

Le concept de collection numérique

FRÉDÉRIC MARTIN
frederic-d.martin@bnf.fr

EMMANUELLE BERMÈS
emmanuelle.bermes@bnf.fr

Bibliothèque nationale de France

Titulaire d'un DEA de lettres modernes et du DCB, Frédéric Martin est chef du service Pôles associés/Gallica de la BnF, après avoir été responsable des techniques documentaires pour les collections numériques. Il est l'auteur d'articles pour diverses revues professionnelles.

Archiviste-paléographe, titulaire du DCB, Emmanuelle Bermès est actuellement chef du service « Prospectives et services documentaires » au sein du département de l'Information bibliographique et numérique, à la Direction des services et des réseaux de la BnF. Elle a écrit de nombreux articles sur les techniques documentaires pour les collections numériques et anime le blog « Figoblog ».

Les bibliothèques sont aujourd'hui plus que jamais interrogées dans leurs pratiques et dans leurs offres de services, face aux métamorphoses rapides d'internet. Au règne de l'hypertexte a succédé le web de données, qui laisse à présent la place aux réseaux sociaux: aux Etats-Unis, la fréquentation de Facebook vient de détrôner celle de Google. Les usages mobiles s'intensifient grâce au développement des *smartphones*. Le grand public consulte ou alimente des plateformes de diffusion de vidéos, de textes, de photos, qu'il choisit ou non de partager avec d'autres internautes. Le monde de la recherche développe des entrepôts de données et des archives ouvertes, en tentant de faire évoluer les modalités de la publication scientifique. Dans cet univers aux facettes toujours changeantes, la notion de « collection numérique », qui repose sur une analogie forte avec le monde physique, a-t-elle un sens? Cette question est lourde d'implication, car nier la validité de ce concept reviendrait à mettre en cause l'idée même d'une présence spécifique et forte des bibliothèques sur internet, tant la collection est inhérente aux bibliothèques, dont elle structure et conditionne l'ensemble de l'activité.

Le web embrasse la totalité des savoirs

Unité discrète de la collection, le document sous forme numérique ou électronique a fait l'objet de nombreuses études et d'une abondante littérature depuis la seconde moitié des années 90, soulignant la radicale singularité du document produit électroniquement par rapport au document publié sur support papier.

Aujourd'hui, il est intéressant de noter que l'expression « document numérique » renvoie moins à la production originale de l'information dans un environnement informatique en réseau qu'à l'idée de dématérialisation d'un support physique, qui en conserverait les principales caractéristiques informationnelles. Le livre électronique entre peu à peu dans les usages, et les industriels comme les éditeurs tablent sur son développement rapide. Les initiatives de numérisation se multiplient dans le monde, qu'elles soient l'œuvre des éditeurs, des institutions culturelles ou des sociétés privées. Internet est conçu comme un support, un réceptacle à cet apport d'information provenant du monde physique. Si le web s'est constitué comme un mode alternatif de production et de diffusion de l'information, il apparaît aujourd'hui comme devoir embrasser la totalité des savoirs, ceux qu'il produit lui-même comme ceux qui sont d'abord produits en dehors de lui. La constitution de collections numériques s'insère dans ce mouvement général.

Une collection comme une autre, en théorie...

Du point de vue théorique, la collection numérique ne semble pas déroger à la définition canonique de la collection (au sens bibliothéconomique), c'est-à-dire un ensemble cohérent de documents, établi en vue d'un usage précis, faisant l'objet d'une gestion. Chacun des objets qui la composent a plus de valeur dans l'entité collective qu'il n'en aurait individuellement. Définition large, qui rend compte de la diversité des fonds conservés dans les bibliothèques, mais qui convient aussi parfaitement à la



Page d'accueil de Gallica. © BnF

pluralité des collections numériques. Cette expression recouvre en effet trois ensembles distincts, partageant un certain nombre de caractéristiques, mais obéissant à des modalités de constitution sensiblement différentes. Ces trois sous-ensembles sont : les abonnements de périodiques électroniques de niveau recherche (et plus largement la publication scientifique éditée sous forme dématérialisée, hors modèle *open access*) ; l'offre culturelle numérique d'*e-books*, de revues, de titres de presse, de VOD (vidéo à la demande), de musique ; et enfin les collections de documents numérisés par les bibliothèques disposant de fonds patrimoniaux. Ces trois catégories¹ de collections numériques répondent à trois types de besoins documentaires, non exclusifs les uns des autres : pour l'enseignement supérieur et la recherche, pour la lecture publique, pour la valorisation patrimoniale.

Les modalités de constitution de la collection numérique, selon qu'elle

relève de tel ou tel sous-ensemble, sont conditionnées par des facteurs différents. L'offre de documentation électronique de niveau recherche (essentiellement des périodiques et des bases de données, mais aussi les *e-books*), la première à s'être imposée, est aujourd'hui fortement concentrée autour d'un petit nombre de distributeurs exclusifs et donc incontournables, de dimension internationale, libres de fixer les prix et les modalités de consultation de leurs documents et de leurs archives. Parallèlement, l'offre culturelle d'*e-books*, de VOD et de musique, se développe plus lentement, mais commence à trouver sa place dans les collections des bibliothèques publiques, sous forme de prêt dématérialisé (téléchargement de documents chrono-dégradables), ou en lien avec le prêt de liseuses électroniques. Quant à la numérisation patrimoniale, elle se partage généralement entre deux objectifs : la mise en valeur d'un patrimoine local (en particulier pour les bibliothèques territoriales) ou la mise à disposition de documents historiques ou de référence dans une discipline précise (notamment pour les bibliothèques universitaires). La

BnF et la bibliothèque municipale de Lyon, engagées dans des programmes de numérisation d'envergure, poursuivent d'autres objectifs : constituer de vastes collections numériques d'imprimés patrimoniaux à partir de leurs fonds anciens. L'ampleur de ces initiatives dépasse largement le cadre seul de la valorisation de corpus ciblés, en créant pour ainsi dire des « hyper-collections ».

Le bibliothécaire dépossédé ?

Pour le professionnel, la collection numérique peut être perçue comme limitative, en comparaison avec les collections physiques, pour plusieurs raisons. En premier lieu, pour les collections « acquises » auprès d'éditeurs (pour la recherche ou la lecture publique), le bibliothécaire verra le plus souvent sa liberté de choix fortement limitée par une offre par « bouquets » ou packages, c'est-à-dire un ensemble de titres déjà constitué par l'éditeur ou le distributeur. L'avantage est réel, il se traduit par un coût global moins élevé, pour un nombre de titres plus grand.

1. À la BnF, s'ajoute à ces trois ensembles le dépôt légal du web, auquel s'appliquent également les principes de gestion détaillés plus loin.

Mais, parmi ces titres, nombreux sont ceux que la bibliothèque n'aurait pas spontanément choisis, et qui se trouvent sous-utilisés, pour ne pas répondre à un besoin réel du public local. Dans ce modèle, la bibliothèque paye pour ce qu'elle veut, mais aussi pour ce qu'elle ne veut pas. De plus, le bibliothécaire se considère dépossédé d'une grande partie de son travail (et de son pouvoir) qui consiste à sélectionner et acquérir les documents entrant dans les collections. Ce sentiment est renforcé par le fait que, bien souvent, la consultation de ces documents s'effectue sur les plateformes des distributeurs, et non au sein d'un système informatique dont la bibliothèque aurait la pleine maîtrise. Cette dépossession est vraie, mais en partie seulement, comme nous le verrons plus tard, car la collection numérique nécessite d'être traitée et gérée, tout autant qu'une collection physique, et sans doute davantage. La numérisation patrimoniale procure d'autres motifs de frustration que sont les contraintes juridiques et techniques, qui obligent à exclure de la numérisation des documents qui, du point de vue du contenu, ou parce qu'ils appartiennent à un corpus donné, mériteraient d'être numérisés. C'est parfois une collection « à trous » qui est mise à la disposition des internautes, qui ne connaissent pas ces contraintes et s'interrogent sur la non-disponibilité de certains titres. Pourtant, qu'il s'agisse des acquisitions électroniques ou de la numérisation patrimoniale, la collection numérique porte en elle les réponses à ses propres limitations. Ces réponses sont l'enrichissement collaboratif et la valorisation *a posteriori*.

Créer virtuellement une collection par des apports multiples

L'une des principales forces de la collection numérique par rapport à la collection physique est qu'elle peut se constituer à partir d'apports multiples. L'interopérabilité entre les gisements de ressources numériques, lorsqu'elle est techniquement possible et pertinente du point de vue de la politique

documentaire que l'on choisit, permet de recréer virtuellement un ensemble cohérent, construit à partir de contributions documentaires dispersées, mais dont le regroupement fait sens. Le concept même de collection reste pleinement opératoire : on réunit des ressources, numérisées par des bibliothèques différentes, dans une discipline donnée, ou pour reconstituer un fonds dont les éléments physiques ont pu être dispersés par les aléas de l'histoire. Pour la BnF, l'offre documentaire de la bibliothèque numérique Gallica doit servir à la création de tels ensembles, en concertation avec d'autres établissements. De même, les bibliothèques disposant de ressources numériques complémentaires de Gallica sont invitées à y présenter leurs documents². Pour ne prendre que quelques exemples, c'est ainsi que la collection numérique en histoire de la médecine est complétée par la collection des monographies de Medic@ (bibliothèque interuniversitaire de médecine et d'odontologie)³, ou que le fonds de partitions de musique baroque est enrichi des documents numérisés par les bibliothèques municipales de Versailles et de Toulouse. Réciproquement, les ressources numérisées par la BnF concernant les Antilles et la Guyane sont référencées sur le portail Manioc⁴, celles concernant le Brésil sur le portail de la Bibliothèque nationale du Brésil⁵, etc. Enfin, l'offre documentaire patrimoniale de Gallica est complétée par plus de 23 000 ouvrages sous droits numérisés par les éditeurs, ouvrages qui font pour la plupart écho aux œuvres anciennes, soit en les prolongeant (ouvrages actuels de fiction), soit en les complétant (études et essais contemporains sur des auteurs ou des thèmes représentés dans la collection du domaine public).

2. Soit par intégration directe dans le programme de numérisation des imprimés (dans ce cas, le document est visualisable dans l'interface de Gallica), soit par indexation des métadonnées suivant le protocole OAI-PMH (dans ce cas, le document est signalé dans la liste de résultats, mais il est consultable sur le site internet du partenaire).

3. www.bium.univ-paris5.fr

4. Bibliothèque numérique Caraïbe, Amazonie, plateau des Guyanes : www.manioc.org

5. www.bn.br

Les collections numériques ne doivent pas être considérées de manière isolée, mais bien dans leurs relations possibles et leurs interactions, l'objectif étant de multiplier les points d'accès aux documents (un document peut appartenir simultanément à plusieurs fonds ou collections numériques, ce qui est rarement le cas dans une bibliothèque physique), et de constituer les ensembles documentaires les plus pertinents possibles en puisant à des sources multiples. Les possibilités de valorisation éditoriale de ces nouveaux ensembles (dossiers, articles en ligne, lettres d'information, expositions virtuelles, outils pédagogiques, animation de communautés sur les réseaux sociaux...) peuvent être également envisagées de manière collaborative, associant une pluralité de partenaires et de contributeurs. Sur une thématique donnée, il pourra être intéressant d'établir des passerelles entre des documents numériques relevant des trois catégories. Cette possibilité est aujourd'hui insuffisamment mise en œuvre, car c'est le portail qui fait seul le lien entre des corpus souvent juxtaposés. Il s'agit d'une voie nouvelle que nous devrions explorer plus largement.

La gestion de la collection numérique

La notion de collection numérique ne peut être abordée sans poser la question centrale de sa gestion (on emploierait en anglais le terme « *curation* », qui malheureusement n'a pas de véritable équivalent en français). Une des caractéristiques inattendues et pourtant essentielles de la collection numérique est en effet de devoir être gérée, c'est-à-dire de faire l'objet d'opérations bibliothéconomiques qui portent sur sa constitution, sa conservation dans le temps, sa communication à un public et sa valorisation.

En ce sens, la gestion de la collection numérique s'inscrit dans le cadre traditionnel des métiers de bibliothèque, dont elle mobilise les savoir-faire existants pour les appliquer à un objet nouveau, dans un contexte différent de celui des collections traditionnelles. Ce contexte diffère à plusieurs titres :

d'une part, parce qu'il porte sur un objet particulier, le numérique, dont les caractéristiques intrinsèques (notamment techniques) doivent être gérées, et d'autre part parce qu'il s'inscrit dans un environnement, celui du web, qui a ses propres usages et ses propres règles, dont la bibliothèque ne saurait faire abstraction. Le premier aspect concerne surtout la collecte et la conservation; le second, plutôt la communication et la valorisation.

Dans ce contexte, la bibliothèque va devoir mobiliser des compétences, des organisations et des forces humaines qui vont lui permettre de mener à bien cette activité de gestion de la collection. C'est dans le domaine de la numérisation patrimoniale que la maîtrise de cette gestion par la bibliothèque est la plus complète, d'un bout à l'autre de la chaîne, tandis que pour les ressources numériques acquises, cette gestion est en partie le fait des éditeurs.

Le numérique exige des compétences nouvelles

Il existe une continuité assez forte, au moins d'un point de vue méthodologique, entre la collection traditionnelle et la collection numérique, puisque les activités de gestion sont presque semblables. Toutefois, les compétences à mobiliser sont très différentes:

- *En termes de sélection* (équivalent numérique de l'acquisition), les ressemblances portent sur la notion de politique documentaire. Mais les contraintes spécifiques au numérique – juridiques en particulier – nécessitent de mobiliser des compétences nouvelles. Il faut connaître les principes du droit de propriété intellectuelle et savoir les appliquer à un contexte particulier.

- *En termes de signalement*, il est toujours nécessaire de décrire les documents, mais cette description ne se limite plus au contenu. Il faut également décrire les caractéristiques techniques du document (format, taille, environnement technique nécessaire pour sa lecture, etc.) Le nombre de métadonnées nécessaire est donc plus important et plus diversifié: métadon-

Numérisation de masse à la BnF

À la BnF, dans le cadre de la numérisation de masse, on a demandé aux agents chargés des collections traditionnelles de sélectionner les ouvrages destinés à la numérisation. Cette sélection met en jeu une combinaison de critères qui leur sont très familiers (intérêt documentaire, état de conservation de l'ouvrage), de critères plus techniques (possibilité d'utiliser un logiciel d'OCR*, «ouvrabilité» de la reliure) et de critères juridiques (l'ouvrage est-il libre de droits?). Autant de compétences à mobiliser, qui ne sont pas directement des compétences numériques, mais qui entrent en jeu dans le contexte spécifique de la constitution d'une collection numérique.

Avec la pratique, il est apparu que ces agents avaient besoin d'une visibilité sur le reste de la chaîne une fois l'ouvrage envoyé à la numérisation. En effet, entre la sélection et la mise en ligne, il y a un «tunnel» de plusieurs semaines, une période durant laquelle on a besoin de savoir à quel stade se situent, d'une part, l'ouvrage physique en attendant son retour dans les collections et, d'autre part, l'ouvrage numérique en attendant sa disponibilité en ligne (pour pouvoir, par exemple, indiquer à un lecteur quand il pourra consulter l'une ou l'autre version).

Des outils de gestion existants ont été adaptés pour donner cette visibilité aux chargés de collection. D'autres outils ont été créés. Mais on s'est également aperçu que ce n'était pas seulement une question d'outil. Une fois l'ouvrage en ligne, le chargé de collection qui vérifie l'état de l'ouvrage numérique et constate un problème de qualité (par exemple, une page manquante, une page dont la qualité d'OCR est moins bonne...) doit connaître suffisamment bien la chaîne pour être capable d'identifier d'où vient le problème. S'agit-il d'un problème de qualité à la source, dans l'ouvrage original? D'une mauvaise manipulation du prestataire de numérisation? D'un problème dans le traitement des fichiers informatiques une fois que ceux-ci ont été reçus par le service informatique? D'un problème à l'affichage dans Gallica?

* Reconnaissance optique de caractère.

nées descriptives, techniques, administratives, de structure. Heureusement, beaucoup de ces métadonnées peuvent être générées automatiquement.

- *En termes de conservation*: c'est peut-être la partie la plus technique, mais pas plus que la conservation des supports traditionnels, qui implique une connaissance des aspects chimiques des encres, papiers, cuirs, moisissures, etc. S'il y a bien une compétence technique à acquérir, cela ne signifie pas que la conservation de la collection numérique ne doit être qu'une affaire d'informaticiens. Les équipes informatiques peuvent fournir le support de l'activité; mais les décisions, elles, relèvent de savoir-faire bibliothéconomiques. Il faut appliquer des méthodes telles que la gestion de risque, la priorisation des actions de préservation (éventuellement en fonction de la valeur des contenus), etc.

- *En termes de diffusion et de valorisation*: comme les collections traditionnelles, les collections numériques doivent faire l'objet d'une médiation. Celle-ci est rendue plus complexe par le fait que le public peut être distant, donc plus difficile à connaître et à saisir. Elle doit prendre toute sa place sur le site de la bibliothèque, mais aussi s'appuyer sur les outils «naturels» des internautes: blogs, réseaux sociaux, wikis, etc.

Une médiation technique indispensable

On voit donc que la gestion de la collection numérique impose de questionner les savoir-faire des bibliothécaires dans une perspective de compétences et d'évolution des métiers. Mais cela pose tout de même un certain nombre d'autres questions.

Quels outils informatiques?

La collection numérique ne peut pas être appréhendée «directement». Elle requiert une médiation technique (c'est-à-dire des outils informatiques) entre le bibliothécaire et la collection, aussi bien qu'entre l'utilisateur et la collection. Or, ici, nous en sommes encore aux balbutiements.

Nous avons consacré beaucoup d'énergie à concevoir des outils (les bibliothèques numériques) de médiation entre la collection numérique et l'utilisateur. Mais ces outils ne sont pas adaptés, ou pas suffisants, pour la gestion de la collection. C'est le même problème qu'avec le catalogue, dont on a voulu faire à la fois un outil de recherche documentaire et un outil de gestion d'une collection physique... Dans le cas de la gestion de la collection numérique, cela se manifeste de façon très problématique, parce que le signalement n'est pas seul en jeu.

Il faut donc définir les outils dont on aura besoin pour gérer la collection, ce qui peut s'avérer complexe car nous manquons encore de visibilité sur la nature des tâches de gestion au quotidien. Dans certains cas, des outils existants peuvent être adaptés. Dans d'autres, il faut créer, inventer de nouvelles solutions.

Des questions en termes d'organisation

Le numérique pose des questions d'organisation accrues, toujours à cause de la médiation technique qui existe entre la collection et le bibliothécaire.

Les frontières sont plus floues entre les métiers strictement techniques (informatiques) et les métiers de bibliothèque. Pour que les acteurs puissent dialoguer, il est nécessaire qu'ils partagent un vocabulaire, des notions, des compétences communes.

Par ailleurs, il existe une transversalité très forte dans la gestion de la collection numérique; il est beaucoup plus difficile de séparer les tâches qu'avec la collection traditionnelle. Pour prendre un exemple trivial, la façon dont on décide d'équiper ou de relier un document n'a que peu d'impact sur la façon dont il sera communiqué en salle de lecture. Au contraire, s'agissant d'un document numérique, la façon dont il est produit (techniquement parlant) détermine fortement les opérations de conservation qu'on sera capable d'effectuer, et les modalités de consultation dans la bibliothèque numérique.

Dans une gestion de collection numérique, on ne peut donc pas mener

une tâche sans avoir la vision d'ensemble de la chaîne de traitement des contenus, de la production à l'accès. Dans le cas de la répartition des tâches sur plusieurs équipes, une contractualisation accrue est nécessaire: les équipes doivent s'entendre sur un vocabulaire commun, des indicateurs qui leur permettent d'évaluer et de manipuler la collection conjointement, et prendre des engagements sur les tâches à effectuer dans un certain délai pour que les autres tâches dépendantes puissent être conduites.

Enfin, il peut y avoir également des interactions entre la collection traditionnelle et la collection numérique au sein d'un même établissement ou, si l'on veut, un effet rétroactif du numérique sur la collection traditionnelle (la façon dont elle est conservée, manipulée, consultée...), et cela doit aussi être pris en compte.

Se former à la culture du numérique

Pour les raisons évoquées précédemment, l'acquisition des compétences spécifiques nécessaires aux gestionnaires de collections numériques est un défi. Une formation à cette gestion ne peut pas se limiter à l'apprentissage concret de tâches. C'est toute une culture professionnelle du numérique qu'il faut élaborer et transmettre, de façon à permettre aux acteurs de s'inscrire dans un environnement global, de comprendre les implications de leurs choix et de leurs actions.

Ce mode de formation très ouvert n'est pas vraiment naturel dans le domaine de la formation continue, qui est généralement plus applicative: prise en main d'un outil, apprentissage d'une tâche particulière quand on change de poste, préparation à un changement de carrière...

De plus, les experts susceptibles de former leurs collègues sont encore peu nombreux. Et dans bien des domaines – comme nous l'avons vu avec les outils –, le sujet est encore trop jeune pour que l'on puisse clairement exprimer et transmettre: on est souvent amené à communiquer sur des actions expérimentales, en cours, qui relèvent d'une expérience empirique.

Beaucoup des personnels qui gèrent les collections numériques aujourd'hui se sont formés «sur le tas».

Il est impératif (voir encadré), pour gérer la collection numérique, de connaître l'ensemble de la chaîne de traitement, y compris, jusqu'à un certain point, dans ses aspects techniques.

L'unicité du concept de collection numérique ne va pas de soi, car il porte sur des réalités différentes, selon que l'on parle de l'offre scientifique, de l'offre culturelle contemporaine ou de l'offre patrimoniale, qui toutes relèvent de problématiques différentes.

Pour autant, le concept de collection numérique reste valide dans la mesure où les pratiques professionnelles lui donnent une cohérence et un sens. Cela est vrai lorsqu'on envisage les possibilités multiples offertes par le numérique d'établir des passerelles entre les gisements documentaires, et par le travail en concertation. Cela est encore vrai à considérer l'activité centrale de gestion de la collection numérique, activité encore jeune mais dont on sait déjà qu'elle mobilise des compétences particulières et nécessite des outils adaptés. Pour autant, il s'agit bien d'une activité qui relève pleinement du métier de bibliothécaire, dans la continuité des savoir-faire existants sur les collections traditionnelles, et en articulation avec eux. Ceci étant posé, de nombreuses voies restent encore à explorer pour donner vie à la collection numérique et en exprimer toutes les potentialités. Sa propre dynamique invite à l'action. ●

Mars 2010