

INTRODUCTION

«La revanche des données»: ce titre résumait la présentation effectuée en 2016 par Gildas Illien¹, alors conservateur à la Bibliothèque nationale de France (BnF)², du principe de la transition bibliographique^{*3}, entamée un an auparavant. Il s'agissait en l'occurrence des données des catalogues de bibliothèques, passés au second plan depuis l'avènement du web, et qui expérimentent ainsi une nouvelle visibilité grâce à l'exposition de leurs données et métadonnées*.

Si revanche il y a, on peut se poser la question de son objet (une revanche suppose une défaite préalable), et d'une prise de conscience tardive: les données n'ont-elles pas toujours fait partie des processus liés à l'information et la connaissance? Alors que nous quittons la galaxie Gutenberg, peut-on considérer que la donnée va être amenée à éclipser le document, à l'instar de l'imprimerie remplaçant à la fin du Moyen Âge l'imagerie sculptée, selon la célèbre expression de Victor Hugo dans *Notre-Dame de Paris* («Ceci tuera cela»)? La révolution numérique a eu, entre autres, pour effet une évolution de la perception du document, vers une vision plus fluide et flexible. Les données se libèrent de leurs barrières et de systèmes souvent clos. En conséquence, la prise de conscience de la valeur des données a connu une accélération importante au début du XXI^e siècle, dans le monde des entreprises comme dans celui de la culture et du savoir.

Qu'en est-il dans le monde des bibliothèques? Au fil des années, les professionnels du secteur ont appris à mettre en place des stratégies de médiation numérique adaptées aux différents publics. À travers de nombreuses actions comme les grands programmes de numérisation, la mise en place de bibliothèques numériques, la dissémination de contenus via des sites portails ou les réseaux sociaux, les bibliothèques ont pu gagner en visibilité et renforcer leur présence sur le web, et plus généralement dans l'espace numérique.

L'évolution vers un modèle centré sur les données peut constituer une nouvelle frontière pour les bibliothèques, au-delà de la seule transition bibliographique. La transformation numérique des processus documentaires amène en effet les professionnels de l'information et des bibliothèques à gérer non seulement des documents imprimés ou numériques, mais aussi des données de toutes natures. L'évolution des technologies et des architectures

1. < <https://www.archimag.com/bibliotheque-edition/2016/03/11/transition-bibliographique-revanche-donnees> >.

2. Pour les sigles et acronymes, voir la liste p. 11.

3. Les termes suivis d'un astérisque (à leur première occurrence) sont définis dans le glossaire en fin d'ouvrage.

informatiques permet désormais de produire, stocker, consulter et traiter des volumes de données de plus en plus importants. L'approche par les données s'accompagne d'une nouvelle granularité* de l'information, comportant différents niveaux.

Les professionnels des bibliothèques sont ainsi de plus en plus concernés par ce nouveau monde de la donnée, et ce à plusieurs égards. Toutes les bibliothèques produisent des données concernant leur activité (fréquentation, prêt, utilisateurs, etc.) qu'il convient de gérer et sécuriser en vertu des réglementations comme le Règlement général sur la protection des données (RGPD)* ou le nouveau Data Act. La transition bibliographique en cours rend les données des catalogues plus visibles sur le web et interopérables avec d'autres données. Les bibliothèques universitaires gèrent, décrivent et stockent les données de la recherche*. De façon plus générale, les bibliothèques sont impactées par le mouvement initié depuis les années 2010 autour de l'ouverture des données publiques (open data*).

L'objectif de cet ouvrage est donc de répondre aux questions que se posent les professionnels des bibliothèques face à ce nouveau monde de la donnée qui ne va cesser de s'étendre, offrant de multiples opportunités.

Il s'agit ainsi de favoriser une acculturation aux données, afin de s'appropriier le vocabulaire, le sens et la valeur des données dans le monde des bibliothèques.

Dans un premier temps, il convient de définir clairement ce que recouvre le concept de la donnée (au-delà de l'expression souvent galvaudée de « big data* »), et d'en dégager les enjeux à la fois techniques, documentaires et juridiques, autour du cycle de vie des données*.

Le premier chapitre permet ainsi de clarifier ces notions complexes concernant information, données et documents, mais également de distinguer les différents formats et de dresser un historique de l'ouverture des données.

L'ouvrage aborde ensuite le principe du web sémantique* et ses enjeux pour les professionnels des bibliothèques. L'univers complexe de la transition bibliographique, basée sur les principes du web de données*, est explicité dans le deuxième chapitre. Plusieurs exemples de réalisations, dont le célèbre site data.bnf.fr, apportent une illustration concrète à ces questions souvent très techniques.

Les bibliothèques produisent elles-mêmes des données, détaillées dans le chapitre 3: la gestion et l'analyse des données d'usage et d'activité* sont essentielles non seulement pour mesurer la valeur des bibliothèques, mais aussi pour mieux connaître les publics et les fidéliser. Les données issues des ressources numériques externes, ou de bibliothèques numériques constituées

à partir de documents patrimoniaux, offrent également un corpus destiné à l'étude et l'analyse.

Pas de données sans métadonnées... Le chapitre 4 présente un panorama de la diversité des métadonnées, ainsi que les nombreux standards utilisés dans le monde des bibliothèques, des archives ou de l'édition numérique.

Les BU sont particulièrement concernées par les données de la recherche, depuis l'avènement de la science ouverte*. Après un rappel de l'impact de l'ouverture des données de la recherche et des avancées de la «Data Driven Science*», le chapitre 5 examine le rôle que peuvent jouer les bibliothécaires auprès des chercheurs en tous domaines. Il s'agit en effet d'accompagner les chercheurs dans leurs travaux en les aidant dans le choix des entrepôts de données*, des formats pérennes ainsi que dans l'élaboration du plan de gestion des données*.

La valeur des données réside en grande partie dans les possibilités d'analyse et de réutilisation. L'analyse de données s'appuie sur plusieurs approches et des langages de programmation ouverts, qui se sont invités au fil des années dans l'univers des bibliothèques. Le chapitre 6 évoque les fondamentaux de l'analyse de données, sans oublier la nécessaire étape de préparation, et détaille les opérations de moissonnage, de fouille de données* et bien sûr la datavisualisation*. Cette dernière approche est particulièrement cruciale pour «faire parler la donnée» et mettre en évidence des tendances ou corrélations. Toutes ces techniques d'analyse s'appuient de plus en plus sur des algorithmes* complexes issus de l'intelligence artificielle*. Plusieurs retours d'expérience, issus notamment de la pratique des humanités numériques*, illustrent ce chapitre.

La protection des données personnelles* constitue un droit fondamental qui protège désormais les citoyens européens via le RGPD et d'autres règlements à venir. Le chapitre 7 détaille l'impact (ainsi que certaines contraintes) de ces nouvelles dispositions en bibliothèque. On y trouve également des éléments concernant les droits d'auteur liés aux données, ainsi qu'un rappel des risques liés aux pertes de données et actes malveillants tels que les cyberattaques, dont certaines bibliothèques ont malheureusement pu être victimes.

Enfin, le chapitre 8 dessine les contours de nouveaux métiers et compétences générés en bibliothèque par l'émergence des données: le métier récent de «data librarian» comporte de multiples facettes, qui dépassent largement les seules compétences techniques. Que ce soit dans le domaine de la gestion, du service ou de la sécurité, les bibliothécaires ont toute légitimité dans la médiation et l'accompagnement autour de la donnée. Un panorama des filières et formations dans ce domaine conclut le chapitre.

Chaque chapitre expose de façon détaillée et pragmatique un aspect du monde de la donnée concernant les bibliothèques, qu'il s'agisse de lecture publique, bibliothèques universitaires (BU) ou spécialisées. En résumé, cet ouvrage se propose de présenter de façon synthétique et opérationnelle ce que les données font aux bibliothèques, et ce que les bibliothèques peuvent faire avec les données. Nous vous en souhaitons bonne lecture.