

tries of Rights information and Orphan Works towards Europeana), financé par la Commission européenne, a été démarré en 2008 pour faciliter cette recherche. Cependant, ARROW ne peut à lui seul identifier l'ensemble des œuvres sous droits et leurs ayants droit légitimes. Aussi, en raison de la réglementation sur le droit d'auteur, il est plus facile aujourd'hui de trouver sur Internet un ouvrage de 1860 que de 1960. Un pan entier de la littérature, celle du XX^e siècle, est en grande part inaccessible en ligne, alors même que la valeur commerciale des documents concernés est faible, voire nulle, et que ceux-ci sont souvent indisponibles – dans le commerce, sous format papier voire même en format électronique – ou orphelins – sans ayant droit connu.

Le programme de numérisation des œuvres indisponibles lancé par l'État en 2013 via la Bibliothèque nationale de France vise justement à numériser 500 000 œuvres du XX^e siècle pour combler ce vide, mais les modalités – notamment de financement – de ce programme sont très critiquées par une partie importante des cercles professionnels. Une loi sur les œuvres indisponibles est en préparation au niveau européen, bénéficiant du soutien des associations de bibliothèques. Celles-ci sont aussi moteur d'un projet de traité international sur les exceptions et les limitations au droit d'auteur en discussion au sein de l'Organisation mondiale de la propriété intellectuelle.

De nouvelles archives à préserver

La pérennité des documents numérisés est aussi au cœur des réflexions. Une erreur fondamentale serait de croire que la préservation de la version physique originale d'un document numérisé serait inutile. D'une part, l'original a une valeur patrimoniale en soi et, d'autre part, la durée de la pérennité des données numériques n'est aujourd'hui de l'ordre que d'une trentaine d'années. Numériser pour faire de la place sur les rayonnages est pour l'heure un danger réel pour la transmission du patrimoine et de la mémoire. Les exemples de pertes des données conservées sur des CD-ROM sont légion, sans parler de la qualité technique des numérisations réalisées il y a quelques années, insuffisante par rapport aux besoins d'aujourd'hui.

Quelques chiffres

130 millions de livres ont été publiés dans le monde depuis l'invention de l'imprimerie, selon *Google*, dont 20 % sont dans le domaine public.

20 % des collections des établissements culturels européens étaient numérisés en 2012, selon l'enquête *Enumerate*.

30 millions de documents sont accessibles via *Google Books*, 26 millions via *Europeana*, 11 millions via *Hathi Trust* et 2,2 millions via *Gallica*.

5 000 euros : c'est le coût annuel moyen de la préservation à long terme d'un téraoctet de données numérisées.

Les bibliothèques ne se sont intéressées à la préservation à long terme des documents numérisés qu'assez récemment, à partir de 2000 pour les plus avancées, et la norme en vigueur – l'OAIS (Système ouvert d'archivage d'information) – n'a été définie qu'en 2002. Si les bibliothèques nationales ont développé leurs propres infrastructures d'archivage (SPAR pour la Bibliothèque nationale de France, E-Depot pour la Bibliothèque royale des Pays-Bas, etc.), les autres bibliothèques recourent à des prestations extérieures d'archivage pour leurs propres collections, soit auprès de leur bibliothèque nationale, soit auprès d'autres organismes, privés ou publics.

En France, dès 2008, le Centre informatique national de l'enseignement supérieur (CINES) a été missionné pour l'archivage pérenne des bibliothèques universitaires par le ministère de l'Enseignement supérieur et de la Recherche (voir l'article de F. Daumas et M. Massol dans ce numéro). Le CINES est notamment au cœur du dispositif envisagé par la Bibliothèque scientifique numérique pour archiver les contenus numériques produits par l'enseignement supérieur national.

Néanmoins, avec un coût estimé aujourd'hui à 5 000 euros par téraoctet et par an, l'archivage numérique représente une charge financière considérable pour les établissements. Ainsi, des politiques documentaires spécifiques ont été définies par les bibliothèques, déterminant des priorités parmi les natures de données à préserver. Toutefois, la question de la préservation à long terme des données numériques reste largement ignorée des porteurs de projets de numérisation. Et l'ampleur des coûts fait peser un risque fort sur la pérennité des résultats de ces projets.

Avec le numérique, les bibliothèques – notamment les bibliothèques universitaires et de recherche – ont été bouleversées en profondeur. Pourtant, leurs missions fondamentales – préserver le patrimoine documentaire, rendre accessible le savoir, accompagner leurs usagers dans la découverte et l'usage de ce savoir noyé dans l'immensité – restent intactes : le numérique leur permet d'accomplir ces missions avec davantage encore de pertinence et d'efficacité. Les professionnels de l'information numérique sont devenus, et deviendront encore davantage, des partenaires incontournables des projets de recherche, ainsi que de l'enseignement à distance, amené à révolutionner l'université dans les prochaines années. ■

■ BIBLIOGRAPHIE

A. Jacquesson, *Google Livres et le futur des bibliothèques numériques. Historique du projet, techniques, documentaires, alternatives et controverses*, Cercle de la librairie, 2012.

Th. Claerr et I. Westeel (sous la dir.), *Manuel de la numérisation*, Cercle de la librairie, 2012.

J.-Ph. Accart et A. Rivier, *Mémento de l'information numérique*, Cercle de la librairie, 2012.

D. Baker et W. Evans (éd.), *Digital Library Economics. An academic perspective*, Chandos Publishing, 2009.

Bibliothèque scientifique numérique : bibliothequescientifique-numerique.fr

Gallica : gallica.bnf.fr

Dictionnaires & encyclopédies à l'ère numérique

Rémi Mathis



L'encyclopédie en ligne *Wikipédia* et le *Wiktionnaire* modifient en profondeur les définitions mêmes de l'encyclopédie et du dictionnaire.

Alors que le livre numérique se développe et se diffuse à une vitesse mesurée, les encyclopédies ont connu, sous l'effet du numérique, des bouleversements majeurs. Et pourraient en subir de nouveaux. L'intérêt d'un support permettant d'accompagner les textes d'un grand nombre d'images, voire de films et de sons, a vite été compris des éditeurs. Entre 1985 et 1995, ils proposent tour à tour des versions sur CD-ROM de leur encyclopédie. L'avantage est évident : plus besoin de manipuler de lourds volumes, ni de les stocker chez soi, possibilité de rechercher un terme ou un concept dans le contenu, ou de rebondir d'un article à l'autre au sein d'une navigation simple et efficace, sans compter l'extrême réduction du coût de fabrication de l'objet en lui-même.

Jusqu'alors, le marché de l'encyclopédie était bien balisé, avec des marques de référence dans chaque pays : *Britannica* en Angleterre, *Universalis* en France, *Brockhaus* en Allemagne. Le changement de support rebat les cartes en attirant sur le marché de nouveaux venus : dès 1993, le géant de l'informatique *Microsoft* propose sa propre

■ L'AUTEUR



Rémi MATHIS est conservateur au Département des estampes de la Bibliothèque nationale de France. Il est aussi président de *Wikimédia France*.

■ BIBLIOGRAPHIE

R. Mathis, *Wikipédia. Une somme originale de copies*, *Medium*, n° 32-33, pp. 356-377, 2012.

R. Mathis, *Wikipédia, espace fluide, espace à parcourir*, *Revue de la BNU*, n° 6, pp. 52-61, 2012.

R. Mathis, *Défiances et production. Les bibliothèques françaises et Wikipédia*, *Bulletin des bibliothèques de France*, vol. 1, pp. 10-13, 2011.

encyclopédie multimédia, *Encarta*, disponible en sept langues. L'entreprise se fonde sur des technologies développées pour d'autres usages (par exemple son logiciel *MapPoint* pour les cartes) afin de rendre la consultation plus efficace et agréable. Elle change aussi complètement le modèle intellectuel de l'encyclopédie.

Une encyclopédie était alors un produit éditorial de luxe, au coût élevé, dont les articles étaient rédigés par les meilleurs spécialistes de la discipline. *Encarta* a une approche différente : *Microsoft* achète des droits non exclusifs de diffusion à des encyclopédies existantes (*Funk&Wagnalls*) et rachète d'autres encyclopédies (*Collier*, *New Merit Scholar*). Le contenu est traduit, mais varie selon les pays, car des partenariats spéciaux sont signés : la version en néerlandais reprend, par exemple, des contenus de l'encyclopédie locale *Winkler Prins*.

Le résultat est un corpus d'environ 40 000 articles (plus ou moins selon les langues), mélanges de textes divers qui ne sont pas signés et ne font pas référence à des sources externes. Le texte est d'abord