.

Géré par le service de traduction de la Commission européenne, Eurodicautom est un dictionnaire multilingue de termes économiques, scientifiques, techniques et juridiques, avec une moyenne de 120.000 consultations quotidiennes. Il permet de combiner entre elles les onze langues officielles de l'Union européenne (allemand, anglais, danois, espagnol, finnois, français, grec, hollandais, italien, portugais, suédois), ainsi que le latin. Fin 2003, Eurodicautom annonce son intégration dans une base terminologique plus vaste regroupant les bases de plusieurs institutions de l'Union européenne. Cette nouvelle base traite non plus douze langues, mais une vingtaine, puisque l'Union européenne s'élargit à l'Est et passe de 15 à 25 membres en mai 2004, pour atteindre 27 membres en janvier 2007. Cette base terminologique voit le jour en mars 2007, sous le nom de IATE (Inter-Active Terminology for Europe), avec 1,4 million d'entrées dans 24 langues.

1997: Convergence multimédia

[Résumé]

La convergence multimédia peut être définie comme la convergence des secteurs de l'informatique, du téléphone et de la radiotélévision dans une industrie de la communication et de la distribution utilisant les mêmes autoroutes de l'information. Cette convergence entraîne l'unification progressive des secteurs liés à l'information (imprimerie, édition, presse, conception graphique, enregistrements sonores, films, etc.) suite à l'utilisation des techniques de numérisation. La convergence multimédia amène-t-elle des emplois nouveaux, comme l'assurent les employeurs, ou bien est-elle source de chômage, comme l'affirment les syndicats? Ce sujet est débattu dès 1997 lors du Colloque sur la convergence multimédia organisé par le Bureau international du travail (BIT) à Genève. Si, dans certains secteurs, ce phénomène entraîne de nouveaux emplois, par exemple ceux liés à la production audiovisuelle, d'autres secteurs sont soumis à d'inquiétantes restructurations. La convergence multimédia a de nombreux revers, par exemple des contrats précaires pour les salariés, l'absence de syndicats pour les télétravailleurs ou le droit d'auteur mis à mal pour les auteurs.

[En détail]

La numérisation permet de créer, d'enregistrer, de combiner, de stocker, de rechercher et de transmettre des textes, des sons et des images par des moyens simples et rapides. Des procédés similaires permettent le traitement de l'écriture, de la musique et du cinéma alors que, par le passé, ce traitement était assuré par des procédés différents sur des supports différents (papier pour l'écriture, bande magnétique pour la musique, celluloïd pour le cinéma). De plus, des secteurs distincts comme l'édition (qui produit des livres) et l'industrie musicale (qui produit des disques) travaillent de concert pour produire des CD-ROM.

La numérisation accélère considérablement le processus matériel de production. Dans la presse, alors qu'auparavant le personnel de production devait dactylographier les textes du personnel de rédaction, les journalistes envoient désormais directement leurs textes pour mise en page. Dans l'édition, le rédacteur, le concepteur artistique et l'infographiste travaillent souvent simultanément sur le même ouvrage. On assiste progressivement à la convergence de tous les secteurs liés à l'information: imprimerie, édition, presse, conception graphique, enregistrements sonores, films, radiodiffusion, etc. C'est ce qu'on appelle la convergence multimédia. On peut aussi la définir comme la convergence de l'informatique, du téléphone, de la radio et de la télévision dans une industrie de la communication et de la distribution utilisant les mêmes autoroutes.

Si certains secteurs voient l'apparition de nouveaux emplois, par exemple ceux liés à la production audiovisuelle, d'autres secteurs sont soumis à d'inquiétantes restructurations. La convergence multimédia a de nombreux revers, à savoir des contrats occasionnels et précaires pour les salariés, l'absence de syndicats pour les télétravailleurs, le droit d'auteur souvent mis à mal pour les auteurs, etc. Et, à l'exception du droit d'auteur, vu l'enjeu financier qu'il représente, il est rare que ces problèmes fassent la Une des journaux.

La convergence multimédia amène-t-elle des emplois nouveaux, comme l'assurent les employeurs, ou bien est-elle source de chômage, comme l'affirment les syndicats? Ce sujet est débattu dès janvier 1997 lors du Colloque sur la convergence multimédia organisé par le Bureau international du travail (BIT) à Genève.

Si elle accélère le processus de production, l'automatisation des méthodes de travail entraîne une diminution de l'intervention humaine et donc un accroissement du chômage. Dans la presse comme dans l'édition, la mise en page automatique permet de combiner rédaction et composition. Dans les services publicitaires aussi, la conception graphique et les tâches commerciales sont maintenant intégrées. L'informatique permet à certains professionnels de s'installer à leur compte, une solution choisie par 30% des salariés ayant perdu leur emploi.

Au Royaume-Uni, les fonctions de correction d'épreuves et de rédaction s'effectuent désormais à domicile, le plus souvent par des travailleurs qui ont pris le statut d'indépendants par suite de fusions d'entreprises, délocalisations ou licenciements. "Or cette forme d'emploi tient plus du travail précaire que du travail indépendant, car ces personnes n'ont que peu d'autonomie et sont généralement tributaires d'une seule maison d'édition", analyse Peter Leisink, professeur associé d'études sociales à l'Université d'Utrecht (Pays-Bas).

A part quelques cas mis en avant par les organisations d'employeurs, la convergence multimédia entraîne des suppressions massives d'emplois. Selon Michel Muller, secrétaire général de la FILPAC (Fédération des industries du livre, du papier et de la communication), les industries graphiques françaises ont perdu 20.000 emplois en dix ans. Entre 1987 et 1996, les effectifs passent de 110.000 à 90.000 salariés. Les entreprises doivent mettre sur pied des plans sociaux coûteux pour favoriser le reclassement des personnes licenciées, en créant des emplois souvent artificiels, alors qu'il aurait été préférable de financer des études fiables sur la manière d'équilibrer créations et suppressions d'emplois quand il était encore temps.

Partout dans le monde, de nombreux postes à faible qualification technique sont remplacés par des postes exigeant des qualifications techniques élevées. Les personnes peu qualifiées sont licenciées. D'autres suivent une formation professionnelle complémentaire, parfois auto-financée et prise sur leur temps libre, et cette formation professionnelle ne garantit pas pour autant le réemploi.

Directeur du géant des télécommunications AT&T aux Etats-Unis, Walter Durling insiste sur le fait que les nouvelles technologies ne changeront pas fondamentalement la situation des salariés au sein de l'entreprise. L'invention du film n'a pas tué le théâtre et celle de la télévision n'a pas fait disparaître le cinéma. Les entreprises devraient créer des emplois liés aux nouvelles technologies et les proposer à ceux qui sont obligés de quitter d'autres postes devenus obsolètes. Des arguments bien théoriques alors que le problème est plutôt celui du pourcentage. Combien de créations de postes pour combien de licenciements?

De leur côté, les syndicats préconisent la création d'emplois par l'investissement, l'innovation, la formation professionnelle aux nouvelles technologies, la reconversion des travailleurs dont les emplois sont supprimés, des conditions équitables pour les contrats et les conventions collectives, la défense du droit d'auteur, une meilleure protection des

travailleurs dans le secteur artistique et enfin la défense des télétravailleurs en tant que travailleurs à part entière.

Malgré tous les efforts des syndicats, la situation deviendra-elle aussi dramatique que celle décrite dans une note du rapport de ce colloque, demandant si "les individus seront forcés de lutter pour survivre dans une jungle électronique avec les mécanismes de survie qui ont été mis au point au cours des précédentes décennies?"

1998: Gallica

[Résumé]

Gallica est le secteur numérique de la Bibliothèque nationale de France (BnF), mis en ligne en octobre 1997. En 1998, le serveur stocke 2.500 livres numérisés en mode image complétés par les 250 volumes numérisés en mode texte de la base Frantext de l'INaLF (Institut national de la langue française). Gallica devient rapidement la plus grande bibliothèque numérique francophone du réseau. Pour des raisons de coût, les documents sont essentiellement numérisés en mode image. En décembre 2006, Gallica propose tous les documents libres de droits du fonds numérisé de la BnF, à savoir 90.000 ouvrages numérisés (fascicules de presse compris) et 80.000 images couvrant une période allant du Moyen-Age au début du 20e siècle, avec plus d'un million de consultations par mois. En 2007, Gallica débute la conversion en mode texte des livres numérisés en mode image pour favoriser l'accès à leur contenu.

[En détail]

Secteur numérique de la Bibliothèque nationale de France (BnF), Gallica est inauguré en octobre 1997 avec des textes et images du 19e siècle francophone, le 19e siècle étant le "siècle de l'édition et de la presse moderne, siècle du roman mais aussi des grandes synthèses historiques et philosophiques, siècle scientifique et technique" (extrait du site web). A l'époque, le serveur stocke 2.500 livres numérisés en mode image complétés par les 250 volumes numérisés en mode texte de la base Frantext de l'INaLF (Institut national de la langue française). Classées par discipline, ces ressources sont complétées par une chronologie du 19e siècle et une synthèse sur les grands courants en histoire, sciences politiques, droit, économie, littérature, philosophie, sciences et histoire des sciences. Le site propose aussi un échantillon de la future iconothèque numérique, à savoir le fonds du photographe Eugène Atget, une sélection de documents sur l'écrivain Pierre Loti, une collection d'images de l'Ecole nationale des ponts et chaussées sur les grands travaux liés à la révolution industrielle en France, et enfin un choix de livres illustrés de la Bibliothèque du Musée de l'Homme.

Gallica se considère moins comme une banque de données numérisées que comme un "laboratoire dont l'objet est d'évaluer les conditions d'accès et de consultation à distance des documents numériques", lit-on sur le site web à la fin de 1997. Le but est d'expérimenter la navigation dans ces collections, en permettant aussi bien le libre parcours du chercheur ou du curieux que des recherches pointues.

Début 1998, Gallica annonce 100.000 volumes et 300.000 images pour la fin 1999, avec un accroissement rapide des collections ensuite. Sur les 100.000 volumes prévus, qui représenteront 30 millions de pages numérisées, plus du tiers concerne le 19e siècle. Quant aux 300.000 images fixes, la moitié appartient aux départements spécialisés de la BnF (Estampes et photographie, Manuscrits, Arts du spectacle, Monnaies et médailles, etc.). L'autre moitié provient de collections d'établissements publics (musées et bibliothèques, Documentation française, Ecole nationale des ponts et chaussées, Institut Pasteur,

Observatoire de Paris, etc.) ou privés (agences de presse dont Magnum, l'Agence France-Presse, Sygma, Rapho, etc.).

En mai 1998, Gallica revoit ses espérances à la baisse et modifie quelque peu ses orientations premières. Dans un article du quotidien Le Figaro du 3 juin 1998, Jérôme Strazzulla, journaliste, écrit que la BnF est "passée d'une espérance universaliste, encyclopédique, à la nécessité de choix éditoriaux pointus". Dans le même article, il interviewe le président de la BnF, Jean-Pierre Angremy, qui rapporte la décision du comité éditorial de Gallica: "Nous avons décidé d'abandonner l'idée d'un vaste corpus encyclopédique de cent mille livres, auquel on pourrait sans cesse reprocher des trous. Nous nous orientons aujourd'hui vers des corpus thématiques, aussi complets que possibles, mais plus restreints. (...) Nous cherchons à répondre, en priorité, aux demandes des chercheurs et des lecteurs." Le premier corpus aurait trait aux voyages en France, avec mise en ligne prévue en 2000. Ce corpus rassemblerait des textes, estampes et photographies du 16^e siècle à 1920. Les corpus envisagés ensuite auront les thèmes suivants: Paris, les voyages en Afrique des origines à 1920, les utopies, et les mémoires des Académies des sciences de province.

En 2003, Gallica donne accès à tous les documents libres de droit du fonds numérisé de la BnF, à savoir 70.000 ouvrages et 80.000 images allant du Moyen-Age au début du 20^e siècle. Mais, de l'avis de nombreux usagers, les fichiers sont très lourds puisque les livres sont numérisés en mode image, et l'accès en est très long. Chose tout aussi problématique, la numérisation en mode image n'autorise pas la recherche textuelle alors que Gallica se trouve être la plus grande bibliothèque numérique francophone en nombre de titres disponibles en ligne. Seule une petite collection de livres (1.117 titres en février 2004) est numérisée en mode texte, celle de la base Frantext de l'ATILF (Analyse et traitement informatique de la langue française, le laboratoire ayant succédé à l'INaLF), intégrée dans Gallica.

En février 2005, Gallica compte 76.000 ouvrages. A la même date, la BnF annonce la mise en ligne prochaine (entre 2006 et 2009) de la presse française parue entre 1826 et 1944, à savoir 22 titres représentant 3,5 millions de pages. Début 2006, les premiers journaux disponibles en ligne sont Le Figaro (fondé en 1826), La Croix (fondée en 1883), L'Humanité (fondée en 1904) et Le Temps (fondé en 1861 et disparu en 1942). En décembre 2006, les collections comprennent 90.000 ouvrages numérisés (fascicules de presse compris), 80.000 images et des dizaines d'heures de ressources sonores. Gallica débute aussi la conversion en mode texte des livres numérisés en mode image pour favoriser l'accès à leur contenu.

1998: Bibliothécaires numériques

[Résumé]

En 1998, Peter Raggett est sous-directeur (puis directeur) de la Bibliothèque centrale de l'OCDE (Organisation de coopération et de développement économiques). Il explique en juin 1998: "L'internet offre aux chercheurs un stock d'informations considérable. (...) A mon avis, les bibliothécaires auront un rôle important à jouer pour améliorer la recherche et l'organisation de l'information sur le réseau. Je prévois aussi une forte expansion de l'internet pour l'enseignement et la recherche. Les bibliothèques seront amenées à créer des bibliothèques numériques permettant à un étudiant de suivre un cours proposé par une institution à l'autre bout du monde. La tâche du bibliothécaire sera de filtrer les informations pour le public. Personnellement, je me vois de plus en plus devenir un bibliothécaire virtuel. Je n'aurai pas l'occasion de rencontrer les usagers, ils me contacteront plutôt par courriel, par téléphone ou par fax, j'effectuerai la recherche et je leur enverrai les résultats par voie électronique."

[En détail]

Voici deux expériences, relatées par Peter Raggett, sous-directeur de la Bibliothèque centrale de l'OCDE, et par Bruno Didier, webmestre de la Bibliothèque de l'Institut Pasteur.

Bibliothèque centrale de l'OCDE

Située à Paris, l'OCDE (Organisation de coopération et de développement économiques) regroupe trente pays membres. Au noyau d'origine, constitué des pays d'Europe de l'Ouest et d'Amérique du Nord, viennent s'ajouter le Japon, l'Australie, la Nouvelle-Zélande, la Finlande, le Mexique, la République tchèque, la Hongrie, la Pologne et la Corée. Réservé aux fonctionnaires de l'organisation, la bibliothèque centrale permet la consultation de quelque 60.000 monographies et 2.500 périodiques imprimés. En ligne depuis 1996, les pages intranet du CDI deviennent une source d'information indispensable pour le personnel.

Peter Raggett est sous-directeur (puis directeur) du centre de documentation et d'information (CDI) de l'OCDE. "Je dois filtrer l'information pour les usagers de la bibliothèque, ce qui signifie que je dois bien connaître les sites et les liens qu'ils proposent, explique-t-il en juin 1998. J'ai sélectionné plusieurs centaines de sites pour en favoriser l'accès à partir de l'intranet de l'OCDE. Cette sélection fait partie du bureau de référence virtuel proposé par la bibliothèque à l'ensemble du personnel. Outre de nombreux liens, ce bureau de référence contient des pages recensant les articles, monographies et sites web correspondant aux différents projets de recherche en cours à l'OCDE, l'accès en réseau aux CD-ROM et une liste mensuelle des nouveaux titres."

Comment voit-il l'avenir de la profession? "L'internet offre aux chercheurs un stock d'informations considérable. Le problème pour eux est de trouver ce qu'ils cherchent. Jamais auparavant on n'avait senti une telle surcharge d'informations, comme on la sent maintenant quand on tente de trouver un renseignement sur un sujet précis en utilisant les moteurs de

recherche disponibles sur l'internet. A mon avis, les bibliothécaires auront un rôle important à jouer pour améliorer la recherche et l'organisation de l'information sur le réseau. Je prévois aussi une forte expansion de l'internet pour l'enseignement et la recherche. Les bibliothèques seront amenées à créer des bibliothèques numériques permettant à un étudiant de suivre un cours proposé par une institution à l'autre bout du monde. La tâche du bibliothécaire sera de filtrer les informations pour le public. Personnellement, je me vois de plus en plus devenir un bibliothécaire virtuel. Je n'aurai pas l'occasion de rencontrer les usagers, ils me contacteront plutôt par courriel, par téléphone ou par fax, j'effectuerai la recherche et je leur enverrai les résultats par voie électronique."

Bibliothèque de l'Institut Pasteur

En 1999, Bruno Didier est bibliothécaire à l'Institut Pasteur (Paris), une fondation privée dont le but est la prévention et le traitement des maladies infectieuses par la recherche, l'enseignement et des actions de santé publique. Séduit par les perspectives qu'offre le réseau pour la recherche documentaire, Bruno Didier crée le site web de la bibliothèque en 1996 et devient son webmestre.

"Le site web de la bibliothèque a pour vocation principale de servir la communauté pasteurienne, relate-t-il en août 1999. Il est le support d'applications devenues indispensables à la fonction documentaire dans un organisme de cette taille: bases de données bibliographiques, catalogue, commande de documents et bien entendu accès à des périodiques en ligne. C'est également une vitrine pour nos différents services, en interne mais aussi dans toute la France et à l'étranger. Il tient notamment une place importante dans la coopération documentaire avec les instituts du réseau Pasteur à travers le monde. Enfin j'essaie d'en faire une passerelle adaptée à nos besoins pour la découverte et l'utilisation d'internet. (...) Je développe et maintiens les pages du serveur, ce qui s'accompagne d'une activité de veille régulière. Par ailleurs je suis responsable de la formation des usagers, ce qui se ressent dans mes pages. Le web est un excellent support pour la formation, et la plupart des réflexions actuelles sur la formation des usagers intègrent cet outil."

Son activité professionnelle a changé de manière assez radicale, tout comme celle de ses collègues. "C'est à la fois dans nos rapports avec l'information et avec les usagers que les changements ont eu lieu, explique-t-il. Nous devenons de plus en plus des médiateurs, et peut-être un peu moins des conservateurs. Mon activité actuelle est typique de cette nouvelle situation: d'une part dégager des chemins d'accès rapides à l'information et mettre en place des moyens de communication efficaces, d'autre part former les utilisateurs à ces nouveaux outils. Je crois que l'avenir de notre métier passe par la coopération et l'exploitation des ressources communes. C'est un vieux projet certainement, mais finalement c'est la première fois qu'on dispose enfin des moyens de le mettre en place."

1998: Web multilingue

[Résumé]

De pratiquement anglophone à ses débuts, le web devient multilingue. Consultant en marketing internet de produits et services de traduction, Randy Hobler écrit en septembre 1998: "Comme l'internet n'a pas de frontières nationales, les internautes s'organisent selon d'autres critères propres au médium. En termes de multilinguisme, vous avez des communautés virtuelles, par exemple ce que j'appelle les "nations des langues", tous ces internautes qu'on peut regrouper selon leur langue maternelle quel que soit leur lieu géographique. Ainsi la nation de la langue espagnole inclut non seulement les internautes d'Espagne et d'Amérique latine, mais aussi tous les Hispanophones vivant aux Etats-Unis, ou encore ceux qui parlent espagnol au Maroc." Professeur d'espagnol en entreprise et traductrice, Maria Victoria Marinetti ajoute en août 1999: «Il est très important de pouvoir communiquer en différentes langues. Je dirais même que c'est obligatoire, car l'information donnée sur le net est à destination du monde entier, alors pourquoi ne l'aurions-nous pas dans notre propre langue ou dans la langue que nous souhaitons lire? Information mondiale, mais pas de vaste choix dans les langues, ce serait contradictoire, pas vrai?»

[En détail]

De pratiquement anglophone à ses débuts, le web devient multilingue.

Consultant en marketing internet de produits et services de traduction, Randy Hobler écrit en septembre 1998: "Comme l'internet n'a pas de frontières nationales, les internautes s'organisent selon d'autres critères propres au médium. En termes de multilinguisme, vous avez des communautés virtuelles, par exemple ce que j'appelle les "nations des langues", tous ces internautes qu'on peut regrouper selon leur langue maternelle quel que soit leur lieu géographique. Ainsi la nation de la langue espagnole inclut non seulement les internautes d'Espagne et d'Amérique latine, mais aussi tous les Hispanophones vivant aux Etats-Unis, ou encore ceux qui parlent espagnol au Maroc."

Bruno Didier, webmestre de la Bibliothèque de l'Institut Pasteur, écrit en août 1999: "Internet n'est une propriété ni nationale, ni linguistique. C'est un vecteur de culture, et le premier support de la culture, c'est la langue. Plus il y a de langues représentées dans leur diversité, plus il y aura de cultures sur internet. Je ne pense pas qu'il faille justement céder à la tentation systématique de traduire ses pages dans une langue plus ou moins universelle. Les échanges culturels passent par la volonté de se mettre à la portée de celui vers qui on souhaite aller. Et cet effort passe par l'appréhension de sa langue. Bien entendu c'est très utopique comme propos. Concrètement, lorsque je fais de la veille, je peste dès que je rencontre des sites norvégiens ou brésiliens sans un minimum d'anglais."

L'anglais reste en effet prépondérant et ceci n'est pas près de disparaître. Comme indiqué en janvier 1999 par Marcel Grangier, responsable de la section française des services linguistiques centraux de l'Administration fédérale suisse, "cette suprématie n'est pas un mal en soi, dans la mesure où elle résulte de réalités essentiellement statistiques (plus de PC par

habitant, plus de locuteurs de cette langue, etc.). La riposte n'est pas de "lutter contre l'anglais" et encore moins de s'en tenir à des jérémiades, mais de multiplier les sites en d'autres langues. Notons qu'en qualité de service de traduction, nous préconisons également le multilinguisme des sites eux-mêmes. La multiplication des langues présentes sur internet est inévitable, et ne peut que bénéficier aux échanges multiculturels."

C'est aussi l'avis de Maria Victoria Marinetti, professeur d'espagnol en entreprise et traductrice, qui écrit en août 1999: "Il est très important de pouvoir communiquer en différentes langues. Je dirais même que c'est obligatoire, car l'information donnée sur le net est à destination du monde entier, alors pourquoi ne l'aurions-nous pas dans notre propre langue ou dans la langue que nous souhaitons lire? Information mondiale, mais pas de vaste choix dans les langues, ce serait contradictoire, pas vrai?"

Professeur en technologies de la communication à la Webster University de Genève, Henri Slettenhaar insiste également sur la nécessité de sites bilingues, dans la langue originale et en anglais. "Les communautés locales présentes sur le web devraient en tout premier lieu utiliser leur langue pour diffuser des informations, écrit-il en décembre 1998. Si elles veulent également présenter ces informations à la communauté mondiale, celles-ci doivent être aussi disponibles en anglais. Je pense qu'il existe un réel besoin de sites bilingues. (...) Mais je suis enchanté qu'il existe maintenant tant de documents disponibles dans leur langue originale. Je préfère de beaucoup lire l'original avec difficulté plutôt qu'une traduction médiocre."

En août 1999, il ajoute: "A mon avis, il existe deux types de recherches sur le web. La première est la recherche globale dans le domaine des affaires et de l'information. Pour cela, la langue est d'abord l'anglais, avec des versions locales si nécessaire. La seconde, ce sont les informations locales de tous ordres dans les endroits les plus reculés. Si l'information est à destination d'une ethnie ou d'un groupe linguistique, elle doit d'abord être dans la langue de l'ethnie ou du groupe, avec peut-être un résumé en anglais."

En été 2000, les usagers non anglophones atteignent puis dépassent la barre des 50%. Ce pourcentage continue ensuite d'augmenter, comme le montrent les statistiques de la société Global Reach, mises à jour à intervalles réguliers. Le nombre d'usagers non anglophones est de 52,5% en été 2001, 57% en décembre 2001, 59,8% en avril 2002, 64,4% en septembre 2003 (dont 34,9% d'Européens non anglophones et 29,4% d'Asiatiques) et 64,2% en mars 2004 (dont 37,9% d'Européens non anglophones et 33% d'Asiatiques).

1999: Format Open eBook

[Résumé]

Créé en septembre 1999, l'OeB (open ebook) est un format de livre numérique basé sur le langage XML (extensible markup language) pour normaliser le contenu, la structure et la présentation des livres numériques. Le format OeB est défini par l'OeBPS (open ebook publication structure), développée par l'Open eBook Forum (OeBF), un consortium industriel international fondé en janvier 2000 pour regrouper constructeurs, concepteurs de logiciels, éditeurs, libraires et spécialistes du numérique (85 participants en 2002). Téléchargeable gratuitement, l'OeBPS dispose d'une version ouverte et gratuite appartenant au domaine public. La version originale est destinée aux professionnels de la publication puisqu'elle doit être associée à une technologie normalisée de gestion des droits numériques, et donc à un système de DRM (digital rights management) permettant de contrôler l'accès des livres numériques sous droits. En avril 2005, l'Open eBook Forum devient l'International Digital Publishing Forum (IDPF), et le format OeB devient le format ePub.

[En détail]

Les années 1998 et 1999 sont marquées par la prolifération des formats, chacun lançant son propre format de livre numérique dans le cadre d'un marché naissant promis à une expansion rapide. Aux formats classiques - format texte, Word, HTML (hypertext markup language), XML (extensible markup language) et PDF (portable document format) - s'ajoutent des formats propriétaires créés par plusieurs sociétés pour une lecture sur leurs propres logiciels: Glassbook Reader, Rocket eBook Reader, Peanut Reader, Franklin Reader, logiciel de lecture Cytale, Gemstar eBook Reader, Palm Reader, etc., ces logiciels correspondant le plus souvent à un appareil donné: Rocket eBook, eBookMan (Franklin), Cybook (Cytale), Gemstar eBook, Palm Pilot, etc.

Inquiets pour l'avenir du livre numérique qui, à peine né, propose presque autant de formats que de titres, certains insistent sur l'intérêt, sinon la nécessité, d'un format unique. A l'instigation du NIST (National Institute of Standards and Technology) aux Etats-Unis, l'Open eBook Initiative voit le jour en juin 1998 et constitue un groupe de travail de 25 personnes (Open eBook Authoring Group). Ce groupe élabore l'OeB (open ebook), un format basé sur le langage XML (extensible markup language) et destiné à normaliser le contenu, la structure et la présentation des livres numériques. Le format OeB est défini par l'OeBPS (open ebook publication structure), dont la version 1.0 est disponible en septembre 1999. L'OeBPS dispose d'une version ouverte et gratuite appartenant au domaine public. La version originale est utilisée uniquement par les professionnels de la publication, puisqu'il doit être associé à un système de gestion des droits numériques (DRM - digital rights management).

Fondé en janvier 2000 pour prendre la suite de l'Open eBook Initiative, l'OeBF (Open eBook Forum) est un consortium industriel international regroupant constructeurs, concepteurs de logiciels, éditeurs, libraires et spécialistes du numérique (85 participants en 2002) dans l'optique de développer le format OeB (open ebook) et l'OeBPS (open ebook publication structure). Le format OeB devient un standard qui sert lui-même de base à de nombreux

formats, par exemple LIT (pour le Microsoft Reader) ou PRC (pour le Mobipocket Reader). En avril 2005, l'Open eBook Forum devient l'International Digital Publishing Forum (IDPF) et le format OeB laisse la place au format ePub.

Outre le format PDF (non basé sur l'OeB puisque créé en 1993) et le format LIT (basé sur l'OeB), le troisième grand format est le format PRC, lisible sur le Mobipocket Reader et destiné en priorité aux appareils mobiles tels que PDA, téléphones et smartphones. Créée en mars 2000 par Thierry Brethes et Nathalie Ting, la société Mobipocket, basée à Paris, se spécialise d'emblée dans la lecture et la distribution sécurisée de livres numériques sur assistant personnel (PDA). Le Mobipocket Reader est "universel", c'est-à-dire utilisable sur tout assistant personnel (Palm Pilot, Pocket PC, eBookMan, Psion, etc.). En avril 2002, la société lance un Mobipocket Reader pour ordinateur. Au printemps 2003, le Mobipocket Reader équipe les premiers smartphones de Nokia et Sony Ericsson. A la même date, le nombre de livres lisibles sur le Mobipocket Reader est de 6.000 titres dans plusieurs langues (français, anglais, allemand, espagnol), distribués soit sur le site de Mobipocket soit dans les librairies partenaires. Le système de gestion des droits numériques est le Mobipocket DRM System. En avril 2005, Mobipocket est racheté par la grande librairie en ligne Amazon.com.

1999: Auteurs numériques

[Résumé]

Jean-Paul, auteur hypermédia et webmestre des Cotres furtifs, relate en août 1999: "L'internet va me permettre de me passer des intermédiaires: compagnies de disques, éditeurs, distributeurs... Il va surtout me permettre de formaliser ce que j'ai dans la tête et dont l'imprimé (la micro-édition, en fait) ne me permettait de donner qu'une approximation." De nombreux genres voient le jour: sites d'écriture hypermédia, oeuvres de fiction hypertexte, romans multimédias, hyper-romans, etc. Jean-Pierre Balpe, chercheur et écrivain, lance le premier mail-roman francophone en 2001. Cette expérience renforce sa "conviction que les technologies numériques sont une chance extraordinaire du renouvellement du littéraire." Le texte fusionne aussi de plus en plus avec l'image et le son.

[En détail]

Voici trois expériences, relatées par Murray Suid, auteur de livres pédagogiques, Jean-Paul, webmestre d'un site hypermédia, et Jean-Pierre Balpe, auteur d'un mail-roman.

Livres pédagogiques

Murray Suid vit à Palo Alto, dans la Silicon Valley, en Californie. Il est l'auteur de livres pédagogiques, de livres pour enfants, d'oeuvres multimédias et de scénarios. Dès septembre 1998, il préconise une solution choisie depuis par de nombreux auteurs. "Un livre peut avoir un prolongement sur le web – et donc vivre en partie dans le cyberspace, explique-t-il. L'auteur peut ainsi aisément l'actualiser et le corriger, alors qu'auparavant il devait attendre longtemps, jusqu'à l'édition suivante, quand il y en avait une. (...) Je ne sais pas si je publierai des livres sur le web, au lieu de les publier en version imprimée. J'utiliserai peut-être ce nouveau support si les livres deviennent multimédias. Pour le moment, je participe au développement de matériel pédagogique multimédia. C'est un nouveau type de matériel qui me plaît beaucoup et qui permet l'interactivité entre des textes, des films, des bandes sonores et des graphiques qui sont tous reliés les uns aux autres."

Un an après, en août 1999, il ajoute: "En plus des livres complétés par un site web, je suis en train d'adopter la même formule pour mes oeuvres multimédias – qui sont sur CD-ROM – afin de les actualiser et d'enrichir leur contenu." Quelques mois plus tard, l'intégralité de ses oeuvres multimédias est sur le réseau, et le matériel pédagogique auquel il contribue est conçu non plus pour diffusion sur CD-ROM, mais directement pour diffusion sur le web. D'entreprise multimédia, la société de logiciels éducatifs qui l'emploie devient une entreprise internet.

Site hypermédia

Jean-Paul, auteur hypermédia et webmestre des Cotres furtifs, relate en août 1999: "L'internet va me permettre de me passer des intermédiaires: compagnies de disques, éditeurs, distributeurs... Il va surtout me permettre de formaliser ce que j'ai dans la tête et

dont l'imprimé (la micro-édition, en fait) ne me permettait de donner qu'une approximation. Puis les intermédiaires prendront tout le pouvoir. Il faudra alors chercher ailleurs, là où l'herbe est plus verte..."

En juin 2000, il se penche sur l'apport de l'hyperlien dans son travail: "La navigation par hyperliens se fait en rayon (j'ai un centre d'intérêt et je clique méthodiquement sur tous les liens qui s'y rapportent) ou en louvoiements (de clic en clic, à mesure qu'ils apparaissent, au risque de perdre de vue mon sujet). Bien sûr, les deux sont possibles avec l'imprimé. Mais la différence saute aux yeux: feuilleter n'est pas cliquer. L'internet n'a donc pas changé ma vie, mais mon rapport à l'écriture. On n'écrit pas de la même manière pour un site que pour un scénario, une pièce de théâtre, etc. (...) Depuis, j'écris (compose, mets en page, en scène) directement à l'écran. L'état "imprimé" de mon travail n'est pas le stade final, le but; mais une forme parmi d'autres, qui privilégie la linéarité et l'image, et qui exclut le son et les images animées. (...)

C'est finalement dans la publication en ligne (l'entoilage?) que j'ai trouvé la mobilité, la fluidité que je cherchais. Le maître mot y est "chantier en cours", sans palissades. Accouchement permanent, à vue, comme le monde sous nos yeux. Provisoire, comme la vie qui tâtonne, se cherche, se déprend, se reprend. Avec évidemment le risque souligné par les gutenbergs, les orphelins de la civilisation du livre: plus rien n'est sûr. Il n'y a plus de source fiable, elles sont trop nombreuses, et il devient difficile de distinguer un clerc d'un gourou. Mais c'est un problème qui concerne le contrôle de l'information. Pas la transmission des émotions."

Mail-roman

Jean-Pierre Balpe est directeur du département hypemédias de l'Université Paris 8, chercheur et écrivain. Il lance le premier mail-roman francophone en 2001. Pendant très exactement cent jours, entre le 11 avril et le 19 juillet 2001, il diffuse quotidiennement un chapitre de Rien n'est sans dire auprès de cinq cents personnes - sa famille, ses amis, ses collègues, etc. - en intégrant les réponses et les réactions des lecteurs. Racontée par un narrateur, l'histoire est celle de Stanislas et Zita, qui vivent une passion tragique déchirée par une sombre histoire politique. "Cette idée d'un mail-roman m'est venue tout naturellement, raconte l'auteur en février 2002. D'une part en me demandant depuis quelque temps déjà ce qu'internet peut apporter sur le plan de la forme à la littérature (...) et d'autre part en lisant de la littérature "épistolaire" du 18e siècle, ces fameux "romans par lettres". Il suffit alors de transposer: que peut être le "roman par lettres" aujourd'hui?"

Jean-Pierre Balpe tire plusieurs conclusions de cette expérience: "D'abord c'est un "genre": depuis, plusieurs personnes m'ont dit lancer aussi un mail-roman. Ensuite j'ai aperçu quantité de possibilités que je n'ai pas exploitées et que je me réserve pour un éventuel travail ultérieur. La contrainte du temps est ainsi très intéressante à exploiter: le temps de l'écriture bien sûr, mais aussi celui de la lecture: ce n'est pas rien de mettre quelqu'un devant la nécessité de lire, chaque jour, une page de roman. Ce "pacte" a quelque chose de diabolique. Et enfin le renforcement de ma conviction que les technologies numériques sont une chance extraordinaire du renouvellement du littéraire."

2000: yourDictionary.com

[Résumé]

yourDictionary.com est co-fondé par Robert Beard en février 2000, dans le prolongement de son ancien site, A Web of Online Dictionaries, créé dès 1995. En septembre 2003, yourDictionary.com, devenu un portail de référence, répertorie plus de 1.800 dictionnaires dans 250 langues, ainsi que de nombreux outils linguistiques: vocabulaires, grammaires, glossaires, méthodes de langues, etc. En avril 2007, le répertoire comprend 2.500 dictionnaires et grammaires dans 300 langues. Soucieux de servir toutes les langues sans exception, le portail propose une section spécifique (Endangered Language Repository) consacrée aux langues menacées d'extinction. D'après l'encyclopédie Ethnologue: Languages of the World, publiée par SIL International et disponible gratuitement sur le web, il existerait 6.912 langues dans le monde.

[En détail]

yourDictionary.com est co-fondé par Robert Beard en février 2000, dans le prolongement de son ancien site, A Web of Online Dictionaries, créé dès 1995. En septembre 2003, yourDictionary.com, devenu un portail de référence, répertorie 1.800 dictionnaires dans 250 langues, ainsi que de nombreux outils linguistiques: vocabulaires, grammaires, glossaires, méthodes de langues, etc. En avril 2007, le répertoire comprend 2.500 dictionnaires et grammaires dans 300 langues. Soucieux de servir toutes les langues sans exception, le site propose une section spécifique (Endangered Language Repository) consacrée aux langues menacées d'extinction.

Robert Beard explique en janvier 2000: "Les langues menacées sont essentiellement des langues non écrites. Un tiers seulement des quelque 6.000 langues existant dans le monde sont à la fois écrites et parlées. Je ne pense pourtant pas que le web va contribuer à la perte de l'identité des langues et j'ai même le sentiment que, à long terme, il va renforcer cette identité. Par exemple, de plus en plus d'Indiens d'Amérique contactent des linguistes pour leur demander d'écrire la grammaire de leur langue et les aider à élaborer des dictionnaires. Pour eux, le web est un instrument à la fois accessible et très précieux d'expression culturelle."

Caoimhín Ó Donnáil est professeur d'informatique à l'Institut Sabhal Mór Ostaig, situé sur l'île de Skye, en Ecosse. Il dispense ses cours en gaélique écossais. Il indique pour sa part en mai 2001: "En ce qui concerne l'avenir des langues menacées, l'internet accélère les choses dans les deux sens. Si les gens ne se soucient pas de préserver les langues, l'internet et la mondialisation qui l'accompagne accéléreront considérablement la disparition de ces langues. Si les gens se soucient vraiment de les préserver, l'internet constituera une aide irremplaçable."

Caoimhín est également le webmestre du site de l'institut, bilingue anglais-gaélique, qui se trouve être la principale source d'information mondiale sur le gaélique écossais. Sur ce site, il tient à jour European Minority Languages, une liste de langues européennes minoritaires elle

aussi bilingue anglais-gaélique, avec classement par ordre alphabétique de langues et par famille linguistique. Il ajoute: "Nos étudiants utilisent un correcteur d'orthographe en gaélique et une base terminologique en ligne en gaélique. (...) Il est maintenant possible d'écouter la radio en gaélique (écossais et irlandais) en continu sur l'internet partout dans le monde. Une réalisation particulièrement importante a été la traduction en gaélique du logiciel de navigation Opera. C'est la première fois qu'un logiciel de cette taille est disponible en gaélique."

Publiée par SIL International (SIL: Summer Institute of Linguistics), l'encyclopédie Ethnologue: Languages of the World existe à la fois en version web (gratuite), sur CD-ROM (payant) et en version imprimée (payante). Barbara Grimes, sa directrice de publication entre 1971 et 2000 (8e-14e éditions), relate en janvier 2000: "Il s'agit d'un catalogue des langues dans le monde, avec des informations sur les endroits où elles sont parlées, une estimation du nombre de personnes qui les parlent, la famille linguistique à laquelle elles appartiennent, les autres termes utilisés pour ces langues, les noms de dialectes, d'autres informations socio-linguistiques et démographiques, les dates des Bibles publiées, un index des noms de langues, un index des familles linguistiques et des cartes géographiques relatives aux langues." En avril 2007, cette encyclopédie répertorie 6.912 langues selon plusieurs critères (pays, nom de la langue, code de la langue attribué par le SIL, famille de langues), avec moteur de recherche.

2000: Bible de Gutenberg en ligne

[Résumé]

En novembre 2000, la version numérique de la Bible de Gutenberg est mise en ligne sur le site de la British Library. Datée de 1454 ou 1455, cette Bible est le premier ouvrage imprimé par Gutenberg dans son atelier de Mayence, en Allemagne. Sur les 180 exemplaires d'origine, 48 exemplaires, dont certains incomplets, existeraient toujours. La British Library en possède deux versions complètes et une partielle. En mars 2000, dix chercheurs et experts techniques de l'Université Keio de Tokyo et de NTT (Nippon Telegraph and Telephone Communications) viennent travailler sur place pendant deux semaines pour numériser les deux versions complètes, légèrement différentes.

2000: Distributed Proofreaders

[Résumé]

Distributed Proofreaders (DP) est lancé en octobre 2000 par Charles Franks pour aider à la numérisation des livres du domaine public. Le site web permet la correction partagée en fragmentant les livres en pages pouvant être relues par des correcteurs différents. Destiné à intensifier la production de livres pour le Projet Gutenberg, grande bibliothèque numérique mondiale au format texte, Distributed Proofreaders en devient rapidement la principale source. A titre indicatif, il est suggéré de relire une page par jour. Cela semble peu, mais une page multipliée par des milliers de volontaires représente un chiffre considérable. 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 était de 300 à 400 relecteurs quotidiens produisant entre 4.000 et 7.000 pages par jour, à savoir quatre pages par minute. Distributed Proofreaders produit 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.

[En détail]

Conçu en octobre 2000 par Charles Franks pour contribuer à la numérisation des livres du domaine public, Distributed Proofreaders (DP) permet la correction partagée en fragmentant les livres en pages pouvant être relues par des correcteurs différents. Destiné à intensifier la production de livres pour le Projet Gutenberg, grande bibliothèque numérique mondiale au format texte, Distributed Proofreaders en devient rapidement la principale source. Il est officiellement affilié au Projet Gutenberg en 2002.

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 produit 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. A cette date, 700 volontaires se connectent chaque jour et 3.000 volontaires se connectent chaque mois. Les volontaires n'ont aucun quota à respecter. A titre indicatif, il est suggéré de relire une page par jour. Cela semble peu, mais une page multipliée par des milliers de volontaires représente un chiffre considérable.

En janvier 2004 est lancé en parallèle Distributed Proofreaders Europe (DP Europe) pour alimenter le site du Projet Gutenberg Europe (PG Europe). Créé à l'initiative du Projet Rastko, basé à Belgrade (Serbie), DP Europe est calqué sur le site original de Distributed Proofreaders, pour gérer la relecture partagée de PG Europe. Dès ses débuts, DP Europe est un site multilingue, qui prend en compte les principales langues nationales. Grâce à des traducteurs volontaires, le site de DP Europe est disponible en 12 langues en avril 2004 et 22 langues en mai 2008. L'objectif à moyen terme est d'atteindre 60 langues, avec prise en compte de toutes les langues européennes. DP Europe comptabilise 100 livres numérisés en

mai 2005, et 500 livres numérisés en octobre 2008.

2000: Public Library of Science

[Résumé]

La Public Library of Science (PLoS) est fondée en octobre 2000 par un groupe de chercheurs des universités de Stanford et de Berkeley, en Californie. Son premier objectif est de contrer les publications spécialisées aux prix prohibitifs en regroupant tous les articles scientifiques et médicaux au sein d'archives en ligne en accès libre. Mais la réponse des éditeurs concernés n'est guère enthousiaste, et ce projet n'aboutit pas. La Public Library of Science met ensuite en oeuvre son deuxième objectif, et devient un éditeur non commercial de périodiques scientifiques et médicaux en ligne, selon un nouveau modèle d'édition en ligne basé sur la diffusion libre du savoir. Le premier numéro de PLoS Biology sort en octobre 2003. PLoS Medicine est lancé en octobre 2004. Trois nouveaux titres voient le jour en 2005: PLoS Genetics, PLoS Computational Biology et PLoS Pathogens. Ils sont suivis par PLoS Clinical Trials en mai 2006 et PLoS Neglected Tropical Diseases en octobre 2007.

[En détail]

A l'heure de l'internet, il paraît assez scandaleux que le résultat de travaux de recherche – travaux originaux et demandant de longues années d'efforts – soit détourné par des éditeurs spécialisés s'appropriant ce travail et le monnayant à prix fort. L'activité des chercheurs est souvent financée par les deniers publics, et de manière substantielle en Amérique du Nord. Il semblerait donc normal que la communauté scientifique et le grand public puissent bénéficier librement du résultat de ces recherches.

Dans le domaine scientifique et médical par exemple, 1.000 nouveaux articles sont publiés chaque jour, en ne comptant que les articles révisés par les pairs. Se basant sur ce constat, la Public Library of Science (PLoS) est fondée en octobre 2000 à San Francisco à l'initiative de Harold Varmus, Patrick Brown et Michael Eisen, enseignants-chercheurs dans les universités de Stanford et de Berkeley, en Californie. Le but de PLoS est de contrer les pratiques de l'édition spécialisée en regroupant tous les articles scientifiques et médicaux au sein d'archives en ligne en accès libre. Au lieu d'une information disséminée dans des millions de rapports et des milliers de périodiques en ligne ayant chacun des conditions d'accès différentes, un point d'accès unique permettrait de lire le contenu intégral de ces articles, avec moteur de recherche multi-critères et système d'hyperliens entre les articles.

Pour ce faire, PLoS fait circuler une lettre ouverte demandant que les articles publiés par les éditeurs spécialisés soient distribués librement dans un service d'archives en ligne, et incitant les signataires de cette lettre à promouvoir les éditeurs prêts à soutenir ce projet. La réponse de la communauté scientifique internationale est remarquable. Au cours des deux années suivantes, la lettre ouverte est signée par 30.000 chercheurs de 180 pays différents. La réponse des éditeurs est nettement moins enthousiaste, mais plusieurs éditeurs donnent leur accord de principe pour une distribution immédiate des articles publiés par leurs soins, ou alors une distribution dans un délai de six mois. Dans la pratique, toutefois, même les éditeurs ayant donné un accord de principe formulent nombre d'objections au nouveau modèle proposé, si bien que le projet d'archives en ligne ne voit finalement pas le jour.

Un autre objectif de la Public Library of Science est de devenir elle-même éditeur. PLoS fonde donc une maison d'édition scientifique non commerciale qui reçoit en décembre 2002 une subvention de 9 millions de dollars US de la part de la Moore Foundation. Une équipe éditoriale de haut niveau est constituée en janvier 2003 pour lancer des périodiques de qualité selon un nouveau modèle d'édition en ligne basé sur la diffusion libre du savoir. Le premier numéro de PLoS Biology sort en octobre 2003, avec une version en ligne gratuite et une version imprimée au prix coûtant (couvrant uniquement les frais de fabrication et de distribution). PLoS Medicine est lancé en octobre 2004. Trois nouveaux titres voient le jour en 2005: PLoS Genetics, PLoS Computational Biology et PLoS Pathogens. PLoS Clinical Trials voit le jour en 2006. PLoS Neglected Tropical Diseases est lancé en octobre 2007 en tant que première publication scientifique consacrée aux maladies tropicales négligées. Ces maladies affectent les populations pauvres aussi bien dans les zones rurales que dans les zones urbaines.

Tous les articles de ces périodiques sont librement accessibles en ligne, sur le site de PLoS et dans PubMed Central, le service public et gratuit d'archives en ligne de la National Library of Medicine (Etats-Unis), avec moteur de recherche multi-critères. Les versions imprimées sont abandonnées en 2006 pour laisser place à un service d'impression à la demande proposé par la société Odyssey Press. Ces articles peuvent être librement diffusés et réutilisés ailleurs, y compris pour des traductions, selon les termes de la licence Creative Commons, la seule contrainte étant la mention des auteurs et de la source. PLoS lance aussi PLoS ONE, un forum en ligne permettant de publier des articles sur tout sujet scientifique et médical.

Le succès est total. Trois ans après les débuts de PLoS en tant qu'éditeur, PLoS Biology et PLoS Medicine ont la même réputation d'excellence que les grandes revues Nature, Science ou The New England Journal of Medicine. PLoS reçoit le soutien financier de plusieurs fondations tout en mettant sur pied un modèle économique viable, avec des revenus émanant des frais de publication payés par les auteurs, et provenant aussi de la publicité, de sponsors et d'activités destinées aux membres de PLoS. PLoS œuvre aussi pour que ce modèle économique incite d'autres éditeurs à créer des revues du même type ou à mettre les revues existantes en accès libre.

2001: Wikipedia

[Résumé]

Lancée en janvier 2001 à l'initiative de Jimmy Wales et Larry Sanger (Larry quitte ensuite l'équipe), Wikipedia est une encyclopédie gratuite écrite collectivement et dont le contenu est librement réutilisable. Elle est immédiatement très populaire. Sans publicité, et financée par des dons, cette encyclopédie coopérative est rédigée par des milliers de volontaires (appelés Wikipédiens), avec possibilité de corriger et de compléter les articles. Les articles restent la propriété de leurs auteurs, et leur libre utilisation est régie par la licence GFDL (GNU free documentation license). En décembre 2004, Wikipedia compte 1,3 million d'articles rédigés par 13.000 contributeurs dans 100 langues. Deux ans après, en décembre 2006, elle compte 6 millions d'articles dans 250 langues, et elle est l'un de dix sites les plus visités du web. En mai 2007, la version francophone fête ses 500.000 articles. A la même date, Wikipedia compte 7 millions d'articles dans 192 langues, dont 1,8 million en anglais, 589.000 en allemand, 260.000 en portugais et 236.000 en espagnol.

[En détail]

Créée en janvier 2001 à l'initiative de Jimmy Wales et Larry Sanger (Larry quitte ensuite l'équipe), Wikipedia est une encyclopédie gratuite écrite collectivement et dont le contenu est librement réutilisable.

Wikipedia est non seulement une encyclopédie mais aussi un wiki. Un wiki - terme hawaïen signifiant: vite, rapide - est un site web permettant à plusieurs utilisateurs de collaborer en ligne sur un même texte. A tout moment, ces utilisateurs peuvent contribuer à la rédaction du contenu, modifier ce contenu et l'enrichir en permanence. Le wiki est utilisé par exemple pour créer et gérer des dictionnaires, des encyclopédies ou encore des sites d'information sur un sujet donné. Le programme présent derrière l'interface d'un wiki est plus ou moins élaboré. Un programme simple gère du texte et des hyperliens. Un programme élaboré permet d'inclure des images, des graphiques, des tableaux, etc.

Wikipedia est immédiatement très populaire. Sans publicité et financée par des dons, cette encyclopédie coopérative est rédigée par des milliers de volontaires - appelés Wikipédiens, et qui s'inscrivent en prenant un pseudonyme - avec possibilité de corriger ou compléter les articles. Les articles restent la propriété de leurs auteurs, et leur libre utilisation est régie par la licence GFDL (GNU free documentation license). En décembre 2004, Wikipedia compte 1,3 million d'articles rédigés par 13.000 contributeurs dans 100 langues. En décembre 2006, elle compte 6 millions d'articles dans 250 langues, et elle est l'un de dix sites les plus visités du web. En avril 2007, Wikipedia publie pour la première fois un CD payant avec une sélection de 2.000 articles en anglais. En mai 2007, la version francophone fête ses 500.000 articles (un CD est également prévu). A la même date, Wikipedia compte 7 millions d'articles dans 192 langues, dont 1,8 million en anglais, 589.000 en allemand, 260.000 en portugais et 236.000 en espagnol.

Fondée en juin 2003, la Wikimedia Foundation gère non seulement Wikipedia mais aussi

Wiktionary, dictionnaire et thésaurus multilingue lancé en décembre 2002, puis Wikibooks (livres et manuels en cours de rédaction) lancé en juin 2003, auxquels s'ajoutent ensuite Wikiquote (répertoire de citations), Wikisource (textes du domaine public), Wikimedia Commons (sources multimédias), Wikispecies (répertoire d'espèces animales et végétales), Wikinews (site d'actualités) et enfin Wikiversity (matériel d'enseignement), lancé en août 2006. La fin 2007 voit le lancement d'un moteur de recherche dénommé Wiki Search, qui utilise le réseau de contributeurs de Wikipedia pour classer les sites en fonction de leur qualité.

Les précurseurs de Wikipedia sont WebEncyclo (disparu depuis) et Britannica.com, lancés tous deux en décembre 1999 sur le web. WebEncyclo, publié par les éditions Atlas, est la première grande encyclopédie francophone en accès libre. La recherche est possible par mots-clés, thèmes, médias (cartes, liens internet, photos, illustrations) et idées. Un appel à contribution incite les spécialistes d'un sujet donné à envoyer des articles, qui sont regroupés dans la section WebEncyclo contributif. Après avoir été libre, l'accès est ensuite soumis à une inscription préalable gratuite. La version web de l'Encyclopaedia Universalis est mise en ligne à la même date, soit un ensemble de 28.000 articles signés par 4.000 auteurs. Si la consultation est payante sur la base d'un abonnement annuel, de nombreux articles sont en accès libre.

Le site Britannica.com - mis en ligne lui aussi en décembre 1999 - propose en accès libre l'équivalent numérique des 32 volumes de la 15^e édition de l'Encyclopaedia Britannica, parallèlement à la version imprimée et à la version CD-Rom, toutes deux payantes. Le site web offre une sélection d'articles issus de 70 magazines, un guide des meilleurs sites, un choix de livres, etc., le tout étant accessible à partir d'un moteur de recherche unique. En septembre 2000, le site fait partie des cent sites les plus visités au monde. En juillet 2001, la consultation devient payante sur la base d'un abonnement mensuel ou annuel. Fin 2008, Britannica.com annonce l'ouverture prochaine de son site à des contributeurs extérieurs, avec inscription obligatoire pour écrire et modifier des articles.

2001: Creative Commons

[Résumé]

Lancée en 2001 à l'initiative de Lawrence Lessig, professeur de droit en Californie, la licence Creative Commons est destinée à favoriser la diffusion d'oeuvres numériques tout en protégeant le droit d'auteur. L'organisme du même nom propose des licences-type, qui sont des contrats flexibles de droit d'auteur compatibles avec une diffusion sur l'internet. Simplement rédigées, ces autorisations non exclusives permettent aux titulaires des droits d'autoriser le public à utiliser leurs créations tout en ayant la possibilité de restreindre les exploitations commerciales et les oeuvres dérivées. L'auteur peut par exemple choisir d'autoriser ou non les reproductions et les rediffusions de ses oeuvres. Ces contrats peuvent être utilisés pour tout type de création: texte, film, photo, musique, site web, etc. Finalisée en février 2007, la version 3.0 des Creative Commons instaure entre autres une licence internationale et la compatibilité avec d'autres licences similaires, dont le copyleft et la GPL (general public license).

[En détail]

Lancée en 2001 à l'initiative de Lawrence Lessig, professeur de droit en Californie, la licence Creative Commons est destinée à favoriser la diffusion d'oeuvres numériques tout en protégeant le droit d'auteur. L'organisme du même nom propose des licences-type, qui sont des contrats flexibles de droit d'auteur compatibles avec une diffusion sur l'internet. Simplement rédigées, ces autorisations non exclusives permettent aux titulaires des droits d'autoriser le public à utiliser leurs créations tout en ayant la possibilité de restreindre les exploitations commerciales et les oeuvres dérivées. L'auteur peut par exemple choisir d'autoriser ou non les reproductions et les rediffusions de ses oeuvres. Ces contrats peuvent être utilisés pour tout type de création: texte, film, photo, musique, site web, etc. Finalisée en février 2007, la version 3.0 des Creative Commons instaure entre autres une licence internationale et la compatibilité avec d'autres licences similaires comme le copyleft et la GPL (general public license).

Qui utilise la licence Creative Commons? O'Reilly Media par exemple. Fondé par Tim O'Reilly en 1978, O'Reilly Media est un éditeur réputé de manuels informatiques et de livres sur les technologies de pointe. O'Reilly dispose d'abord d'une formule de "copyright ouvert" pour les auteurs qui le souhaitent, ou alors pour des projets collectifs. A partir de 2003, il privilégie le Creative Commons Founders' Copyright permettant d'offrir des contrats flexibles de droit d'auteur à ceux qui veulent également diffuser leurs oeuvres sur le web. La Public Library of Science (PLoS) utilise elle aussi la licence Creative Commons. Les articles de ses périodiques en ligne peuvent être librement diffusés et réutilisés ailleurs, y compris pour des traductions, la seule contrainte étant la mention des auteurs et de la source.

Une licence Creative Commons est utilisée pour un million d'oeuvres en 2003, 4,7 millions d'oeuvres en 2004, 20 millions d'oeuvres en 2005, 50 millions d'oeuvres en 2006, 90 millions d'oeuvres en 2007 et 130 millions d'oeuvres en 2008. En complément, Science Commons est fondé en 2005 pour définir les stratégies et les outils nécessaires à la diffusion sur le web de

la recherche scientifique, et ccLearn est fondé en 2007 dans le même but, mais pour l'enseignement.

2002: Cours du MIT gratuits

[Résumé]

Le MIT (Massachusetts Institute of Technology) décide de publier le contenu de ses cours en ligne, avec accès libre et gratuit, pour les mettre à la disposition de tous, enseignants, étudiants et autodidactes. L'initiative est menée avec le soutien de la Hewlett Foundation et de la Mellon Foundation. Mise en ligne en septembre 2002, la version pilote du MIT OpenCourseWare (MIT OCW) offre en accès libre le matériel d'enseignement de 32 cours représentatifs des cinq facultés du MIT. Ce matériel d'enseignement comprend les textes des conférences, les travaux pratiques, les exercices et corrigés, les bibliographies, les documents audio et vidéo, etc. Le lancement officiel du site a lieu en septembre 2003, avec accès à quelques centaines de cours. En mars 2004, les 500 cours disponibles couvrent 33 disciplines. En mai 2006, les 1.400 cours disponibles couvrent 34 disciplines. La totalité des cours dispensés par le MIT, soit 1.800 cours, est disponible en novembre 2007, avec actualisation régulière.

[En détail]

Basé comme son nom l'indique dans le Massachusetts, un Etat des Etats-Unis, le MIT (Massachusetts Institute of Technology) a toujours été à la pointe de la recherche dans de nombreux domaines. En avril 1997, par exemple, ce sont des chercheurs du Media Lab du MIT qui créent la société E Ink pour développer une technologie d'encre électronique, elle aussi appelée E Ink.

Professeur à l'Université d'Ottawa (Canada), Christian Vandendorpe salue en mai 2001 "la décision du MIT de placer tout le contenu de ses cours sur le web d'ici dix ans, en le mettant gratuitement à la disposition de tous. Entre les tendances à la privatisation du savoir et celles du partage et de l'ouverture à tous, je crois en fin de compte que c'est cette dernière qui va l'emporter." Le MIT décide en effet de publier le contenu de ses cours en ligne, avec accès libre et gratuit, une initiative menée avec le soutien financier de la Hewlett Foundation et de la Mellon Foundation.

Lancée en septembre 2002, la version pilote du MIT OpenCourseWare (MIT OCW) offre en accès libre le matériel d'enseignement de 32 cours représentatifs des cinq facultés du MIT. Ce matériel d'enseignement comprend des textes de conférences, des travaux pratiques, des exercices et corrigés, des bibliographies, des documents audio et vidéo, etc. Le lancement officiel du site a lieu un an plus tard, en septembre 2003, avec accès à quelques centaines de cours. En mars 2004, les 500 cours disponibles couvrent 33 disciplines. En mai 2006, les 1.400 cours disponibles couvrent 34 disciplines. La totalité des 1.800 cours dispensés par le MIT est en ligne en novembre 2007, avec actualisation régulière. Parallèlement, certains cours sont traduits en espagnol, en portugais et en chinois avec l'aide d'autres organismes.

Le MIT espère que cette expérience de publication électronique - la première du genre - va permettre de définir un standard et une méthode de publication, et inciter d'autres universités à créer un "opencourseware" pour la mise à disposition gratuite du contenu de

leurs propres cours. Un "opencourseware" peut être défini comme la publication électronique en accès libre du matériel d'enseignement d'une université donnée. A cet effet, le MIT lance l'OpenCourseWare Consortium (OCW Consortium) en décembre 2005, avec accès libre et gratuit au matériel d'enseignement de cent universités dans le monde un an plus tard.

2004: Project Gutenberg Europe

[Résumé]

Dans la lignée du Projet Gutenberg, grande bibliothèque numérique de livres du domaine public, le Projet Gutenberg Europe est lancé en janvier 2004 par le Projet Rastko (Belgrade, Serbie), en même temps que Distributed Proofreaders Europe (DP Europe), calqué sur le site original de Distributed Proofreaders (qui opère aux Etats-Unis). Le concept est un site web qui permet la correction partagée en fragmentant les livres en pages pouvant être relues par des correcteurs différents. La présence de plusieurs langues reflète la diversité linguistique prévalant en Europe. La norme utilisée pour définir le domaine public est l'équation "décès de l'auteur + 50 ans", selon le copyright en vigueur en Serbie. Quand il aura atteint sa vitesse de croisière, le Projet Gutenberg Europe devrait se répartir en plusieurs bibliothèques numériques nationales et/ou linguistiques, avec respect du copyright en vigueur dans le pays donné. 100 livres sont numérisés en juin 2005, et 500 livres en octobre 2008.

[En détail]

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 (Serbie), pour soutenir la création du Projet Gutenberg Europe et de Distributed Proofreaders Europe.

Quand il aura atteint sa vitesse de croisière, le Projet Gutenberg Europe devrait alimenter plusieurs bibliothèques numériques nationales et/ou linguistiques, par exemple le Projet Gutenberg France pour la France. Chaque pays ou région aurait son propre accès réseau autorisé (respectant la législation en vigueur dans le pays donné), qui serait 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).

Basé à Belgrade, en Serbie, le Projet Rastko 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.

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. Le Projet Gutenberg 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 Projet Gutenberg atteint 100 livres en mai 2005 et 500 livres en octobre 2008.

2004: Google Books

[Résumé]

En octobre 2004, Google lance la première partie de son programme Google Print, établi en partenariat avec les éditeurs pour consulter à l'écran des extraits de livres, puis commander les livres auprès d'une librairie en ligne. En décembre 2004, Google lance la deuxième partie de son programme Google Print, cette fois à destination des bibliothèques, le but étant de numériser 15 millions de livres, à commencer par ceux des bibliothèques de plusieurs universités partenaires (Harvard, Stanford, Michigan, Oxford) et de la ville de New York. La version bêta de Google Print est mise en ligne en mai 2005. En août 2005, le programme est suspendu pour cause de conflit avec les associations d'auteurs et d'éditeurs de livres sous droits. Il reprend en août 2006 sous le nom de Google Books (Google Livres). La numérisation des fonds de grandes bibliothèques se poursuit, tout comme le développement de partenariats avec des éditeurs. En octobre 2008, Google clôt le conflit avec les associations d'auteurs et d'éditeurs en signant un accord avec eux.

[En détail]

En 2004, le moteur de recherche Google met son expertise au service du livre. En octobre 2004, Google lance la première partie de son programme Google Print, établi en partenariat avec les éditeurs pour consulter à l'écran des extraits de livres, puis commander les livres auprès d'une librairie en ligne. En décembre 2004, Google lance la deuxième partie de son programme Google Print, cette fois-ci à destination des bibliothèques. Il s'agit d'un projet de bibliothèque numérique de 15 millions de livres, qui consisterait à numériser en quelques années les livres de plusieurs grandes bibliothèques partenaires. Les premières bibliothèques participantes sont celles des universités du Michigan (dans sa totalité: 7 millions d'ouvrages), de Harvard, de Stanford et d'Oxford, et la New York Public Library. Le coût estimé se situe entre 150 et 200 millions de dollars US (environ 10 dollars par livre) et la durée prévue est de dix ans. La version bêta de Google Print est mise en ligne en mai 2005. En août 2005, ce programme est suspendu pour un temps indéterminé pour cause de conflit avec les associations d'auteurs et d'éditeurs de livres sous droits. Le programme reprend en août 2006 sous le nouveau nom de Google Books (Google Livres).

Google Book Search, le moteur de recherche de Google Books, permet chercher les livres par date, titre ou éditeur. La numérisation des fonds de grandes bibliothèques se poursuit, en étant cette fois axée sur les livres libres de droit, et sur le développement de partenariats avec les éditeurs qui le souhaitent. Les livres libres de droit sont consultables à l'écran et leur texte copiable, avec possibilité d'impression page à page. Ils sont téléchargeables sous forme de fichiers PDF et imprimables dans leur entier.

A l'exception de la New York Public Library, les bibliothèques participantes sont des bibliothèques universitaires (Harvard, Stanford, Michigan, Oxford, California, Virginia, Wisconsin-Madison, Complutense de Madrid). S'y ajoutent début 2007 les bibliothèques des Universités de Princeton et du Texas (Austin), ainsi que la Biblioteca de Catalunya (Catalogne, Espagne) et la Bayerische Staatsbibliothek (Bavière, Allemagne). En mai 2007, Google

annonce la participation de la première bibliothèque francophone, la Bibliothèque cantonale et universitaire (BCU) de Lausanne (Suisse), rejointe ensuite par la Bibliothèque municipale de Lyon (France). Google scannerait 3.000 livres par jour, ce qui représenterait un million de livres par an.

Pour les livres sous droits, Google fournit la fiche du livre et des extraits incluant les mots-clés recherchés, en invoquant le droit de citation. De ce fait, le conflit avec les éditeurs et les auteurs se poursuit lui aussi, puisque Google continue de numériser des livres sous droits sans l'autorisation préalable des éditeurs, en invoquant là aussi le droit de citation qu'il procure en aval. L'Authors Guild et l'Association of American Publishers (AAP) invoquent pour leur part le non respect de la législation relative au copyright pour attaquer Google en justice et réitérer leurs plaintes pendant deux ans. Ce conflit prend fin en octobre 2008 avec la signature d'un accord prévue en 2009 entre Google et les parties plaignantes.

2005: Open Content Alliance

[Résumé]

Lancé en octobre 2005 à l'instigation de l'Internet Archive, l'Open Content Alliance (OCA) est un projet public et coopératif de bibliothèque numérique mondiale. Le but est de créer un vaste répertoire libre et multilingue de livres numérisés et de documents multimédia pour consultation et téléchargement sur n'importe quel moteur de recherche. L'OCA regroupe de nombreux partenaires: bibliothèques, universités, organisations gouvernementales, associations à but non lucratif, organismes culturels, sociétés informatiques (Adobe, Hewlett Packard, Microsoft, Yahoo!, Xerox, etc.). Les premiers participants sont les bibliothèques des universités de Californie et de Toronto, l'European Archive, les Archives nationales du Royaume-Uni, O'Reilly Media et Prelinger Archives. L'OCA souhaite s'inspirer de l'initiative de Google tout en évitant ses travers, à savoir la numérisation des livres sous droits sans l'accord préalable des éditeurs, tout comme la consultation et le téléchargement impossibles sur un autre moteur de recherche.

[En détail]

Lancé en octobre 2005 à l'instigation de l'Internet Archive, l'Open Content Alliance (OCA) est un projet public et coopératif de bibliothèque numérique mondiale. Le but est de créer un vaste répertoire libre et multilingue de livres numérisés et de documents multimédia pour consultation et téléchargement sur n'importe quel moteur de recherche. L'OCA regroupe de nombreux partenaires: bibliothèques, universités, organisations gouvernementales, associations à but non lucratif, organismes culturels, sociétés informatiques (Adobe, Hewlett Packard, Microsoft, Yahoo!, Xerox, etc.). Les premiers participants sont les bibliothèques des universités de Californie et de Toronto, l'European Archive, les Archives nationales du Royaume-Uni, O'Reilly Media et Prelinger Archives. L'OCA souhaite s'inspirer de l'initiative de Google tout en évitant ses travers, à savoir la numérisation des livres sous droits sans l'accord préalable des éditeurs, tout comme la consultation et le téléchargement impossibles sur un autre moteur de recherche.

L'Open Content Alliance (OCA) franchit la barre des 100.000 livres numérisés en décembre 2006, avec un rythme de 12.000 nouveaux livres par mois. Ces livres sont disponibles dans la collection Text Archive de l'Internet Archive. A la même date, l'Internet Archive reçoit une subvention d'un million de dollars de la Sloan Foundation pour numériser cinq collections historiques appartenant à des établissements réputés (Metropolitan Museum of Art, Boston Public Library, Getty Research Institute, John Hopkins University, Université de Californie à Berkeley). La barre des 200.000 livres numérisés est franchie en mai 2007. La barre du million de livres numérisés est franchie en décembre 2008.

2006: WorldCat gratuit

[Résumé]

WorldCat, grand catalogue collectif mondial, voit le jour dès 1971. A cette date, l'association OCLC (Online Computer Library Center) lance un catalogue collectif permettant un catalogage partagé entre les bibliothèques universitaires de l'Ohio, un Etat des Etats-Unis. En 2006, 73 millions de notices provenant de 10.000 bibliothèques dans 112 pays permettent de localiser un milliard de documents. Une notice type contient non seulement la description du document mais aussi des informations sur son contenu: table des matières, résumé, couverture, illustrations et courte biographie de l'auteur. Toujours en 2006, WorldCat migre progressivement sur le web, tout d'abord en rendant la consultation des notices possible par le biais de plusieurs moteurs de recherche (Yahoo!, Google et d'autres), puis avec le lancement en août 2006 de la version web (bêta) de WorldCat en accès libre. Les bibliothèques membres y proposent non seulement leur catalogue mais aussi un accès direct (gratuit ou payant) à leurs documents électroniques: livres du domaine public, articles, photos, vidéos, musique et livres audio.

[En détail]

L'ancêtre de WorldCat est créé dès 1971 par l'association OCLC (Online Computer Library Center) pour permettre un catalogage partagé entre les bibliothèques universitaires de l'Ohio, un Etat des Etats-Unis. Renommé OCLC Online Union Catalog puis WorldCat, il devient au fil des ans l'un des deux grands catalogues collectifs mondiaux, l'autre étant le RLG Union Catalog (RLG: Research Library Group). En 1998, WorldCat est disponible sur abonnement payant et comprend 38 millions de notices en 370 langues, auxquelles s'ajoutent 2 millions de nouvelles notices par an. WorldCat utilise huit formats bibliographiques (livres, périodiques, documents visuels, cartes et plans, documents mixtes, enregistrements sonores, partitions, documents informatiques). En 2005, 61 millions de notices bibliographiques produites par 9.000 bibliothèques et centres de documentation sont disponibles dans 400 langues, avec translittération des notices pour les caractères non romains des langues JACKPHY (japonais, arabe, chinois, coréen (Korean), persan, hébreu et yiddish).

En 2006, 73 millions de notices provenant de 10.000 bibliothèques dans 112 pays permettent de localiser un milliard de documents. Une notice type contient non seulement la description du document mais aussi des informations sur son contenu: table des matières, résumé, couverture, illustrations et courte biographie de l'auteur. A la même date, WorldCat migre progressivement sur le web, tout d'abord en rendant la consultation des notices possible par le biais de plusieurs moteurs de recherche (Yahoo!, Google et d'autres), puis avec le lancement en août 2006 de la version web (bêta) de WorldCat en accès libre. Les bibliothèques membres y proposent non seulement leur catalogue mais aussi un accès direct (gratuit ou payant) à leurs documents électroniques: livres du domaine public, articles, photos, vidéos, musique et livres audio.

Deux ans auparavant, le catalogue RedLightGreen fait figure de pionnier en tant que premier

catalogue collectif mondial librement disponible sur le web. Lancé au printemps 2004 par le RLG (Research Libraries Group) suite à une phase pilote débutée en automne 2003, RedLightGreen est issu du RLG Union Catalog, lancé dès 1980, et qui est le deuxième grand catalogue collectif mondial après WorldCat. La mise en ligne de RedLightGreen inaugure une ère nouvelle. C'est en effet la première fois qu'un catalogue de cette importance est mis en accès libre sur le web. RedLightGreen est particulièrement destiné aux étudiants du premier cycle universitaire, préparant à la licence. Il comprend 130 millions de notices (livres, cartes, manuscrits, films, bandes sonores, etc.), avec des liens vers des informations spécifiques aux bibliothèques d'un campus donné (cote, version en ligne si celle-ci existe, etc.). RedLightGreen cesse après trois ans d'activité, en novembre 2006, et les usagers sont invités à utiliser WorldCat, dont la version web (bêta) est en accès libre depuis août 2006. A la même date, le RLG est intégré à OCLC.

2007: Citizendium

[Résumé]

Citizendium (qui se veut l'abrégé de "The Citizens' Compendium") est une grande encyclopédie collaborative en ligne conçue en novembre 2006 et lancée en mars 2007 (en version bêta) par Larry Sanger, co-fondateur de Wikipedia en janvier 2001, mais qui quitte ensuite l'équipe de Wikipedia suite à des problèmes de qualité de contenu. Citizendium est basé sur le même modèle que Wikipedia (collaborative et gratuite) tout en évitant ses travers (vandalisme et manque de rigueur). Les auteurs signent les articles de leur vrai nom et les articles sont édités par des experts ("editors") titulaires d'une licence universitaire et âgés d'au moins 25 ans. De plus, des "constables" sont chargés de la bonne marche du projet et du respect du règlement. Le jour de son lancement officiel le 25 mars 2007, Citizendium comprend 1.100 articles, 820 auteurs et 180 éditeurs. 9.800 articles sont disponibles en décembre 2008.

2007: Encyclopedia of Life

[Résumé]

Projet débuté en mai 2007, l'Encyclopædia of Life sera une vaste encyclopédie collaborative en ligne rassemblant les connaissances existantes sur toutes les espèces animales et végétales connues (1,8 million), y compris les espèces en voie d'extinction, avec l'ajout de nouvelles espèces au fur et à mesure de leur identification (il en existerait de 8 à 10 millions). Il s'agira d'une encyclopédie multimédia permettant de ressembler textes, photos, cartes, bandes sonores et vidéos, avec une page web par espèce, et permettant aussi d'offrir un portail unique à des millions de documents épars en ligne et hors ligne. Outil d'apprentissage et d'enseignement pour une meilleure connaissance de notre planète, elle sera à destination de tous: scientifiques, enseignants, étudiants, scolaires, médias, décideurs et grand public. Ce projet collaboratif est mené par plusieurs grandes institutions (Field Museum of Natural History, Harvard University, Marine Biological Laboratory, Missouri Botanical Garden, Smithsonian Institution, Biodiversity Heritage Library).

[En détail]

Projet débuté en mai 2007, l'Encyclopædia of Life sera une vaste encyclopédie collaborative en ligne rassemblant les connaissances existantes sur toutes les espèces animales et végétales connues (1,8 million), y compris les espèces en voie d'extinction, avec l'ajout de nouvelles espèces au fur et à mesure de leur identification (il en existerait de 8 à 10 millions). Il s'agira d'une encyclopédie multimédia permettant de ressembler textes, photos, cartes, bandes sonores et vidéos, avec une page web par espèce, et permettant aussi d'offrir un portail unique à des millions de documents épars en ligne et hors ligne. Outil d'apprentissage et d'enseignement pour une meilleure connaissance de notre planète, elle sera à destination de tous: scientifiques, enseignants, étudiants, scolaires, médias, décideurs et grand public. Ce projet collaboratif est mené par plusieurs grandes institutions (Field Museum of Natural History, Harvard University, Marine Biological Laboratory, Missouri Botanical Garden, Smithsonian Institution, Biodiversity Heritage Library).

A la date de son lancement, ce projet est estimé à 100 millions de dollars US, sur une durée de dix ans, avant de pouvoir s'autofinancer. Le financement initial est assuré par la MacArthur Foundation (10 millions de dollars) et la Sloan Foundation (2,5 millions de dollars). La réalisation des pages web débute courant 2007. L'encyclopédie devrait faire ses débuts à la mi-2008, être opérationnelle dans trois à cinq ans et être complète (c'est-à-dire à jour) dans dix ans. En tant que consortium des dix plus grandes bibliothèques des sciences de la vie (d'autres suivront), la Biodiversity Heritage Library a d'ores et déjà débuté la numérisation des 2 millions de publications des bibliothèques partenaires. En mai 2007, on compte 1,25 million de pages traitées dans les centres de numérisation de Londres, Boston et Washington DC, et disponibles sur le site de l'Internet Archive.