



HAL
open science

Les chargés de collections numériques au travail : une analyse des identités professionnelles émergentes à la Bibliothèque nationale de France (BnF)

Géraldine Camile, Marianne Clatin, Louise Fauduet, Clément Oury,
Jean-Philippe Tramoni

► To cite this version:

Géraldine Camile, Marianne Clatin, Louise Fauduet, Clément Oury, Jean-Philippe Tramoni. Les chargés de collections numériques au travail : une analyse des identités professionnelles émergentes à la Bibliothèque nationale de France (BnF). 2015. hal-03609759

HAL Id: hal-03609759

<https://hal-bnf.archives-ouvertes.fr/hal-03609759>

Preprint submitted on 15 Mar 2022

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



Distributed under a Creative Commons Attribution| 4.0 International License

Les chargés de collections numériques au travail : une analyse des identités professionnelles émergentes à la Bibliothèque nationale de France (BnF)

Traduction française de l'article *Digital curators at work: analyzing emerging professional identities at the Bibliothèque nationale de France (BnF)*

Traduction assurée par **Géraldine Camile**

Département du dépôt légal, Bibliothèque nationale de France, Paris, France
geraldine.camile@bnf.fr

Marianne Clatin

Département de l'information bibliographique et numérique, Bibliothèque nationale de France, Paris, France

Louise Fauduet

Département de l'audiovisuel, Bibliothèque nationale de France, Paris, France
louise.fauduet@bnf.fr

Clément Oury

Département du dépôt légal, Bibliothèque nationale de France, Paris, France
clement.oury@bnf.fr

Jean-Philippe Tramoni

Département du système d'information, Bibliothèque nationale de France, Paris, France
jean-philippe.tramoni@bnf.fr



Copyright © 2014 par Géraldine Camile, Marianne Clatin, Louise Fauduet, Clément Oury et Jean-Philippe Tramoni. Ce travail est disponible sous licence Creative Commons Attribution 3.0:
<http://creativecommons.org/licenses/by/3.0/>

Résumé :

La gestion de collections numériques, ou *digital curation*, est un concept global qui inclut tous les aspects du travail sur les objets culturels numérisés et nés numériques: de l'entrée du document à la description des données ou l'indexation, le stockage, la diffusion et la préservation à long terme. C'est une activité en pleine expansion, dont les règles et les meilleures pratiques ne sont pas encore établies à l'échelle mondiale. La BnF a choisi de s'appuyer sur des observations sur le terrain pour comprendre comment son personnel, face au besoin urgent de collecter et de préserver un patrimoine numérique en rapide croissance, développe de nouvelles tâches et compétences. Un groupe dédié, appelé ORHION, acronyme français pour Observatoire des Organisations et Ressources Humaines sous l'Impact Organisationnel du Numérique, a lancé une série d'études sur les identités professionnelles particulièrement impactées par l'accroissement des activités numériques. Il a jusqu'ici terminé l'analyse de deux sujets : la gestion des métadonnées numériques et les entrées

numériques. Afin d'évaluer les pratiques de la BnF, ORHION s'appuie également sur l'expérience d'autres institutions. Par exemple il a organisé un atelier sur la gestion des archives du web en mai 2014, lors de l'assemblée générale du Consortium international pour la préservation de l'internet (*International internet preservation consortium* ou IIPC).

Ce document se concentre sur les résultats de l'analyse des trois cas d'utilisation suivants : le passage du catalogage à la gestion des métadonnées pour la description et l'accès ; les processus et les agents impliqués dans les entrées numériques ; et l'organisation, les compétences et la formation des équipes chargées des collections d'archives du web. Sur la base de ces trois expériences, ce document vise à identifier certaines tendances clé de la gestion des collections numériques : la notion de collection numérique elle-même n'est pas évidente et doit être repensée ; les identités professionnelles sont remises en cause par l'apparition de nouvelles tâches et l'émergence de nouveaux acteurs ; les personnes concernées peinent à comprendre leurs propres rôles et responsabilités dans des chaînes de traitement plus complexes. La formation du personnel et le partage de la charge de travail entre les premiers pionniers et le reste des équipes sont nécessaires pour impliquer tous les professionnels dans la gestion des collections numériques.

Mots-clés : gestion de collections numériques, compétences et formation, évolution des identités professionnelles, ressources humaines, organisations.

1 L'approche de la BnF et l'observatoire ORHION

La gestion de collections numériques est une activité en pleine expansion au sein des bibliothèques et des autres institutions patrimoniales. Ce concept inclut tous les aspects du travail sur les objets numérisés et nés numériques : de l'entrée du document à la description des données ou l'indexation, le stockage, la diffusion, et la préservation à long terme. Dans un certain sens, c'est une adaptation aux documents numériques de la chaîne de traitement traditionnelle utilisée pour des supports physiques analogues. A la Bibliothèque nationale de France, la gestion de collections numériques est apparue dès que la bibliothèque a commencé à manipuler des documents numériques : à la fin des années 90 quand sa bibliothèque numérique, Gallica, a été lancée. Le périmètre des documents numériques s'est étendu rapidement. Il s'est étendu « horizontalement » : la bibliothèque a pris la responsabilité de nouveaux types d'objets numériques : les archives du web, les contenus audiovisuels nés numériques, les livres numériques... et « verticalement » : de la création ou la réception de contenus numériques à l'accès et la préservation à long terme – l'entrepôt numérique de la BnF, SPAR, a été lancé en 2008 [1].

Cependant, le besoin de considérer l'impact de ces nouvelles activités sur les professionnels de la bibliothèque, sur les « chargés de collections numériques », a émergé plus tard. Quand la bibliothèque a lancé ses programmes de « numérisation de masse » en 2007-2008, des centaines de bibliothécaires et d'autres acteurs de la BnF ont été impliqués dans des tâches liées à la gestion des documents numériques. Il est alors devenu nécessaire d'identifier les questions auxquelles ces professionnels étaient confrontés, les formations qui étaient requises, et les défis organisationnels rencontrés. Tout au long de 2008 et 2009, l'idée d'un groupe informel de bibliothécaires intéressés par le partage de leur expérience sur les défis que les collections numériques posent dans leurs activités quotidiennes a été mise en pratique. Ce groupe a été appelé ORHION, acronyme français pour Observatoire des Organisations et Ressources Humaines sous l'Impact Organisationnel du Numérique ; il s'est progressivement

développé pour devenir un groupe sur le long terme officiellement approuvé par la direction générale de la BnF en 2010 [2 et 3].

Le groupe opère à l'extérieur des structures existantes de la BnF, puisque son rôle est d'observer et d'articuler les changements dans les pratiques et les compétences des bibliothécaires, et non d'élaborer des stratégies pour la bibliothèque. Cependant il joue un rôle important dans la sensibilisation grâce à quatre types d'action : des entretiens, des groupes de travail, des ateliers et des réunions d'information pour l'encadrement.

Pendant ses premières années, ORHION a traité de questions telles que la définition de collection(s) numérique(s), leur gestion dans des entrepôts numériques, et leur valorisation auprès des lecteurs et des usagers de l'internet. En 2013, l'accent a été mis sur l'aspect humain de la question : ORHION a lancé une série d'études sur les identités professionnelles qui sont particulièrement impactées par l'accroissement des activités numériques. L'encadrement de la BnF a proposé quatre domaines de recherche pour la période 2013-2015 : la gestion des métadonnées, les entrées numériques, l'interaction entre les bibliothécaires et les informaticiens, et les nouvelles formes de gestion de projet. ORHION a jusqu'à présent terminé ses analyses sur les deux premiers sujets.

Afin d'évaluer les pratiques de la BnF, ORHION s'appuie également sur l'expérience d'autres institutions. Il a organisé par exemple un atelier sur la gestion des archives du web cette année lors de l'assemblée générale du Consortium international pour la préservation de l'internet, en mai 2014.

Les observations d'ORHION n'ont pas pour but de donner une définition de la gestion des collections numériques. Cependant à travers des études de cas, elles ont apporté une analyse pratique sur les activités des chargés de collections numériques. C'est pourquoi cet article s'intéressera spécifiquement aux résultats de trois enquêtes récemment réalisées : le passage du catalogage à la gestion des métadonnées pour la description et l'accès, les processus et les agents impliqués par les entrées numériques, et l'organisation, les compétences et la formation des équipes chargées des collections d'archives du web.

2 Le passage du catalogage à la gestion des métadonnées pour la description et l'accès

Étudier la gestion des métadonnées pour la description suppose non seulement d'examiner le travail des catalogueurs, mais également les différents métiers et compétences nécessaires pour rendre un document identifiable soit dans les catalogues, soit dans la bibliothèque numérique Gallica. L'accent a été mis sur trois questions spécifiques : la granularité de la description, les métadonnées ne provenant pas de la bibliothèque, et les interactions entre les multiples emplois et compétences requises pour créer des métadonnées liées au processus de numérisation.

Les questions de granularité

La question de la granularité dans le catalogue est posée depuis un certain temps, en particulier en matière de publications en série. Cette question touche également la bibliothèque numérique, dans la mesure où une grande quantité de documents disponibles dans Gallica est de la presse. Sans table des matières ni index, le signalement de ces collections suppose d'identifier les différents fascicules d'un journal. Même si cette tâche immense n'a jamais été faite auparavant dans le catalogue, les compétences nécessaires pour cela existent déjà dans la bibliothèque : il est en effet comparable au travail réalisé par des magasiniers lors de la préparation des documents pour le microfilmage.

D'autres types de collections numérisées posent des questions de granularité, comme les images de grands photographes, collectés précédemment sur support « diapositive » et maintenant au format numérique. Dans ce cas, leur signalement a été réalisé par les mêmes catalogueurs qui décrivaient déjà des photographies analogiques. Alors que le catalogage d'un lot de photographies dans le catalogue est de fait suffisant, dans la mesure où l'on peut se contenter de communiquer aux lecteurs l'ensemble d'un lot 71 diapositives prises lors d'un festival de théâtre, cela ne l'est plus dans une bibliothèque numérique où les lecteurs espèrent trouver une image spécifique sur un sujet précis – sans doute trouvée dans différentes séries de diapositives. Le fait que l'on décide désormais de décrire ces collections au niveau de chaque image prouve que les critères de description sont maintenant influencés par les conditions d'accès aux documents dans la bibliothèque numérique, et non plus seulement par le catalogue. Cependant tout le monde s'est accordé pour dire que les améliorations apportées par la bibliothèque numérique devraient aussi profiter au catalogue original lui-même, qui est toujours considéré comme le principal entrepôt de métadonnées descriptives sur lequel les nouveaux services doivent être construits dans le futur.

Les métadonnées descriptives créées à l'extérieur de la bibliothèque

En tant qu'agence bibliographique nationale, la BnF est responsable de la description des documents physiques dans le cadre de la bibliographie nationale (produite en France ou par des éditeurs français). Mais qu'en est-il des autres types de documents ?

De même que les autres bibliothèques dérivent les notices des catalogues de la BnF, les catalogueurs de la BnF dérivent les notices de documents étrangers d'autres réservoirs, comme WorldCat. Le temps économisé grâce à ça peut être utilisé pour d'autres tâches, comme la sélection pour la numérisation ou le dépôt légal du web. Cependant comme c'est l'activité même de catalogage qui est concernée, ce processus remet aussi en question l'identité des catalogueurs. Choisir la meilleure notice disponible parmi les doublons dans WorldCat peut être considéré comme une tâche moins intéressante que de créer de toute pièce une notice. À l'inverse, certains catalogueurs considèrent qu'économiser du temps pour le signalement leur permet de se concentrer sur des tâches demandant un autre type d'analyse, comme choisir des vedettes sujets, créer des liens aux autorités, ou même créer de nouvelles notices d'autorité. La même question se posera bientôt quand la BnF importera les descriptions des éditeurs en ONIX pour les livres numériques français reçus par dépôt légal.

Un autre défi portant à la fois sur les compétences et l'organisation est venu des livres numériques acquis individuellement. L'achat consiste ici à sélectionner une référence dans la base de données internationale EBSCO. Quand le document n'est pas encore référencé (ce qui est souvent le cas pour les documents qui ne sont pas écrits en anglais), les bibliothécaires créent eux-mêmes une courte description, selon les critères déjà définis dans la base de données EBSCO. Cependant certains considèrent cette description insuffisante pour un catalogue de bibliothèque (pas d'indexation matière, pas de lien aux notices d'autorité...). La question suivante se pose alors : quelles sont les compétences en matière de catalogage nécessaires dans de tels cas ? Les magasiniers chargés de contrôler régulièrement l'accès aux documents en dehors du catalogue ne pourraient-ils pas être aussi responsables de la création de ces quelques champs de métadonnées nécessaires à ces outils distincts ? Et quand des dizaines de milliers de livres numériques sont concernés, ces professionnels ne devraient-ils pas être répartis dans les différents départements de la bibliothèque plutôt que dans un département unique d'« acquisitions électroniques » ?

Des interactions entre les compétences professionnelles pour la description à des fins de numérisation

Quand Gallica a été ouvert, certains ont considéré que la description deviendrait rapidement moins utile, puisque les lecteurs accéderaient directement aux documents sans utiliser le catalogue. Cependant une bibliothèque numérique ne peut fonctionner sans un minimum de métadonnées (des métadonnées descriptives, techniques, et même de préservation). La numérisation des collections s'est révélée par ailleurs assez difficile quand elles n'avaient pas du tout été cataloguées. Un cas récent a donné à la BnF l'occasion d'améliorer sa capacité de création et d'adaptation de ces processus.

Quand le département des monnaies de médailles a décidé d'accélérer son programme de numérisation de plus d'une centaine de milliers de monnaies anciennes grecques non cataloguées, il est apparu que créer des notices de catalogue complètes dans le catalogue n'était pas réalisable. Un nouveau processus de métadonnées a été testé à cette occasion, considéré maintenant comme un modèle pour les prochains programmes de numérisation similaires : il a consisté à donner un modèle de description « allégé » qui ne pourrait être suffisant pour des recherches scientifiques (ex sans une description iconographique complète) mais qui serait suffisant pour distinguer les documents les uns des autres dans la bibliothèque numérique. Les catalogueurs ont alors décrit les documents dans un document CSV plutôt que dans le catalogue en MARC, qui n'a pas été conçu pour autoriser la répétition d'informations dans des notices distinctes pour des milliers de pièces identiques. Cependant utiliser un autre outil que le catalogue a permis à beaucoup d'autres acteurs, en plus des catalogueurs, de participer au processus afin d'inclure enfin les descriptions dans la bibliothèque numérique : les chargés de collections experts sur ce type de documents spécifiques ; les coordonnateurs bibliographiques ; les informaticiens impliqués dans le traitement des données de masse dans le catalogue ; les professionnels chargés du processus de photographie lui-même ; les responsables fonctionnels des entrepôts OAI ; et les développeurs de la bibliothèque numérique. L'ensemble des contributeurs ont souligné l'importance et l'enrichissement professionnel d'une telle coopération entre les différentes compétences et personnes qui n'ont pas la possibilité habituellement de travailler ensemble à la bibliothèque.

3 Les processus et les acteurs impliqués dans les entrées numériques

Après s'être intéressé aux spécialistes des métadonnées, il a été décidé de se pencher sur le cas des professionnels responsables de la manipulation de documents numériques. Un groupe de travail a été constitué pour délimiter plus précisément le domaine à étudier. Il a été décidé de mener une enquête sur cinq chaînes différentes :

- le processus de numérisation de différents documents sur support ;
- la collecte de sites web (et en particulier de la presse en ligne) dans le cadre du dépôt légal ;
- le dépôt légal des documents audiovisuels dématérialisés ;
- le dépôt légal des livres numériques ;
- l'abonnement à des périodiques dématérialisés.

La chaîne d'entrée des objets physiques a été retenue comme référence à des fins de comparaison. L'enquête s'est concentrée sur les processus et les acteurs impliqués dans ces cinq chaînes, de l'entrée initiale jusqu'au stockage à long terme, en utilisant la même

méthodologie que dans les études d'ORHION précédentes. Le catalogage et la communication ont été exclus de la portée de cette enquête.

Des flux non-linéaires

Les premiers résultats ont confirmé certaines divergences fondamentales entre les processus physiques et numériques.

La chaîne d'entrée des objets physiques se caractérise par la linéarité du processus d'entrée. On trouve des tâches communes : les processus sont de fait globalement similaires pour tous les supports physiques, des livres, périodiques ou cartes, jusqu'aux CD. Dans ce processus, les tâches de vérification (par exemple la vérification de l'état physique de l'objet) sont souvent implicites. Enfin, la manipulation d'objets physiques est un type de tâche évidemment spécifique et affectée à des acteurs spécialisés.

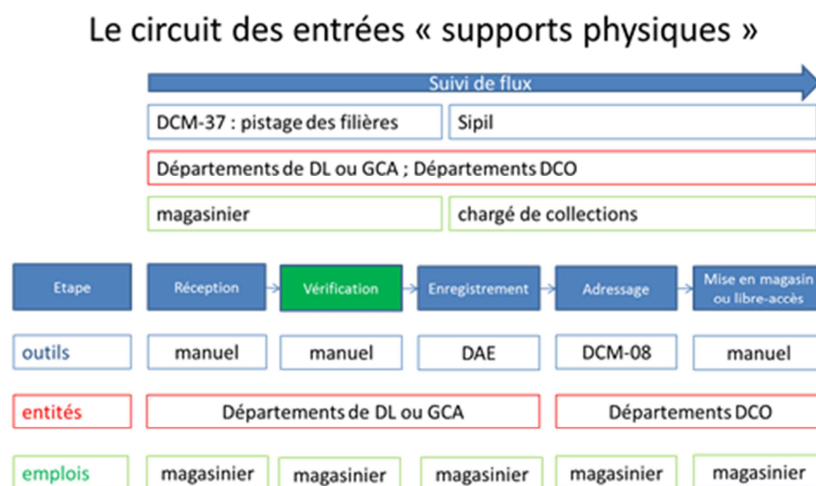


Figure 1: Schéma de la chaîne d'entrée des objets physiques

Au contraire, les chaînes d'entrée numériques sont des processus beaucoup moins séquentiels. Alors que le contrôle qualité du document reçu était implicite dans la chaîne d'entrée physique (en manipulant les documents, les professionnels vérifient s'ils sont endommagés ou pas), il devient obligatoire de la rendre explicite avec les entrées numériques, car les étapes sont automatisées. Les tâches de vérification ou de contrôle peuvent être répétées et créer donc une boucle dans le processus d'entrée.

En outre, une chaîne spécifique est nécessaire pour chaque type de document numérique : aucun modèle unique de processus n'est encore apparu. Enfin, cela implique des tâches différentes, avec des objectifs différents, des méthodes et des acteurs différents.

La chaîne d'entrée pour les documents numérisés (dans le schéma ci-dessous) représente un exemple typique de ces observations. Il y a beaucoup de contrôles et d'étapes de vérification, ce qui implique différents retours en arrière ; cela contraste avec la linéarité de la chaîne d'entrée physique. En outre, cette chaîne doit gérer deux flux parallèles. Par conséquent, une gestion plus explicite des flux de documents est nécessaire pour faire face à cette complexité.

Le circuit des entrées de numérisation

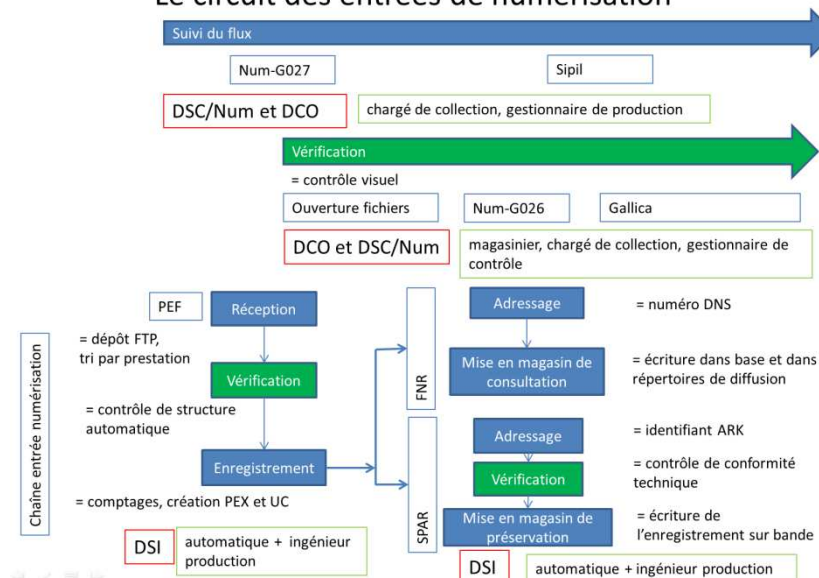


Figure 2: Schéma de la chaîne d'entrée de la numérisation

Une évolution de la distribution des rôles et des responsabilités

L'étude a montré une évolution dans le partage des rôles et des responsabilités. Dans la chaîne d'entrée physique, les magasiniers sont les acteurs clés du processus – ce n'est pas le cas pour les documents numériques. Par exemple, pour la chaîne d'entrée de la presse en ligne, les responsabilités sont réparties entre les ingénieurs qui gèrent le flux des opérations informatiques et des bibliothécaires en amont et en aval, sous la supervision de l'équipe du dépôt légal numérique. Cependant, le processus n'est pas entièrement automatisé ; et il est nécessaire de définir précisément ce qui doit être contrôlé, comment, par qui, et où dans le processus.

Le circuit des entrées DL web/presse en ligne

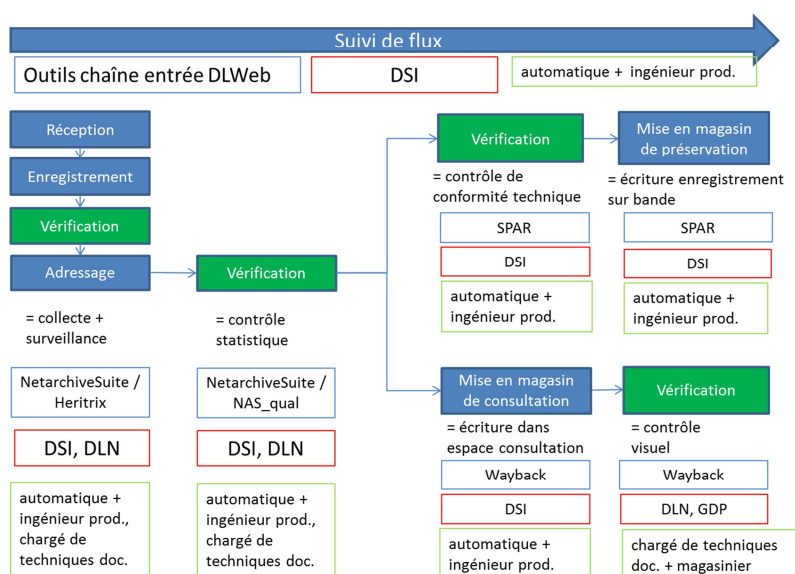


Figure 3: schéma de la chaîne d'entrée de la presse en ligne

Pour ce processus d'entrée comme pour les autres processus étudiés, le rôle des informaticiens a donc changé. Leur rôle est non seulement d'appuyer les bibliothécaires en leur fournissant des outils, mais aussi de gérer directement la collection - cette situation peut être comparée à celle de la préservation numérique, où l'administration directe de l'entrepôt est assurée par des ingénieurs (qui mettent en œuvre les exigences et les spécifications exprimées par les bibliothécaires).

Même si le rôle de magasinier a changé, il n'est pas moins important, car ceux-ci doivent vérifier la qualité des collections à chaque étape du processus. Cependant, l'échantillonnage est souvent nécessaire en raison de l'énorme quantité de documents. Pour définir le travail de magasinier lui-même et ses responsabilités, il devient nécessaire d'établir une typologie de ces tâches : contrôle visuel, vérification de la conformité ou de l'exhaustivité, audit ...

Des problèmes de définition d'identité professionnelle

Les entretiens avec des professionnels qui travaillent sur les cinq chaînes sélectionnées ont mis en évidence certaines préoccupations qui ont plus tard été débattues lors d'un séminaire. Outre les deux questions qui ont été déjà identifiées (émergence de nouveaux acteurs et transformation du travail des magasiniers), deux autres ont été spécifiquement abordées : le sentiment de perdre le contact avec le document, et l'absence de perception du processus d'entrée dans son ensemble.

Le sentiment de perte de contact avec le document est particulièrement fort pour les acteurs qui contrôlent le flux de documents numériques grâce à des données agrégées - certains d'entre eux travaillent même essentiellement sur des statistiques financières et administratives. Ces acteurs peuvent percevoir leur travail comme trop abstrait. De l'autre côté, le contrôle qualité via le contrôle visuel des documents, en dépit de son côté fastidieux, permet un sentiment d'appropriation de la collection.

Pour compenser cette perte de sens, les gestionnaires doivent concevoir comment répartir le traitement des supports physiques et des objets numériques entre les équipes. En divisant la charge de travail et en l'attribuant à des groupes distincts d'acteurs, on peut donner l'impression que certaines tâches ou chaînes sont mieux considérées que les autres : il existe alors un risque de créer une distance entre les équipes au sein de l'organisation. En outre, on perçoit que la réutilisation de la compétence et de l'expertise acquises sur le traitement des entrées physiques (par exemple, la connaissance du système de publication des périodiques) peut être avantageuse et plus efficace pour contrôler les entrées numériques (par exemple, la collecte de la presse en ligne dans le cadre du dépôt légal).

Enfin, en raison de ces évolutions dans le processus d'entrée, il devient difficile pour le personnel d'acquérir une perception de ce processus dans son ensemble. Pour réduire ce manque de vision, une solution mise en œuvre a été l'élaboration de programmes de formation spécifiques, non seulement envers les personnes directement impliquées dans les activités numériques, mais aussi vers les gestionnaires et le personnel d'autres départements.

4 Les chargés de collections d'archives du web : organisation, compétences et formation

L'assemblée générale du Consortium international pour la préservation de l'internet ayant lieu à Paris en 2014, les membres d'ORHION ont jugé que ce serait une excellente occasion d'examiner et de comparer les pratiques de la BnF avec celles d'autres institutions. Comme l'observatoire s'intéresse spécifiquement à l'étude du changement des identités professionnelles, il a proposé un atelier sur les chargés de collections d'archives du web. Ils

furent définis à cette occasion comme les personnes participant au développement et à l'utilisation des archives du web qui ne sont ni les techniciens en charge des opérations et des évolutions des outils matériels et logiciels, ni les chercheurs constituant de la connaissance à l'aide de ces archives. Ce sont ceux qui font le lien entre les techniciens et les chercheurs, et sont souvent un peu des deux¹.

Les collections d'archives du web

Le débat a porté avant tout sur la nature des collections d'archives du web et du rôle de ceux qui sont chargés de les gérer. La constitution de collections fait partie des activités les plus vénérables du métier de bibliothécaire ; mais quand il s'agit de constituer une collection d'archives du web, la façon de s'y prendre a-t-elle changé ?

Quand les institutions bénéficient d'un mandat pour réaliser un dépôt légal de l'internet, il est souvent lié à une aire géographique nationale ou régionale. Pourtant le principe de collecte des documents selon leur aire de publication est mis à mal dans l'environnement du web, défini par des hyperliens infinis et internationaux. Les lecteurs pourraient avoir du mal à recomposer l'ensemble de la documentation qui les intéresse à partir des collections de plusieurs institutions, puisqu'elles doivent souvent en limiter à l'accès à leurs salles de lectures, ou choisissent de le faire.

D'autres participants ont souligné le besoin de prendre en compte les spécificités techniques d'une collecte du web dès le départ, et de les intégrer comme critères dans les demandes de collectes soumises par les chargés de collections. La fréquence et la profondeur de la collecte peuvent être laissées à leur appréciation, et les limites du logiciel de collecte doivent être comprises et prises en compte au moment de la sélection (par exemple, selon le dispositif en place, certains types de contenus vidéo, ou de contenus protégés par mot de passe, ne peuvent être collectés).

Que les sites susceptibles de disparaître rapidement soient à préserver en priorité a aussi été sujet à débat. Ce critère semblait peser dans les choix au début de la plupart des programmes, mais il a perdu de son importance. L'objectif premier est de construire une collection qui a du sens, cohérente en soi et complétant les forces des collections existantes. Les bibliothèques de l'université Columbia² par exemple ont constitué une archive du web sur les droits humains, une collection sur la préservation et l'aménagement urbain lié à la Avery Library, une collection sur les religions de la ville de New York liée à la Burke Library, et ainsi de suite.

La plupart des institutions ont tâché de s'appuyer sur leurs domaines d'excellence traditionnels et font travailler des chargés de collections sur des collectes thématiques. Cependant l'adéquation entre leurs domaines d'expertise et leurs sélections de sites web n'est pas simple à garantir, car la classification des sites n'est pas toujours évidente. La bibliothèque du Congrès a tenté de créer une collection de « mélanges web », dans l'espoir d'élargir les sélections et d'améliorer la participation des chargés de collections, en espérant qu'au fil du temps des ensembles raisonnés émergeraient naturellement, mais ce ne fut pas le cas. En effet, les chargés de collections éprouvent surtout des difficultés à intégrer cette nouvelle tâche à leurs responsabilités existantes, faute de temps, de ressources ou d'intérêt. Même quand ils sont motivés et n'ont pas peur de la technique, ils tendent à travailler par courtes périodes de grande activité lors de la sélection initiale sur un thème, puis leur participation s'amenuise au fil du temps.

Puisque même dans un contexte de dépôt légal, les contraintes techniques forcent à sélectionner seulement ce qui peut être collecté, la question des intérêts et des suggestions des

¹ Le programme et les présentations de l'Assemblée Générale sont disponibles à l'adresse <http://netpreserve.org/general-assembly/2014/overview>.

² https://library.columbia.edu/bts/web_resources_collection.html

lecteurs prend de l'importance. Cependant les lecteurs sont encore peu nombreux : les chercheurs et le grand public découvrent progressivement l'existence de ce type de collections, dont une part majeure de leur contenu est encore disponible en ligne, et les outils d'accès et d'analyse sont en cours de perfectionnement. Les chargés de collections doivent deviner ce qui sera utile aux générations futures, ce qui n'est pas un processus habituel pour les bibliothèques dépositaires du dépôt légal. Certaines institutions se demandent si des lecteurs pourraient directement impliqués dans la procédure ; Archive-it, fournisseur de services pour la fondation à but non lucratif Internet Archive, a mis au point un tel programme avec des enfants et des adolescents ³.

Les partenaires des chargés de collections web

Les choix de constitution de ces collections web ont été formalisés avec un degré de précision plus ou moins important et communiqués au public⁴. Ils influencent la façon dont les institutions sont organisées pour réaliser la capture, la préservation et l'accès aux archives de sites web. Les équipes au cœur de l'archivage du web font souvent partie des services dédiés au numérique, en particulier quand les programmes sont encore en développement. Ils peuvent aussi faire partie des services de dépôt légal (Bibliothèque nationale de France, Bibliothèque royale du Danemark...), des services de collections (Bibliothèque royale des Pays-Bas, British Library...) ou des services d'acquisition (Bibliothèque et Archives Canada). Certains sont des bibliothécaires experts en informatique et/ou des informaticiens, et collaborent avec des informaticiens, d'autres sont chargés de collections numériques au sein d'un réseau de sélectionneurs, et/ou des chercheurs

Les chargés de collections d'archives du web adoptent de multiples stratégies pour faire connaître leurs collections auprès des chercheurs : cela consiste souvent à guider les chercheurs pour interroger de manière pertinente les archives du web dans les domaines des sciences sociales, d'histoire ou de littérature. Ils sont des intermédiaires entre les lecteurs et les informaticiens qui peuvent développer des outils d'analyse, alors que les usages des collections du web évoluent de la recherche de documents et de contenus sur des sites web au *data mining*.

La formation des chargés de collections web

Les chargés de collections travaillant sur les archives du web sont rarement en contact direct avec le grand public des salles de lecture, et comptent sur leurs collègues pour en parler et pour aider les lecteurs. Cependant il n'y a pas de cours spécifiques pour concevoir ou gérer des archives du web. La formation est alors hiérarchisée : l'équipe principale, pour qui une connaissance technique approfondie est nécessaire, a besoin d'être en contact avec les collègues de la cinquantaine d'institutions également engagées dans l'archivage du web, grâce à des consortiums internationaux ou des événements tels qu'IIPC⁵ ou iPRES. A leur tour, ils forment les correspondants associés à la constitution de collections web et coordonnent leurs activités au fil du temps. La Bibliothèque du Congrès a par exemple intégré les archives du web dans la formation de base obligatoire des chargés de collections. Pour permettre à l'ensemble du personnel d'être sensibilisé aux archives du web et de répondre aux questions des lecteurs, des sessions d'information générale doivent aussi être organisées.

³ K-12 Web Archiving, <https://archive-it.org/k12/>.

⁴ Certains exemples de politique documentaire de collections web sont disponibles sur le site d'IIPC: <http://netpreserve.org/collection-development-policies>.

⁵ La BnF, par exemple, a organisé un atelier parrainé par IIPC sur le programme d'intégration de l'archivage du web dans son organisation en novembre 2012. Le rapport est disponible à l'adresse : <http://netpreserve.org/sites/default/files/resources/Putting%20it%20all%20together.pdf>.

5 Les enseignements de ces études en matière d'identités professionnelles

Chacune des trois études suivantes illustre un aspect différent – à un stade différent – de la gestion de collections numériques : la deuxième étude porte sur les entrées, la première sur le signalement bibliographique, la troisième sur les étapes initiale et finale de la chaîne : la sélection d'un côté et de l'autre côté la valorisation. De même, différentes catégories professionnelles ont été examinées : des magasiniers, des catalogueurs et des chargés de collections numériques. Et pourtant, est-il possible d'obtenir une vue plus large et de dépasser les spécificités de ces cas d'utilisation, sans définir plus précisément ce qu'est la gestion de collections numériques, pour identifier au moins quelques tendances principales ?

Une première conclusion de ces études est que la notion de « collection numérique » n'est pas évidente. Avec les documents numériques, il est nécessaire de s'interroger sur la notion de la granularité de la collection : chaque sous-ensemble d'un document (une diapositive dans un lot d'images, une page web dans un site archivé...) est un document en lui-même, qui peut avoir besoin d'une description spécifique. Les professionnels ont tendance à perdre la notion même de document : quand l'activité d'une personne consiste essentiellement à produire des indicateurs et des rapports, est-ce cela peut encore être considéré comme de la gestion de collections de bibliothèque ? Comme le reconnaissent les bibliothécaires s'efforçant de valoriser les archives du web, les lecteurs peuvent également avoir du mal à comprendre à quel genre de documents ils doivent faire face lorsqu'ils se confrontent avec des collections numériques. Enfin la grande diversité des documents numériques (numérisés ou nés numériques, acquis par la bibliothèque à titre temporaire ou pour le long terme) rend difficile la compréhension de l'unité des collections numériques au sein d'une même institution.

Il y a ensuite des questions sur les rôles, les responsabilités, et même les identités professionnelles des personnes qui sont chargées de gérer les collections numériques. Il faut préciser que même lorsqu'il s'agit de documents numériques, la connaissance « traditionnelle » des bibliothécaires est toujours précieuse : par exemple, le catalogage conserve une part essentielle dans l'activité des bibliothécaires (du moins dans une bibliothèque nationale comme la BnF). Les connaissances acquises par les magasiniers en matière de modèle de publication de la presse, lorsqu'ils gèrent des publications en séries, sont tout aussi précieuses pour la gestion de leur équivalent numérique. Cependant la distribution typique des rôles entre les profils professionnels est désormais remise en cause : le traitement des métadonnées descriptives n'est plus confié aux catalogueurs ; les ingénieurs remplacent les magasiniers au cœur du processus d'entrée des documents. Les chargés de collections travaillant sur les archives du web continuent de se demander quel niveau de connaissance technique est nécessaire à acquérir pour remplir correctement leur mission. Comme ce dernier exemple le montre, les organigrammes sont souvent remis en cause par l'émergence de nouvelles activités et l'évolution de la répartition des responsabilités.

En fait, les chaînes de traitement ont tendance à être plus longues, plus complexes, et demandent plus d'intervenants. Comme elles sont plus récentes, ces chaînes sont moins bien connues et moins intelligibles pour les différents acteurs. Par conséquent il existe le risque que les acteurs ne reconnaissent plus leur place dans le flux de travail, et qu'ils ne comprennent pas pourquoi on leur demande d'accomplir telle ou telle tâche. Les programmes de formation semblent donc être importants pour aborder cette question. La gestion de collections numériques nécessite bien sûr une formation technique sur les spécificités des documents numériques. On peut toutefois considérer que la gestion de collections numériques

nécessite moins une connaissance technique qu'une compréhension globale des chaînes de production.

Enfin plusieurs préoccupations ont été soulevées sur la manière d'articuler le travail entre les documents physiques analogiques et les documents numériques. Quel que soit le domaine (la description de métadonnées, les entrées numériques, l'archivage du web), l'activité a été d'abord expérimentale ; avec des pionniers très motivés. Impliquer toutes les équipes dans l'activité a alors été aussi difficile que de la lancer : cependant, le partage de la charge de travail entre les premiers pionniers et le reste du personnel apparaît comme une clé pour éviter de créer un fossé entre les missions de la bibliothèque sur les collections physiques et numériques.

Remerciements

Les auteurs souhaitent remercier les membres d'ORHION et ceux qui ont eu l'amabilité de relire ce document.

Bibliographie

[1] Derrot S., Fauduet L., Oury C., et Peyrard S. 2012. Preservation is Knowledge: A community-driven preservation approach. In *Proceedings of the 9th International Conference on Preservation of Digital Objects* (Toronto, Canada, Octobre 2012).

[<https://ipres.ischool.utoronto.ca/sites/ipres.ischool.utoronto.ca/files/iPres%202012%20Conference%20Proceedings%20Final.pdf>].

[2] Bermès E et, Fauduet L. The Human Face of Digital Preservation: Organizational and Staff Challenges, and Initiatives at the Bibliothèque nationale de France. In *International Journal of Digital Curation*, 2011, vol. 6, n. 1

[<http://www.ijdc.net/index.php/ijdc/article/view/175/244>].

[3] Clatin M. et Fauduet L. Watching the library change, making the library change? An observatory of digital influence on organizations and skills at the Bibliothèque nationale de France. In *Proceedings of the 78th IFLA Conference* (Helsinki, Finlande, août 2012).

[<http://conference.ifla.org/past-wlic/2012/150-clatin-en.pdf>].