

## **Cours d'initiation aux Humanités Numériques** **Session 2020-2021**

**Fiche de lecture : Thierry Claerr et Isabelle Westeel,**  
*Manuel de numérisation ; Manuel de constitution des bibliothèques numériques*

Éditions du Cercle de la Librairie, 2011 – 2013

Comment amener les Humanités au numérique, quand les supports de recherche en histoire, littérature et arts est d'abord celui du papier des archives, des manuscrits et des ouvrages de référence ? Faut-il transposer les usages existants des documents physiques aux documents numériques, ou concevoir *ex nihilo* une architecture intellectuelle radicalement différente, et ce avec les moyens techniques, financiers et humains dont l'on dispose au sein de son établissement ? C'est à cette question que répondent les deux ouvrages dirigés par Thierry Claerr et Isabelle Westeel, *Manuel de numérisation* (2011) et *Manuel de constitution des bibliothèques numériques* (2013). Ils sont destinés aux responsables de la conservation en bibliothèques, premiers acteurs dans la production et l'organisation de données à des fins de recherche. Tous deux conservateurs eux-mêmes, les auteurs sont spécialistes de la numérisation patrimoniale et de sa valorisation via des projets en collaboration avec des laboratoires de recherche, qu'ils enseignent en université et à l'ENSSIB.

Le contexte de parution de ces manuels est celui de l'accélération à l'échelle nationale de la numérisation des fonds patrimoniaux des bibliothèques publiques ; débutée en 1993 sur des initiatives locales, elle prend une toute autre ampleur en 2008 avec le Plan de développement de l'économie numérique, qui acte la prise de conscience de l'omniprésence du numérique dans les pratiques culturelles des générations les plus jeunes ; ainsi, en 2010, les bibliothèques peuvent bénéficier de la Dotation générale de décentralisation pour subventionner leurs projets de numérisation, ce qui ouvre les institutions plus modestes au numérique. Il importe donc de créer un ouvrage de référence, afin de détailler tous les volets d'un projet de numérisation patrimoniale (technique, juridique, managérial) à destination d'agents encore peu formés sur la question, et de présenter un vaste panel d'outils pour couvrir les besoins d'institutions de toutes tailles.

Cette politique de numérisation de masse est alors motivée par la valorisation patrimoniale et scientifique des fonds plus que par leur sauvegarde ; la qualité des documents numériques dépend du caractère accessible, intelligible et interactif du site qui les publie. Toutefois, les manuels se focalisent sur la création des contenus en ligne, et non sur les méthodes de recherche qui en résultent. Bien qu'à vocation exhaustive, il ne s'agit pas non plus de traités sur le numérique ; ils visent à développer la maîtrise de ses outils par les conservateurs, afin qu'ils soient en mesure d'en questionner eux-mêmes le rôle au sein de la patrimonialisation et de la recherche.

Conçus comme un diptyque (production des documents, puis constitution des sites et conservation des données) les manuels sont organisés pour suivre pas à pas chaque étape de la création d'une bibliothèque numérique, ce qui totalise douze parties très finement détaillées et exemplifiées autour de cinq thèmes centraux.

Les manuels sont d'abord introduits par un récapitulatif de l'histoire de la numérisation des fonds des bibliothèques françaises, qui associe directement sauvegarde, valorisation du patrimoine et politiques publiques à travers le détail des plans nationaux et rapports ministériels. Même si aucun texte de loi ne l'assume, un parti est pris en faveur de la diffusion des collections comme service public, à protéger des intérêts commerciaux des autres instances en mesure de proposer des marchés de numérisation de masse. Ainsi, le cas de la numérisation des fonds patrimoniaux de la bibliothèque de la Part Dieu par Google (en 2008, pour Google Books) est détaillé, mais rejeté comme option favorable à long terme, notamment du fait que Google est rendu propriétaire d'une partie des documents numériques produits. Du reste, l'ambition des projets de numérisation reste conditionnée aux moyens de chaque bibliothèque, qui doit alors choisir entre traiter un petit corpus de documents complexes (clarifiés par les possibilités de zoom et de commentaires) ou un plus important, mais dont la qualité visuelle et descriptive ne servira que la compréhension de la dynamique d'ensemble.

En deuxième lieu sont développées les questions juridiques préliminaires à la création de bibliothèques numériques, relatives aux droits d'auteur et à la diffusion des données. L'accent est mis sur les droits moraux des œuvres, incessibles, dont dépend la divulgation ; sur les conditions de passage dans le domaine public (notamment les pièges tels que les prorogations des droits d'auteur, et les droits spéciaux liés à un monopole d'éditeur) ; enfin sur la manière dont les bibliothèques peuvent cerner leur marge de manœuvre sur la base de leur *régime d'exception* (les œuvres du domaine public, ou devant être sauvegardées sans but commercial, peuvent être numérisées sans autorisation préalable.) Il est toutefois précisé que les bibliothèques n'ont pas encore de définition légale, et que l'application des *exceptions* peut être sujette à débat.<sup>1</sup> La mise en ligne implique citations et réutilisations par les usagers ; il revient alors aux licences ouvertes d'en définir les conditions (comme Etalab ou Creative Commons, qui partagent l'obligation de mentionner l'origine de l'information et de ne pas dénaturer le contenu original) ainsi qu'aux lois sur l'exception *pour fins d'enseignement et de recherche* – les travaux produits devant alors faire objet de déclaration sur le site du CFC ; rien ne peut être laissé au hasard, le principe de *fair use* par défaut n'étant pas légalisé en France. Les bases de données, quant à elles, sont également soumises au droit d'auteur, et ne peuvent être réutilisées que partiellement (le flou étant entretenu sur la limite entre exploitation partielle et substantielle).

Plusieurs chapitres sont consacrés aux questions organisationnelles de formation des agents à la gestion de projet et aux marchés publics. En effet, il ne va pas de soi que tous les établissements aient expérimenté ce mode de fonctionnement ; les acquisitions et déclassements (désherbage) sont des actions continues sur budget annuel, sans recours à des prestataires externes. Il en résulte un grand panel de conseils relatifs à la définition d'objectifs sur un temps donné (diagramme de Gantt), des équipes de pilotage, des plannings, des études préliminaires à mener, des cahiers des charges (figurant prestations et performances attendues) ainsi que des formations à dispenser. Le recours aux vacataires et CDD pour la durée du projet, et la coordination d'agents d'horizons professionnels divers, sont autant de contraintes dont ni usagers chercheurs ni tutelles n'ont d'emblée conscience.

L'enjeu principal des manuels demeure la description des procédés techniques de numérisation et des logiciels de diffusion et de gestion en ligne. Une mise au point concernant les fondamentaux de l'optique, de la photographie numérique, du codage de la résolution d'image et de leur compression est

---

<sup>1</sup>. D'où l'actuel projet de loi porté par la sénatrice Sylvie Robert sur la définition des équipements et missions des bibliothèques.

incontournable, dans l'optique de former le bibliothécaire moyen du tout au tout. Les manuels se présentent comme de véritables catalogues comparatifs de logiciels, classés selon les usages attendus, dont il émerge plusieurs préoccupations. Premièrement, la nécessité de normaliser les métadonnées associées à chaque document numérique, et ce dès la numérisation, afin d'éviter la confusion d'informations lors des mises à jour ; secondement, le débat entre choix de logiciels propriétaires et logiciels libres, dont avantages et contraintes sont soigneusement évoqués à part égales ; si le logiciel libre, protéiforme, ne bride pas la bibliothèque dans l'évolution des formats et de ses collections numériques, le logiciel propriétaire la décharge de mener cette évolution elle-même (le prestataire se chargeant d'innover sur son produit), et donc de pâtir du turnover de ses agents qualifiés. Enfin, la gestion de la *pérennisation* des collections est conçue comme une extension de la mission des bibliothèques publiques de conservation des collections physiques, car le numérique est lui-même, de fait, une collection de supports physiques variés (CD, disque dur, bande magnétique) soumis à leur propre fragilité, aux risques techniques (l'incendie du data center d'OVH en étant désormais l'exemple) voire aux risques *budgétaires* : une politique d'économies pouvant se traduire par l'utilisation de formats plus légers et moins onéreux, ou par l'externalisation des serveurs d'hébergement – les conservateurs n'ayant alors plus la main sur les mises à jour.

*Last but not least*, le second manuel en particulier s'emploie à développer des pistes concernant la diffusion et la valorisation des collections numériques auprès des usagers (qu'elles soient pédagogiques, scientifiques ou simplement ludiques.) Or cette étape dépend de facteurs complexes. La bibliothèque numérique ne se conçoit pas indépendamment des usages qu'on lui imagine, et repose alors sur la double anticipation des pratiques futures du Web et des évolutions technologiques : la multiplication des smartphones et tablettes nécessitant de constants ajustements des interfaces des sites, dont l'attractivité est cruciale pour la visibilité et l'exploitation des collections. Pour arriver spontanément sur le chemin de l'internaute, les métadonnées associées aux documents doivent être conçues selon les standards établis par des consortiums des bibliothécaires afin que les catalogues des différents établissements puissent se citer mutuellement ; plus les métadonnées seront riches, plus le document sera décrit en détail et visible sur le Web. Or le volume croissant de ces derniers tend à rendre leur enrichissement impossible à réaliser manuellement par les bibliothécaires ; il faudrait alors récupérer de manière automatique les métadonnées produites par les éditeurs, mais qui ne sont pas dans le même format d'exploitation. C'est là un exemple de la complexité de l'*interopérabilité* des données numériques, condition *sine qua non* à leur réutilisation : les données doivent pouvoir être reconnaissables entre machines, fournisseurs et catalogues, avec des standards d'écriture et des protocoles d'échange normalisés, et ce d'autant plus vite que les internautes ne recherchent plus un document, mais une information ; soit directement les données portées par une œuvre. Cela nécessite de structurer les données en réseau, pour les rendre interrogeables toutes à la fois sans les cloisonner dans des notices propres à chaque document : enjeux de la Transition Bibliographique actuelle, trop récente pour être développée dans les manuels.

Même s'ils ne se destinent pas aux métiers des bibliothèques, la lecture de ces manuels peut être profitable aux jeunes chercheurs de SHS pour appréhender la plus-value des collections numériques : comme l'OCR, fondement des recherches en plein texte et du *text mining* ; les traducteurs immédiats ; la visualisation des données d'un corpus sur carte ou sur frise chronologique ; les nuages de *tags* ajoutés par les usagers pour indexer le sujet d'un document dans un vocabulaire simple, similaire à celui que

connaissent les grands moteurs de recherche. Plus encore, la connaissance des logiques d'indexation des articles scientifiques et de leurs standards (le langage Rameau) permet de définir une stratégie vis-à-vis de leurs propres travaux.

Les deux manuels souffrent néanmoins d'obsolescence programmée : constitués en majorité d'exemples, et de plusieurs exemples pour chaque option évoquée, ils se heurtent à l'évolution rapide des technologies des logiciels, des projets d'envergure des institutions (tels que la Transition Bibliographique<sup>2</sup>) voire des politiques de subvention publiques et textes de lois eux-mêmes. Or ces exemples sont trop nombreux pour être vérifiés un à un lors de la lecture, ce qui conduit à se demander continuellement quels logiciels ont encore cours ou non, tout en constatant qu'aucun autre ouvrage de bibliothéconomie ne traite d'autant de problématiques concrètes. Bien conscients de ces évolutions constantes, les manuels soulignent que certaines constantes apparaissent *a contrario* : ainsi la stabilité des formats de métadonnées descriptives tels que Dublin Core, MARC, MODS et METS est réelle tant qu'elle repose sur une utilisation massive par les institutions, et sur le recours à des identifiants uniques et pérennes (*Archival Resource Key, Digital Object Identifier*, ou identifiants personnels tels que VIAF pour les auteurs). Les contributeurs insistent sur l'importance de s'intéresser de près aux technologies et médiations innovantes, sans développer de méthodologie sur la question ; les seuls exemples institutionnels pris sont des laboratoires de bibliothèques, qu'ils reconnaissent eux-mêmes à la traîne sur d'autres, paradoxe quand il s'agit d'innovation.

L'évocation des recherches en Humanités numériques est elle-même datée ; tout en partant du constat que l'accès à d'immenses corpus exploitables directement n'a jamais été aussi facile, les contributeurs regrettent un manque manifeste de formation des doctorants vis-à-vis des usages numériques. Le Web n'est alors pas utilisé comme source d'étude, mais comme réservoir d'articles de littérature secondaire, en présentant les difficultés des accès restreints par les abonnements et les authentifications – signe que les bases de données ne sont pas si facilement accessibles qu'il le faudrait. Si quelques programmes menés conjointement par bibliothèques et laboratoires sont listés à l'heure de l'écriture du manuel, les bibliothèques traitent encore leurs silos de données seules (dont le catalogue de données data.bnf est un exemple), et en définitive peinent à comprendre les besoins des chercheurs. Huit ans de créations de parcours d'humanités numériques (PSL) de formations au numérique en bibliothèques universitaires (Paris VIII) et de développement de projets conjoints entre centres patrimoniaux et laboratoires, depuis 2013, nuancent ces propos.

Ces deux manuels visent donc à enseigner la maîtrise des outils numériques afin d'encourager les agents à se les approprier et les dépasser, en guettant auprès de leurs publics de nouvelles demandes et de nouveaux usages, tout en garantissant la conservation de leurs collections et leur ancrage dans une politique documentaire cohérente à long terme. Les usagers sont les prescripteurs des nouvelles pratiques numériques en environnement culturel ; c'est alors en mutualisant leurs observations et leurs moyens d'action que les institutions – bibliothèques, archives, musées, laboratoires – modernisent l'accès à leurs collections et le traitement scientifique qui s'ensuit, poursuivant la dynamique de regroupement et d'harmonisation des données à toutes les échelles.

---

<sup>2</sup>. Détaillée, ainsi que les questions d'interopérabilité et d'intégration au Web sémantique, dans un ouvrage un peu plus récent : *Vers de nouveaux catalogues*, sous la direction d'Emmanuelle Bermès, édition du Cercle de la Librairie, 2017.