

Diplôme de conservateur de bibliothèques

Mémoire d'étude - DCB26 / Mars 2018

Les coopérations entre chercheurs et bibliothécaires dans le cadre des projets de numérisation de corpus documentaires

Graziella PASTORE

Sous la direction de Nicolas DI MÉO
Conservateur des bibliothèques – Université de Strasbourg

Remerciements

Je souhaite remercier mon directeur de mémoire, Nicolas di Méo, pour la confiance et la patience qu'il m'a témoignées à chaque étape de mon travail.

Un grand merci à tous les chercheurs et les professionnels des bibliothèques qui ont bien voulu me recevoir et répondre à mon enquête.

Pour m'avoir aidée à nourrir ma réflexion, ma gratitude va aussi à : Marie-Françoise Bois-Delatte (Bibliothèque municipale de Grenoble), Emmanuelle Bermes (Bibliothèque nationale de France), Alix Lamé-Bergis (Bibliothèque nationale de France), Amelia Laurenceau (Bibliothèque Interuniversitaire de la Sorbonne), Yves Le Guillou (Bibliothèque nationale de France), Jean-Philippe Moreux (Bibliothèque nationale de France); Florent Palluault (Médiathèque François-Mitterrand, Poitiers), et Jean-Claude Ragot (Réseau Maisons d'écrivain en Nouvelle-Aquitaine).

Je remercie enfin mes amis et mes collègues, Mathilde, Nicolas et Benjamin, pour les relectures et les encouragements.

Résumé: Depuis quelques années, le milieu de la recherche, des bibliothèques et des institutions culturelles porte une attention de plus en plus marquée à l'étude de corpus documentaires, à leur numérisation et à leur exploitation. Les projets de numérisation de corpus reposent sur l'établissement de partenariats, de collaborations ou de contacts plus ou moins étroits entre institutions patrimoniales et de recherche (laboratoires de recherche, archives, bibliothèques, musées, maisons de sciences de l'homme, etc.) et s'adressent à un public varié (chercheurs, étudiants, « grand public »). S'appuyant sur un état des lieux préliminaire et sur les résultats d'une double enquête menée auprès des bibliothécaires et des chercheurs, ce travail souhaite examiner cette situation composite, présenter comment ces projets se construisent et évoluent, les possibles coopérations entre chercheurs et professionnels des bibliothèques, ainsi que le point de vue des chercheurs et des bibliothécaires, afin de repérer les bonnes pratiques à encourager ou à développer, mais aussi les pièges à éviter.

Descripteurs: Corpus (recherche), Exploration de données, Humanités digitales, Numérisation, Bibliothèques numériques

Abstract: Since a few years now, researchers, libraries and cultural institutions are starting to establish partnerships in order to investigate digital corpus and collections. Based on a preliminary state-of-the-art and on the results of a double survey of librarians and researchers, this work aims to examine this collaborations, how they are constructed and evolve, as well as the possible collaborations between researchers and library professionals. It presents the point of view of researchers and librarians in order to identify some good practices to encourage and pitfalls to avoid.

Keywords: Text and Data Mining, Digitization, Digital preservation, Digital Humanities, Digital Libraries

Ce document a été rédigé avec LaTeX, à l'aide du logiciel TeXShop. BibDesk a servi pour la gestion des références bibliographiques.

Toutes les adresses des sites Internet ont été vérifiées le 1^{er} mars 2018.

Droits d'auteurs Droits d'auteur réservés.

Toute reproduction sans accord exprès de l'auteur à des fins autres que strictement personnelles est prohibée.
--

Sommaire

SIGLES ET ACRONYMES	9
INTRODUCTION	11
Présentation du sujet	11
Périmètre et méthodologie de l'enquête	13
Intérêt et enjeux de l'étude de corpus	17
QU'EST CE QU'UN CORPUS ?	19
Du latin au français	19
Le point de vue des chercheurs	19
Le point de vue des professionnels des bibliothèques	23
Quelles typologies de corpus ?	24
CONSTITUTION ET ACCÈS AUX CORPUS : LES ENJEUX DE LA NUMÉRISATION ET DU TRAITEMENT DES DONNÉES	33
Les enjeux didactiques et scientifiques	34
Du travail en salle de lecture à l'approche distante	34
Les Humanités numériques et la création de corpus intelligents	36
Les bibliothèques numériques vers de nouveaux usages ?	42
La gestion et la valorisation des fonds patrimoniaux	42
Une offre numérique en cours d'évolution	44
Les services d'accompagnement aux chercheurs et les infrastructures de recherche	46
Des services d'accompagnement <i>ad hoc</i> ?	46
Le développement des infrastructures de recherche	47
Autour de la numérisation de corpus : la feuille de route de la coopération	51
DE L'INITIATIVE À LA MISE EN PLACE DE LA COOPÉRATION	53
D'où vient la demande de numérisation ? À quel moment ?	53
Quel mode projet ?	56
La question des droits	57
L'établissement de la convention	59
DU DÉROULEMENT À LA VALORISATION DU PROJET	61
Les choix documentaires et scientifiques	61
La sélection des documents, le catalogage et l'enrichissement des notices	63
Quel set de métadonnées ? Quel modèle conceptuel ?	64
Les choix techniques de numérisation et d'archivage	67
Qui numérise et comment ?	68
Le stockage et la pérennisation : méthodes et outils	70
La diffusion et la mise en ligne de corpus	70
Les dispositifs de valorisation et de soutien	72

La numérisation et l'étude de corpus : un terrain de rencontre stimulant	75
LES JALONS D'UNE COLLABORATION INTER-MÉTIERS	77
Les bonnes pratiques nationales et internationales	77
Quel équilibre entre médiation aux données et formation numérique des chercheurs ?	79
Expertise scientifique et technique : une question de légitimité ou de complémentarité ?	81
Les institutions patrimoniales sont-elles un guichet de données ?	83
Chercheurs et professionnels : une temporalité différente ?	84
BIBLIOTHÈQUES TERRITORIALES ET BIBLIOTHÈQUES DE L'ESRI : DES SERVICES, DES RESSOURCES ET DES PRATIQUES DIFFÉRENTES ?	87
Rejoindre les chercheurs par des services adaptés	87
Les dispositifs de financement	89
LES RÉPONSES DE LA BNF	93
Gallica Marque Blanche	93
Le projet CORPUS	95
CONCLUSION	97
GLOSSAIRE	111
ANNEXES	113
TABLE DES ILLUSTRATIONS	157
TABLE DES MATIÈRES	159

SIGLES ET ACRONYMES

- ABES : Agence bibliographique de l'enseignement supérieur
- ABF : Association des bibliothécaires de France
- ADBDP : Association des directeurs des bibliothèques départementales de prêt
- ADBGV : Association des directeurs des bibliothèques municipales et intercommunales des grandes villes
- ADBS : Association des professionnels de l'information et de la documentation
- ADBU : Associations des directeurs et des personnels de direction des bibliothèques universitaires et de la documentation
- ADDNB : Association pour la diffusion des documents numériques en bibliothèque
- AFNOR : Agence française de normalisation
- ANR : Agence nationale de recherche
- BDP : Bibliothèque départementale de prêt
- BM : Bibliothèque municipale
- BNU : Bibliothèque nationale et universitaire de Strasbourg
- BPI : Bibliothèque publique d'information
- BU : Bibliothèque universitaire
- CADIST : Centre d'acquisition et de diffusion de l'information scientifique et technique
- CG3P : Code général de la propriété des personnes publiques
- CINES : Centre informatique national de l'enseignement supérieur
- CNL : Centre national du livre
- CNRS : Centre national de la recherche scientifique
- CPER : Contrat de projet État-Région
- DGD : Dotation générale de décentralisation
- DH : *Digital Humanities*
- DR : Directeur de recherche
- DRAC : Direction régionale des affaires culturelles
- EAD : *Encoded Archival Description*
- ECIA : *European Council of Information Associations*
- FR(R)AB : Fonds régional d'acquisition (et de restauration) des bibliothèques
- FRBR : *Functional Requirements of Bibliographic Records*
- HN : Humanités numériques
- IFLA : *International Federation of Library and Institutions*
- IGE : Ingénieur d'études
- IGR : Ingénieur de recherche
- INHA : Institut national de l'histoire de l'art
- INRIA : Institut national de recherche en informatique et en automatique
- IRHT : Institut de recherche et d'histoire des textes
- IUT : Institut universitaire de technologie
- JPEG : *Joint Photography Experts Group*
- LIBER : Ligue des bibliothèques européennes de recherche
- LOD : *Linked Open Data*
- MISTRID : Mission de l'information scientifique et du réseau documentaire
- MCF : Maître de conférence

- MSH : Maison des sciences de l’homme
- OPER : Observatoire du patrimoine écrit en région
- PAPE : Plan d’action pour le patrimoine écrit et graphique
- PMR : Patrimoine musical en région
- PU : Professeur des universités
- RNBFD : Répertoire national des bibliothèques et des fonds documentaires
- SCD : Service commun de la documentation
- SHS : Sciences humaines et sociales
- SLL : Service du livre et de la lecture
- SUDOC : Système universitaire de la documentation
- TEI : *Text Encoding Initiative*
- UFR : Unité de formation et de recherche
- URI : *Uniform Resource Identifier*
- URL : *Uniform Resource Locator*
- XML : *EXtensible Markup Language*
- TIFF : *Tagged Image File Format*

INTRODUCTION

Présentation du sujet

Depuis quelques années, les colloques et les journées d'étude en sciences humaines et sociales, ainsi que les rencontres professionnelles dans le milieu des bibliothèques et des institutions culturelles, semblent porter une attention de plus en plus marquée à l'étude de corpus documentaires, à leur numérisation et à leur exploitation. Même si la notion de corpus est parfois équivoque, ne recouvrant pas la même signification et n'évoquant pas les mêmes attentes dans toutes les communautés scientifiques et professionnelles, elle ne manque pas de capter l'intérêt, tant des chercheurs que des professionnels. On peut citer plusieurs exemples, comme le séminaire « Pratiques de corpus numériques : enjeux et méthodes disciplinaires » et le colloque « Corpus de textes : composer, mesurer, interpréter » tenus respectivement le 29 mars 2012 et les 17 et 18 juin 2013 à l'ENS de Lyon; la journée d'études du 14 novembre 2014 à l'Université de Savoie Mont-Blanc consacrée à plusieurs corpus numérisés et illustrant, sur une durée de presque dix ans, le cycle de vie de projets ayant « des héritages et des ambitions différents » [Pety, 2015, Introduction, p. 9]; la journée d'étude « Que faire des corpus (une fois) numérisés ? L'exemple du discours littéraire » organisée par l'Université de Caen en juin 2013¹; le colloque « Les corpus de textes numériques comme objets de recherche » du 19 avril 2016 à l'IHRIM²; la journée d'étude organisée à l'IRHT en novembre 2016 « Qu'est-ce qu'un corpus ? », dont l'objectif était de s'interroger sur la définition d'un corpus pour les historiens médiévistes³; l'atelier du consortium CAHIER « Les dispositifs numériques : des corpus aux usages » à la Maison de Sciences de l'Homme d'Aquitaine en juin 2017⁴; l'intervention d'Emmanuelle Bermès « Fournir des corpus à destination de la recherche à des fins de TDM : le projet CORPUS, BnF » lors de la journée d'étude « Les bibliothécaires, acteurs de la recherche » durant le 47^e congrès de l'ADBU⁵; la journée d'étude « Corpus numériques et bases de données (littérature et linguistique) » du 20 janvier 2018 à l'Université Sorbonne-Nouvelle Paris III⁶; ainsi que plusieurs initiatives à venir⁷.

S'agit-il d'un simple effet de mode ? L'attention portée à l'étude de corpus n'est pas récente : ces initiatives s'inscrivent à la suite d'une tradition d'études qui, dans certains domaines, est consolidée de longue date; elles témoignent également d'une

1. Voir : http://www.unicaen.fr/recherche/mrsh/document_numerique/959.

2. Voir : <http://ihrim.ens-lyon.fr/manifestations/article/les-corpus-de-textes-numeriques-comme-objets-de-recherche>.

3. Pour le programme, voir : <http://cosme.hypotheses.org/291>; pour le compte rendu, voir [Magnani, 2017].

4. Voir : <http://obvil.paris-sorbonne.fr/actualite/les-dispositifs-numeriques-des-corpus-aux-usages/lun-26062017-0000>.

5. Pour le retour sur la journée et pour les vidéos de présentation, voir : <http://adbu.fr/retour-sur-la-journee-detude-du-congres-adbu2017-les-bibliothecaires-acteurs-de-la-recherche/>.

6. Voir : <http://www.univ-paris3.fr/journee-d-etudes-corpus-numeriques-et-bases-de-donnees-litterature-et-linguistique--464834.kjsp>

7. « Des corpus pour l'histoire à l'âge du numérique. Éditions électroniques d'actes royaux et princiers », École nationale des chartes (Paris, 16 mars 2018) : <http://www.chartes.psl.eu/fr/actualite/corpus-histoire-age-du-numerique-editions-electroniques-actes-royaux-princiers-moyen-age>; colloque « Corpus et textes de représentation », LabEx OBVIL (Avignon, 14-15 juin 2018).

« vulgarisation » et d'une diffusion de l'emploi du terme « corpus » dans les sciences humaines et sociales (SHS) datant au moins du début des années 2000, comme l'attestent, entre autres, le colloque d'Albi « Corpus en Lettres et Sciences sociales » [Duteil-Mougel et Foulquié, 2006] et les appels à projet ANR « Corpus et outils de la recherche en SHS » depuis 2006.

Quel que soit le sujet choisi, la numérisation de corpus documentaires semble se situer à la croisée de plusieurs phénomènes qui, surtout depuis les années 2010, intéressent le monde de la recherche et de l'information, et qui sont liés à l'idée que le numérique n'est pas simplement un outil mais aussi un sujet produisant de la culture, ainsi qu'une nouvelle façon de voir le monde [Doueihy, 2011]. Nous pensons plus particulièrement aux évolutions des pratiques informatiques individuelles et de l'offre numérique des bibliothèques, au *big data* et à l'émergence des données massives [Calderan *et al.*, 2014, p. 9-30], à l'ouverture des données, ainsi qu'au développement progressif des Humanités numériques et à la création de profils de *digital librarians* et de *data scientists*. C'est peut-être cette position médiane, à l'intersection de plusieurs champs disciplinaires et de plusieurs métiers, qui contribue à la fortune de ce sujet.

L'intérêt pour la numérisation de corpus et pour leur valorisation s'inscrit ainsi dans une réflexion plus générale [Chevry Pébayle, 2012] qui interpelle chercheurs, enseignants-chercheurs, bibliothécaires et ingénieurs sur les usages des technologies numériques dans la création d'un espace culturel partagé [Juanals et Minel, 2017]. Une réflexion qui engage surtout les établissements patrimoniaux pour comprendre et anticiper les évolutions des logiques d'usage des chercheurs et le rôle des institutions au sein de la nouvelle « République des Lettres promise par le numérique » [Roustan *et al.*, 2016, Introduction de Sylvaine Tarsot-Gillery, p. 9].

Tout en s'appuyant sur ces thèmes d'actualité, notre étude les interroge en partant d'un cas de figure précis et selon une approche aussi pragmatique que possible. De ce fait, en dehors d'une entrée en matière apportant des précisions de périmètre concernant la définition de corpus, nous n'insisterons pas sur la constitution des bibliothèques et des collections numériques⁸, ni sur les enjeux des bibliothèques à l'heure du *big data*⁹, ni sur le mouvement de libre accès aux données scientifiques¹⁰ ou de dépôt de la production scientifique des chercheurs dans les archives ouvertes, non plus que sur l'ensemble des rapports entre chercheurs et bibliothécaires¹¹.

Les projets de numérisation de corpus documentaires reposent sur l'établissement de partenariats, de collaborations ou de contacts plus ou moins étroits entre plusieurs institutions patrimoniales et de recherche (laboratoires de recherche, archives, bibliothèques, musées, Maisons de Sciences de l'Homme, Maisons d'écrivains, etc.) et s'adressent à un public varié (chercheurs, étudiants, « grand public »); en partant de cette situation composite, cette étude s'attache à présenter la façon dont ces projets se construisent et évoluent, les possibles coopérations entre chercheurs et professionnels des bibliothèques, ainsi que les réflexions et les actions que ceux-ci peuvent mener ensemble.

Ce travail soulève ainsi plusieurs questions. D'où vient l'initiative de la numérisation de corpus ? Les différents interlocuteurs ont-ils un véritable dialogue, et se comprennent-ils entre eux ? Comment ce dialogue peut-il être constructif au vu de l'établissement d'une

8. Sur ce sujet, voir par exemple les mémoires DCB de [Pérésan-Roudil, 2014] et de [Le Coz, 2017].

9. Sur ce sujet, voir le mémoire DCB de [Gillium, 2016].

10. Sur ce sujet, voir le mémoire DCB de [Gaillard, 2014].

11. Sur les échanges entre chercheurs et bibliothécaires, voir par exemple le mémoire DCB [Arènes, 2015].

politique numérique, ou du projet scientifique et culturel de l'établissement (PSC) d'un côté, et des projets des chercheurs de l'autre? Quelles conséquences les relations entre chercheurs et professionnels des bibliothèques ont-elles sur les services et les infrastructures qui vont être mises en place autour de la numérisation des corpus? Le rôle des institutions patrimoniales doit-il s'arrêter au signalement et à la mise à disposition des documents, ou à leur numérisation? Doivent-elles aller plus loin? Jusqu'à quel point? Nous n'aurons malheureusement pas la possibilité de répondre à toutes ces questions, mais nous fournirons quelques pistes destinées à en comprendre les enjeux.

S'appuyant sur un état des lieux préliminaire et sur les résultats d'une double enquête menée auprès des bibliothécaires et des chercheurs, notre travail souhaite mettre en avant les principaux enjeux de ce type de projet, ainsi que le point de vue des chercheurs et des bibliothécaires, afin de repérer les bonnes pratiques à encourager ou à développer, mais aussi les pièges à éviter.

En partant de la définition de l'objet d'étude, et notamment de la manière dont les chercheurs et les professionnels des bibliothèques interprètent l'idée de corpus, la première partie de ce travail présente les grands enjeux de la numérisation, tant du point de vue des apports qu'elle peut fournir à la recherche scientifique et à la didactique, que de l'évolution de l'offre numérique proposée par les bibliothèques; cette partie propose également un point sur les services et les infrastructures permettant d'établir des échanges et des collaborations entre les différents acteurs. La deuxième partie s'intéresse plus précisément à la nature des collaborations entre chercheurs et professionnels; elle est organisée autour de la feuille de route de la numérisation d'un corpus, de l'initiative du projet à sa diffusion et à sa valorisation. La conception du projet, les choix documentaires, les différentes étapes de l'expertise technique et scientifique sont évoqués, afin de réfléchir aux difficultés susceptibles de se manifester, ainsi qu'aux avantages liés à l'apport conjoint de chercheurs et de professionnels dans le cadre du bon déroulement du projet. La troisième partie fait le point sur la collaboration entre chercheurs et professionnels en abordant les éléments qui ont émergé dans le cadre de notre enquête, notamment en termes de légitimité et de positionnement professionnel des acteurs des projets. Elle examine enfin les réponses, parfois différentes, qui peuvent être apportées par les bibliothèques territoriales, les bibliothèques de l'enseignement supérieur et la Bibliothèque nationale de France. Les annexes rassemblent d'une part les questionnaires soumis aux chercheurs et aux bibliothécaires, d'autres part les résultats de l'enquête.

Périmètre et méthodologie de l'enquête

Ce mémoire se nourrit des résultats de deux questionnaires administrés en ligne (voir Annexes, p. 114), un premier questionnaire destiné aux chercheurs, un autre adressé aux bibliothécaires, complétés par des entretiens individuels semi-directifs et par un certain nombre d'échanges informels dans le cadre de séminaires et de journées d'études - notamment la journée d'étude de l'ADEMEC (Association des diplômés et des étudiants de master de l'École des chartes) du 14 octobre 2017, et l'un des ateliers du projet CORPUS de la BnF suivi le 16 octobre 2017. Les données et les extraits issus de l'ensemble de cette enquête figurent tout au long de ce travail et notamment dans la deuxième et la troisième partie.

Les deux questionnaires ont été réalisés avec l'outil Google Forms et diffusés en ligne entre fin septembre et mi-décembre 2017. Prévoyant à la fois des questions ouvertes et des

questions fermées, les deux sondages portaient sur les projets de numérisation de corpus menés à partir de 2012, à travers les différentes étapes d'une hypothétique feuille de route, de l'initiative du projet à sa diffusion et à sa valorisation. Ils étaient conçus de manière quasi symétrique, afin de pouvoir comparer les réponses fournies par les deux groupes de personnes interrogées. Les questionnaires ont d'abord été envoyés individuellement à des listes d'adresses, puis, dans un second temps, diffusés *via* les réseaux sociaux (à partir de nos comptes Twitter et LinkedIn) et les listes de diffusion des communautés concernées : la liste de diffusion DH pour le questionnaire aux chercheurs ; les listes Bibliopat et ADBU pour le questionnaire aux bibliothécaires.

Le lien vers le questionnaire adressé aux chercheurs a été directement envoyé à une liste d'environ 400 adresses, repérées à la suite d'un état des lieux des projets en cours que nous avons mené au préalable ; nous avons interprété la notion de « chercheur » au sens large, en y incluant l'ensemble des acteurs de la recherche : doctorants, post-doctorants, enseignants du secondaire, ingénieurs d'études/de recherche, chargés/directeurs d'études, maîtres de conférence, professeurs des universités, etc. Pour restreindre l'enquête, nous avons privilégié les chercheurs impliqués dans des programmes dont l'initiative émanait d'une institution française ou située dans un pays francophone, ce qui n'exclut évidemment pas la participation de chercheurs étrangers. Étant donné que notre première liste comprenait inévitablement une large majorité de chercheurs en sciences humaines et sociales, nous avons essayé d'élargir notre réseau afin d'y inclure des chercheurs en sciences du vivant et en sciences « dures », dans le but d'atteindre un échantillon suffisamment significatif par rapport à la population-mère, qui s'est finalement stabilisé à environ 20 % des destinataires de l'enquête. Nous nous sommes réjouis de recevoir également quelques réponses de la part des professionnels des bibliothèques ayant des missions de recherche. Pour plus de commodité dans la gestion des réponses, nous avons distingué les domaines de recherche en fonction de la grille de l'Unesco¹², bien que celle-ci soit assez peu adaptée aux sciences humaines et sociales ; néanmoins, nous avons également laissé aux chercheurs la possibilité de préciser leur domaine d'intérêt (voir les questions n^{os} 34 et 35). Nous avons obtenu 57 réponses au questionnaire en ligne et nous avons engagé une dizaine d'échanges de courriels avec d'autres chercheurs afin d'obtenir des précisions supplémentaires. Les réponses ont été enregistrées de manière anonyme : ainsi, dans les extraits qui figurent en annexe, nous avons veillé à supprimer toute donnée personnelle.

D'autre part, le lien vers le questionnaire adressé aux bibliothécaires, a été directement envoyé à une liste aussi exhaustive que possible comprenant, en France, les directeurs, les directeurs adjoints, les responsables des services numériques et les responsables des fonds patrimoniaux des bibliothèques des grands établissements, des bibliothèques de grandes écoles ou d'instituts, des bibliothèques universitaires et inter-établissements, et des bibliothèques municipales classées. Dans ce cas aussi, les réponses ont été enregistrées de façon anonyme ; néanmoins, dans la rédaction de ce mémoire, nous les rapportons à l'établissement de rattachement du répondant, du moins lorsque cela s'avère nécessaire pour comprendre la nature du projet ou de la collaboration dont il est question. Nous avons reçu un total de 87 réponses (correspondant à 82 établissements) et engagé quelques échanges en dehors du questionnaire.

Pour compléter et mieux interpréter les résultats des questionnaires, nous avons également mené cinq entretiens semi-directifs d'une durée d'environ une heure : trois avec des conservateurs (travaillant à la Bibliothèque nationale de France, dans une bibliothèque

12. Voir : <http://skos.um.es/unesco6/00/html>.

municipale classée, et dans une bibliothèque universitaire) et deux avec des chercheurs (un avec un chercheur en sciences humaines et sociales ; un avec un chercheur en sciences « dures »).

Si ce taux de réponse positif témoigne de l'intérêt et de l'actualité de la numérisation de corpus, nous n'avons pas l'ambition d'en faire ressortir des données scientifiques ; l'échantillon d'enquête reste modeste, même si le public des chercheurs, souvent difficile à atteindre, s'est montré ici relativement disponible. Nous avons également remarqué que certaines questions ont posé des difficultés aux chercheurs ou aux bibliothécaires, ce qui a compliqué ou empêché leur exploitation, du moins d'un point de vue strictement quantitatif ; ainsi, par exemple, nous avons regretté, dans le cas du questionnaire adressé aux bibliothécaires, de ne pas avoir prévu une option de réponse explicite pour les bibliothèques qui ne sont pas impliquées dans des projets de numérisation de corpus : si certaines nous l'ont signalé par courriel, pour d'autres, nous ne savons pas si l'absence de réponse correspond à l'absence de projets ou, plus simplement, à un oubli. Du point de vue du traitement et de l'interprétation des données brutes, nous avons, pour une partie des réponses, proposé des options, mais aussi laissé ouverte la possibilité de saisie : le nettoyage des résultats nous a donc demandé un certain effort critique ; pour faciliter cette phase, qui est « la plus compliquée et souvent la plus longue de l'enquête » [Gaudet, 2011, p. 27], nous avons eu recours à l'outil OpenRefine¹³, un logiciel qui permet de regrouper et d'enrichir des données tout en gardant une trace de chaque opération.

Mises à part ces quelques limites, cette enquête conjointe sur les projets de numérisation de corpus nous a permis de réunir de nombreux et riches témoignages dans des domaines professionnels et de recherche variés. Elle nous a permis de cibler au mieux les principaux enjeux de ce type de projet, ainsi que les difficultés éventuelles et les attentes des professionnels des bibliothèques et des chercheurs.

13. Voir : <http://openrefine.org/>.

Intérêt et enjeux de l'étude de corpus

QU'EST CE QU'UN CORPUS ?

Du latin au français

Attesté en ancien français seulement pour indiquer le « corps du Christ », l'« hostie », l'emprunt français du substantif latin *corpus* (*ōris*, neutre ; d'où le nominatif et accusatif pluriels *corpora*)¹⁴ ne date que du début du XIX^e siècle. Délaissant désormais la dimension physique latine associée au « corps », l'emploi français est forgé sur le sens figuratif, déjà présent en latin, qui renvoie à « ensemble », « tout », également employé pour indiquer le « corps d'ouvrage » ; c'est le cas, par exemple, du *corpus iuris civilis*, la compilation de droit romain réalisée sous le règne de Justinien (527-565). C'est précisément dans cette dernière acception, au sens de « nom d'une collection de droit romain » que le mot est attesté en 1826 dans la traduction française de l'*Introduction à l'étude du droit romain* de Ferdinand Mackeldey et en 1845 dans le *Dictionnaire national ou Dictionnaire universel de la langue française* de Louis-Nicolas Bescherelle. À vrai dire, *corpus* avait déjà fait une timide apparition en 1809, en concurrence avec *corps*, pour indiquer une « collection de textes réunis en vue de leur étude scientifique »¹⁵. En 1855, on retrouve le mot *corpus* pour désigner un « ensemble de textes établi selon un critère thématique en vue de leur étude linguistique » dans le *Bulletin du bibliophile belge*, édité à Bruxelles par F. Heussner ; en 1968, dans *Supplément 1* du *Grand Larousse encyclopédique*, un *corpus* est un « ensemble de documents servant de base à la description ou à l'étude d'un phénomène ». Désormais francisé, le mot *corpus* est le plus souvent employé de façon invariable tant au singulier qu'au pluriel. De nos jours, notamment dans le domaine de la philologie et des sciences humaines, le mot sert pour indiquer un « recueil réunissant ou se proposant de réunir, en vue de leur étude scientifique, la totalité des documents disponibles d'un genre donné, par exemple épigraphiques, littéraires, etc. »¹⁶.

Néanmoins, si son étymologie est claire, le sens donné au mot *corpus* ne fait pas l'unanimité dans l'ensemble des communautés scientifiques et professionnelles et sa définition peut fortement varier d'une discipline à l'autre. La volumétrie du *corpus*, la nature des documents qui le composent, les formats source, leurs traitements et leurs finalités font partie des éléments qui participent de sa définition.

Le point de vue des chercheurs

Il n'est pas étonnant qu'une partie des chercheurs que nous avons interviewés à propos de la numérisation de *corpus* ait eu quelques incertitudes à identifier le sujet de notre enquête. Si pour certains le mot *corpus* évoque immédiatement des scénarios de recherche, pour d'autres, ce mot ne fait pas partie du lexique de travail quotidien. C'est pourquoi

14. Voir le lemme correspondant dans le dictionnaire latin-français Gaffiot, <http://www.lexilogos.com/latin/gaffiot.php?q=corpus>.

15. Cette attestation est tirée de l'ouvrage de Charles Villers, *Coup-d'œil sur l'état actuel de la littérature ancienne et de l'histoire en Allemagne. Rapport fait à la troisième Classe de l'Institut de France par Charles Villers, Correspondant de cette Classe, Membre de la Société royale des Sciences de Göttingue*, Amsterdam/Paris, Bureau des Arts et de la Littérature/Treuttel & Würtz, 1809, disponible sur Google-Books : <https://lc.cx/gENo>.

16. Pour l'ensemble des données concernant l'histoire du mot en français, voir la fiche disponible dans le *Trésor de la langue française*, en ligne : <http://www.cnrtl.fr/etymologie/corpus#>.

certaines chercheurs en répondant à nos questions ont pensé plutôt aux services de numérisation à la demande pour convenances personnelles (par exemple pour travailler sur un document difficilement accessible), ou à la constitution de répertoires ou bases de données numériques, alors que d'autres ont pensé à des collections d'articles ou de textes au format électronique, et d'autres encore, plus simplement, à des groupes de documents. Quant aux chercheurs qui ont couramment affaire à la notion de corpus, bien qu'il soit difficile de les rassembler dans des compartiments étanches, nous pouvons tout au moins observer quelques tendances et approches différentes.

Dans le domaine des sciences « dures », le mot corpus semble rarement employé. Lorsqu'il est utilisé, il semble tout d'abord évoquer des répertoires numérisés et surtout des dépôts de publications scientifiques, avec les politiques favorisant le libre accès aux résultats de la recherche¹⁷ ; c'est le cas par exemple des amples gisements documentaires mis à la disposition de l'ESRI dans le cadre de la politique nationale d'acquisition confiée à ISTEEX [Melzac, 2017] [Mazens, 2017]¹⁸. Un corpus indique alors le plus souvent un ensemble d'articles et de ressources actualisés.

Dans le domaine des sciences humaines et sociales (SHS) l'idée de corpus est plus largement répandue et déclinée sous plusieurs angles¹⁹.

« Le corpus de textes constitue l'un des lieux les plus favorables à l'observation des réalisations de la langue pour la linguistique ; en témoigne une de ses principales, sinon sa principale branche méthodologique, la linguistique justement dite "de corpus" » [Comby *et al.*, 2016, p. 7]

Originaire des pays anglophones, la linguistique de corpus a connu son essor depuis les années 1980 : en s'appuyant sur les progrès de l'informatique, et notamment sur les possibilités d'analyse des données à grande échelle, elle cherche à appréhender la langue de façon plus objective et plus vaste. Chez les chercheurs en linguistique, les corpus (parfois sous le pluriel latin *corpora*), peuvent être considérés comme des collections de textes de taille importante²⁰, ou suffisamment importante pour permettre des analyses, rassemblés selon des critères spécifiques de représentativité et dans un but de cohérence, aujourd'hui collectés sous format électronique²¹. Néanmoins, « tout ensemble de texte ne constitue pas un corpus » et une « archive ne correspond ni aux qualités ni aux objectifs intrinsèques aux corpus » : l'agrégation des textes doit vérifier des conditions de signifiante, d'acceptabilité et d'exploitabilité²². Deux approches linguistiques

17. Parmi les projets les plus récents qui nous ont été rapportés, nous signalons le projet allemand «Projekt DEAL» (<https://www.projekt-deal.de>), une coalition d'environ 200 universités, hautes écoles spécialisées, instituts de recherche et bibliothèques d'État cherchant à négocier des accords de licence à l'échelle nationale avec trois grands éditeurs universitaires pour l'ensemble de leur portefeuille de revues électroniques.

18. Le projet ISTEEX a pour ambition d'offrir un accès aux collections rétrospectives de la littérature scientifique numérique aux membres de la communauté de l'ESRI; voir le portail : <https://dl.istex.fr/>; voir aussi <http://www.abes.fr/Ressources-electroniques2/Acquisitions/Licences-nationales-ISTEX>.

19. Rappelons que les chercheurs en SHS constituent le public cible de notre enquête.

20. Comme, par exemple, le *British National Corpus* (BYU-BNC) initialement constitué par la Oxford University Press : <https://corpus.byu.edu/bnc/>

21. Voir aussi l'introduction disponible sur le site de l'UFR EILA de l'Université Paris Diderot Paris 7 : <http://www.eila.univ-paris-diderot.fr/enseignement/lea/cours/linguistique/appli-corpus/slides-intro>.

22. Voir [Comby *et al.*, 2016, Introduction, p. 7-17] et notamment les passages tirés de la thèse de doctorat de Bénédicte Pincemin, *Diffusion ciblée automatique d'informations : conception et mise en œuvre*

différentes s'appuient sur deux conceptions différentes des corpus : l'approche *corpus-based*, qui considère le corpus comme « un recueil d'exemples ou d'attestations utilisé de façon déductive pour vérifier des hypothèses ou étayer des théories linguistiques » [Comby *et al.*, 2016, p. 7]; et l'approche *corpus-driven*, qui perçoit le corpus comme « un outil d'exploration à part entière, susceptible de faire émerger de manière inductive des savoirs linguistiques » [Comby *et al.*, 2016, p. 7].

La constitution d'un corpus fait toujours l'objet d'une discussion dont les critères de sélection des données varient en fonction de l'objet de la recherche [Pinon, 2016, p. 25]; parmi les critères importants pourront alors figurer : le type de support et le format des documents (numérique, discours à transcrire, etc.), la visée synchronique ou diachronique (données récentes, données relevant d'une période bien définie, etc.), la variété diatopique ou diastratique de la langue (standard, dialectale, littéraire, etc.), le genre de document (presse, littérature, textes techniques, oralité, etc.), la taille du corpus (représentativité de l'échantillon choisi), ainsi que d'autres informations d'ordre démographique ou socio-culturel liées aux auteurs des documents (formation, profession, âge, etc.). Les corpus peuvent ainsi être appréhendés pour différentes études linguistiques, par exemple pour dégager des régularités ou irrégularités dans l'emploi d'un verbe à une époque et dans une langue donnée (on procédera alors à l'étiquetage morpho-syntaxique, à l'identification des concordances, à l'étude des occurrences produites par des locuteurs, etc.), pour étudier des états différents de la langue (en synchronie ou en diachronie), avec ou sans locuteurs²³; on pourra appliquer des programmes et techniques informatiques pour le traitement automatique des langues (TAL), par exemple pour étudier sémantiquement le contenu d'un texte ou observer les représentations sociales et politiques dans un discours de presse; ou encore, s'intéresser à la « parole » orale (prosodie, phonologie), ou au fonctionnement du dialogue et aux règles d'interactions homme-machine²⁴.

Ces intérêts de recherche ont forcément conduit les linguistes à suivre de près les évolutions des disciplines informatiques et à développer une plus grande familiarité avec les outils et les méthodes liés au numérique, plus poussée que celle d'autres chercheurs en sciences humaines et sociales; enrichissant le champ interdisciplinaire de la linguistique informatique, ils ont progressivement développé des méthodes propres, pouvant également profiter à d'autres chercheurs, par exemple dans les disciplines liées à l'histoire ou à la sémantique historique [Guerreau, 2011]. En sociolinguistique et sociologie, la mutation numérique offrant désormais de nouveaux corpus (ou « très grand corpus ») de sources de données au-delà des recensements, des sondages et des questionnaires [Boullier, 2015], nombreux sont les chercheurs qui analysent et exploitent des données collectées sur les réseaux numériques, comme des corpus web [Le Béhec et Alloing, 2016] ou des fils Twitter autour de « phénomènes de haute vibration collective », ciblant un ou plusieurs *hashtags* précis; c'est le cas par exemple des tweets collectés par l'équipe DLWeb de l'INA,

d'une linguistique textuelle par la caractérisation des destinataires des documents, Université Paris IV-Sorbonne, 1999.

23. « Travailler sur une langue sans locuteurs rend les textes indispensables : les données attestées constituent en effet le seul objet possible d'investigation, et l'établissement de fréquences permet d'affiner des descriptions sinon approximatives » [Prévoist, 2015, p. 23].

24. Pour les outils et les pratiques de traitement de corpus écrits et oraux, nous renvoyons au site du consortium CORLI, labellisé en janvier 2016, porté par l'Institut de Linguistique Française (CNRS, FR 2393), et né du rapprochement des deux consortiums de linguistique : « Corpus Écrits » et « Corpus Oraux et Multimodaux » : <http://explorationdecorpus.corpusecrits.huma-num.fr/>; pour la constitution de corpus de langue parlé en interaction, voir CORINTE <http://www.icar.cnrs.fr/sites/corinte/>.

exploités par les projets REAT et ASAP dans le cas d'une collecte d'urgence lors des attentats de Paris en 2015²⁵.

Chez les historiens de la langue, les lexicologues et les philologues, le mot corpus peut évoquer à la fois les grands corpus électroniques issus de la numérisation des éditions imprimées (*Base de français médiéval*²⁶, *Corpus de la littérature médiévale* de Garnier²⁷, FRANTEXT²⁸, *Nouveau Corpus d'Amsterdam*²⁹, etc.) ainsi que des ensembles d'éditions nativement numériques³⁰, quoique encore rares [Duval, 2017], qui permettent de dépasser les limites des éditions imprimées, par exemple en superposant plusieurs couches de codage XML/TEI pour pouvoir afficher, selon le cas, différentes versions éditoriales (version facsimilaire, diplomatique, etc.)³¹.

La disponibilité de plus en plus accrue de textes numérisés et la constitution de corpus ont également exercé un fort impact sur les études d'histoire littéraire. Pour Franco Moretti, l'application des méthodes quantitatives à l'analyse littéraire, permettrait de produire une histoire littéraire « plus rationnelle »; à travers « une conception matérialiste de la forme » [Moretti, 2005, p. 127]³², qui peut être explicitée à travers des graphes, des cartes ou des arbres illustrant les taxonomies et les évolutions littéraires, elle peut par exemple permettre de visualiser les formes hégémoniques romanesques à une époque donnée, les emplacements des protagonistes des romans parisiens, l'utilisation du discours indirect libre dans le récit moderne, etc. Pour comprendre ce que la littérature est devenue dans le nouvel espace des laboratoires littéraires, il suffira de regarder les études du groupe de chercheurs du Stanford Literary Lab, dont par exemple le projet de création d'un algorithme informatique capable de reconnaître un genre littéraire [Moretti, 2016, p. 19-54] ou les recherches de stylistique permettant de distinguer les romans jacobins des romans gothiques [Moretti, 2016, p. 55-92]. Il n'est pas simplement question d'aborder un objet d'étude plus vaste : la taille change la nature de l'objet et c'est donc la méthodologie critique qui est mise en discussion :

« quand nous travaillons sur 200 000 romans plutôt que 200, nous ne faisons pas la même chose en mille fois plus grand; nous faisons autre chose. La nouvelle échelle modifie notre rapport à l'objet d'étude; plus encore, *elle modifie l'objet lui-même* » [Moretti, 2016, p. 266]

En effet, si la plupart des chercheurs en littérature s'accordent à dire que les corpus de textes littéraires constituent un terrain de choix pour les études statistiques

25. Voir [DLWeb INA, 2017] et le site internet <http://inatheque.hypotheses.org/60>.

26. Voir : <http://bfm.ens-lyon.fr/>.

27. Voir : <https://classiques-garnier.com/corpus-de-la-litt%C3%A9rature-m%C3%A9di%C3%A9vale.html>.

28. Voir : <http://www.frantext.fr/>.

29. Voir : <http://www.uni-stuttgart.de/lingrom/stein/corpus/#nca>

30. Voir par exemple la collection ELEC de l'École nationale des chartes : <http://elec.enc.sorbonne.fr/> ainsi que la section Corpus : <http://corpus.enc.sorbonne.fr/>; voir aussi les sources en ligne sur le site Presses Universitaires de Caen, comme le *Corpus Descartes*, édition en ligne des œuvres et de la correspondance de Descartes : <http://www.unicaen.fr/puc/sources/prodescartes/>.

31. Pour des exemples d'après l'édition de la *Queste del Saint Graal* et pour des réflexions à propos de la philologie numérique, voir [Guillot *et al.*, 2013]; pour les pratiques philologiques en France, voir [Duval, 2006].

32. Pour la version française, réalisée depuis l'édition anglaise, Verso 2005, voir F. Moretti, *Graphes, cartes et arbres : modèles abstraits pour une autre histoire de la littérature*, traduit de l'anglais par Etienne Dobenesque, préface de Laurent Jeanpierre. Paris : Les Prairies Ordinaires, 2008.

[Évrard, 2002, p. 90], l'approche computationnelle ne manque pas de susciter des critiques [Moretti, 2016, p. 7-17] :

« La plupart des littéraires qui utilisent des techniques quantitatives espèrent trouver grâce à elles, de manière quasi automatique, la solution de problèmes propres à l'histoire littéraire : attribution, authenticité, interpolations ou remaniements, etc. C'est un espoir vain » [Évrard, 2002, p. 94]

Si les véritables profits de l'application d'une telle méthodologie font donc l'objet de discussion, la spécificité de l'étude de corpus en tant que matériau de la recherche et son apport de connaissance par rapport à la lecture directe des textes ne semblent pas poser problème :

« Un texte est "événement de communication" : écrit par quelqu'un, dans des circonstances particulières, en vue de transmettre une signification donnée. Personne n'écrit de corpus ; il ne s'agit pas d'"événements de communication" - ce ne sont même pas des événements du tout, mais des objets artificiels qu'un chercheur crée par la mise entre parenthèses de tous les contextes. Un texte est pensé comme s'adressant à nous ; il nous "parle", il essaie de nous dire quelque chose. Les corpus ne nous parlent pas ; autrement dit, *dans l'acception usuelle du terme, ils n'ont aucun sens* » [Moretti, 2016, p. 268]

Bien qu'il s'agisse d'une sélection partielle, et qui ne prend pas en compte le point de vue des chercheurs en Humanités numériques, dont nous traiterons plus tard, ces quelques exemples nous montrent la variété d'horizons de recherche que le mot corpus peut évoquer chez les chercheurs ; il va de soi que ces différentes approches conditionnent le choix des documents, des outils et des méthodes, ainsi que le type de requêtes que les chercheurs pourront avancer dans le cadre d'un projet de numérisation de corpus.

Le point de vue des professionnels des bibliothèques

Si l'idée de corpus figure dans le lexique des chercheurs en SHS depuis quelques décennies, elle occupe une place bien plus maigre, quoique en progressive expansion, chez les professionnels des bibliothèques, tant en France qu'au niveau international. Si le terme corpus figure dans le vocabulaire ISO 1087-1 (3.6.9) de 2000, on pourra constater qu'il n'apparaît pas dans les principaux dictionnaires du métier³³, à quelques rares exceptions près³⁴. L'utilisation du terme chez les professionnels, souvent concurrencée par celui, bien plus employé, de collection, semble d'introduction plus récente. On notera que le terme est de plus en plus employé dans les bibliothèques numériques pour indiquer une collection numérisée à laquelle peuvent être associés différents types de jeux de données et de métadonnées. Quand les professionnels font référence à un « corpus à numériser », il peut s'agir par exemple d'une sélection de documents opérée sur la base de critères matériels et de conservation, afin de garantir un accès à des documents en état critique ou fragilisés :

33. Nous n'avons pas trouvé la définition de corpus (ou de *corpora*) dans les lexiques et dictionnaires suivants : [Strauch et Rehm, 2007], [Keenan et Johnston, 2000], [Balle, 2006], [Cacaly, 2008], [Feather et Sturges, 2003].

34. « Corpus - "Raccolta di dati linguistici da analizzare (UNI ISO 1087-1, 3.6.9)". "Ingegneria del linguaggio". Raccolta sistematica di testi leggibili da una macchina [...] o di parti di testo preparate, codificate e memorizzate secondo regole definite » [Alberani, 2008, p. 84]. « Insieme di fonti documentarie, relative ad un determinato settore, usate specialmente per studi terminologici e linguistici » [Capitani, 2005, p. 49].

« La sélection des corpus à numériser peut obéir à des critères variables selon les documents. D'une manière générale, ce sont les documents les plus rares, les plus précieux ou les plus fragiles qui sont privilégiés, parfois après avoir fait l'objet d'une restauration » [Daire, 2016]

C'est surtout à partir des années 2010 avec l'expansion des bibliothèques numériques, et plus tard avec les mouvements liés à l'ouverture des données et au traitement de données massives, que le mot corpus se fait de plus en plus courant dans la production scientifique des professionnels de l'information et des bibliothèques ; on parle alors de traitement linguistique de corpus homogènes dans le cadre de la numérisation de l'image fixe [Thompson *et al.*, 2011, p. 186], de collecte de corpus de données nativement numériques³⁵, de « corpus numérisés » (tant pour indiquer des sources primaires ou des sources secondaires) tout particulièrement au vu de l'impulsion du mouvement des Humanités Numériques [Claerr et Westeel, 2013a, p. 42-44], jusqu'à reconnaître le rôle important des bibliothèques dans la constitution de corpus pour le *data mining* et le *text mining* :

« La mission des bibliothécaires sera donc de constituer des corpus correspondant à ces besoins afin de répondre à cette demande en forte progression, et de permettre aux chercheurs, non seulement de lire et d'analyser les documents classiques, mais aussi d'exploiter un ensemble de données et de documents souvent hétérogènes dans leurs formats » [Mesguich, 2017, p. 157]

Quelles typologies de corpus ?

Quels sont les corpus documentaires demandés par les chercheurs et ceux mis à disposition par les institutions ? Est-ce qu'il est possible de dégager des tendances quant aux typologies de sujets ou à la nature des documents qui les composent ?

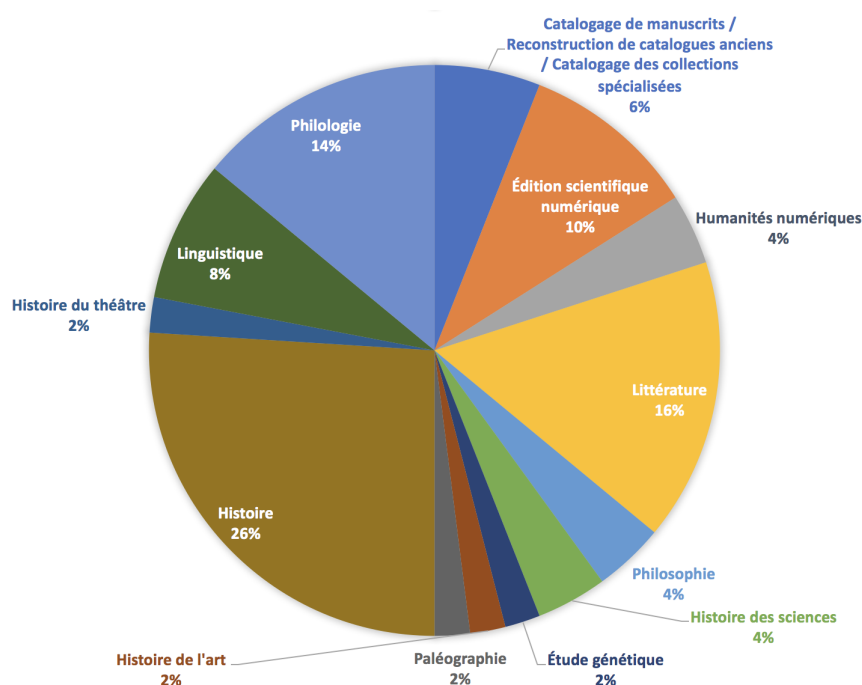


FIGURE 1 – Questionnaire aux chercheurs - question n° 10

35. Voir le projet des Archives du climat porté par les Archives nationale <http://archivesduclimat.meteofrance.fr/>.

Dans le cadre de l'enquête conjointe que nous avons proposée aux chercheurs et aux bibliothécaires, nous avons indiqué quelques macro catégories, selon une sélection certes partielle mais utile pour simplifier la gestion des réponses (voir Annexes, p. 114).

En termes de typologie, les options communes aux deux enquêtes étaient les suivantes : production d'un même auteur ; production de plusieurs auteurs sur un thème cible ; production de plusieurs auteurs sur une période cible ; documents pédagogiques servant à l'histoire d'une discipline ; documents d'archive. La nature des documents était la suivante : cartes et plans ; dessins ; documents d'archive ; estampes incunables ; livres imprimés avant/après 1810 ; manuscrits ; photographies.

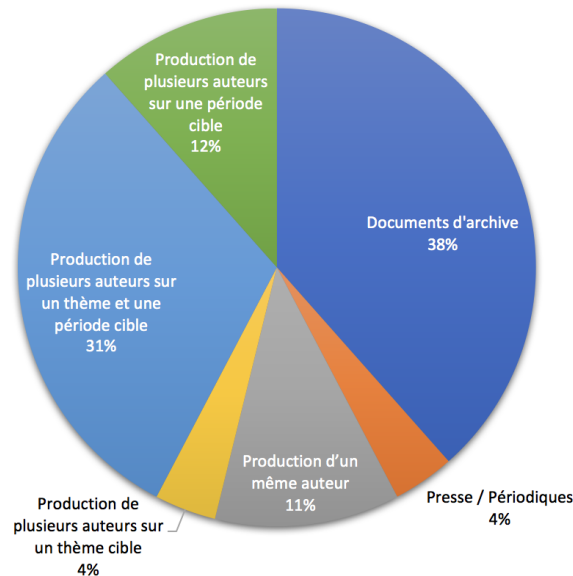


FIGURE 2 – Questionnaire aux chercheurs - question n° 8

Les réponses des chercheurs au sujet de leur dernière demande de numérisation de corpus ont mis en évidence une forte attention pour le document d'archive (38 %), comme dans le cadre du projet concernant le fonds Élie Cartan à l'Académie des Sciences³⁶, ainsi que pour la production de plusieurs auteurs autour d'un thème ou d'une période cible (31 %), comme dans le cadre du projet *Digital Alfieri* soutenu par l'Institut des Textes et Manuscrits Modernes (ITEM)³⁷ ou du projet *France-Angleterre, 700-1200 : manuscrits médiévaux de la Bibliothèque nationale de France et de la British Library, un programme de la Fondation Polonsky*³⁸, ou encore du projet *Idt - Les idées du théâtre* concernant les textes du théâtre de la Renaissance et de l'âge classique³⁹. Signalons également quelques projets portant sur la production d'un seul auteur, comme le projet d'édition des pensées de Montesquieu *Montedite*⁴⁰ et celui des écrits de l'abbé de Saint-Pierre⁴¹, ainsi qu'un projet sur les corpus de presse, *Mauriac en ligne*⁴².

De manière plus générale, quant à la nature des documents, les chercheurs ont principalement mis en avant les demandes suivantes : numérisation de corpus de manuscrits,

36. Voir : <http://eliecartanpapers.ahp-numerique.fr/>

37. Voir : <https://cahier.hypotheses.org/digital-alfieri>.

38. Voir : <https://manuscripta.hypotheses.org/241>.

39. Voir : <http://www.idt.paris-sorbonne.fr/> et [Vuillermoz, 2015].

40. Voir : <https://www.unicaen.fr/services/puc/sources/Montesquieu/>

41. Voir : <http://cahier.hypotheses.org/castel-de-saint-pierre>.

42. Voir : <http://mauriac-en-ligne.u-bordeaux-montaigne.fr/>

comme dans le cadre du projet d'édition numérique de la *Queste del saint Graal*⁴³, du projet *ArchiZ*⁴⁴, du projet *The Values of French Language and Literature in the European Middle Ages*⁴⁵, du projet *Medieval Francophone Literary Culture Outside France*⁴⁶, du projet *À la recherche des manuscrits de Chartres. Étude et renaissance virtuelle d'un fonds de manuscrits sinistré*⁴⁷ ou du projet *La santé en région Centre-Val de Loire au Moyen Âge et à la Renaissance*⁴⁸; numérisation de textes imprimés avant 1810, comme dans le cadre du projet « Philosophie cl@ndestine. Les manuscrits philosophiques clandestins »⁴⁹, ou du projet en cours MHARS *Mémoires et Histoire de l'Académie Royale des Sciences*; et numérisation de documents d'archive. Plusieurs chercheurs évoquent une pluralité de supports manuscrits et imprimés, comme dans le cadre du programme de recherche *Bibliothèques virtuelles humanistes* (CESR, Tours)⁵⁰, du projet *Hyperdonat*⁵¹, du projet en cours *HyperFronde* portant sur la numérisation et l'édition critique de Mazarinades, du projet concernant l'édition critique de la correspondance de Pierre Bayle⁵², ou du projet MAP *Montréal, l'avenir du passé*⁵³.

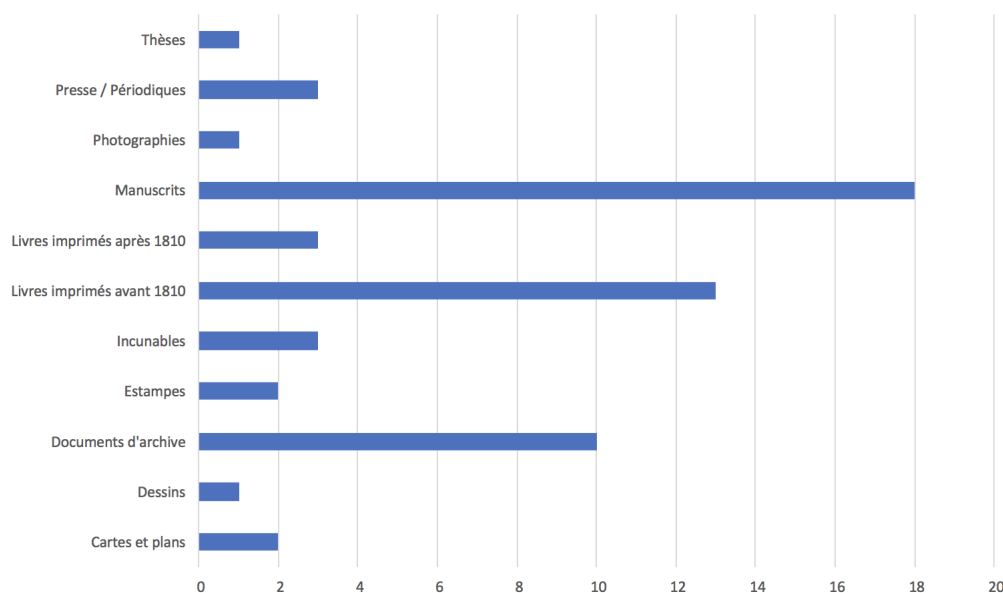


FIGURE 3 – Questionnaire aux chercheurs - question n° 9

Du côté des bibliothèques, les numérisations de corpus effectuées depuis 2012 semblent principalement orientées vers les corpus thématiques. Comme nous le verrons plus tard, une partie de ces projets a simplement pour vocation à amplifier les collections patrimoniales des bibliothèques numériques, alors que d'autres projets proposent des solutions explicitement destinées aux chercheurs. Si certaines institutions sont porteuses de

43. Voir http://catalog.bfm-corpus.org/qgraal_cm.

44. Voir <http://www.archives-zoliennes.fr> et [Macke, 2015].

45. Voir : <http://www.tvof.ac.uk/>.

46. Voir : <http://www.medievalfrancophone.ac.uk/>.

47. Voir : <http://www.manuscrits-de-chartres.fr/>.

48. Voir : <https://intelligencedespatriimoines.fr/sarc-2/>.

49. Voir : <http://philosophie-clandestine.huma-num.fr/>.

50. Voir : <http://www.bvh.univ-tours.fr> et <http://bvh.hypotheses.org>

51. Voir : <http://hyperdonat.huma-num.fr/> et [Ambrosino *et al.*, 2015].

52. Voir : <http://bayle-correspondance.univ-st-etienne.fr/>.

53. Voir : <http://www.mun.ca/mapm/>.

plusieurs projets ⁵⁴, le nombre de retours enregistrés dans le cadre de notre enquête prouve que l'attention portée aux corpus est forte dans l'ensemble des institutions patrimoniales.

Pour ne citer que quelques exemples, nous pouvons distinguer en fonction de la typologie du corpus :

- la production de plusieurs auteurs sur un thème cible (28 %), dont par exemple les corpus thématiques produits par la Bibliothèque du Patrimoine de Clermont Auvergne Métropole ⁵⁵, le corpus sur l'histoire industrielle et l'histoire du travail en Alsace au XIX^e siècle ⁵⁶, les corpus de traités sur les bioagresseurs des cultures agricoles de l'INRA ⁵⁷, le corpus de Langue, littérature et civilisation celtique de Rennes 2 ⁵⁸, ainsi que des fonds thématiques ou historiquement cohérents, comme par exemple le Patrimoine médical de l'Université de Bordeaux ⁵⁹, la correspondance de Flaubert en partenariat entre la Bibliothèque de l'Institut de France et le Centre Flaubert de l'Université de Rouen ⁶⁰, ou encore la correspondance du mathématicien Montessus de Ballore pour alimenter le corpus « Histoire des mathématiques » de la Bibliothèque patrimoniale de l'UPMC ⁶¹ ;
- la production de plusieurs auteurs sur une période cible (11 %), comme par exemple, le corpus iconographique issu du fonds 14/18 de la Bibliothèque du Grand Verdun ⁶², ou le fonds Jules Legras dans le cadre de la commémoration des révolutions russes de 1917 par la Bibliothèque municipale de Dijon ⁶³ ;
- les documents d'archive (20 %), comme par exemple les fonds d'artistes, architecture et design du Centre Pompidou ⁶⁴ ou les documents concernant les élections françaises sous la V^e République ⁶⁵ ; ou encore, la production d'un même auteur (16 %), comme les manuscrits de Jules Verne conservés à Nantes ⁶⁶, le corpus des

54. C'est le cas, par exemple, de : la Bibliothèque nationale Universitaire de Strasbourg (BNUS), voir : <http://www.bnu.fr/numistral/projets-en-cours> ; NuBis, la bibliothèque numérique de la Bibliothèque Interuniversitaire de la Sorbonne (BIS), voir : <https://nubis.univ-paris1.fr/le-corpus> ; Bibliothèque Mazarine, voir : <http://mazarinum.bibliotheque-mazarine.fr/a-propos/> ; Bibliothèque municipale de Lyon, notamment pour la numérisation des journaux des canuts pour alimenter *L'Echo de la fabrique*, voir : <http://echo-fabrique.ens-lyon.fr/document.php?id=1090>, ainsi que les numérisations thématiques dans le cadre de Numelyo, voir : <http://numelyo.bm-lyon.fr/> ; Tolosana, la bibliothèque numérique patrimoniales des universités toulousaines, voir : <http://tolosana.univ-toulouse.fr/fr/corpus> ; Argonaute (BDIC), voir : <https://argonnaute.parisnanterre.fr/> ; Lyon 1 : <https://portaildoc.univ-lyon1.fr/les-collections/fonds-patrimoniaux/patrimoine-720975.kjsp> ; Odysée (Aix-Marseille Université) : <https://odyssee.univ-amu.fr/galleries>.

55. Voir : <https://www.bibliotheques-clermontmetropole.eu/overnia/>.

56. Voir : <http://www.phototheque-scd.uha.fr>.

57. Voir : <https://archive.org/details/inra>.

58. Voir : <http://bibnum.univ-rennes2.fr/collections/show/3>.

59. Voir : <http://www.babordnum.fr/collections/show/1>.

60. Voir : <http://flaubert.univ-rouen.fr/correspondance/>.

61. Voir : <http://jubilotheque.upmc.fr/>.

62. Voir : <http://www.bibliotheques-discotheque-verdun.fr/Fonds-14-18/Fonds-14-18>.

63. Voir : http://patrimoine.bm-dijon.fr/pleade/ead.html?id=FR212316101_legras.

64. Voir : <http://archivesetdocumentation.centrepompidou.fr>.

65. Voir : <https://archive.org/details/archiveselectoralesducevipof>.

66. Voir : <https://bm.nantes.fr/home/espaces-dedies/patrimoine/patrimoine-numerise.html>.

manuscrits et œuvres de Flaubert⁶⁷, ou l'inventaire des dessins de Charles Percier en partenariat entre la Bibliothèque de l'Institut de France, la RMN-GP et l'INHA⁶⁸ ;

- d'autres projets (7 %) portent sur la numérisation de la presse (locale, ancienne) et des périodiques⁶⁹, notamment à la suite de l'appel à projets 2013 du Plan national de numérisation financé par le MCC, ou des opérations effectuées en partenariats entre la Bibliothèque nationale de France et ses « pôles associés régionaux » (PAR)⁷⁰ ;
- d'autres projets encore impliquent des documents de natures diverses, comme par exemple des cartes postales à la Bibliothèque municipale de Nantes ou des cartes postales et des revues en langue tahitienne dans le cadre de l'Université de la Polynésie française⁷¹, des corpus de dessins⁷² ou de thèses d'Ancien Régime à la BIU Santé, les cours des Écoles nationales d'Arts et Métiers⁷³, l'historique de la production institutionnelle (annuaires, programmes des cours, etc.) à l'Université Bordeaux Montaigne, ou le corpus relatif à l'histoire de l'Université de Perpignan⁷⁴.

67. Voir : <http://www.bovary.fr/>.

68. Voir : http://agorha.inha.fr/inhaprod/jsp/referenc...reference=INHA__METADONNEES__51.

69. Pour la presse locale ancienne, voir par exemple : http://archives.valdoise.fr/archive/fonds/FRAD095_00477, ou la numérisation de la presse alsacienne à la BNU : <http://www.bnu.fr/numistral/projets-en-cours> ; pour la presse quotidienne régionale, voir : <http://www.kiosque-lorrain.fr>.

70. Voir le site BnF consacré à la presse locale ancienne : <http://presselocaleancienne.bnf.fr/accueil> ainsi que le rapport d'Aline Girard [Girard, 2014] ; voir également l'ensemble des conférences de la bibliothèque municipale de Lyon au sujet de la mise en ligne et de la valorisation de la presse XIX-XXI^e s. : https://www.bm-lyon.fr/spip.php?page=video&id_video=442.

71. Voir : <http://anaite.upf.pf>.

72. Voir : <http://www.biusante.parisdescartes.fr/histoire/medica/dessins.php>.

73. Voir : <http://patrimoine.ensam.eu>.

74. Voir : <https://estudi.univ-perp.fr/collections/browse>.

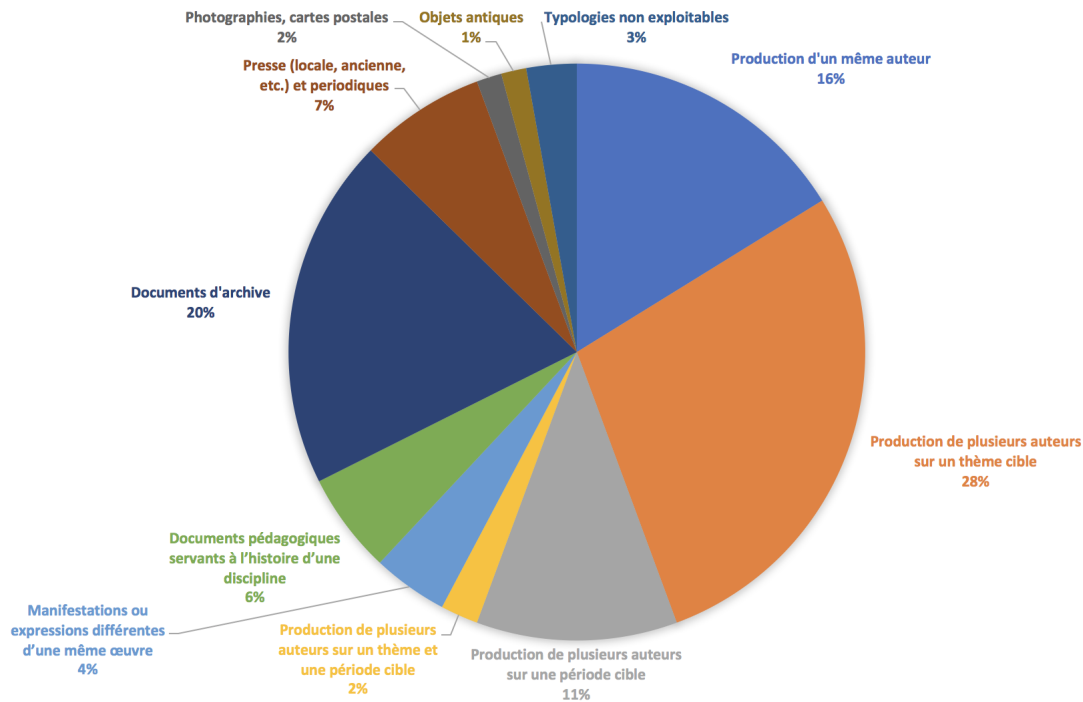


FIGURE 4 – Questionnaire aux bibliothécaires - question n° 4

Pour ce qui est de la nature des documents composant les corpus récemment numérisés, il s'agit principalement de : manuscrits (14 %), comme à la Bibliothèque municipale de Chalon sur Saône, à la Bibliothèque multimédia intercommunale d'Epinal, à la Bibliothèque du Grand Verdun, à la Bibliothèque municipale d'Autun et à la Bibliothèque patrimoniale de Pau, numérisés principalement dans le cadre du programme de numérisation de l'Institut de Recherche et d'Histoire des Textes (IRHT) ; de livres imprimés après 1810 (14 %), comme le corpus en cours au sujet du droit colonial du SCD de l'Université de Poitiers, celui concernant les médecins et pharmaciens montpelliérains ⁷⁵, le corpus en histoire des sciences et des techniques de Lille 1 ⁷⁶ ; de livres imprimés avant 1810 (12 %), comme le corpus concernant la jurisprudence provençale à Aix-Marseille Université, celui d'histoire du droit de l'Université Paris Sud, ou celui des livres d'architecture du XVI^e siècle du Prytanée national militaire ; de documents d'archives (12 %) ; de photographies (11 %), comme celles de Bains de mer et de la ville du Havre avant destruction.

75. <http://www.biu-montpellier.fr/florabium/jsp/redir.jsp?redir=folia>.

76. Voir : <http://iris.univ-lille1.fr>.

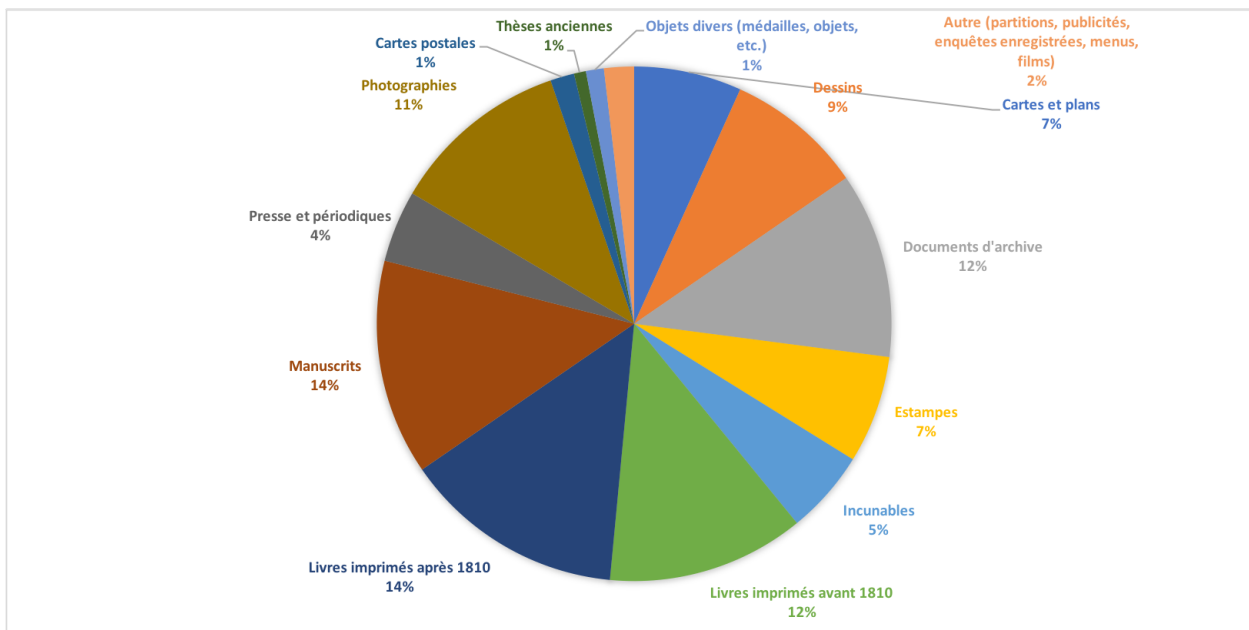


FIGURE 5 – Questionnaire aux bibliothécaires - question n° 5

Ces corpus sont principalement issus des collections propres de chaque institution, plus rarement constitués dans le cadre de projets de partenariat ; c'est le cas, par exemple, de la numérisation de 72 manuscrits de la BU Médecine de Montpellier du fonds Clairvaux en lien avec la Médiathèque du Grand Troyes⁷⁷, du projet en cours concernant le Corpus juridique lorrain, mené en collaboration entre la BU de l'Université de Lorraine, le Centre Lorrain d'Histoire du Droit de l'Institut François GénY, la Bibliothèque municipale de Nancy et la BnF⁷⁸, du projet concernant les ligériennes du Prytanée national militaire dans le cadre de *Bibliissima*⁷⁹, ainsi que des projets menés par la BnF : le projet *COLLECTA Actualisation numérique de la collection de François-Roger de Gaignières et création d'un CLOM pour les Humanités numériques* en partenariat avec la Fondation de la Maison des sciences de l'homme⁸⁰ ; le projet *Deloynes Écrits sur l'art : fonds Deloynes du département des Estampes et de la Photographie de la BnF* en partenariat avec le Centre André Chastel⁸¹ ; *VESPERA Versailles en perspectives. Plans des Archives Nationales et modélisation*, mené avec le Centre de Recherche du château de Versailles, les Archives nationales et l'Équipes Traitement de l'Information et Systèmes⁸².

La typologie et la nature des documents des corpus documentaires qui ressortent de notre enquête sont sans doute très hétérogènes. Principalement orienté vers les ressources patrimoniales, le choix des documents relève à la fois des intérêts de recherche des équipes des chercheurs, des collections que les institutions ont décidé de et ont pu numériser, ainsi que des éventuels projets et partenariats qui ont permis la constitution de

77. Voir : <https://www.bibliotheque-virtuelle-clairvaux.com/>.

78. Voir : <https://bu.univ-lorraine.fr/nos-bibliotheques/droit-et-sciences-economiques/pulsar-projet-de-numerisation-patrimonial>.

79. Voir : <http://www.bibliissima-condorcet.fr/fr/appels-a-projets/projets-retenus/bibliotheques-humanistes-ligeriennes-bhli>.

80. Voir : <http://c.bnf.fr/gcP>.

81. Voir : <http://c.bnf.fr/gcM>.

82. Voir : http://actions-recherche.bnf.fr/BnF/anirw3.nsf/IX01/A2015000134_verspera-versailles-en-perspectives-plans-des-archives-nationales-et-modelisation.

projets communs. La numérisation de corpus réclame des compétences et des ressources qu'il est souvent difficile de retrouver au sein de l'équipe qui est à l'initiative du projet : l'ingénierie des métadonnées ou l'archivage pérenne ne figurent pas forcément parmi les compétences ou les champs d'intérêt des chercheurs en SHS, de même que les actualités de la recherche ou les préoccupations méthodologiques des chercheurs en SHS ne circulent que partiellement dans le monde des professionnels de l'information et des bibliothèques. Comme nous le verrons par la suite, la création d'un réseau de métiers structuré permettant la planification des chantiers à mener, et le partage des compétences et des ressources figurent sans doute parmi les principaux facteurs de réussite d'un projet concernant les corpus.

CONSTITUTION ET ACCÈS AUX CORPUS : LES ENJEUX DE LA NUMÉRISATION ET DU TRAITEMENT DES DONNÉES

La numérisation de corpus recoupe sans doute une bonne partie des enjeux de toute opération de numérisation documentaire. Les pratiques de production et de diffusion du patrimoine littéraire et culturel sur internet, auxquelles des acteurs traditionnels tels que les bibliothèques et les institutions culturelles sont habitués depuis la mise en œuvre des premières bibliothèques numériques, jouent un rôle important dans la chaîne de traitement et de mise à disposition de corpus. Néanmoins, les évolutions technologiques et surtout l'acculturation croissante au numérique ont permis la naissance de nouveaux usages et entraîné une modifications des hiérarchies dans la production et dans le traitement des contenus culturels - deux éléments qui ne sont pas sans conséquences sur la constitution et l'utilisation de corpus.

La diffusion des Humanités numériques, ainsi que celle des technologies de reproduction de documents, ont laissé émerger, à côté des acteurs traditionnels, des nouveaux contributeurs (chercheurs et lecteurs) qui ont parfois des compétences et des exigences techniques très pointues et qui sont de plus en plus directement impliqués dans la production et dans le traitement des données patrimoniales et culturelles. Le débat autour de ces transformations, particulièrement vif, est étroitement lié à la production et à l'usage de corpus. En dehors d'un indéniable effet de mode, ces nouveaux outils et pratiques du numérique influencent et transforment notre rapport aux objets culturels et permettent de nouvelles possibilités d'exploitation des contenus. Pour certains, l'on pourrait parler d'un nouvel « humanisme numérique » [Doueïhi, 2011], ainsi que d'une nouvelle dimension pour le livre, entre l'objet fétiche [Doueïhi, 2010] et ses manifestation numériques « dé-livrées » de l'objet matériel [Vinck et Clivaz, 2014]⁸³. Des congrès et des journées d'études ont également été consacrés à cette sorte de « démocratisation », non seulement de l'accès aux contenus culturels, mais aussi de leur production⁸⁴. Par conséquent, les institutions culturelles et de recherche nationales et internationales s'interrogent sur cette modification dans la perception des contenus patrimoniaux, sur le rôle du lecteur/chercheur, sur les compétences nécessaires pour les professionnels des bibliothèques voulant s'impliquer dans des projets numériques⁸⁵, ainsi que sur les services à proposer à l'ère du numérique.

83. Comme le souligne Roger Chartier, avec le monde électronique, c'est désormais l'écran et non plus le livre qui permet de lire les textes, « est ainsi rompue la relation qui, dans toutes les cultures écrites antérieures, liait étroitement des objets, des genres et des usages [...]. Le monde numérique est un monde de fragments décontextualisés, juxtaposés, indéfiniment recomposables, sans que soit nécessaire ou désirée la compréhension de la relation qui les inscrit dans l'œuvre dont ils ont été extraits » [Chartier, 2012], en ligne <https://www.cairn.info/revue-le-francais-aujourd-hui-2012-3-page-11.htm>.

84. Voir la journée d'étude organisée par l'Université Paris Descartes en 2014 "Numérisation des patrimoines culturels pour qui? Pour quoi?", https://www.fabula.org/actualites/numerisation-des-patrimoines-culturels-pour-qui-pour-quoi_63840.php; voir aussi les actes du congrès de 2011 à l'Université de Lausanne [Clivaz *et al.*, 2012].

85. Sur ce thème, voir les résultats de l'enquête « HUBIC » de Olivier Le Deuff [Le Deuff, 2015].

Les enjeux didactiques et scientifiques

Du travail en salle de lecture à l'approche distante

C'est un fait reconnu que la numérisation et la mise à disposition de bibliothèques numériques sont à l'origine d'un fort intérêt de la part des chercheurs et des enseignants du secondaire, parfois même plus importante que la fréquentation des « murs » des bibliothèques [Urfist Nice, 2011]. En effet, comme le soulignent Mélanie Roustan et Anne Monjaret, la « démultiplication des objets institués en patrimoine, de leurs répliques numériques et des modalités de leurs accès élargit la disponibilité de sources pour les chercheurs en sciences humaines et sociales, amateurs et professionnels » [Roustan *et al.*, 2016, p. 11]; cette « démultiplication » produit un « renouveau du rapport à la matérialité des savoirs » qui se répercute directement sur les méthodes de recherche et sur la manière dont les chercheurs organisent leur temps de travail dans les bibliothèques, les archives et les musées.

L'accès distant à des collections numérisées a l'avantage de rendre accessibles des documents patrimoniaux à tout utilisateur à n'importe quel moment :

« Etant moi-même chercheur, avant d'avoir rejoint le monde des bibliothèques, j'ai pu faire quelques bonnes et mauvaises expériences de collaboration chercheurs/bibliothécaires. Si certains bibliothécaires sont heureux de faire découvrir et étudier leurs fonds, qu'ils mettent facilement à la disposition des chercheurs, cela est parfois plus difficile dans les grandes bibliothèques et dans les centres d'archives majeurs (BnF, Archives nationales, etc.). En effet, la limitation du nombre de documents consultables par jour et les systèmes parfois complexes de réservation des documents empêchent parfois les (jeunes) chercheurs d'accéder convenablement aux documents dont ils ont besoin pour leurs travaux. Dans ce cas, les numérisations de corpus peuvent être importantes pour permettre l'accès à des documents patrimoniaux y compris à distance, et à n'importe quelle heure du jour et de la nuit... » [conservateur]

Si des partenariats impliquent la possibilité d'étudier des collections dispersées ou difficiles d'accès⁸⁶, outre l'aspect pratique lié au fait de pouvoir effectuer des recherches sans se déplacer, les chercheurs mettent surtout l'accent sur les « effets de savoir inédits », comme le montrent les résultats de l'enquête menée autour du projet *European Regia* : « pouvoir faire des comparaisons sur écran, des mises en relation, au-delà des collections et des domaines institutionnellement ou académiquement définis » [Bouvier-Ajam *et al.*, 2011]. De même, les stratégies de lecture ont changé : la mise à disposition massive de corpus de textes en ligne, ainsi que « les vertus de l'effet de sérendipité », ont favorisé « un nouveau type de lecture extensive des textes » et des nouvelles possibilités de contextualisation [Bouqui, 2015, p. 57]. Les lecteurs, et plus particulièrement les chercheurs, ne cherchent plus seulement des stratégies pour accéder au document numérique avec des modalités proches de celles qu'ils utiliseraient pour le document original, mais ils cherchent à prendre du recul par rapport à l'intégralité de la collection, pour en tirer d'autres informations. Le dépassement des limites et des cloisons de la lecture sur place à l'aide des solutions informatiques évoque immédiatement les réflexions de Franco Moretti à propos de nouvelles stratégies de lecture à l'ère numérique,

86. Un exemple est constitué par le programme international *A Comprehensive Edition of Tocharian Manuscripts*, lancé en 2011 par l'Université de Vienne en collaboration avec plusieurs institutions de conservation et visant à la constitution d'un corpus de textes tokhariens [Le Masne de Chermont, 2016].

ainsi que la dichotomie, désormais largement exploitée dans le cadre des études sur les données massives, entre lecture rapprochée *close reading* » et lecture distante « *distant reading* », cette dernière permettant justement d’appréhender un ensemble de documents sous un nouveau regard⁸⁷. Quant aux enseignants du secondaire, ils signalent que « le support numérique permet des usages variés (en cours par l’enseignant ou à la maison par l’élève) et des travaux collaboratifs qui suscitent aussitôt des idées d’échanges au-delà de la "classe" » ; dans le cas précis de Europeana Regia, la collection numérique permet de rendre visible aux élèves « la notion de patrimoine commun à l’échelle européenne » [Bouvier-Ajam *et al.*, 2011].

Bien évidemment, les deux solutions d’accès aux documents ne semblent pas s’exclure, le livre et sa version numérique pouvant être complémentaires, comme c’est le cas de l’édition critique des œuvres de Molière dans le cadre du projet *Molière 21*⁸⁸. Dans certains domaines, le document original constitue toujours un support de recherche incontournable, le pendant matériel de la ressource numérique apporte une valeur ajoutée et ouvre de nouvelles pistes de recherche ; dans d’autres, ce sont l’environnement de conservation du document et les services offerts qui constituent la valeur ajoutée. C’est pourquoi la recherche d’un bon équilibre entre l’espace physique et l’espace numérique, les deux lieux incontournables du travail et de la recherche, ne cesse d’interroger les professionnels des bibliothèques, tant sur le plan de la configuration des espaces que sur celui des services, notamment pour les bibliothèques universitaires [Gasc, 2010] et de recherche⁸⁹.

Les professionnels de l’information savent que le chercheur, toutes disciplines confondues, « est aujourd’hui en quête d’une efficacité accrue du temps passé en bibliothèque » [Brault, 2016, p. 29]. Les modalités selon lesquelles les chercheurs repèrent et étudient les documents, ainsi que celles à travers lesquelles ils produisent et diffusent du savoir, sont liées à la numérisation massive des collections et aux outils et aux compétences numériques dont ils disposent. Il n’est donc pas étonnant que des chercheurs ayant besoin de travailler sur un ensemble de documents transforment leur séance de travail en salle de lecture en une séance de prises de vue de documents qu’ils étudieront, avec plus de calme, dans un deuxième temps⁹⁰. Cela correspond aussi à la pratique de certains des chercheurs que nous avons interviewés : tout au moins dans la phrase de dépouillement, ils utilisent leur appareil photographique numérique pour se fabriquer une base d’images à utiliser pour leurs recherches ou dans le cadre de séminaires et conférences ; par ailleurs, ils ne sont pas forcément informés des services de numérisation offerts par les institutions de conservation. Contre toute attente, il ne s’agit pas uniquement de jeunes chercheurs (doctorants ou post-doctorants) ou de chercheurs travaillant sur des petits corpus documentaires : d’après notre enquête, ceux qui ne savent pas à qui s’adresser pour mettre en place un projet de numérisation sont des chercheurs expérimentés (4 maîtres de conférence et

87. Le concept « *distant reading*, énoncé par Franco Moretti en 2000 [Moretti, 2000], est repris et précisé dans ses conférences à Berkley au printemps 2002, qui constituent les bases de ses « *Modelli astratti per la storia letteraria* » [Moretti, 2005].

88. Voir : <http://www.moliere.paris-sorbonne.fr/> et [Bouqui, 2015].

89. Nous pensons en particulier au projet concernant le quadrilatère Richelieu, dont une première partie a ouvert ses portes en décembre 2016, qui réunit les départements spécialisés de la BnF, la bibliothèque de l’INHA et celle de l’ENC : « ce campus d’un genre unique se veut un outil documentaire, scientifique et culturel dédié à l’art, à son histoire et au patrimoine. Il ambitionne d’être un cadre de travail exceptionnel pour les chercheurs et, dans le même temps, ouvert à un public plus large désireux de découvrir ce site et ses collections » [Dégez-Selves et Buxtorf, 2017, p. 22].

90. Pour des réflexions à ce sujet, plus précisément dans le domaine des archives, voir par exemple [Artières, 2016].

1 ingénieur de recherche), sans compter ceux qui manifestement une légère méfiance sur le service de numérisation, notamment envers le temps de traitement de la demande :

« idée (peut-être fausse) que ce serait en vain, ou si long que je n’aurais plus besoin des documents quand ils finiraient par être en ligne » [DR]

« je sais qu’un service de numérisation existe. Nous verrons plus tard, pour le moment c’est plus rapide si je numérise moi-même ce qui m’intéresse » [MCF]

Ces corpus numériques « fabriqués à la main » sont alors surtout un aide-mémoire, comme des photocopies ou des corpus préparatoires qui présupposent, dans un deuxième temps, le recours à des solutions numériques plus performantes. Néanmoins, cela n’exclut pas que des projets puissent s’appuyer sur des images de basse qualité prises par des chercheurs, afin de simplifier les opérations de collecte.

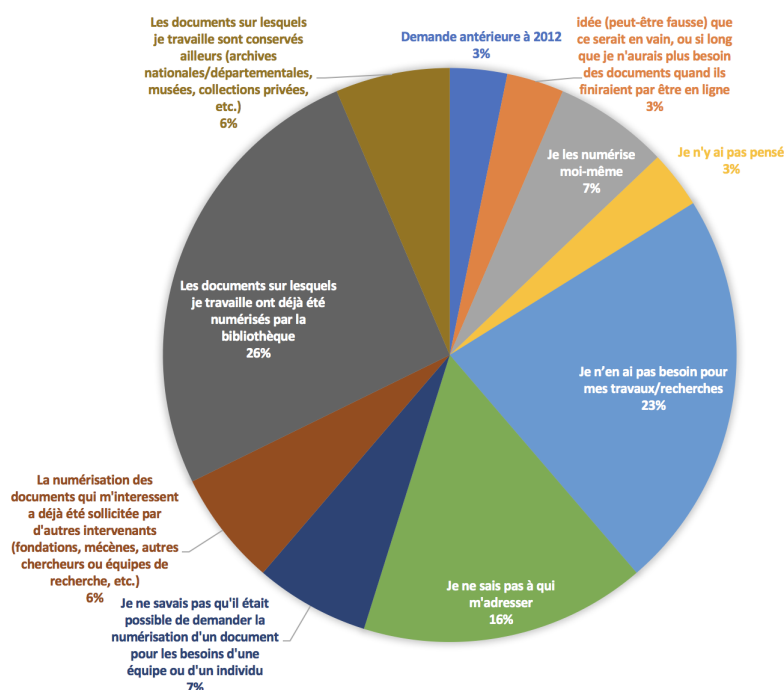


FIGURE 6 – Questionnaire aux chercheurs - question n° 22

Les Humanités numériques et la création de corpus intelligents

Discipline à part entière, intradiscipline ou transdiscipline⁹¹, les *Digital Humanities* (DH), ou Humanités Numériques (HN), constituent une branche de recherche de plus en plus d’actualité. Si les contours du terme restent encore difficiles à préciser⁹², il ne fait pas de doute que ces pratiques gagnent en popularité en France, tout au moins depuis le début des années 2010⁹³. Bien qu’il ne soit pas simple de s’affranchir de l’hégé-

91. « Les *digital humanities* désignent une transdiscipline, porteuse des méthodes, des dispositifs et des perspectives heuristiques liés au numérique dans le domaine des Sciences humaines et sociales » [Dacos, 2011].

92. Parmi les nombreuses contributions sur ce thème, voir [Berra, 2012], [Berra, 2015] et [Granjon et Magis, 2016].

93. Pour l’histoire des Humanités numériques en France, voir le *Manifeste des Humanités Numériques*, [Dacos, 2011], le volume d’introduction sous la direction de Pierre Mounier [Mounier, 2012], ainsi que la bibliographie en cours de réalisation par Jérôme Valluy : <http://www.hnp.terra-hn-editions.org/TEDI/article423.html>.

monie du monde anglophone [Fiormonte, 2016], les Humanités numériques acquièrent progressivement leur place dans le domaine des SHS et réunissent désormais une communauté francophone active⁹⁴, qui organise nombre de séminaires et de conférences, anime des projets⁹⁵ et organise des formations spécifiques⁹⁶. Les Humanités numériques ne sont pas non plus exemptes de critiques ; comme le relève Aurélien Berra, lors d'un récent colloque, l'une des questions qu'elles suscitent concerne leur fondement critique : « Y-a-t il suffisamment de réflexivité dans la démarche des Humanités numériques ? » [Berra et Mounier, 2016]. Si elles sont capables d'apporter un nouveau regard, on peut leur reprocher d'avoir du mal à atteindre la hauteur de vue des disciplines « traditionnelles » des SHS, ou des disciplines informatiques et des sciences de l'ingénieur : cela se traduit par un positionnement professionnel encore incertain⁹⁷. Si, pour certains, les HN sont un lieu d'interdisciplinarité, de collaboration, « principalement une pratique de recherche » [Berra, 2012], et « peut-être ne sont-elles pas non plus une innovation majeure » [Glorieux, 2015]⁹⁸, un de leurs mérites semble être de participer au « brouillage des frontières scientifiques » [Granjon et Magis, 2016] : les Humanités numériques sont souvent vues comme une remise en cause politique des méthodes académiques traditionnelles, déployant des nouveaux systèmes de communication et mettant en scène de nouveaux acteurs et de nouvelles possibilités de financements. À travers la création et la gestion de bases de données, la conception d'algorithmes et d'interfaces de programmation (API), les Humanités numériques semblent porteuses de la promesse d'une fructueuse rencontre entre l'informatique et les sciences humaines et sociales, dans laquelle le travail autour de corpus est presque structurel :

« Les Humanités numériques ou digitales sont le résultat de ce dialogue de sciences : sciences interprétatives et sciences objectives se rejoignent alors pour créer des corpus numériques "intelligents", formidables outils de connaissance. Ces corpus ne sont pas statiques, mais évolutifs, ce sont des infrastructures numériques au service de la connaissance » [Bensamoun et Latil, 2015, Préface, p. 9].

Ces « corpus numériques intelligents », qui constituent la base de travail des Humanités numériques, peuvent être considérés comme des collections numériques aux performances accrues, qui associent aux traditionnels enjeux scientifiques de la numérisation des collections de niveaux et de possibilités supérieurs de connaissance. Par exemple,

94. Pour en suivre les activités, voir le site de l'institution francophone des Humanités numériques *Humanistica* : <http://www.humanistica.ca/>, et la liste de diffusion DH : <https://groupes.renater.fr/sympa/info/dh>.

95. Pour un répertoire des projets en ligne, voir : <http://www.humanitesnumeriques.fr/bases-et-projets-en-ligne/>.

96. Pour ne citer que quelques exemples de formation : le master « Humanités classiques et humanités numériques » de l'Université Paris Nanterre, créé en 2014, et les masters de l'École nationale des chartes « Technologies numériques appliquées à l'histoire », créé en 2008, et « Humanités numériques », créé en 2017 en collaboration avec l'École normale supérieure et l'École Pratique des Hautes Études dans le cadre de l'Université PSL.

97. Les jeunes chercheurs soulignent la précarité de l'emploi des spécialistes des Humanités numériques et le manque de perspectives professionnelles claires : « Il devrait y avoir des perspectives à long terme pour les chercheurs, ingénieurs et bibliothécaires qui s'engagent dès le début de leur carrière dans la voie des Humanités numériques » [IHA, 2013].

98. « Les Humanités numériques ne sont pas une mode passagère. Peut-être ne sont-elles pas non plus une innovation majeure. Elles apparaissent avant tout comme une nouvelle manière de voir qu'il s'agit d'appropriation : il faut diffuser et développer une compréhension critique des algorithmes, pour en lire les intentions cachées, pour en déconstruire les effets idéologiques, mais aussi pour mettre au point de nouveaux instruments de connaissance » [Glorieux, 2015].

ce seront des collections d'images numérisées associées à différentes couches de métadonnées (Dublin Core, METS, XML/BDD, XML/TEI, XML/EAD, etc.) moissonnables à travers un entrepôt OAI-PMH, des textes numérisés disponibles au format OCR et encodés en XML/TEI, des bases de données XML téléchargeables *via* une API (*Application Programming Interfaces*), du code ou des ressources mises à disposition sur un entrepôt GitHub, etc. Différents des collections numériques traditionnelles, simplement consultables à distance depuis les sites des institutions patrimoniales, ces corpus naissent, souvent sous l'impulsion des chercheurs, avec la vocation d'être exploités par d'autres chercheurs, en s'inscrivant ainsi dans une logique de partage des connaissances et d'ouverture des données caractéristiques de l'ère du *big data* et du *Linked Open Data* (LOD).

C'est sur la base de ces pratiques et outils de connaissance que nombre d'initiatives de recherche se sont construites autour de la création de corpus et de la fouille de textes⁹⁹, principalement dans le but de permettre des échanges méthodologiques et de partager un soutien technique et juridique. Il serait difficile de dresser une liste exhaustive de toutes les initiatives et de leurs relations mutuelles, mais nous pouvons tout au moins citer les grands réseaux fédérant plusieurs projets :

- Consortium CAHIER (Corpus d'Auteurs pour les Humanités : information, édition, recherche)¹⁰⁰ : proposé en 2011, il s'agit d'un consortium interdisciplinaire de projets numériques en accès libre qui permet de tester des prototypes d'édition et de traitement des données (en collaboration avec la TGIR Huma-Num)¹⁰¹. Le consortium est composé de 52 projets membres ainsi que de 10 projets associés : *Bibliothèques Virtuelles Humanistes (BVH)*¹⁰², *Corpus Électronique de la Première Modernité (CEPM)*¹⁰³, *Manuscrits de Stendhal*¹⁰⁴, *Base de français médiéval (BFM)*¹⁰⁵, *Archives Numériques de l'École Lvov-Varsovie (e-LV)*¹⁰⁶, *Études sur la Renaissance d'Horace (ERHO)*¹⁰⁷, *Schola Rhetorica*¹⁰⁸, *Les dossiers de Bouvard et Pécuchet*¹⁰⁹, *Projets Ichtya et Malaterra*¹¹⁰, *Montédite – Édition électronique des Pensées de Montesquieu*¹¹¹, *Manuscrits francophones*¹¹², *Archive Numérique Desanti*¹¹³, *Montesquieu, Œuvres complètes en ligne et Bibliothèque virtuelle Montesquieu*¹¹⁴, *Bibliothèques Privées à l'Âge Moderne (BI-*

99. Sur ce point, voir l'introduction d'Isabelle Tellier (Université Sorbonne Nouvelle Paris 3) : http://www.lattice.cnrs.fr/sites/itellier/poly_fouille_textes/fouille-textes.pdf.

100. Voir : <http://cahier.hypotheses.org/le-consortium>.

101. Voir le dossier consacré à la constitution du consortium, <https://humanum.hypotheses.org/203>.

102. Voir : <http://www.bvh.univ-tours.fr>.

103. Voir : <http://cepm.paris-sorbonne.fr>.

104. Voir : <http://manuscrits-de-stendhal.org/>.

105. Voir : <http://txm.bfm-corpus.org/>.

106. Voir : <http://www.elv-akt.net/>.

107. Voir : <http://www.univ-paris3.fr/horace>.

108. Voir : <http://schola-rhetorica.org/>.

109. Voir : <http://www.dossiers-flaubert.fr/>.

110. Voir : <https://www.unicaen.fr/puc/sources/malaterra/>.

111. Voir : <http://www.unicaen.fr/services/puc/sources/Montesquieu/>.

112. Voir : <http://www.item.ens.fr/index.php?id=579252>.

113. Voir : <http://desanti.huma-num.fr/>.

114. Voir : <http://montesquieu.huma-num.fr/accueil>.

PrAM)¹¹⁵, *Hyperdonat, édition numérique de commentaires anciens*¹¹⁶, *Projets numériques du Groupe D'Alembert*¹¹⁷, *Sociorama. Littérature panoramique internationale du XIX^e siècle*¹¹⁸, *LAFABREV (La Fabrique de la Révolution)*¹¹⁹, *Journaux d'Alexandre Dumas*¹²⁰, *Édition électronique de la correspondance de Flaubert*¹²¹, *CHISPA : Création d'outils pour l'exploitation numérique de corpus de manuscrits HISPANiques*¹²², *Les Écrits de l'abbé de Saint-Pierre*¹²³, *Correspondance de Pierre Bayle*¹²⁴, *Projets IG-Louvre et E-STAMPAGES*¹²⁵, *Édition électronique des œuvres de Louis de Boissy*¹²⁶, *Corpus 14*¹²⁷, *Correspondance de Condorcet*¹²⁸, *Fonds numérique Jean-Luc Lagarce*¹²⁹, *Édition génétique numérique du «Robinson» de Paul Valéry*¹³⁰, *Correspondance électronique d'Émile Zola (CorrELEZ)*¹³¹, *Archives des Traducteurs et des Écrivains de la Littérature Italienne : Éditions et Recherches (ATELIER)*¹³², *Les deux évasions de Benoîte Groult (E2E)*¹³³, *Mauriac en ligne*¹³⁴, *Édition électronique de contes populaires français (E-Conte)*¹³⁵, « *Renan Source* » *Une édition génétique numérique des manuscrits d'Ernest Renan*¹³⁶, « *Testaments de Poilus* » – *Une plateforme de transcription collaborative au service du patrimoine manuscrit*¹³⁷, *Basnage*¹³⁸, *Édition de la correspondance du géomètre Gaspard Monge 1795-1799*¹³⁹, *Écritures savantes au siècle des Lumières. La correspondance et les carnets de visiteurs de Jean-François Séguier*¹⁴⁰, *DYRIN*¹⁴¹, *Thresors de la Re-*

115. Voir : <http://www.enssib.fr/les-poles-thematiques/histoire-des-bibliotheques/dossiers/e-collections-et-collectionneurs/bipram>.

116. Voir : <http://hyperdonat.huma-num.fr/editions/index.html>.

117. Voir : <http://dalembert.academie-sciences.fr/>.

118. Voir : <http://promethee.philo.ulg.ac.be/sociorama/index.html>.

119. Voir : <http://thomas-lallier.fr/axmaleo/web/>.

120. Voir : <http://alexandredumas.org/>.

121. Voir : <http://flaubert.univ-rouen.fr/correspondance/edition/>.

122. Voir : <http://chispa.hypotheses.org/>.

123. Voir : http://www.unicaen.fr/mrsh/crhq/_index.php?page=equipes/cultures-programmes-3&suite=.

124. Voir : <http://bayle-correspondance.univ-st-etienne.fr/>.

125. Voir : <http://www.hisoma.mom.fr/recherche-et-activites/programmes-cofinances/programme-E-STAMPAGES>.

126. Voir : <http://www.licorn-research.fr/Boissy.html>.

127. Voir : <http://corpus14.ortolang.fr/>.

128. Voir : <http://inventaire-condorcet.com/fra/page/home>.

129. Voir : <http://fanum.univ-fcomte.fr/lagarce/>.

130. Voir : <http://www.item.ens.fr/index.php?id=13898>.

131. Voir : <http://www.archives-zoliennes.fr>.

132. Voir : <http://cahier.hypotheses.org/archives-des-traducteurs-et-des-ecrivains-de-la-litterature-italienne-editions-et-recherches-atelier>.

133. Voir : <http://www.espace-transcription.org/corpus/description/3>.

134. Voir : <http://mauriac-en-ligne.u-bordeaux-montaigne.fr/>.

135. Voir : <http://las.ehess.fr/index.php?2326>.

136. Voir : <http://www.item.ens.fr/projet-renan-source/>.

137. Voir la présentation <https://cahier.hypotheses.org/testaments-de-poilus>, et le site pour la transcription collaborative : <https://testaments-de-poilus.huma-num.fr>.

138. Voir : <http://www.licorn-research.fr/Basnage.html>.

139. Voir : <https://eman.hypotheses.org/187>.

140. Voir : <https://www.seguier.org/>.

141. Voir : <https://cahier.hypotheses.org/DYRIN>.

- naissance*¹⁴², *Projet La Réticence*¹⁴³, *Marc-Michel Rey*¹⁴⁴, *André Frénaud numérique*¹⁴⁵, *NaviLog (AnaLog + VariaLog)*¹⁴⁶, *CORR-Proust*¹⁴⁷, *Création de corpus de livrets d'opéra sous l'ancien régime, l'Académie royale de musique de Paris (1671-1791)*¹⁴⁸, *Nouvelle Édition Numérique de Fac-similés de Référence (NENUFAR)*¹⁴⁹, *Digital Alfieri*¹⁵⁰, *BIBLINDEX*¹⁵¹, *Digital Edition of the Roman de Florimont (DigiFlor)*¹⁵², *L'invention du théâtre antique dans l'Europe de la première modernité – Commentaires et paratextes (ITHAC)*¹⁵³; les projets associés (ne permettant pas le partage du corpus sous forme de données libres et interopérables) : *André Breton*¹⁵⁴, *Corpus Descartes (ProDescartes)*¹⁵⁵, *Corpus Boris Tchitchérine*¹⁵⁶, *Correspondance Nalèche*¹⁵⁷, *Cours d'Antoine Desgodets*¹⁵⁸, *EDITEF*¹⁵⁹, *Guides de Paris*¹⁶⁰, *Projet Ponge*¹⁶¹, *Projets de la section Humanisme de l'IRHT*¹⁶², *Routes du livre italien ancien en Normandie*¹⁶³
- LabEx OBVIL (Observatoire de la vie littéraire)¹⁶⁴ : actif dans le domaine des Humanités numériques, ses recherches s'appuient sur une bibliothèque électronique composée de corpus d'œuvres et de commentaires dans le domaine de la littérature française. Ce laboratoire accueille actuellement 21 projets¹⁶⁵ : *Caraïbes*, *Mercurie galant*¹⁶⁶, *Autorités en partage*¹⁶⁷, *Haine du théâtre*¹⁶⁸, *Storytelling*, *HyperPaulhan*¹⁶⁹, *Discours sur la danse*¹⁷⁰, *HyperApollinaire*¹⁷¹, *Historiographie*

142. Voir : <https://eman.hypotheses.org/198>.

143. Voir : <https://cahier.hypotheses.org/projet-la-reticence>.

144. Voir : <http://rey.huma-num.fr/home>.

145. Voir : <https://cahier.hypotheses.org/andre-frenaud-numerique>.

146. Voir : <http://forell.labo.univ-poitiers.fr/>.

147. Voir : <https://cahier.hypotheses.org/corr-proust>.

148. Voir : <http://fanum-txm.univ-fcomte.fr/txm/>.

149. Voir : <https://cahier.hypotheses.org/nenufar> et <https://nenufar.hypotheses.org/>.

150. Voir : <https://cahier.hypotheses.org/digital-alfieri>.

151. Voir : <http://www.biblindex.info/>.

152. Voir : <https://digiflorimont.hypotheses.org/>

153. Voir : <https://cahier.hypotheses.org/ithac>.

154. Voir : http://www.andrebretton.fr/card/andre_bretton.

155. Voir : <https://prodescartes.greyc.fr/>.

156. Voir : <http://wiki-tchitcherine.ens-lyon.fr/index.php/Accueil>.

157. Voir : <https://naleche.hypotheses.org/>.

158. Voir : <http://www.desgodets.net/>.

159. Voir : <http://www.editef.univ-tours.fr/>.

160. Voir : <http://passes-present.eu/fr/guides-de-paris-les-historiens-des-arts-et-les-corpus-numeriques-363>.

161. Voir : <http://obvil.paris-sorbonne.fr/corpus/ponge/>.

162. Voir : <https://www.irht.cnrs.fr/?q=fr/recherche/sections/humanisme>.

163. Voir : <https://www.unicaen.fr/recherche/mrsh/rdli/>.

164. Voir : <http://obvil.paris-sorbonne.fr/>.

165. Pour l'ensemble des projets, voir <http://obvil.paris-sorbonne.fr/projets>.

166. Voir : <http://obvil.paris-sorbonne.fr/corpus/mercurie-galant/>.

167. Voir : <http://obvil.paris-sorbonne.fr/corpus/mythographie/>.

168. Voir : <http://obvil.paris-sorbonne.fr/corpus/haine-theatre/>.

169. Voir : <http://obvil.paris-sorbonne.fr/corpus/paulhan/>; voir aussi [Barthélemy *et al.*, 2017].

170. Voir : <http://obvil.paris-sorbonne.fr/corpus/danse/>.

171. Voir : <http://obvil.paris-sorbonne.fr/corpus/apollinaire/>.

*théâtrale*¹⁷², *Molière*¹⁷³, *Góngora*¹⁷⁴, *Oraliser la littérature, Challe, Ponge*¹⁷⁵, *Fabula Numerica*¹⁷⁶, *Valéry, Facéties, Mercure de France, littérature italienne*¹⁷⁷, *Sainte-Beuve*¹⁷⁸, *L'écrivain, l'institution scolaire et la littérature. L'écrivain face aux modèles scolaires (1840-1940)*¹⁷⁹, *Patrimoine sonore de la poésie, Apollinaire/Marie Curie*.

- « Usages Des Patrimoines Numérisés » (UDPN)¹⁸⁰ : créé en 2013 en réponse à un appel à projets interdisciplinaires émanant de la COMUE « Sorbonne Paris Cité » et financé par les crédits IDEX [Ferjoux et Dupond, 2017], ce réseau compte actuellement 16 projets de recherche autour de quatre grands types de matériau : corpus textuels, iconographiques, audiovisuels et des arts vivants ; la liste des projets comprend : *ArchiZ*¹⁸¹, *De la trace à l'archive, ECFOLI*¹⁸², *EHGEL, LAFA-BREV - La fabrique de la Révolution, FANAS, Méthodes de collectes et usages des archives et collections, Modélisation et simulation du patrimoine, Outils de mise en ligne de corpus, Patrimoine numérique et création, Plateforme 14. Une famille dans la Grande Guerre*¹⁸³, *Technological Uncanny*¹⁸⁴, *TLHUB - Translation and Literary Hub*¹⁸⁵, et *Visualisation de données textométriques*.

On notera que ces corpus ne sont pas simplement destinés à l'usage des chercheurs : les usages pédagogiques et didactiques des archives numérisées sont également mis en évidence, comme dans le cas du corpus consacré à Emile Zola¹⁸⁶, notamment à propos de la méthode d'écriture zolienne [Macke, 2015, p. 92].

Comme le montrent les collaborations évoquées par les chercheurs qui ont répondu à notre enquête, de nombreuses autres d'initiatives rattachées au « Programme Investissements d'avenir » (PIA)¹⁸⁷ sont également à l'origine de la mise en ligne de corpus organisés autour d'un auteur ou d'une thématique ; grâce à l'institution de partenariats, ces programmes permettent de faire dialoguer plusieurs acteurs : les unités de recherche, les différents professionnels de l'information scientifique et technique et les institutions patrimoniales ; c'est le cas, par exemple, du LabEx *Les passés dans le présent* (partenariats avec le Musée du quai Branly, la Bibliothèque de Documentation Internationale Contemporaine, le Musée d'Archéologie Nationale, la Bibliothèque nationale de France)¹⁸⁸, du LabEx *Hastec* (partenariat avec les Archives Nationales et le CNAM Musée des tech-

172. Voir : <http://obvil.paris-sorbonne.fr/corpus/historiographie-theatre/>.

173. Voir : <http://obvil.paris-sorbonne.fr/corpus/moliere/>.

174. Voir : <http://obvil.paris-sorbonne.fr/corpus/gongora/>.

175. Voir : <http://obvil.paris-sorbonne.fr/corpus/ponge/>.

176. Voir : <http://obvil.paris-sorbonne.fr/corpus/fabula-numerica/>.

177. Voir : <http://obvil.paris-sorbonne.fr/corpus/mdf-italie/>.

178. Voir : <http://obvil.paris-sorbonne.fr/corpus/sainte-beuve/>.

179. Voir : <http://obvil.paris-sorbonne.fr/corpus/ecole/>.

180. Voir : <http://udpn.fr/>.

181. Voir : <http://www.archives-zoliennes.fr/> et <http://journals.openedition.org/genesis/1658>.

182. Voir : <http://udpn.fr/spip.php?article101>.

183. Voir : <http://udpn.fr/spip.php?article115> et <http://www.plateforme1418.com/introduction>.

184. Voir : <http://udpn.fr/spip.php?article88>.

185. Voir : <http://udpn.fr/spip.php?article92>.

186. Voir : <http://udpn.fr/spip.php?article120>.

187. Voir : <http://www.enseignementsup-recherche.gouv.fr/cid55892/comprendre-le-programme-investissements-d-avenir.html>.

188. Voir : <http://passes-present.eu/>.

niques)¹⁸⁹ ou de l'EquipEx *Biblissima*¹⁹⁰. Dans le cadre de *Biblissima*, signalons que les « projets partenariaux *Biblissima* », sélectionnés sur la base d'un appel à manifestation d'intérêt, dont le 6^e est en cours¹⁹¹, prévoient des « opérations conjointes de recherche, de documentation, de numérisation portant sur des collections historiques de manuscrits et d'imprimés anciens, associant au moins un établissement de conservation et un établissement d'enseignement et/ou de recherche, quels qu'ils soient ». Signalons aussi des dispositifs de partage de méthodologies dans le champs des Humanités numériques, comme MutEC (Mutualisation pour les éditions critiques et les corpus numériques)¹⁹², attentif aux projet de collaboration autour de corpus [Ingarao et Saïdi, 2010]. D'autres réseaux, comme celui des Maisons de Sciences de l'Homme, figurent également parmi les exemples de collaborations fructueuses évoquées par les chercheurs. Des projets comme « Fonte Gaia » constituent déjà un fructueux réseau d'échanges entre chercheurs et bibliothécaires autour de projets numériques¹⁹³ ; aussi, d'autres retour nous témoignent de l'intérêt des ces collaborations :

« Cette collaboration est indispensable pour répondre aux besoins des chercheurs. Triptyque gagnant pour les humanités numériques : chercheur / bibliothécaire ou documentaliste / informaticien » [Université de Nice, SCD]

Toute équipe ayant un projet de mise en ligne de corpus organisé autour d'un auteur ou d'une thématique aura ainsi intérêt à se rapprocher de ces réseaux, à la fois parce qu'ils offrent la possibilité d'échanges scientifiques et parce qu'ils disposent de possibilités de financement. De même, les institutions patrimoniales auront intérêt à tisser des partenariats avec ces interlocuteurs, dont le réseau et l'expertise pourront constituer un levier pour la valorisation des fonds ; quel qu'il soit l'approche choisi, les Humanités numériques seront un vecteur important pour les futurs services de bibliothèque :

« No matter which approaches to supporting the digital humanities you opt to take, keep in mind that what we call *The Digital Humanities* today will soon be considered *The Humanities* » [Schaffner et Erway, 2014]¹⁹⁴

Les bibliothèques numériques vers de nouveaux usages ?

La gestion et la valorisation des fonds patrimoniaux

Comme l'écrit Dominique Maillet :

« Deux objectifs essentiels président à la constitution d'un fonds numérique : la gestion et la valorisation des collections. Les deux sont liés : pour les collections iconographiques particulièrement, la gestion et la valorisation des collections passent par la conservation des documents originaux » [Maillet, 2011, p. 123]

« Aide à la préservation », capable de protéger les documents fragiles de la manipulation directe, mais aussi de les manipuler à l'écran en s'adaptant à différents types

189. Voir : <https://labexhastec-psl.ephe.fr/>.

190. Voir : <http://www.biblissima-condorcet.fr/>.

191. Voir : <http://www.biblissima-condorcet.fr/fr/actualites/sixieme-appel-manifestation-interet-biblissima-2018>.

192. Voir : <https://www.ish-lyon.cnrs.fr/mutec>.

193. <http://fontegaia.hypotheses.org/projet-fonte-gaia-2>.

194. « Quelles que soient les approches pour soutenir les Humanités numériques que vous choisissiez de suivre, gardez à l'esprit que ce que nous appelons aujourd'hui "Les Humanités numériques" sera bientôt considéré Humanités ».

de lecteurs, « la numérisation permet de se constituer une vitrine, tout en facilitant le rayonnement des fonds et leur accessibilité » [Jacquet, 2012, p. 139]. Les campagnes de numérisation sont l'occasion de mettre en valeur les collections de la bibliothèque, notamment les fonds d'intérêt patrimonial et ceux qui portent sur l'histoire et les cultures locales¹⁹⁵. C'est dans cette logique que des expositions virtuelles se constituent en tant que mode permanent de communication afin de susciter l'intérêt du public autour des collections, l'enjeu étant aussi de stimuler des travaux de recherche sur les documents qui ont pu être numérisés¹⁹⁶ ; cela peut amener à la constitution de « numérisations enrichies » par les jeux de données, qui peuvent s'ajouter aux documents physiques, finissant par constituer un véritable patrimoine numérique¹⁹⁷, tel qu'il est reconnu par la *Charte sur la conservation du patrimoine numérique* de l'UNESCO [Unesco, 2003]¹⁹⁸.

La numérisation des collections patrimoniales peut également s'appuyer sur le « fort pouvoir de viralité et de lien émotionnel » de l'image numérique [Mesguich, 2017, p. 49] pour publier des contenus sur des plateformes comme *Instagram*, *Pinterest*, ou *Snapchat* qui connaissent une forte expansion depuis les années 2000¹⁹⁹, ou susciter des initiatives pleines d'humour comme le Tumblr des images extraites des collections numérisées de la Bodleian Library²⁰⁰, Gallica LOL, conçu lors du Hackathon 2016 de la BnF²⁰¹ et les innovations proposées par les « Gallicanautes » recensées dans Gallica Studio²⁰², les Gifs animés de Numelyo²⁰³, ou encore pour enrichir l'offre des services commerciaux de reproduction. Si la manière dont les usagers utilisent les collections numériques est très importante, il est encore plus important, comme le souligne Philippe Chevalier, de multiplier les occasions d'échange avec le public, de passer « de la connaissance par l'observation à la connaissance par la participation », afin d'adapter et de corriger l'offre :

« La participation du public est le meilleur moyen d'empêcher que les bibliothèques numériques ne deviennent de nouvelles boîtes noires, dont on ne comprend ni les questions qu'elles nous posent ni les réponses qu'elles nous donnent » [Chevallier, 2016, p. 56]²⁰⁴

La numérisation autorise ainsi à construire de nouveaux objets de recherche participatifs :

195. Pour une synthèse concernant la notion de patrimoine, voir l'encadré de Bernard Huchet « Le patrimoine : approches conceptuelles et collections concernées » [Alix, 2013, p. 179-180] et [Haquet, 2012].

196. Voir par exemple le cas de la numérisation dans les bibliothèques de l'ENS : <http://www.bib.ens.fr/Numerisation-dans-le.912.0.html>.

197. La distinction entre patrimoine numérique et patrimoine numérisé s'impose : « Le patrimoine numérique et le patrimoine numérisé ne sont pas à confondre. Le patrimoine numérique est tout à la fois un objet numérique et l'outil qui sert à le faire fonctionner. [...] Le patrimoine numérisé renvoie à l'existence d'un objet original existant dans une matérialité différente et dont la numérisation permet d'en sauvegarder une image. Pourtant, ils mettent en jeu tous les deux la question de l'usage de la mémoire sociale dans la qualification du patrimoine et le statut même de ce qui fait patrimoine » [Tardy et Dobedei, 2015].

198. Art. 1 : « Les documents numériques comprennent, parmi un large éventail de formats électroniques qui ne cesse de se diversifier, des textes, des bases de données, des images fixes et animées, des documents sonores et graphiques, des logiciels et des pages Web. Ils sont souvent éphémères, et leur conservation nécessite des mesures volontaires d'entretien et de gestion dès leur création » [Unesco, 2003].

199. Outre la section concernant ces plateformes de diffusion des images, il sera profitable de réfléchir à l'adéquation entre les contenus à valoriser et les différents canaux [Mesguich, 2017, p. 30-31].

200. Voir : <http://bodleianlibs.tumblr.com/>.

201. Voir : <https://gallica.lol>.

202. Voir : <http://gallicastudio.bnf.fr/>.

203. Voir : http://numelyo.bm-lyon.fr/collection/BML:BML_00COL00101Gifanim.

204. Sur les usages de Gallica, voir aussi [Chevallier et Amar, 2014].

« Il faut insister sur le caractère collectif de cette entreprise ; en effet, chaque chercheur qui consultera ce corpus peut, s'il le souhaite, laisser un commentaire, proposer une hypothèse, voire ajouter une pièce jointe à l'archive-image. Ainsi c'est avec la numérisation outillée une forme nouvelle de recherche qui se met en place, plus proche en cela des manières de faire recherche des chercheurs en biologie ou en physique, mais aussi très innovante et très stimulante car pouvant être au niveau international non tributaire de l'obtention de financement et de visas » [Artières, 2016, p. 25]

Une offre numérique en cours d'évolution

À la suite des projets pionniers menés surtout par la BnF et par la BPI à partir des années 1980, l'offre numérique des bibliothèques et des institutions culturelles n'a fait qu'augmenter. En suivant de près l'évolution des technologies informatiques, elle s'est de plus en plus enrichie et différenciée, en cherchant à trouver un positionnement par rapport aux acteurs marchands²⁰⁵ et à s'adapter aux pratiques et aux besoins de nouveaux publics²⁰⁶.

Cette offre numérique est déclinée selon différents types de réservoirs et d'outils dont la distinction, tout au moins du point de vue terminologique, est parfois difficile à cerner. L'appellation de « bibliothèque numérique » par exemple, peut être utilisée pour indiquer des contenus et des réalisations techniques assez différents : portails documentaires, sites web dédiés, archives, dépôts numériques, collections patrimoniales numérisées, collections d'ebooks, banques d'images généralistes ou spécialisées, etc. De ce fait, comme le montrent les résultats de notre enquête, la numérisation et le traitement de corpus viennent intégrer cette offre en se positionnant à l'intérieur de solutions déjà en place, aux contours parfois flous ; ou, plus rarement, ils tracent une nouvelle voie. Ainsi, qu'on les retrouve sous l'étiquette de « collections numériques » ou de « corpus numérisés », ces documents et ces données (et métadonnées) peuvent être localisés à différents endroits dans l'arborescence des ressources d'une institution. Ils peuvent être simplement signalés parmi les projets des numérisations effectuées et intégrées à la bibliothèque numérique (accessibles seulement sur place ou également à distance), ils peuvent figurer en tant que dossier à part en relation avec une exposition ou une activité événementielle, ils peuvent constituer une section de la banque d'images, ou encore, ils peuvent disposer d'une section réservée et être associés à des services supplémentaires par rapport à d'autres collections numériques.

205. Peu après leur création, des grandes bibliothèques numériques ont été confrontées aux enjeux du marketing, à l'émergence d'acteurs privés et d'entreprises commerciales disposant de ressources financières importantes et dont le but principal est la commercialisation des données sur Internet ; sur les réactions de la Bibliothèque nationale de France, de la British Library et de l'Österreichische Nationalbibliothek à l'annonce de Google de la numérisation de fonds d'éditeurs et de collections de bibliothèques, voir [Béquet, 2014] ; la conclusion de partenariats public/privé semble l'une des pistes possibles, comme le montre la création de la filière de droit privé BnF-Partenariat en 2012, voir : http://www.bnf.fr/fr/acces_dedies/bnf_partenariats/s.bnf_partenariats_actualite.html ; sur ces thèmes, voir aussi le rapport sur la numérisation du patrimoine écrit [Tessier, 2010].

206. Pour retracer l'histoire de la création de cette offre depuis les premiers projets français, voir [Collart, 2011, p. 87-99] ; pour le catalogue en ligne du patrimoine culturel numérisé, voir <http://www.numerique.culture.fr/pub-fr/index.html> ; pour une liste des bibliothèques patrimoniales en ligne, voir : <https://bibandclac.wordpress.com/2017/04/22/bibnum-patrimoniales-quelle-visibilite-sur-le-site-de-leur-bibliotheque/>. Pour la constitution des bibliothèques numériques, voir la boîte à outil [Claerr et Westeel, 2013b].

L'essor de la dimension collaborative, à travers l'implication des usagers, qui s'engagent dans la production de contenus autour de corpus numérisés, est également à prendre en compte dans l'évolution de l'offre numérique. En effet, au delà de la mise à disposition des images numériques, ces initiatives nécessitent la mise en place de plateformes adaptées. Ces projets peuvent par exemple porter sur la possibilité de commenter et d'analyser les documents numérisés²⁰⁷, sur la correction collaborative des documents numérisés²⁰⁸, ou sur la transcription de documents; ce dernier cas est illustré par deux projets récents : le projet TRANSCRIRE, qui permet de contribuer à la transcription de documents manuscrits numérisés issus de fonds d'archives de scientifiques provenant des collections de plusieurs bibliothèques de recherche en France²⁰⁹; ainsi que du projet « Testaments de Poilus », qui vise à produire l'édition électronique de testaments des Poilus de la Première Guerre mondiale conservés par les Archives nationales et les Archives départementales des Yvelines²¹⁰. Par ailleurs, on remarquera que la conception de ces plateformes et leur intégration impliquent de se confronter également aux questions d'ergonomie et de facilité d'accès des interfaces proposées, et de se demander comment les lecteurs et les chercheurs accèdent aux documents et aux jeux de données concernés²¹¹.

Ce sont surtout les évolutions de ces dernières années, le développement des Humanités numériques, des technologies et des langages du Web sémantique et de la culture de l'*Open Data*, qui semblent à la fois apporter un nouveau regard sur les bibliothèques numériques²¹² et contribuer à leur évolution. La tendance à opter pour la diffusion des contenus en Licence Ouverte, particulièrement développée dans les bibliothèques de l'ESRI (BNU de Strasbourg, BDIC, BIU Santé, BabordNum, et plus récemment, l'INHA) mais aussi en territoriale (BM de Lyon, Les Tablettes rennaises, La bibliothèque numérique du Limousin, etc.)²¹³ semble un des moteurs de l'innovation et de la mise à jour des plateformes, ainsi que le point de départ d'un meilleur dialogue avec les chercheurs²¹⁴.

Parmi ces moteurs d'innovation des bibliothèques numériques, signalons notamment le IIIF (*International Image Interoperability Framework*), une communauté internationale et un ensemble de spécifications techniques définissant un cadre d'interopérabilité pour la diffusion d'images haute résolution sur le Web. Partant du constat qu'aujourd'hui chaque institution développe son propre visualiseur d'images et que l'utilisateur est contraint de s'adapter à des environnements différents et parfois peu ergonomiques, ce projet vise à créer un cadre technique commun pour permettre l'interopérabilité des bibliothèques numériques et favoriser les interactions avec les usagers²¹⁵.

207. C'est le cas de *Rosalipédie*, dans le cadre de la bibliothèque numérique de Toulouse : <https://lc.cx/Q9fY>.

208. Pour un état de l'art en matière de *crowdsourcing*, voir : [Moirez *et al.*, 2013].

209. Voir : <http://transcrire.huma-num.fr/>.

210. Voir : <https://testaments-de-poilus.huma-num.fr>.

211. Pour quelques exemples, voir [Brault, 2016]; sur les interfaces de consultation et sur les questions de design dans le cadre des Humanités numériques, voir [Burdick, 2012] et [Vial, 2016].

212. Voir, par exemple, l'analyse de la « représentativité » de Gallica par rapport au catalogue de la BnF [Langlais, 2017], ainsi que les contributions de la journée d'études du 3 mai 2017 *L'évolution des usages des collections patrimoniales numérisées à travers l'exemple de Gallica*, <https://lc.cx/ASQt>.

213. Sur le domaine public dans les bibliothèques numériques patrimoniales, voir le mémoire DCB [Le Coz, 2017].

214. Concernant l'*Open Data* culturel, voir le rapport du MCC [Domange, 2013], et les réflexions de [Calimaq, 2017].

215. Les cas d'usage de l'IIIF, publiés sur le site de l'EquipEx Bibliissima, mettent en évidence comment cette solution technique peut avantager plusieurs chercheurs : <http://doc.bibliissima-condorcet.fr/introduction-iiif> : « Exemples de cas d'usage de IIIF : 1) un paléographe sou-

Les services d'accompagnement aux chercheurs et les infrastructures de recherche

Le chercheur n'est pas un simple usager de la bibliothèque : pouvant participer à la production des données et menant des activités avec les organismes de recherche et les institutions nationales et internationales, il joue un rôle actif dont les implications scientifiques, juridiques et financières doivent être prises en compte dans la mise en place d'un projet numérique. Il n'est pas anodin de rappeler que la nécessité de relever le défi numérique et d'améliorer les compétences numériques dans les bibliothèques académiques et de recherche européennes fait partie des préoccupations de LIBER (Ligue des Bibliothèques Européennes de Recherche), dont les axes stratégiques 2018-2022, comme le souligne Cécile Swiatek [Swiatek, 2017], se décomposent en trois directions principales visant à faire des bibliothèques des « plateformes pour une communication scientifique innovante », des « hubs de compétences et de services numériques », et des « partenaires de l'infrastructure de recherche »²¹⁶.

Des services d'accompagnement *ad hoc* ?

Comme cela est apparu clairement lors du dernier congrès de l'ADBU (Lille, 2017), les services aux chercheurs font actuellement l'objet d'une réflexion importante au sein de la communauté des professionnels des bibliothèques, notamment dans les bibliothèques universitaires et de recherche²¹⁷. Sans que ces services soient omniprésents (environ 40 % des établissements qui ont répondu à notre enquête disposent de services aux chercheurs), plusieurs institutions proposent des services dédiés, tels que, par exemple, des conditions particulières pour la réservation d'espaces et de documents, une aide à la diffusion en libre accès de la production scientifique, des formations aux doctorants et aux chercheurs concernant la recherche documentaire, la veille, la gestion de la bibliographie, etc.²¹⁸.

haitant inspecter et comparer les mains de plusieurs manuscrits issus de différents entrepôts numériques ; 2) un enseignant en art et architecture cherchant à composer une collection pédagogique d'images issues de plusieurs sources ; 3) un chercheur souhaitant annoter une carte numérisée disponible sur un site web, mais au moyen d'un outil d'annotation déployé sur un autre site web ; 3) un conservateur souhaitant exposer non pas seulement les métadonnées de ses collections numérisées mais aussi les images elles-mêmes à un agrégateur de contenus du domaine patrimonial ».

216. Voir notamment DARIAH et CLARIN, <http://libereurope.eu/strategy/digital-skills-services/digitalhumanities/>.

217. Sur les rapports que les enseignants-chercheurs entretiennent avec la bibliothèque universitaire, voir le mémoire de [Paret, 2012].

218. Pour ne citer que quelques exemples, voir les services de la BIS <https://www.biu.sorbonne.fr/biu/spip.php?rubrique132>, de l'Université de Versailles St-Quentin-en-Yvelines : <http://www.bib.uvsq.fr/services-aux-chercheurs>, de l'Université Grenoble Alpes <https://bibliotheques.univ-grenoble-alpes.fr/services/services-aux-chercheurs/services-aux-chercheurs-176168.kjsp> et de Sciences Po <http://www.sciencespo.fr/bibliotheque/fr/enseignants-chercheurs/enseignants-sciencespo>. Pour les formations, voir notamment l'Université de Lorraine : <https://bu.univ-lorraine.fr/formations/formations-doctorants-chercheurs> ; Université de Lille : <https://www.univ-lille3.fr/bibliotheques/chercheurs/> ; l'Université Paris 8 : <https://www.bu.univ-paris8.fr/services-aux-chercheurs> ; l'Université Lyon 2 : <https://www.univ-lyon2.fr/bibliotheques/formations/> ; voir aussi la section « Concevoir et rédiger des écrits scientifiques, produire et gérer des données de recherche » des Urfist de Bordeaux <http://weburfist.univ-bordeaux.fr/>.

Notre enquête montre que la gestion des demandes de numérisation est principalement rattachée soit au pôle de la numérisation, soit à celui du patrimoine, soit à celui des services aux chercheurs lorsqu'il existe. En fonction des moyens techniques à disposition et de la taille du corpus à numériser, les demandes peuvent être gérées en interne ou à travers un prestataire externe. Des missions et des ressources humaines peuvent à l'occasion être mises à disposition pour la gestion de corpus :

« Il ne s'agit pas d'un service à proprement parler, un certain nombre de collègues étant engagés à divers titres dans des projets collectifs de recherche (Université Paris Lumière, Labex *Les passés dans le présent* etc.). L'établissement emploie une chargée de mission pour la recherche, qui coordonne les projets de recherche auxquels l'établissement est lié et les publications de recherche (revue *Matériaux pour l'histoire de notre temps*, Collection « Sources et travaux de la BDIC » aux Presses universitaires de Nanterre (Actes de JE et colloques organisés par la BDIC, travaux de thèse réalisés à partir des fonds de l'établissement) » [BDIC]

Au-delà de la gestion des demandes de numérisation, qui ne sont pas toujours prioritaires, d'autres services sont impliqués.

« La numérisation de corpus documentaires n'est pas la priorité en regard des autres services que souhaitent les chercheurs » [Université de Franche-Comté, SCD]

Lorsqu'ils constituent des « produits de la recherche » (par exemple dans le cas des corpus numérisés enrichis), ces travaux nécessitent d'être évalués, à la fois pour participer aux appels à projet et dans le cadre de la progression de carrière des chercheurs. C'est pourquoi, parmi les services qui vont se développer, notamment dans les bibliothèques universitaires²¹⁹, figurent ceux qui concernent l'aide à la réalisation de plans de gestion de données²²⁰. En effet, dans le cadre des projets de recherche, il est demandé aux chercheurs d'explicitier dans le détail le cycle de vie des données qu'ils vont produire pendant leur projet : leur nature, leur statut juridique, leur licence, leurs modalités de pérennisation et de diffusion, etc. Dans la mesure où ces précisions permettent d'assurer la faisabilité des projets et leur pérennité à long terme, elles deviennent un atout important pour recevoir des financements.

« Les demandes de numérisation ne sont (de loin) pas le principal service que nous rendons quantitativement parlant. Parmi nos autres "services" aux chercheurs, nous intervenons dans l'aide à la conception de projets pour : indexation de corpus, conception de bases de données ou d'interfaces de diffusion de documents, aide au signalement et à l'identification des publications et données produites, et d'une façon générale tout ce qui touche à la diffusion des publications de recherche et à la gestion des données de recherche » [Université Bordeaux Montaigne]

Le développement des infrastructures de recherche

Dans le cadre d'un projet de recherche, des infrastructures nationales et internationales peuvent intervenir pour apporter une instrumentation avancée, des ressources humaines et financières, ainsi qu'une expertise scientifique et technique. De ce fait, elles

219. Pour la création de la bibliothèque numérique Anpersana de l'Université de Pau <https://anpersana.iker.univ-pau.fr/>.

220. Sur ce point, voir les dossier pour la préparation d'un plan de gestion de données réalisé par l'Université Paris Descartes <https://hal-descartes.archives-ouvertes.fr/page/data-management-plan>, ainsi que le diaporama [Puren, 2017].

peuvent figurer parmi les interlocuteurs des chercheurs et des professionnels dans les différentes phases d'un projet orienté sur la numérisation de corpus.

Les infrastructures internationales

Depuis sa création en 2002, le Forum Européen consacré aux infrastructures de Recherche (*European Strategic Forum for Research Infrastructures* ESFRI)²²¹ permet de coordonner les collaborations entre les différentes structures des États membres dans une volonté de rayonnement des projets et de partage de bonnes pratiques. Les Grandes infrastructures de recherche nationales sont répertoriées sur le site de l'ESFRI, dans le portail MERIL (*Mapping of the European Research Infrastructure Landscape*)²²², et participent de leur feuille de route nationale ainsi que de la feuille de route stratégique européenne (*Roadmap* ESFRI).

Dans le domaine des sciences humaines et sociales, des infrastructures internationales soutiennent le développement de la recherche, la diffusion de pratiques communes et l'utilisation de formats standard. C'est le cas de l'infrastructure DARIAH-EU (*Digital Research Infrastructure for the Arts and Humanities*)²²³, un Consortium Européen pour une infrastructure de recherche (ERIC)²²⁴ créé en novembre 2014 et rassemblant 17 pays membres, dont la France²²⁵; la phase de développement de DARIAH a été parallèlement soutenue par le projet HaS (*Humanities at Scale*)²²⁶ destiné à promouvoir la coopération des chercheurs à travers des initiatives et des services (aide à la recherche de financements, écoles d'été, échanges, etc.). Une autre infrastructure ERIC, CLARIN (*Common Language Resources and Technology Infrastructure*)²²⁷ est consacrée plus précisément au langage; elle met à disposition des outils de traitement qui permettent par exemple de baliser et annoter des corpus en ligne sur des serveurs décentralisés en garantissant la pérennité des données et des méthodes; elle soutient aussi la création de registres de catégories de données pour les concepts pertinents de la linguistique comme ISOcat, passé depuis 2015 au CCR (*CLARIN Concept Registry*)²²⁸. Construit autour des infrastructures DARIAH et CLARIN, PARTHENOS (*Pooling Activities, Resources and Tools for Heritage E-research Networking, Optimization and Synergies*)²²⁹ est un projet financé dans le cadre du programme Horizon 2020 de l'Union européenne, composé de 15 partenaires européens et visant à renforcer la cohésion de la recherche dans les secteurs des études linguistiques, des sciences humaines, du patrimoine culturel, de l'histoire, de l'archéologie et des autres domaines connexes à travers la mise en relation des infrastructures de la recherche, l'harmonisation des pratiques, la définition de normes communes et l'utilisation de formats standard.

221. Voir : <https://ec.europa.eu/research/infrastructures/index.cfm> et le programme Horizon2020 : <http://www.horizon2020.gouv.fr/cid72594/les-infrastructures-recherche.html>.

222. Voir : <https://portal.meril.eu/meril/>.

223. Voir : <https://www.dariah.eu/>.

224. Pour le cadre juridique, voir la fiche sur EUR-Lex : <http://eur-lex.europa.eu/legal-content/FR/TXT/?uri=LEGISSUM:ri0005>.

225. Pour une présentation de la participation de la France dans DARIAH, voir : <https://humanum.hypotheses.org/2675>.

226. Voir : <http://has.dariah.eu/>.

227. Voir : <https://www.clarin.eu/>.

228. Voir : <http://www.datcatinfo.net/> ainsi que le *Concept Registry Browser* de CLARIN <https://concepts.clarin.eu/ccr/browser/>.

229. Voir <http://www.parthenos-project.eu/>.

En partenariat avec plusieurs institutions et projets européens (APEF, CLARIN, Europeana, E-RIHS ; Iperion-CH, PARTHENOS), l'équipe DARIAH travaille actuellement à la *Cultural Heritage Data Reuse Charter*, une charte visant à simplifier les interactions entre institutions patrimoniales, chercheurs, hébergeurs de données, et équipements scientifiques, autour de la réutilisation des données issues des institutions patrimoniales (photographies, modèles 3D, transcriptions, bases de données, etc.)²³⁰.

Les infrastructures nationales

À l'échelle française, la feuille de route du MENESR pour les infrastructures de recherche, mise à jour en 2016, se décline en quatre modèles : les organisations internationales (OI), les très grandes infrastructures de recherche (TGIR), les infrastructures de recherche (IR) et les projets [MENESR, 2016, p. 7]. Accompagnée par un comité de pilotage numérique, dont une des composantes est la BSN (Bibliothèque scientifique numérique)²³¹, et par la loi pour une République numérique²³², « la politique nationale de l'information scientifique et technique est fondée sur la volonté d'ouvrir l'accès aux résultats de la recherche » [Mazens, 2017]²³³. Cette feuille de route, constituant un outil de pilotage stratégique de l'État pour structurer le paysage de la recherche, met en avant l'attention portée au traitement de corpus dans le domaine des sciences humaines et sociales :

« Les infrastructures de recherche en SHS doivent permettre de constituer et de manipuler des corpus volumineux et très hétérogènes, de nature qualitative ou quantitative, susceptibles d'ouvrir de nouvelles voies de recherche et de favoriser l'interdisciplinarité » [MENESR, 2016, p. 12]

Plusieurs infrastructures composent la branche française dans les domaines des Sciences humaines et sociales et de l'Information scientifique et technique. La TGIR Huma-Num (Humanités Numériques)²³⁴, née en mars 2013 de la fusion du TGE (Très Grand Equipement) Adonis et du Corpus IR (Consortium CAHIER). Soutenue par l'Alliance Athena, accompagne les équipes de recherche de l'ESRI en mettant à leur disposition « des outils et services de stockage, de traitement, d'interopérabilité, de diffusion, d'exposition, de signalisation et d'archivage de données numériques ». La TGIR Progedo (PROduction et GEstion de DONnées)²³⁵, a pour but principal, quant à elle de promouvoir et de structurer une politique publique des données, notamment pour la recherche en droit, économie, géographie, gestion, histoire, sciences politiques et sociologie.

L'IR RnMSH (Réseau national Maison des Sciences de l'Homme)²³⁶, a, entre autres, une mission de fédération des projets des 23 MSH sur le territoire national et de déploiement-relais des dispositifs nationaux des TGIR auprès des collectivités territoriales.

230. Sur la charte, voir [Romary *et al.*, 2016] et [Baillot *et al.*, 2017].

231. Voir : <http://www.bibliothequescientificquenumerique.fr/presentation/> et notamment le volet n° 5 portant sur la numérisation.

232. LOI n° 2016-1321 du 7 octobre 2016 pour une République numérique, voir le site Legifrance <https://lc.cx/QRNR>.

233. Voir aussi : <http://www.enseignementsup-recherche.gouv.fr/pid25384/strategie-nationale-des-infrastructures-de-recherche.html>.

234. Voir : <https://www.huma-num.fr/>.

235. Voir : <http://www.progedo.fr/>.

236. Voir : <http://www.msh-reseau.fr/>.

Le projet ERIHS-FR2 (*European Research Infrastructure for Heritage Science*)²³⁷, est une infrastructure de recherche dédiée à l'étude des matériaux du patrimoine culturel et naturel (collections des musées et muséums, ensembles monumentaux, sites archéologiques, fonds d'archives, de bibliothèques, etc.).

L'IR OpenEdition (Édition électronique ouverte en Sciences humaines et sociales)²³⁸ encourage la diffusion en libre accès de la production scientifique éditée. À cette dernière infrastructure s'ajoutent les trois autres plateformes de l'IST : HAL²³⁹ pour la gestion des pré-prints et de l'archivage de la production scientifique en libre accès, NUMEDIF²⁴⁰ pour la création d'une plateforme mutualisée de diffusion-distribution multicanaux dans le domaine de l'édition, et COLLEX-Persée²⁴¹ pour l'identification et la consolidation de grandes collections scientifiques accessibles en ligne, ainsi que pour la mise en œuvre d'une politique nationale de numérisation pour l'ESRI. Signalons que, via Collex-Persée, un questionnaire a été diffusé à la rentrée 2017 par le Service de la Coordination des Stratégies de l'Enseignement Supérieur et de la Recherche afin de « disposer d'un panorama des programmes de numérisation réalisés sur les dix dernières années, d'effectuer un bilan de la diffusion des ressources numérisées et d'identifier les besoins des porteurs de projets et des chercheurs en termes de services et d'accompagnement »²⁴².

237. Voir : <http://www.erihs.fr/>.

238. Voir : <https://www.openedition.org/>.

239. Voir : <https://hal.archives-ouvertes.fr/>.

240. Voir : <http://www.numedif.fr/>.

241. Voir : <http://www.collex.eu/>.

242. L'introduction du questionnaire précise que : « Un programme de numérisation est compris comme une démarche scientifique coordonnée de reproduction numérique de documents, concernant un ou plusieurs corpus, dotée de moyens humains et financiers dédiés, et ayant produit des résultats. Les programmes de numérisation sont souvent dotés d'une logique thématique (ex : « Les sources du droit ») ; ils peuvent aussi être réalisés dans le cadre d'un projet de recherche (ex : opération de numérisation menée par un laboratoire dans le cadre d'un projet ANR) » - Enquête sur les programmes de numérisation au service de la recherche (2006-2016) : <https://lc.cx/QRYG>.

Autour de la numérisation de corpus : la feuille de route de la coopération

DE L'INITIATIVE À LA MISE EN PLACE DE LA COOPÉRATION

« La mise en œuvre de ces missions nécessite la prise en compte des contraintes intellectuelles, techniques, juridiques liées au statut du document numérique » [Science Po]

En suivant le fil des questions que nous avons proposées dans nos deux enquêtes, ce chapitre s'articule autour de la feuille de route d'un hypothétique projet de collaboration entre chercheurs et bibliothécaires autour de la numérisation de corpus. Bien qu'il soit difficile de prévoir une démarche qui s'adapte aux différents contextes et besoins, nous essaierons de mettre en avant les éléments principaux à prendre en compte dans le but d'établir une coopération fructueuse²⁴³.

D'où vient la demande de numérisation ? À quel moment ?

À la question n° 1,

« Dans le cadre de vos activités de recherche, avez-vous déjà sollicité une bibliothèque pour numériser un corpus de documents ? Si oui, combien de fois ? »

une bonne moitié des chercheurs (54,4 %) nous a répondu par la négative.

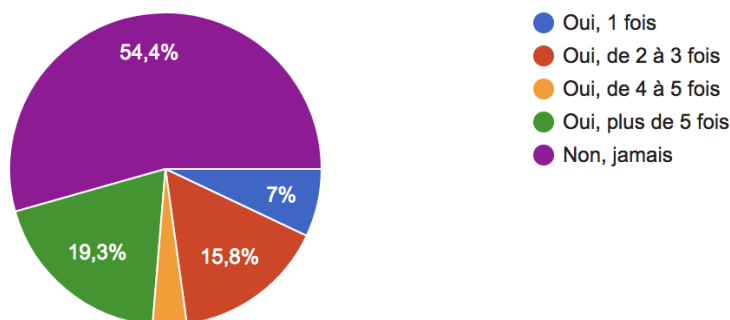


FIGURE 7 – Questionnaire aux chercheurs - question n° 1

En prenant en compte les explications qu'ils fournissent peu après, concernant leur éventuelle participation à un projet qui s'appuie sur un corpus numérisé (question n° 22), on observe que si 8 chercheurs n'ont pas besoin de corpus pour leurs travaux ou recherches, la plupart d'entre eux (20 sur 31), sans être à l'origine de l'initiative, collaborent dans des projets dont l'initiative de la numérisation dépend d'une bibliothèque ou d'autres intervenants ; ainsi, environ 82 % parmi les chercheurs répondants étaient impliqués, à différents titres, dans l'étude d'un corpus numérisé.

De leur côté, les bibliothèques, surtout en l'absence d'une coopération établie, ne distinguent pas forcément les demandes de numérisation des chercheurs de celles d'autres particuliers, la répartition se faisant le plus souvent entre « particuliers » et « professionnels ». Signalons d'ailleurs que les chercheurs ne constituent pas les seuls demandeurs de

243. Signalons que des guides de bonnes pratiques pour la numérisation de corpus sont disponibles, notamment celle préparé par le Mutec [Ingarao et Saïdi, 2010].

	Question n° 1 - Initiative	Question n° 22 - Collaboration
Bibliothécaire		1
Chargé de collection	1	
Chargé de recherche		3
Chercheur post-doctoral	4	2
Conservateur		1
Directeur de recherche	1	1
Enseignant du secondaire	1	
Ingénieur d'études	1	1
Ingénieur de recherche	2	3
IUF	2	
Maître de conférence	5	3
Professeur émérite		1
Professeur des universités	9	4
Retraitée	1	

FIGURE 8 – Questionnaire aux chercheurs - Synthèse de la participation aux projets de numérisation de corpus

corpus : la BnF reçoit de nombreuses commandes de la part de sociétés commerciales, effectuées par exemple par un éditeur et destinées au grand public, comme récemment dans le cas d'une édition en fac-similé de certains journaux de la Grande Guerre. Par conséquent, l'initiative des chercheurs ne ressort pas de manière évidente des résultats quantitatifs de notre enquête.

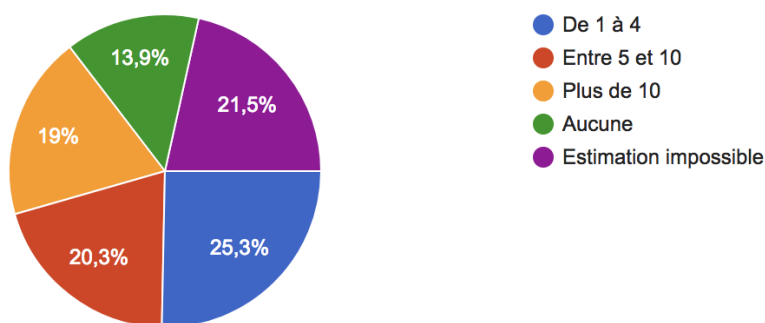


FIGURE 9 – Questionnaire aux bibliothécaires - question n° 7

Néanmoins, la situation devient plus claire quand on examine les retours que les bibliothécaires ont ajoutés pour préciser l'origine des initiatives :

« Étudiants/chercheurs, éditeurs, particuliers. [...] Il existe un service de numérisation à la demande, mais les corpus de numérisation sont rarement sollicités par les chercheurs, ils correspondent à la politique d'établissement » [BM de Bordeaux]

« La sélection des corpus émane avant tout de la bibliothèque sans concertation avec les acteurs de la recherche. Il arrive parfois que certains documents isolés soient numérisés à la demande de chercheurs. Mais dans l'absolu les corpus numérisés répondent à une demande qui n'est pas forcément exprimée » [BM de Dijon]

« Définition d'un corpus par un chercheur (e.g. mss. musicaux médiévaux); définition d'un corpus en partenariat entre la BM et un ou des chercheurs (e.g. projet BHLi avec le CESR ou projet des Premiers récits imprimés des Guerres de Vendée au sein du comité scientifique d'une exposition); définition d'un corpus directement par la BM à l'adresse d'une communauté de chercheurs éclatée (correspondance David d'Angers sollicitée par plusieurs chercheurs parfois un peu rivaux) » [BM d'Angers]

« Souvent à la sollicitation du chercheur, la BNU a répondu à l'attente. Pour certains projets, l'initiative est venue de la BNU directement » [BNU]

« Comme expliqué plus tôt, l'initiative vient des laboratoires et donc in fine, des chercheurs. Il n'y a pas de politique de numérisation par ailleurs » [Université Sorbonne Paris 4, SCD²⁴⁴]

« La demande de numérisation vient directement des chercheurs qui voulaient mettre en ligne des ouvrages anciens de leur bibliothèque. Le schéma est plutôt inverse à ce que vous proposez dans vos questions » [Université Paris-Sud]

D'après l'ensemble des réponses, il apparaît que l'initiative se répartit surtout entre bibliothèques territoriales, dont les projets relèvent principalement du projet d'établissement, et chercheurs dans le contexte des bibliothèques de l'Enseignement supérieur et de la Recherche. Bien sûr, nous avons également enregistré des initiatives de la part des bibliothèques universitaires (portant surtout sur l'histoire de l'université et sur les thèses anciennes), mais en moindre quantité. Concernant les initiatives provenant des chercheurs, signalons que les demandes autour des corpus ne touchent pas uniquement à la numérisation des collections propres de la bibliothèque sollicitée : en particulier quand ils s'adressent aux bibliothèques relevant de l'ESRI, les chercheurs peuvent poser des questions sur des données que la bibliothèque ne possède pas, mais sur lesquelles ils souhaitent travailler. C'est ainsi que ces initiatives s'insèrent dans une demande d'expertise plus large, que nous avons déjà évoquée précédemment (voir p. 46) ; dans ce cas, la bibliothèque est interpellée et associée à l'équipe projet pour accompagner les chercheurs sur les questions touchant à l'indexation, à la rédaction de cahiers des charges concernant la numérisation et la diffusion des documents, à la recherche de documents numérisés préexistants afin de rassembler une collection numérisée pertinente. S'inscrivent dans cette demande d'expertise sur des collections conservées ailleurs les collaborations menées par les SCD de l'Université Bordeaux Montaigne, dans le cadre des projets « Mauriac en ligne »²⁴⁵ et « MédiaBd »²⁴⁶. En outre, comme l'indique Julien Baudry [Baudry, 2017], la bibliothèque peut également être sollicitée par des « chercheurs qui ont eux-mêmes des fonds, notamment les fonds documentaires d'un laboratoire ou d'un centre de recherche, et qui s'interrogent sur la manière d'en faire une diffusion numérique » : c'est notamment le cas du fonds de photographies de monuments et d'œuvres d'art du Moyen-Âge appartenant au professeur Jacques Lacoste²⁴⁷.

Quelle que soit l'origine de l'initiative, il est opportun qu'elle soit préparée suffisamment à l'avance, tant du côté des chercheurs que des professionnels. Il conviendra de considérer, dès la conception du projet, quels sont les objectifs visés : conservation, diffusion, exploitation, etc. Une phase de réflexion initiale conditionnera alors l'orientation

244. En janvier 2018, l'université Paris-Sorbonne a depuis fusionné avec l'UPMC pour devenir Sorbonne-Université.

245. Voir : <http://mauriac-en-ligne.u-bordeaux-montaigne.fr/>.

246. Voir : <http://collections.citebd.org/in/faces/homeInBook.xhtml>.

247. Voir : <http://1886.u-bordeaux-montaigne.fr/collections/show/13>.

générale et l'ampleur du projet, en permettant d'estimer les ressources nécessaires en termes financiers, humains et techniques.

D'une part, en répondant à la question n° 4,

« À quel moment de votre projet avez-vous sollicité la collaboration d'un professionnel des bibliothèques ? »

les chercheurs qui sont à l'initiative d'un projet de numérisation nous disent en grande majorité demander la collaboration d'un bibliothécaire dès la conception du projet.

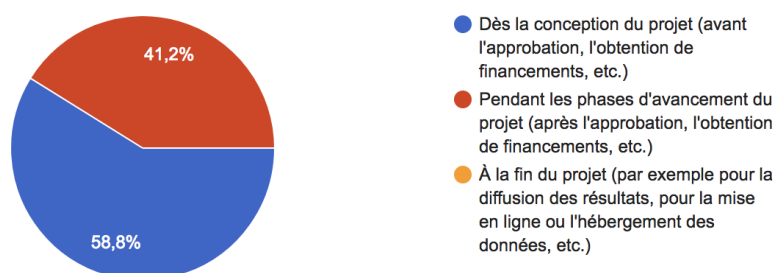


FIGURE 10 – Questionnaire aux chercheurs - question n° 4

De l'autre, plusieurs professionnels remarquent le contraire, et soulignent l'importance d'associer promptement les bibliothèques, ce qui permettrait notamment de mieux évaluer les coûts du projet et de tenir compte des aspects de préservation numérique et physique des documents :

- « 1. Associer les bibliothèques dès la conception des projets scientifiques ;
- 2. Intégrer la dimension "coûts de stockage et préservation des données" dans le budget du projet, et au-delà de la durée du projet ;
- 3. Ne pas oublier la dimension patrimoniale : certains corpus documentaires partent en miette ; il est tout aussi urgent de les numériser pour les préserver que de participer à des projets scientifiques » [BNU]

Quel mode projet ?

Concernant les modalités d'idéation et de mise en route du projet, tant les chercheurs que les bibliothécaires signalent l'importance de réunions et d'échanges. Sans pouvoir tirer d'informations détaillées quant à l'application d'une méthode ou d'une démarche projet spécifique²⁴⁸, nous observons qu'en fonction de la taille et de la nature des projets, ces modes de collaboration peuvent prendre la forme de réunions informelles, de conseils scientifiques, de partenariats, de conventions, de réponses à des appels à projet avec ou non la création d'un comité de pilotage :

« Pour chaque collection thématique, nous constituons un comité scientifique composé de scientifiques et spécialistes du thème, ainsi que de deux personnes de l'équipe de la bibliothèque. Nous organisons des réunions de travail à travers ces

248. Dans le cas des institutions relevant des collectivités territoriales, signalons le volet scientifique du PCSES (Projet culturel, scientifique, éducatif, et social), dispositif visant à développer une « démarche projet dans les bibliothèques », dont un guide à la rédaction a été récemment publié par le MCC [MCC, 2017] ; sur ce sujet, voir aussi le mémoire DCB [Morel, 2017].

comités scientifiques, au cours desquelles nous travaillons sur les bornes temporelles de la collection (si besoin), les contenus (les différents thèmes), ainsi que sur les collections de documents qui y rentreront (prise en compte des droits d'auteurs, etc.). Certains nous aident également à la construction de petites expositions virtuelles [...] » [Bibliothèque du Patrimoine de Clermont Auvergne Métropole] »

« Échanges physiques ou courriels, création de conseils scientifiques spécifiques, rédactions d'articles ou de billets de blog » [MMSH Phonothèque (AMU - CNRS)]

« Définition conjointe bibliothécaires/chercheurs d'une politique scientifique de développement de la collection numérique [...] » [Université de Lille 1, Sciences et technologies]

« Des collaborations plus rapprochées (expertise, choix de corpus en fonction de projets de recherche) sont prévues pour les prochains corpus et certaines pistes ont déjà été ébauchées avec le directeur d'un labo pour la numérisation d'archives en archéologie (fonds Burnouf). La politique de numérisation, et le positionnement des BU Lorraines par rapport aux chercheurs sur ce sujet, sont en cours de renouvellement profond depuis février 2017. Les nouvelles modalités envisagées sont encore à l'état de travail. Cette partie ne relève pas du Réseau d'appui à la recherche » [Université de Lorraine]

Si parfois la collaboration est simplement évoquée pour se garantir l'accès aux documents, les chercheurs nous expliquent que les tâches « ont été distribuées en fonction des compétences de chaque équipe » ; ils soulignent également l'importance de l'accompagnement de la part des professionnels :

« Pour le moment, j'ai vu deux fois les conservateurs, au moment de monter mon dossier de financement, pour pouvoir justifier de leur collaboration. Ils m'ont aussi averti des écueils possibles, des projets déjà en cours, de leurs propres entreprises en lien avec leur bibliothèque numérique *** » [MCF]

« Partenariat avec l'établissement. Le responsable du fond ancien est impliqué dans toutes les étapes. De la conception du projet jusqu'à la mise en ligne » [IGE]

« Plusieurs réunions en amont du projet et dans la phase d'élaboration. Les collaborateurs des bibliothèques avaient un rôle de proposition et de relecture pour correction des *draft* du projet. Les réunions avaient lieu surtout en phase d'élaboration pour la discussion collective. Ensuite, le travail se faisait par voie électronique et une réunion finale de validation était ensuite organisée » [PU/IUF]

« Accompagnement sur la création du fichier de récolement, accompagnement sur le cahier des charges de numérisation et intermédiaire avec le prestataire privé en charge de la numérisation, accompagnement sur le choix du CMS (Omeka). » [IGR]

La question des droits

« Tout projet culturel ou scientifique produisant des données et/ou des contenus doit s'interroger sur les conditions juridiques de la mise à disposition de ces objets, sous peine que ces questions ne se posent ensuite *a posteriori*, en provoquant alors souvent difficultés et blocages pour ne pas avoir été suffisamment anticipées » [Maurel, 2017b, p. 179].

En prenant l'exemple d'un « projet type », portant sur la numérisation d'un corpus de livres anciens et sur sa valorisation via une édition en ligne, ce projet amènera à la

production de données et de métadonnées différentes. Il s'agira de fichiers d'images réalisés à partir des ouvrages, de fichiers OCR bruts et corrigés, de métadonnées bibliographiques associées aux documents, de fichiers d'encodage TEI des textes numérisés, de contenus destinés à la diffusion du projet sur un site Web, de contenus scientifiques pour des revues, etc. Quels droits sont-ils associés à ces contenus ? La phase de préparation d'un projet de collaboration sera l'occasion de préciser les droits associés aux différents contenus, de manière à équilibrer le droit des auteurs et les droits de la collectivité en terme de liberté d'expression, de créativité et de recherche, conformément au Code de la propriété intellectuelle²⁴⁹. Il faudra également réfléchir aux modalités d'expression des droits, notamment pour les données produites par les chercheurs, de façon à ce qu'elles puissent être citées et évaluées dans leur production scientifique. En effet, l'une des questions qui se posent est notamment celle de la « redocumentarisation ». Comment peut-on gérer la question des droits par exemple pour les chercheurs/ingénieurs qui réalisent des données sur des données culturelles dont ils n'ont pas la propriété ? Par exemple, lorsque des chercheurs renseignent des données de provenance à propos des documents d'un corpus. Comment peut-on faire communiquer les données de la recherche et les données des collections tout en leur reconnaissant leurs droits respectifs ?

Ces nombreux points de vigilance expliquent que, dans les projets numériques, les questions juridiques suscitent l'inquiétude de part et d'autre. La numérisation de contenus impose aux institutions culturelles de faire face à des choix difficiles en matière juridique. Dans la chaîne de numérisation, les documents sous droits doivent être identifiés et signalés, afin de pouvoir procéder au marquage ou à la modification de la qualité de l'image ; les conditions de diffusion, duplication et impression doivent également être précisées [Maillet, 2011, p. 130-1]. Si certains documents et données doivent faire l'objet de limitations, d'autres peuvent être diffusés en leur associant des licences adaptées ; les conditions de réutilisation, de modification et de diffusion des données ont des retombées importantes dans le cadre des projets s'appuyant sur le modèle de l'*Open Data* ou du *Linked Open Data*²⁵⁰, et, plus généralement, sur la capacité à valoriser les collections et à toucher des usagers potentiels²⁵¹.

Comment faire concernant des corpus documentaires composés de documents récents, encore soumis au droit d'auteur ? Sous quelles conditions les chercheurs peuvent-ils avoir accès à ces corpus ? Quel est le régime juridique appliqué en France pour la fouille de textes (*TDM Text and Data Mining*) ? Si, encore aujourd'hui, « le droit d'auteur entrave les *digital humanities* » [Marino, 2015, p. 66], l'introduction d'une exception autorisant la fouille de textes et de données, promulguée le 7 octobre 2016 par la « loi pour une

249. Code de la propriété intellectuelle (CPI) : <https://www.legifrance.gouv.fr/affichCode.do?cidTexte=LEGITEXT000006069414> ; voir notamment l'art. L. 122-5 pour l'exception de courte citation, l'exception pédagogique et l'exception de conservation et consultation ; sur ces exceptions, voir aussi [Ego, 2015] et [Dupuis, 2015].

250. Sur ce sujet, voir [Bauer et Kaltenböck, 2012] ainsi que la présentation [Courtin, 2014].

251. Les Humanités numériques s'inscrivent dans le débat juridique concernant la propriété intellectuelle : « Face à l'idée d'une science réservée dont les travaux et les résultats devraient être de plus en plus protégés et brevetés au nom du respect de l'auctorialité, ils opposent l'idée d'une science qui pourrait s'avérer nettement plus ouverte et collaborative en se détachant du principe de paternité des œuvres (*authorship*) » [Granjon et Magis, 2016]. En 2015, une analyse sur comment constituer et exploiter des corpus numériques sans porter atteinte aux droits de propriété littéraire et artistique a été effectuée par le CERDI (Centre d'Études et de Recherche en Droit de l'Immatériel) et par le LabEx OBVIL ; signalons en particulier les contributions de [Bensamoun et Latil, 2015] et [Léger, 2015].

République numérique »²⁵², constitue une avancée majeure. Cette exception doit respecter trois contraintes : la fouille de données ne peut s'effectuer que sur un corpus dont les contenus ont été acquis de façon licite ; elle ne peut être pratiquée que pour les besoins de la recherche publique ; elle exclut toute finalité commerciale. Les conclusions de l'étude comparative pilotée par l'ADBU sur la compétitivité de la recherche en France et au Royaume-Uni soulignent que le TDM « favorise la diffusion de la performance de la recherche scientifique » ; en fournissant des recommandations à la communauté des bibliothèques, l'étude insiste sur le fait que cette pratique nécessite d'être accompagnée par de nombreuses améliorations, notamment en termes d'infrastructures, de compétences et de mesures de sensibilisation²⁵³ :

« Le TDM requiert un bon niveau de culture numérique. Les experts de la fouille de textes, les départements informatiques et les bibliothèques peuvent tous jouer un rôle dans la montée en compétence des chercheurs » [Research Consulting, 2017]

Les bibliothèques, et particulièrement la Bibliothèque nationale, sont ainsi sollicitées pour le développement de services de soutien et pour la mise en place de mécanismes de suivi des expériences menées par les chercheurs. En effet, si cette exception constitue une ouverture, la fouille de contenus doit toujours être faite dans le respect du droit d'auteur, ce qui peut impliquer des dispositifs de verrouillage technique des données, ou le fait qu'elles ne peuvent être consultées qu'à l'intérieur de l'institution, comme dans le cas de Gallica *intra muros*. D'ailleurs, si des restrictions juridiques empêchent les chercheurs d'accéder à ou de créer des copies de certains corpus, une réflexion s'impose sur les infrastructures que les bibliothèques peuvent mettre à disposition.

L'établissement de la convention

La phase de conception du projet est normalement suivie par la mise en place d'un accord formel. L'examen de quelques conventions de collaboration déjà existantes peut se révéler utile pour préciser les grands aspects à ne pas oublier avant de démarrer un nouveau projet. En prenant comme point de départ la convention établie entre la ville de Grenoble et l'Université Stendhal (équipes Traverses 19-21 EA3748 et LIDILEM EA609) au sujet de la numérisation de la consultation de la base documentaire des manuscrits de Stendhal [Grenoble, 2008], et la convention de coopération de recherche entre

252. LOI n° 2016-1321 du 7 octobre 2016 pour une République numérique, art. 38 : « Le code de la propriété intellectuelle est ainsi modifié : 1° Après le second alinéa du 9° de l'article L. 122-5, il est inséré un 10° ainsi rédigé : « 10° Les copies ou reproductions numériques réalisées à partir d'une source licite, en vue de l'exploration de textes et de données incluses ou associées aux écrits scientifiques pour les besoins de la recherche publique, à l'exclusion de toute finalité commerciale. Un décret fixe les conditions dans lesquelles l'exploration des textes et des données est mise en œuvre, ainsi que les modalités de conservation et de communication des fichiers produits au terme des activités de recherche pour lesquelles elles ont été produites ; ces fichiers constituent des données de la recherche ; » 2° Après le 4° de l'article L. 342-3, il est inséré un 5° ainsi rédigé : « 5° Les copies ou reproductions numériques de la base réalisées par une personne qui y a licitement accès, en vue de fouilles de textes et de données incluses ou associées aux écrits scientifiques dans un cadre de recherche, à l'exclusion de toute finalité commerciale. La conservation et la communication des copies techniques issues des traitements, au terme des activités de recherche pour lesquelles elles ont été produites, sont assurées par des organismes désignés par décret. Les autres copies ou reproductions sont détruites ».

253. Sur la TDM et le rôle des bibliothèques, voir le billet de blog de Catherine Muller [Muller, 2017] et [Oudenhoven, 2017].

la ville et l'Université de Rouen concernant le « Programme d'étude *Madame Bovary* » [Rouen, 2003], nous pourrions articuler une convention sur les axes suivants :

1. la description générale du programme de collaboration (sujet, durée, participants, moyens, etc.);
2. l'indication des responsables scientifiques, répartis en fonction du projet entre chercheurs, professionnels de l'information et des bibliothèques ;
3. la fréquence des réunions, les modalités de documentation du projet et la nature des livrables ;
4. les caractéristiques de la base documentaire : l'endroit où seront stockées les images et leurs copies de sécurité (« La bibliothèque d'images sera hébergée sur un serveur de la ville de Grenoble ; la base documentaire et les données de transcription seront hébergées sur un serveur de l'Université » [Grenoble, 2008]) ; l'institution en charge de la gestion et de la maintenance du site Internet ; les contenus du site, avec des indications sur la qualité des images, les formats, le type de métadonnées, les champs d'information documentaire et le public visé ;
5. les conditions de propriété matérielle et intellectuelle des documents et des données associées (par exemple, pour les images, pour les transcriptions, pour les notices descriptives), ainsi que les mentions à indiquer pour préciser les droits ; on veillera également à préciser les conditions d'exploitation et de réutilisation des différents contenus numériques ;
6. les conditions de communication et de publication : les participants à la convention devront être indiqués et, de leur côté, ils doivent s'engager « à mentionner clairement leur collaboration dans leurs publications et communications scientifiques et de vulgarisation » [Grenoble, 2008] ; on pourra aussi signaler les éventuelles conditions et tarifs pour l'exploitation des images de la part des éditeurs ou des acteurs commerciaux ;
7. les engagements mutuels de confidentialité concernant les informations scientifiques, techniques ou économiques des partenaires, ainsi que la durée de ces engagements ;
8. les conditions d'exploitation des résultats, « chacune des parties conservera l'intégralité des droits d'exploitation pour les éléments dont elle a eu la responsabilité de réalisation, programmes informatiques compris » [Rouen, 2003] ;
9. les termes de validité de la convention : la date de prise d'effet, les modalités de reconduction et de modification ;
10. les conditions de résiliation en cas de divergences et de litiges : les modalités de résolution des divergences (à l'amiable ou à travers le tribunal en fonction de la persistance du désaccord), ainsi que les dispositions qui resteront valides même en cas de résolution de la convention ;
11. les conditions pour apporter des modifications ou des adjonctions à la convention.

Des annexes pourront également préciser le plan de recherche, le programme des travaux et les conséquences attendues au plan scientifique, technique, etc.

DU DÉROULEMENT À LA VALORISATION DU PROJET

Une fois que les modalités de travail et de collaboration ont été définies, un projet de numérisation de corpus documentaire induit une série de choix et de tâches scientifiques et techniques qui articulent la chaîne de constitution du corpus et qui préparent sa diffusion et sa valorisation. Si, à plusieurs reprises, les bibliothécaires soulignent l'importance de mener des collaborations, car « tout projet mené en collaboration avec les chercheurs a plus de chance d'aboutir », cette forte demande d'implication ne semble pas correspondre au sentiment qui émerge d'après les réponses des chercheurs ; s'agit-il, comme certains nous ont fait remarquer, d'un manque de communication autour des possibles dispositifs de collaboration ? Quoi qu'il en soit, quelle est l'effective implication des chercheurs ? Comment les chercheurs et les professionnels se répartissent les rôles et les tâches dans un projet de numérisation de corpus ?

Les choix documentaires et scientifiques

À la question n° 17,

« Depuis 2012, votre institution a-t-elle sollicité des chercheurs pour une expertise technique ou scientifique dans le cadre de nouvelles numérisations de corpus documentaires ou pour des corpus documentaires précédemment numérisés ? »

environ 70 % des professionnels de bibliothèque qui ont répondu à notre enquête déclare avoir demandé la collaboration d'un chercheur, pour des corpus déjà numérisés mais aussi pour de nouvelles numérisations :

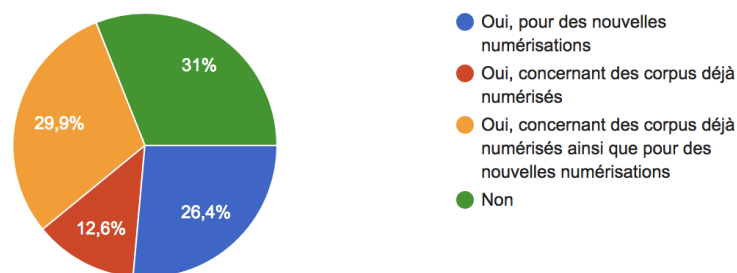


FIGURE 11 – Questionnaire aux bibliothécaires - question n° 17

Contre toute attente, confrontés à la une question analogue (question n° 28),

« Avez-vous déjà été sollicité(e) par une bibliothèque pour une expertise scientifique ou technique à propos d'un corpus de documents numérisé ou à numériser ? »

la plupart des chercheurs répond par la négative (71,9 %) : seulement une dizaine a été sollicitée pour une expertise scientifique, cinq pour une expertise scientifique et technique, un seul pour une expertise technique.

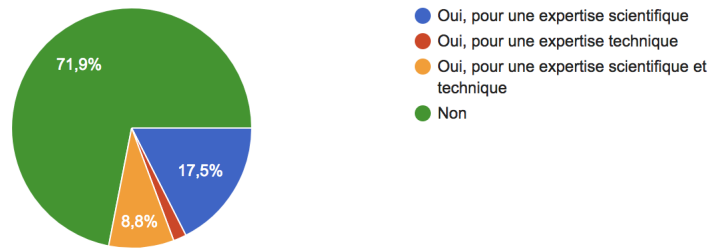


FIGURE 12 – Questionnaire aux chercheurs - question n° 28

En revanche, les deux groupes s'accordent quant aux sujets d'intervention des chercheurs : de part et d'autre, les cas les plus souvent évoqués concernent surtout la sélection des documents à numériser, la création des métadonnées correspondantes, la valorisation à travers un travail de recherche (article, thèse, conférence, etc.) et la collaboration dans le cadre de la préparation d'expositions :

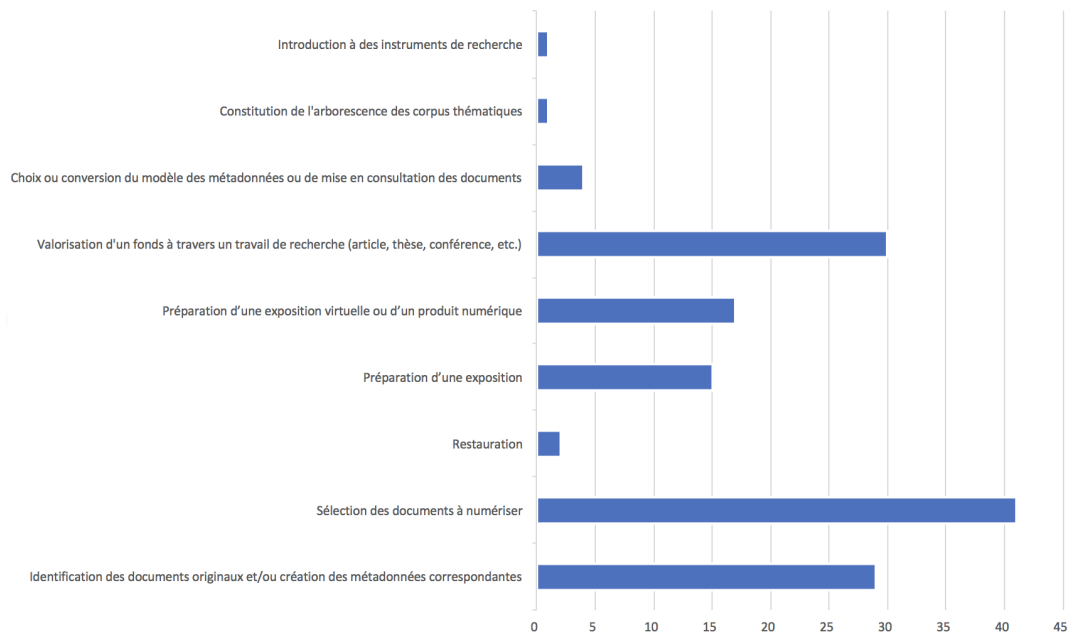


FIGURE 13 – Questionnaire aux bibliothécaires - question n° 18

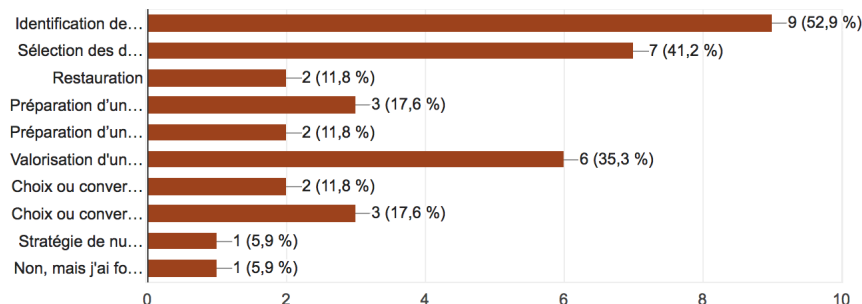


FIGURE 14 – Questionnaire aux chercheurs - question n° 29

La sélection des documents, le catalogage et l'enrichissement des notices

Dans le cadre d'une coopération autour d'un corpus encore à numériser, les chercheurs sont le plus souvent sollicités pour participer d'abord à la sélection des documents, et, dans un deuxième temps, à l'enrichissement des notices descriptives :

« Le programme principal de numérisation de la bibliothèque repose sur une sélection effectuée par deux enseignants-chercheurs en histoire du droit. La sélection date de 2007, mais les presque 900 000 pages correspondantes n'ont pas encore été complètement numérisées. Par ailleurs, la bibliothèque a entrepris un travail de refonte de l'interface de consultation de sa bibliothèque numérique, avec l'intégration d'une éditorialisation. Dans cette optique, les mêmes enseignants-chercheurs ont été de nouveau sollicités [...] » [BIU Cujas]

« Dans le cadre de la numérisation des manuscrits médiévaux, un musicologue a piloté le projet en lien avec le conservateur de la bibliothèque, les manuscrits sélectionnés étant tous des manuscrits notés du Moyen Âge. Ensuite, plusieurs chercheurs ont été sollicités selon leurs compétences respectives pour rédiger les notices complètes des manuscrits. Certains ont également publié des articles ou donné des conférences à ce sujet » [Bibliothèque du Grand Verdun]

« Sur le projet de recherche en cours : travail collaboratif de sélection des documents et de conception du site. Travail sur l'enrichissement des contenus, etc. » [Université Nice Sophia Antipolis, SCD]

Si le document ne dispose pas encore d'une notice, l'étape initiale de signalement sera suivie par des travaux de catalogage, dans lesquels, pour les professionnels, la valeur ajoutée des chercheurs est évidente :

« La valeur ajoutée de la mise en ligne de ce corpus en histoire provient de l'implication des chercheurs dans le catalogage des documents et dans la création des tables de matières servant à naviguer dans ces documents non océrisables » [Université Paris-Sud]

Cette étape, dont la charge n'est pas à sous-estimer (surtout pour les fonds décrits à la pièce et non pas par recueil) est à inscrire dans le planning de réalisation du projet. Soulignons qu'il s'agit d'un préalable indispensable à la numérisation, car tout document numérisé sans avoir été identifié par une notice « est un document inaccessible, donc inexploitable et, par conséquent, perdu » [Maillet, 2011, p. 131].

« Il faudrait toujours associer la numérisation à une entreprise claire et importante de CATALOGAGE des documents numérisés : *keine Digitalisierung ohne Tieferschliessung* » [chercheur/chargé de collection]

Le catalogage implique de transposer les enrichissements scientifiques selon les règles du codage standardisé ; comme le synthétise Jean-Marie Pierrel :

« les conditions de mutualisation des ressources prennent (ici) une importance toute particulière, notamment autour des trois aspects précis : les formats, les référentiels de vocabulaires utilisés ainsi que la mise en place d'une interconnexion simple entre nos entrepôts de données » [Kunstmann *et al.*, 2013, p. 7]

C'est autour de ces aspects que le dialogue entre les chercheurs et les institutions patrimoniales devient plus difficile, menaçant l'interopérabilité du projet.

Quel set de métadonnées ? Quel modèle conceptuel ?

Si le monde des bibliothèques, des archives et des musées produit des efforts importants en termes de normalisation pour garantir l'interopérabilité des jeux de données, des vocabulaires et des pratiques²⁵⁴, on ne peut pas en dire autant des équipes de recherche, notamment en sciences humaines et sociales.

De manière générale, malgré les initiatives en cours au niveau international (voir p. 48), les unités de recherche ne connaissent que très sommairement les pratiques métier des institutions patrimoniales ; en dehors de la modélisation des données de plus en plus sollicitée à travers les Humanités numériques, les modèles conceptuels tels que CIDOC-CR²⁵⁵ ou FRBR²⁵⁶, les formats des métadonnées bibliographiques et descriptives (les formats MARC), et les vocabulaires d'indexation contrôlée (RAMEAU, Dewey, etc.), sont méconnus des chercheurs, voire méprisés.

Un tableau lucide des difficultés de communication entre chercheurs et bibliothécaires apparaît dans l'une des réponses des chercheurs :

[...] « Alors que les bibliothécaires partagent des formats de structuration de l'information implémentés depuis quelques décennies (notamment autour de la famille MARC), les chercheurs (en sciences humaines), par leur parcours métier, élaborent souvent des modèles *ad hoc* qui restent étrangers aux pratiques de la bibliothéconomie. Depuis quelques années, ils s'intéressent aux modèles construits dans le monde des archives, des bibliothèques et des musées (que ce soit FRBRoo/ RDA ou CIDOC-CRM). Avec les besoins de numérisation de corpus, les notions de métadonnées et d'encodage de l'information font leur chemin petit à petit ; pourtant, au sein même des bibliothèques, les campagnes de numérisation ne sont pas toujours accompagnées de la création de métadonnées structurées selon les bonnes pratiques de l'EAD ou de la TEI ce qui rend très fastidieuse la récupération et l'exploitation des données. Les numérisations devraient être menées en concertation avec des équipes de chercheurs qui élaborent les métadonnées avec les conservateurs/ bibliothécaires (alors que, jusqu'ici, pour un même manuscrit, on peut trouver 4-5 bases de données qui décrivent différemment et partiellement son contenu en plus de sa description dans le catalogue de la bibliothèque qui le conserve)» [chercheur/IGE]

Si certains de ces formats posent des limites, notamment les formats MARC²⁵⁷, la transition bibliographique des catalogues vers le « Web des données » est en cours de route²⁵⁸. Par ailleurs, dans le cadre des concertations avec les professionnels, des chercheurs nous signalent que la connaissance des pratiques métier et de l'historique de leurs évolutions à l'intérieur des institutions patrimoniales tient un rôle important, davantage que les contraintes techniques ; en effet, des pratiques locales accumulées dans le temps

254. Signalons, par exemple, le modèle HADOC et le projet GINCO en développement au ministère de la Culture depuis 2013 [Briatte, 2014], ainsi que le Référentiel général d'interopérabilité publié par la Direction Interministérielle des Systèmes d'Information et de Communication [DISIC, 2015].

255. Voir : <http://www.cidoc-crm.org/> ; sur l'identification des ontologies utiles en SHS et l'utilisation de RDFa pour les corpus, voir [Pouyllau, 2012].

256. Voir la fiche pratique de l'Esssib : <http://www.enssib.fr/bibliotheque-numerique/documents/65520-comprendre-le-modele-frbr-et-ses-extensions.pdf>.

257. « Les formats Marc, créés dans les années 1960, participent à l'isolement des catalogues car leurs données bibliographiques ne sont pas interopérables avec d'autres données du web [Mesguich, 2017, p. 87] ; quant aux bonnes pratiques pour améliorer le référencement, voir le chapitre « Rendre les catalogues de bibliothèques plus visibles sur le Web » [Mesguich, 2017, p. 77-103].

258. <https://www.transition-bibliographique.fr/>.

peuvent être la conséquence de choix politiques et institutionnels et constituer un frein à l'introduction de modèles ouverts et collaboratifs ; dans la mesure où ces contraintes peuvent avoir des retombées sur le choix du modèle des données pour la description des sources du projet, ainsi que sur les possibilités d'intégration des données produites par les chercheurs, la connaissance de l'organisation et des pratiques de travail se révèle indispensable.

D'après les résultats de nos enquêtes, nous constatons que, chez les bibliothécaires, l'association de métadonnées structurées aux documents numérisés reste encore relativement limitée. Lorsque cette pratique existe, le choix des formats porte surtout sur METS, EAD, Dublin Core et, en moindre mesure, TEI (question n° 15).

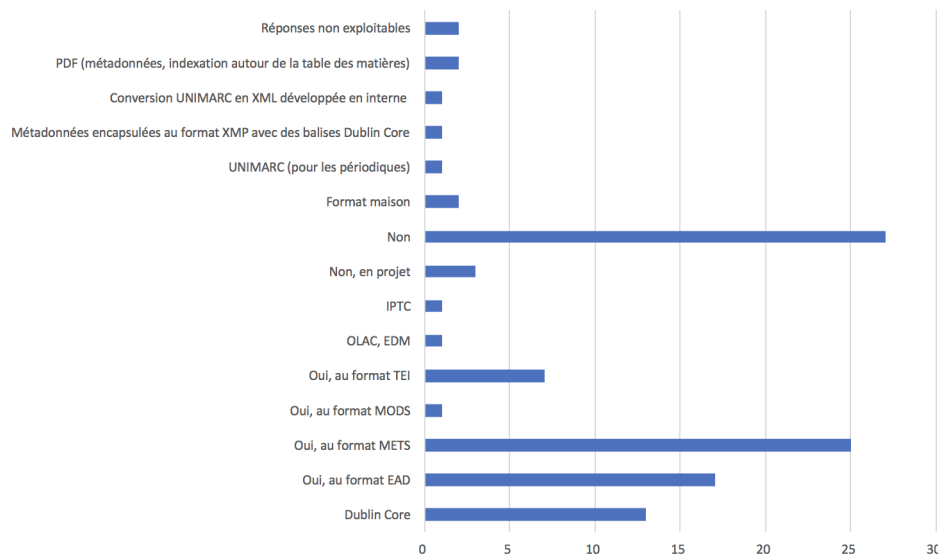


FIGURE 15 – Questionnaire aux bibliothécaires - question n° 15

Certaines institutions proposent également la conversion de certains documents en mode texte, principalement les fichiers texte issus de l'océrisation ou des fichiers au format ALTO (*Analyzed Layout and Text Object*) :

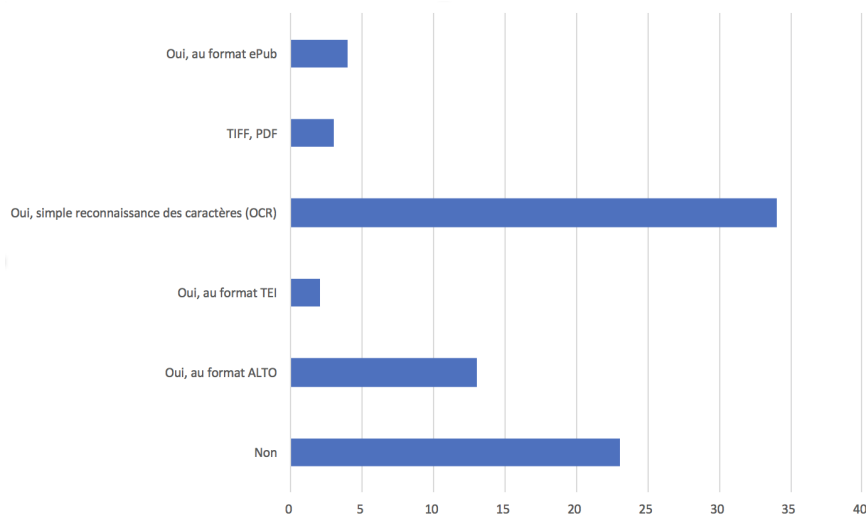


FIGURE 16 – Questionnaire aux bibliothécaires - question n° 14

D'autre part, interrogés sur l'utilisation des métadonnées relatives aux documents du corpus dont ils ont demandé la numérisation (question n° 12), moins de la moitié des chercheurs déclarent avoir utilisé les métadonnées produites par la bibliothèque.

Si ce faible résultat est à nuancer du fait qu'une partie de ces mêmes chercheurs expliquent, par la suite, que l'utilisation de métadonnées n'était pas prioritaire dans le cadre de leur projet, il n'empêche que les retours sur la qualité des métadonnées des bibliothèques sont parfois critiques. Plusieurs chercheurs constatent que les données fournies par les institutions patrimoniales ne sont pas toujours de bonne qualité : elles ont parfois été saisies trop rapidement ou avec des règles métier qui ne sont pas transparentes. C'est pourquoi ils se sentent directement impliqués dans la production de métadonnées fiables :

« Dans la plupart des cas, les bibliothèques ne disposaient pas de métadonnées (fiables). L'un des enjeux du projet était précisément l'enrichissement des catalogues des bibliothèques et la création des métadonnées » [chercheur postdoctoral]

« C'est nous qui avons créé les métadonnées (DC) et non la bibliothèque » [IGR]

« Je pense que la détermination des métadonnées en question relève du travail du chercheur (i.e. quelqu'un qui connaît parfaitement l'ensemble du corpus de l'intérieur, car des éléments des métadonnées sont souvent implicites) » [PU]

Signalons que les métadonnées que les chercheurs indiquent avoir récupérées sont le plus souvent au format Dublin Core, format simplifié à l'extrême, dont le succès semble lié à la multiplication des ressources électroniques²⁵⁹ ; ils récupèrent plus rarement des fichiers EAD ou TEI, utilisés pour les notices descriptives ou pour l'édition.

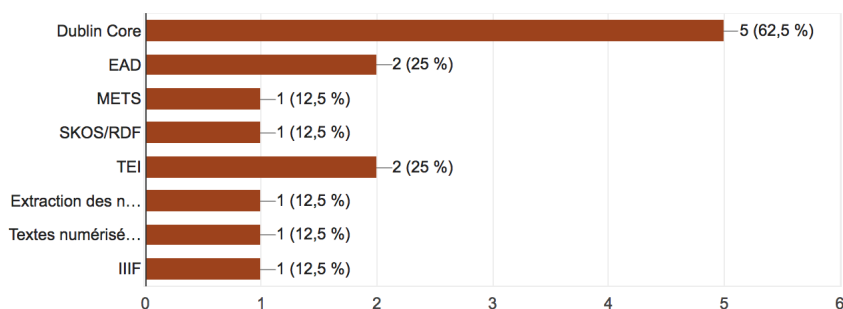


FIGURE 17 – Questionnaire aux chercheurs - question n° 13

En revanche, lorsqu'on leur demande s'ils ont produit des métadonnées susceptibles d'intégrer celles produites par la bibliothèque qui conserve les documents, l'encodage TEI arrive avant le Dublin Core.

259. « Les descriptions bibliographiques sont encodées dans des formats qui appartiennent tous à la famille des formats Marc. Le développement des techniques du Web et l'émergence du "mouvement des métadonnées" remettent en cause ce monopole. La multiplication des ressources numériques a entraîné le développement d'un format simplifié à l'extrême, le Dublin Core. Le langage XML ouvre la voie à des formats permettant de rendre les données bibliographiques - mais aussi archivistiques - compatibles avec le développement du Web » [Thompson *et al.*, 2011, p. 157].

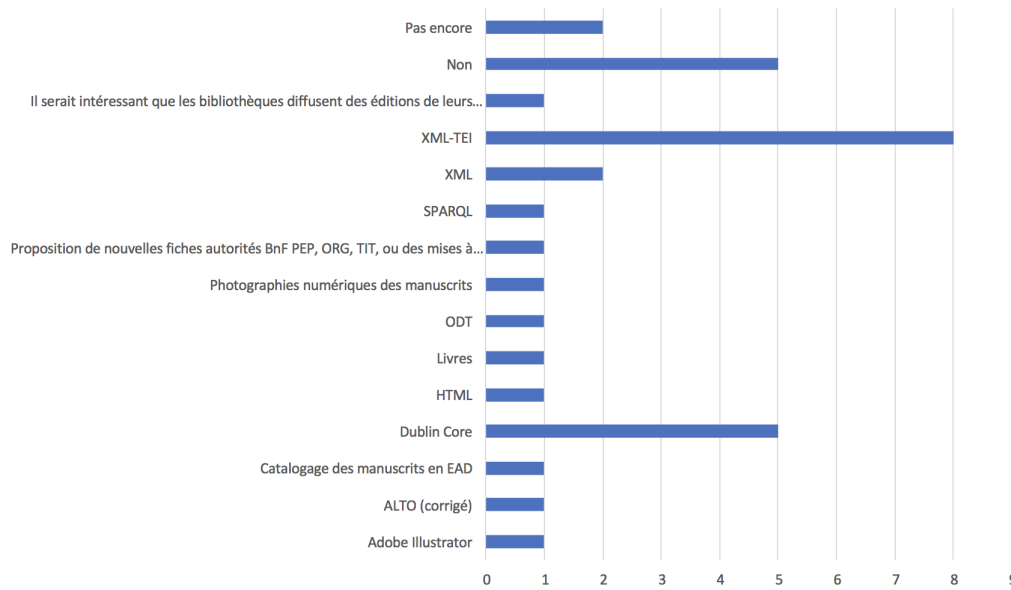


FIGURE 18 – Questionnaire aux chercheurs - question n° 16

Par ailleurs, l'un des souhaits des chercheurs serait justement que les bibliothèques accompagnent la diffusion des fichiers d'image par leur édition :

« Il serait intéressant que les bibliothèques diffusent des éditions de leurs manuscrits à côté des photos, mais pour l'instant elles semblent peu intéressées » [MCF]

Les choix techniques de numérisation et d'archivage

Une fois que la sélection des documents, le catalogage et les choix des métadonnées associées ont été assurés, la chaîne de numérisation se poursuit avec les étapes suivantes : la préparation des documents à numériser (dépeussierage, restauration, dépliage, gommage, etc.), la définition des caractéristiques de la numérisation désirée (qualité) et des règles de prises de vue afin d'assurer la cohérence des traitements, l'exécution des travaux de numérisation (à travers un atelier de reproduction interne, à travers un partenariat avec d'autres institutions, ou en s'appuyant sur un prestataire qui traitera les documents dans ces locaux ou dans ceux de l'institution), et le contrôle de qualité sur les documents numérisés. Il est également nécessaire de définir les modalités de versement et de stockage des documents numériques, ainsi que celles d'accès. Signalons que l'externalisation n'affranchit pas de toutes les charges de travail pour le personnel en interne. Cette chaîne de traitement et la rédaction d'un cahier des charges adapté ne sont toujours pas évidents pour les chercheurs. L'évaluation de ces opérations, en terme de coûts, de temps de traitement et de ressources humaines doit être claire attentivement prise en compte par les différentes composantes de l'équipe projet²⁶⁰. Bien que la différence de vocabulaires et de pratiques métier puisse occasionner de longs temps de concertation, la mixité de l'équipe (chercheurs, bibliothécaires, spécialistes de la photographie et de la numérisation, informaticiens spécialistes de l'archivage, etc.) est reconnue comme un élément susceptible de favoriser la réussite du projet.

260. Des bonnes pratiques à ce sujet sont résumées dans [Maillet, 2011, p. 132-137]

« [...] La conception d'un projet de numérisation interne débute par un questionnaire quant aux limites du projet (conditions de sa pertinence), avant de travailler à sa mise en œuvre (conditions de sa réalisation) » [Sciences Po]

Qui numérise et comment ?

Comme le montrent les résultats de notre enquête, les techniques de numérisation s'articulent principalement autour de la prise de vue photographique, et, en moindre mesure, du balayage :

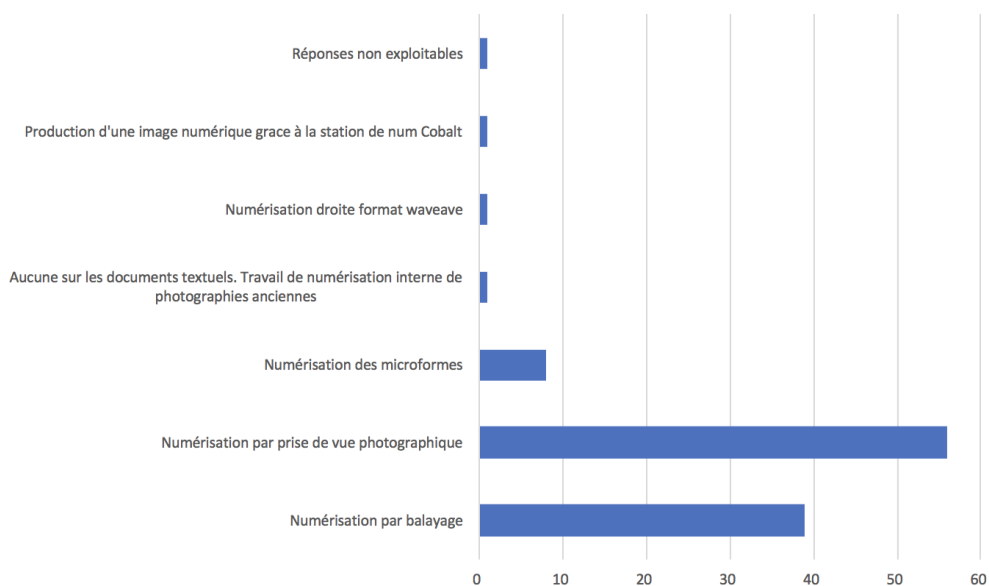


FIGURE 19 – Questionnaire aux bibliothécaires - question n° 13

Si les fonds sont le plus souvent numérisés par un acteur autre que l'institution de conservation,

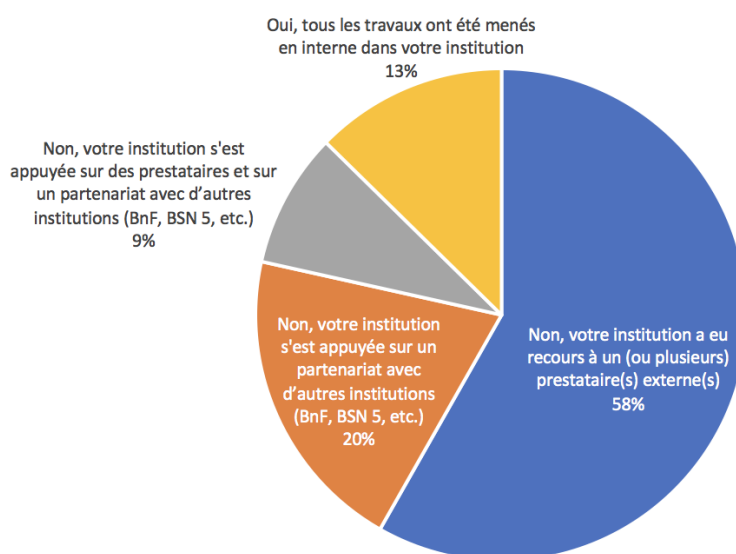


FIGURE 20 – Questionnaire aux bibliothécaires - question n° 12

en examinant les réponses de plus près, on constatera que les pratiques sont mixtes et diversifiées. Dans notre échantillon, plusieurs bibliothèques s'appuient sur un partenariat avec la BnF ; pour d'autres, les modalités de numérisation varient en fonction de la taille et de la nature du corpus à traiter :

« En interne et dans le cadre d'un partenariat avec la BnF. Les démarches sont complémentaires et ne concernent pas les mêmes corpus ni les mêmes typologies documentaires » [Université de Caen Normandie]

« Les ouvrages postérieurs à 1830 sont numérisés en interne. Les ouvrages antérieurs à 1830 sont confiés à un prestataire de numérisation. Par ailleurs, dans le cadre d'un partenariat avec la BnF, la bibliothèque envoie des ouvrages de ses collections à numériser sur les chaînes de traitement de la BnF » [BIU Cujas]

« Les solutions sont cumulatives : numérisation par différents types de prestataires suivant les marchés, ou numérisation en interne suivant la nature des documents » [BM de Lyon]

« Pour les corpus, nous avons recours à des prestataires extérieurs. Pour les images isolées, nous numérisons en interne » [Observatoire de Paris]

Signalons aussi que la numérisation peut être menée par un laboratoire de recherche et financé par mécénat :

« Les projets de numérisation ne sont pas menés par le SCD mais par les laboratoires (organisation liée au fonctionnement spécifique de Paris-Sorbonne où les UFR sont très indépendants). Il s'agit d'une forme de mécénat pour numérisation (qui n'a pas encore été faite). Le SCD a financé partiellement l'achat du scanner [...] » [Université Sorbonne Paris 4, SCD]

De manière plus générale, signalons aussi que la numérisation peut être menée selon différents choix techniques et procédures, en fonction de nature de la demande ; des accords ponctuels peuvent ainsi être conclus entre l'institution et le chercheur :

« 2 agents assurent la numérisation des images [***] En interne, nous collons à l'actualité culturelle (sacres et cathédrale pour le 800tenaire en 2011, presse ancienne, plaques de verre et photographies Deneux pour 1914-1918). Nous numérisons pour les animations du réseau de la BM et pour d'autres services culturels de la ville. En externe, il s'agit des demandes de chercheurs. Service assuré dans un délai de 2 mois maximum en fonction de la quantité d'images à numériser. En règle générale, nous numérisons en TIFF (400 dpi pour du texte, 600 à 1200 dpi pour de l'image) pour la conservation. Puis,

- cas n° 1 : demande à usage privé, nous fournissons une images watermarked en JPEG moyenne ou basse définition.
- cas n° 2 : demande en vue d'une éventuelle publication, nous fournissons les images commandées avec le watermark en JPEG moyenne ou basse définition et demandons au commanditaire d'opérer une sélection pour publication ; nous fournissons ensuite la ou les hautes définition.
- cas n° 3 : demande en vue d'une publication certaine, nous fournissons des images haute définitions en TIFF.

Dans tous les cas, nous facturons l'intégralité des images numérisées. Les établissements culturels bénéficient de la gratuité dans le cadre d'une coopération culturelle (ainsi nous nous assurons un service de qualité en cas de sollicitation desdits établissements pour des prêt d'expo ou autres...). Pour les cas 2 et 3, nous rédigeons

une autorisation de publication, voire un contrat passé entre la ville et le commanditaire lorsque nous souhaitons obtenir plusieurs livres à titre de justificatifs » [BM de Reims]

Le stockage et la pérennisation : méthodes et outils

La sauvegarde et la mise à disposition à long terme des fichiers numériques et des métadonnées produites dans le cadre d'une collaboration autour d'un corpus numérique nécessitent une infrastructure spécifique. Cette infrastructure doit être calibrée en fonction des dimensions et de la nature du corpus, mais aussi en fonction de ses moyens d'accès : un corpus de documents numérisés peut impliquer la gestion de lourds fichiers sources (les photographies des documents, parfois sous plusieurs tailles et formats), l'ensemble des métadonnées associées et les travaux qui accompagnent les sources (des transcriptions encodées sous différents formats, par exemple), ainsi que des logiciels capables d'interroger les documents ou l'ensemble des données et métadonnées du projet. Par conséquent, il faudra prévoir des systèmes de stockage adaptés et des stratégies de pérennisation pour garantir l'accès aux données en évitant l'obsolescence des outils et des formats. Il faudra évaluer attentivement si ces éléments sont disponibles chez l'un des partenaires, ou bien se tourner vers d'autres infrastructures dotées d'une politique de préservation du patrimoine numérique ou offrant des infrastructures garanties suffisantes. On pensera alors à SPAR (Système de préservation et d'archivage réparti) de la BnF ; ou au CINES pour l'archivage des données et documents numériques produits par l'ESRI²⁶¹ ; signalons aussi, du côté des archives, le programme VITAM (Valeurs Immatérielles Transférées aux Archives pour Mémoire)²⁶².

La pérennisation des données numériques devra répondre à des modèles communs, comme OAIS (*Open Archival Information System*)²⁶³, un modèle conceptuel de référence pour l'archivage numérique, et prévoir des métadonnées répondants aux principaux standards²⁶⁴. Signalons aussi, dans le cadre de la diffusion des ressources numériques sur le Web, la nécessité de réfléchir au choix et à la mise en place d'un système d'identifiants pérennes, comme le système ARK (*Archival Resource Key*)²⁶⁵.

La diffusion et la mise en ligne de corpus

Concernant les modalités de mise en ligne du corpus (question n° 16),

« Votre institution dispose-t-elle d'un outil pour la mise en ligne des corpus numérisés ? Si oui, lequel (CMS, système de gestion documentaire, etc.) ? »

les bibliothèques nous ont le plus souvent signalé l'outil employé dans le cadre de leur bibliothèque numérique, plus rarement un outil externe ponctuellement conçu pour gérer

261. Voir : <https://www.cines.fr/archivage/>.

262. Il s'agit d'un programme conjoint des ministères des Affaires étrangères et Développement international, de la Culture et Communication et de la Défense, visant à la réalisation d'une solution mutualisée d'archivage numérique, <http://www.programmevitam.fr/>.

263. Voir la fiche descriptive sur le site du CINES : <https://www.cines.fr/archivage/un-concept-des-problematiques/le-modele-de-referance-loais/>.

264. Sur les questions de la conservation des données numériques, voir le rapport [Hourcade *et al.*, 2010] et le billet de blog <http://www.lespetitescases.net/la-conservation-des-donnees-numeriques-n-est-ni-une-revolution-ni-inatteignable>.

265. Pour plus de détails, voir le site de la BnF : <https://lc.cx/AspQ>.

la mise en ligne d'un corpus suite à une collaboration avec des chercheurs. Quoi qu'il en soit, la solution la plus diffusée semble le système de gestion de contenus Omeka²⁶⁶, développé par le *Roy Rosenzweig Center for History and New Media* (RRCHNM) de l'Université George Mason ; ce logiciel semble obtenir la faveur tant des chercheurs que des bibliothécaires. Particulièrement adaptée à la gestion des bibliothèques numériques, cette solution est aussi soutenue par la TGIR Huma-Num et proposée « aux programmes de recherche souhaitant débiter dans la diffusion de données éditorialisées »²⁶⁷.

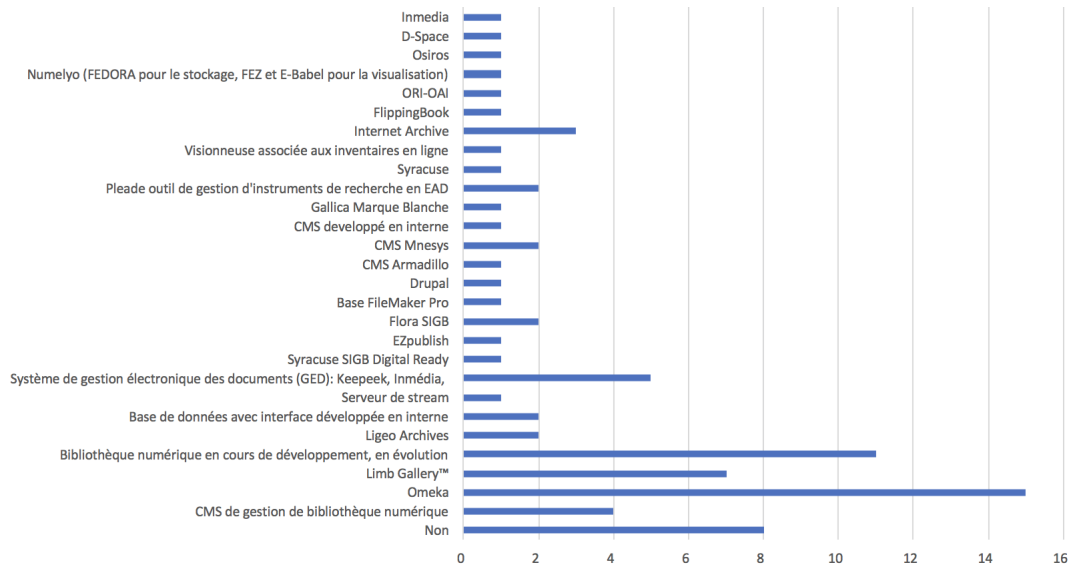


FIGURE 21 – Questionnaire aux bibliothécaires - question n° 16

En l'occurrence, les résultats recueillis sont trop génériques pour en tirer des conclusions quant aux fonctionnalités mises en production : la plupart des logiciels évoqués se composent de modules permettant de configurer de multiples services que les institutions peuvent façonner en fonction de leurs besoins et de leurs publics. En observant les solutions en ligne, plusieurs fonctionnalités peuvent être intégrées au système de publication. Selon le type de corpus documentaire, il pourra s'agir d'une interface permettant la transcription des documents, d'un système offrant différentes couches d'affichage pour un corpus d'éditions en ligne²⁶⁸, d'un moteur destiné à établir les concordances sur un corpus de documents, de solutions de géolocalisation des documents et des possesseurs, ou encore d'interfaces permettant d'interroger les documents et de les télécharger sous différents formats.

La solution technique et les modalités de consultation du corpus en ligne peuvent également faire l'objet d'échanges avec les chercheurs. Comme l'explique Julien Baudry [Baudry, 2017], au sujet de la numérisation d'anciennes cartes de Madagascar, intégrées

266. Voir : <https://omeka.org/>.

267. Huma-Num propose le service d'exposition de données NAKALA et le « pack de diffusion des données de la recherche » Nakalona, géré et administré par Huma-Num en mode SAAS (*Software as a Service*), voir : <https://www.huma-num.fr/services-et-outils/diffuser>.

268. « Il devient dès lors crucial pour une édition critique en ligne, non seulement de répondre immédiatement à la demande de l'utilisateur, mais également de lui offrir l'information qu'il demande *lorsqu'*il la demande, et de la *façon* dont il la demande. Et il s'agit là d'un véritable défi, les nouvelles possibilités d'affichage de l'information qu'offre une édition en ligne permettent de le relever » [Schuwey, 2015, p. 75].

dans la bibliothèque numérique et valorisées à travers une exposition, le travail du chercheur ne se limite pas seulement au travail sur la donnée ; le chercheur « va aussi réfléchir à la façon dont on peut la présenter, la mettre en avant : dans quel cadre et face à quel public »²⁶⁹.

Les dispositifs de valorisation et de soutien

La valorisation par la recherche est l'une des formes privilégiées de valorisation des collections patrimoniales, tant pour les archives [MCC Archives, 2014] que pour les bibliothèques. « L'attribution de bourses de recherche valorise les collections tout en améliorant la connaissance scientifique du fonds » [Jacquet, 2012, p. 140] :

Bilan « La plupart du temps, très positif. C'est l'un des publics privilégiés de la numérisation patrimoniale, il est très important que leurs travaux s'appuient sur nos collections et les mettent en valeur » [BIU Santé]

Bonnes pratiques : « La Deutsche Forschungsgemeinschaft fait accompagner le travail des centres de catalogage des manuscrits par un conseil scientifique international, qui aide dans les projets de valorisation » [PU]

C'est pourquoi les bibliothèques disposant de collections patrimoniales ont intérêt à prévoir des espaces et des dispositifs pour les chercheurs. Par exemple, la Cinémathèque française a mis en place plusieurs dispositifs pour favoriser les échanges, autour de ses collections, avec la communauté scientifique ; une collaboration stricte est assurée par le Conservatoire des techniques cinématographiques, par l'institution d'un « espace chercheur » au sein de la Bibliothèque du film, par des aides spécifiques aux jeunes chercheurs, et par la participation à des projets internationaux [Daire, 2016, p. 131-136]. De son côté, la BnF prévoit plusieurs dispositifs de valorisation autour de ses collections, comme celui des « Chercheurs invités » ; signalons que le dernier appel à chercheurs 2017-2018, destiné aux masters et aux doctorants, prévoit notamment un volet « Innover avec les humanités numériques » et une attention particulière portée aux collections numériques²⁷⁰.

Il n'empêche que les formes de restitution d'une recherche peuvent poser problème : un travail de thèse pourra être sauvegardé dans les archives de l'institution et opportunément signalé, mais sa vulgarisation auprès d'un public de non-spécialistes comporte un travail de communication supplémentaire ; faudrait-il alors demander aux doctorants et aux chercheurs de produire également des livrables en ligne avec les missions de la bibliothèque ? Quel type de retour les bibliothèques attendent-elles de la part des chercheurs ?

Dans la phase de valorisation d'un projet de numérisation de fonds, la solution de collaboration qui semble le mieux répondre aux intérêts des chercheurs et des professionnels est celle de la préparation d'une exposition, *in situ* ou virtuelle :

« Autour de l'exposition : "Aux origines du surréalisme, cendres de nos rêves" (Château des ducs de Bretagne, février-mai 2017). Commissariat d'exposition partagé entre un chercheur et la responsable du service Patrimoine ; édition d'un livre prévue à l'automne par le chercheur » [BM Nantes]

269. Voir : <http://1886.u-bordeaux-montaigne.fr/> et [Baudry, 2016].

270. Voir : <https://bnf-aac-2017-18.sciencesconf.org/resource/page/id/11>.

« Travail éditorial voir expo Francis Jammes Poète ou encore Houbigant sur PIRE-NEAS. Contrat passé avec les auteurs des textes » [Bibliothèque patrimoniale Pau]

« Projet 2013 sur les livres d'heures : exposition physique et exposition virtuelle, avec commentaire de chercheurs sur chaque document mis en ligne » [Médiathèque François-Mitterrand, Poitiers]

« La collaboration avec les chercheurs est très riche, en particulier sur les projets d'exposition [...] », « Pour les expositions virtuelles, travail conjoint : recherche des sources et images par les bibliothécaires, conception du projet et du parcours par les chercheurs. Sur la mise en forme, les bibliothécaires jouent le rôle d'éditeurs - vérification de l'adéquation des textes au public visé » [BU]

« Pour les expositions virtuelles, travail conjoint : recherche des sources et images par les bibliothécaires, conception du projet et du parcours par les chercheurs. Sur la mise en forme, les bibliothécaires jouent le rôle d'éditeurs - vérification de l'adéquation des textes au public visé [Université de Lille 1, Sciences et technologies]

La numérisation et l'étude de corpus : un terrain de rencontre stimulant

LES JALONS D'UNE COLLABORATION INTER-MÉTIER

Le « cycle de vie de la recherche » à l'ère numérique, ainsi que nombre de savoirs et de compétences à mobiliser dans la constitution et la mise à disposition d'un projet de corpus numérisés invite à s'interroger sur la disponibilité concrète de ces éléments dans l'ensemble des institutions patrimoniales, ainsi que sur les modalités permettant de les intégrer ou de les repérer à l'extérieur²⁷¹. Impliquant de mobiliser un réseau de disciplines et de ressources théoriques, méthodologiques et techniques issues de plusieurs domaines, ces projets constituent un défi, y compris pour les grandes institutions. Face à la demande croissante de corpus (et d'expertises concernant les corpus) de la part des chercheurs, faut-il créer à l'intérieur des institutions des équipes dédiées, composées de bibliothécaires et d'experts informatiques ? Faut-il que des tâches de préparation de corpus et de guide en matière de fouille de textes soient intégrés aux missions des professionnels ? Faut-il prévoir des formations spécifiques pour permettre aux étudiants et aux chercheurs d'interpréter et de travailler sur les jeux de données de l'institution ?

Les bonnes pratiques nationales et internationales

La numérisation de corpus implique une collaboration interdisciplinaire et inter-métiers à l'intérieur des institutions patrimoniales, d'autant plus lorsqu'il s'agit de travailler sur de grands volumes de documents :

« On soulignera que ces numérisations imposent la nécessité aux archivistes et aux chercheurs, mais aussi aux informaticiens, de travailler en collaboration. Cet agencement est nouveau et souvent producteur de résultats inédits [...]. Cette possibilité est plus remarquable si le corpus étudié compte plusieurs centaines de documents. Le passage par un outillage n'est pas une possibilité mais une nécessité » [Artières, 2016, p. 24].

Si, à l'international, les mêmes questions se posent, les conditions de départ de la réflexion ne sont pas les mêmes. Les stratégies de curation de contenus (*Digital Curation*) [Higgins, 2011] - c'est-à-dire la création, la gestion et le partage de collections numériques sur un sujet spécifique à travers une véritable politique documentaire²⁷² -, le développement des pratiques de *Digital Scholarship* - qui font de la bibliothèque un « centre d'expertise sur les questions du numérique au service de la recherche » [Claerr et Westeel, 2013a, p. 42-44] -, constituent des sujets de réflexion encore relativement nouveaux en France et en Europe²⁷³, tout au moins lorsqu'ils sont associés à la préparation de corpus numérisés. De même, des profils professionnels se situant à l'intersection de plusieurs disciplines, comme les profils anglo-saxons de « bibliothécaire

271. « Les professionnels des bibliothèques doivent connaître les normes et standards des disciplines pour décrire les collections de jeux de données, la manière dont elles sont produites et analysées, les cadres juridiques auxquelles elles sont soumises pour leur partage et leur réutilisation, les principes de base de leur conservation et les aspects techniques liés à leur préservation pérenne. D'autres aspects, d'ordre organisationnel et financier mais aussi éthique, sont également cruciaux » [Neuroth, 2015].

272. Sur la pratique de la curation, voir [Mesguich *et al.*, 2012].

273. Parmi les initiatives européennes comportant un dialogue entre chercheurs et développeurs autour de collections numériques, signalons le KB Lab de la Bibliothèque nationale des Pays-Bas (voir <http://lab.kb.nl/>), et le HUMlab de la Bibliothèque royale du Danemark (voir <https://kub.kb.dk/humlab>).

spécialisé » (*Subject Librarian*)²⁷⁴ ou de *Humanities Subject Librarian*²⁷⁵, semblent encore minoritaires dans le panorama professionnel en France, bien qu'encouragés par la publication de la « Loi pour une République numérique » et par la création d'un service public de la donnée²⁷⁶ :

« Malheureusement, en Europe, nous ne disposons pas encore d'une connaissance commune des compétences essentielles à la gestion des données, ni de cursus appropriés, associés à des diplômes spécialisés (*data librarian, data scientist, etc.*). Nous devons aussi établir de nouveaux profils de postes, accompagnés de parcours professionnels identifiés et d'une large intégration à l'univers des communautés scientifiques. Ce n'est qu'à la condition de coopérer avec les chercheurs que les professionnels des bibliothèques seront aptes à répondre aux questions que les données vont poser » [Neuroth2015]

Le contexte international semble plus accoutumé aux collaborations interdisciplinaires, tout au moins dans les bibliothèques universitaires, comme le montrent les investissements dédiés à la création des centres consacrés aux Humanités numériques à la Columbia University Library²⁷⁷ ou à l'Université de Rochester²⁷⁸ ; néanmoins, il ne semble pas que ce modèle soit à encourager dans toutes les institutions. Un dossier de OCLC (*Online Computer Library Center*) invite à réfléchir aux modalités et aux risques de ces investissements : l'analyse attentive de l'environnement de l'institution et de ses collaborations dans les infrastructures de recherche nationales et internationales est un prérequis nécessaire pour pouvoir raisonnablement situer un service destiné aux Humanités numériques à l'intérieur de l'institution (rattaché à la bibliothèque, à un laboratoire de recherche, au service informatique, etc.) ou en commun avec d'autres institutions de recherche [Schaffner et Erway, 2014].

Si les collaborations inter-métiers semblent mieux consolidées à l'international, les chercheurs nous signalent également de bonnes pratiques en France (« Essentiellement dans les MSH») et en Europe, comme le Bodmer Lab de Genève²⁷⁹, fondé sur la collaboration entre chercheurs appartenant à plusieurs départements d'études et divisions informatiques de l'Université de Genève et les professionnels de l'information et de la conservation de la Fondation Martin Bodmer.

« J'ai trouvé en France, aux USA et en Norvège d'excellentes pratiques pour mettre à la disposition des chercheurs des documents numérisés » [PU émérite]

« Nous avons collaboré avec beaucoup de succès avec des collègues à la British Library et à la Cambridge University Library » [PU]

L'absence de services définis et structurés présente évidemment des inconvénients. D'une part, les projets semblent reposer surtout sur les compétences internes, sur la disponibilité et la bonne volonté des professionnels, avec le risque que les collaborations puissent s'interrompre avec le départ des agents impliqués :

274. Pour une introduction à la pluralité des métiers de bibliothécaire dans le monde anglo-saxon, on consultera encore [Law, 2004].

275. Sur l'introduction des Humanités numériques en bibliothèque, sur les défis et les opportunités pour les bibliothécaires, voir en particulier [Hartsell-Gundy *et al.*, 2015].

276. Loi n° 2016-1321 du 7 octobre 2016 pour une République numérique.

277. Voir : <http://library.columbia.edu/locations/dhc.html>.

278. Voir : <http://dslab.lib.rochester.edu/>.

279. Voir : *The Bodmer Lab. A Digital Word Literature*, <http://bodmerlab.unige.ch/>.

« Le partenariat existant actuellement entre le SCD et le laboratoire de recherche C***, ainsi qu'avec la bibliothèque municipale et les archives municipales, s'est révélé très fructueux. Mais le dispositif implique peu de personnes sur qui repose l'essentiel des projets et dont le remplacement en cas de départ peut s'avérer difficile » [BU]

D'autre part, l'inadéquation des moyens est une source de frustration pour le chercheur :

Bilan « Aspects positifs : disponibilité des conservateurs ; intérêt partagé pour la valorisation et diffusion (selon les cas : mise à disposition des portails de diffusion des bibliothèques). Aspect négatifs : manque des instrumentations sur place (d'où la nécessité de faire appel à un prestataire, avec augmentation des coûts) ; difficultés logistiques ; [...] » [chercheur postdoctoral]

Bilan « Positif : être sûre d'être dans les clous pour les questions d'ingénierie documentaire. Négatif : le service de bibliothèque ne se pose pas trop la question des usages numériques (autre que la consultation de documents) et service surchargé de boulot... » [IGR]

Quel équilibre entre médiation aux données et formation numérique des chercheurs ?

« Le déficit de formation en informatique des chercheurs en sciences humaines et sociales est flagrant. Leur formation est nettement insuffisante. Elle peut inclure de la programmation, pas tant pour coder que pour apprendre les logiques informatiques » [Clavert *et al.*, 2012]

Si la diffusion des Humanités numériques a aidé à la diffusion de profils aux compétences techniques accrues, se situant dans la continuité entre les métiers d'accompagnement de la recherche et les chercheurs, il n'empêche que les chercheurs qui demandent de travailler sur des corpus peuvent faire preuve de compétences informatiques très hétérogènes et parfois inadaptées pour interpréter les modèles conceptuels et les jeux de données qu'ils souhaitent explorer et étudier. Faut-il former des chercheurs autonomes ? Si oui, à qui revient l'organisation de cette formation ?

« Le chercheur doit-il savoir coder ? L'ingénieur doit-il être au fait des thématiques abordées par les chercheurs ? Ou chacun doit-il garder ses compétences respectives et apprendre à mieux communiquer ? Quelle est donc la distance à adopter par rapport aux données et comment appréhender cette distance ? » [Le Deuff et Berra, 2012]

La formation informatique d'un public qui a intérêt à gérer et manipuler d'importantes quantités de textes et de données est au cœur du questionnement pédagogique des sciences humaines et sociales : une formation adaptée aux usages de recherche de tous les étudiants en SHS semble profitable, non pas seulement pour le sous-groupe qui s'intéresse aux Humanités numériques [Le Deuff, 2014, p. 126 sq.]²⁸⁰. Cela passe par la création d'ateliers pour apprendre à coder et par l'amplification du dialogue entre les bibliothèques et les disciplines métier.

280. Il s'agit de former les étudiants et les chercheurs au traitement de fichiers numériques, de leur donner des conseils de méthodes et logiciels, tout en sachant qu'il n'existe pas de solution passe-partout et qu'il faut « bien comprendre le sens et la structure des textes d'un côté, et les capacités exactes des divers outils disponibles » [Guerreau, 2012].

Le manque de familiarité avec les outils informatiques peut être à l'origine d'incompréhensions susceptibles de se manifester surtout à propos des formats des données, les formats utilisés par les bibliothèques ou par les institutions patrimoniales n'étant pas suffisamment familiers et accessibles aux chercheurs, aux enseignants et aux étudiants. Signalons par exemple, le texte d'une annonce parue récemment dans la liste de diffusion *Humanistica* concernant un *workshop* prévu pour fin mai 2018, « Fabriquer des jeux de données en art », ouvert aux enseignants-chercheurs, aux bibliothécaires et aux ingénieurs de recherche, dont l'objet « sera la constitution de jeux de données qui puissent servir dans un cadre pédagogique, en particulier dans le cadre d'une initiation des étudiants à la *Digital Art History* » :

« En effet, nous avons constaté la difficulté d'accès à de tels corpus. Très peu d'institutions ont mis à disposition leurs données de manière facilement accessible au format CSV ou encore sous forme de corpus d'une taille « raisonnable » de 2000 objets environ afin de permettre des manipulations aisées pour les étudiants. L'utilisation des API fournis par les institutions, ou encore la collecte depuis des bases de données en ligne sont également des freins pour les enseignants. Le *workshop* sera l'occasion de prendre en main les outils de collecte sur le web, le maniement des API et le nettoyage de données. Il permettra également d'identifier des gisements de données pouvant être utiles aux enseignants et aux étudiants dans les différentes disciplines artistiques. Les jeux de données constitués à l'issue de ce *workshop* seront mis à disposition de la communauté ».

Dans la logique du partage des données, faut-il adapter les pratiques métier aux pratiques de la recherche, afin de permettre la récupération de jeux de données « plus compréhensibles » ? Faut-il prédisposer des API adaptées en fonction du public ? Et dans ce cas, selon quelles grilles effectuer les transformations entre les « données brutes » de la bibliothèque et les « données maniables » par les différents chercheurs ?

La mise à disposition d'une infrastructure adaptable et capable d'anticiper les besoins d'un public hétérogène est un défi important et difficilement envisageable dans tous les établissements. En revanche, la mise à disposition des données selon les logiques de l'*Open Data*, ainsi que la publication du modèle des données de l'institution, accompagnée d'un travail de médiation vers le public, restent des objectifs plus facilement atteignables. Dans ce sens, signalons des initiatives pour sensibiliser et réfléchir aux enjeux des données culturelles, comme la programmation collaborative des *hackathons*²⁸¹, ou à la manipulation des données et des formats bibliographiques - par exemple le nouveau portail « BnF API et jeux de données »²⁸² ; ce dernier, en plus de mettre à disposition des jeux de données pré-constitués, décrit et documente l'ensemble des interfaces de programmations permettant d'interroger les métadonnées des catalogues et les collections numérisées de la BnF.

D'autre part, l'appropriation des nouvelles technologies est un sujet délicat, aussi bien pour les chercheurs que pour les institutions patrimoniales, surtout pour les institutions de petite taille ; les projets de collaboration seront l'occasion pour encourager l'acquisition de nouvelles compétences :

281. Par exemple : le *Hackathon Culture et Tourisme*, organisé par le MCC, <http://www.culturecommunication.gouv.fr/Actualites/Hackathon-Culture-Tourisme>; le *Hackathon BnF*, http://www.bnf.fr/fr/evenements_et_culture/auditoriums/f.hackathon_2017.html?seance=1223928195323.

282. Voir : <http://api.bnf.fr/>

Bilan « Le plus souvent, acquisition de nouvelles compétences liées au numérique pour les personnels des bibliothèques impliquées » [IGE]

Si la médiation relative au modèle de données de la bibliothèque reste un atout important, à prendre en compte depuis le début du projet, la construction d'une véritable stratégie de recherche autour des données peut s'élaborer dans l'espace du laboratoire, grâce à une synergie entre chercheurs en sciences humaines, ingénieurs et informaticiens, sans que cela se limite à de simples collaborations utilitaires ; comme le souligne Julien Longhi :

« Dans le cadre des Humanités numériques, l'interdisciplinarité entre SHS et sciences et techniques/sciences informatiques est importante, sinon soit l'informatique et les technologies sont des « prestataires » pour SHS, sans que les chercheurs participent au développement des outils, soit les SHS sont une « caution » pour le développement d'outils sans que les chercheurs ne se préoccupent fondamentalement de la nature sémiotique des données par exemple. Le terme interdisciplinarité est donc assumé, même si nous sommes conscient des difficultés inhérentes à de telles interactions » [Longhi, 2017].

Expertise scientifique et technique : une question de légitimité ou de complémentarité ?

Les bibliothèques et les institutions patrimoniales disposent d'une expertise avérée quant à la gestion de données. Le rôle des professionnels des bibliothèques quant à la valorisation, la diffusion, la pérennité et l'interopérabilité des données numériques, mais aussi dans la constitution des normes internationales, est bien reconnu et mis en évidence par le bilan positif fait par les chercheurs :

Bilan « Validité des solutions techniques » [PU]

Bilan « Les aspects positifs consistent dans : la richesse de la documentation conservée par la BnF ; les multiples compétences mises à disposition par la BnF pour le programme ; le rôle pivot de la BnF dans le réseau international [...] » [chercheur/chargé de collection]

Bilan « Une collaboration fructueuse et une pérennisation qui est assurée (sauf erreur de ma part) conjointement par la bibliothèque et par l'université porteuse du projet » [PU]

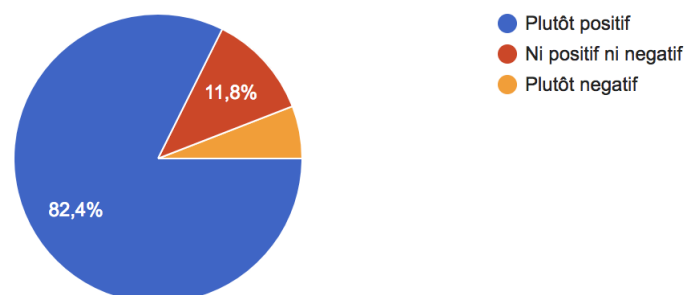


FIGURE 22 – Questionnaire aux chercheurs - question n° 6

Par contre, lorsqu'il s'agit de l'étude scientifique des documents, il apparaît que « le patrimoine n'a plus que jamais besoin des chercheurs »²⁸³ et que les bibliothécaires ne sont que des exécutants :

« La collaboration est absolument nécessaire [***]. Les chercheurs connaissent les textes qui mériteraient d'être numérisés et les bibliothèques ont les moyens et le personnel pour exécuter ces travaux. Il faudrait développer des projets communs » [PU]

« Sans ces comités scientifiques, nous ne serions pas capables, nous bibliothécaires, de proposer des collections "crédibles" au public, nous ne sommes pas des spécialistes de Blaise Pascal, de la Grande Guerre, de la cartographie etc. Ces comités scientifiques sont un échange entre le personnel de la bibliothèque qui en retire un projet solide et les spécialistes (universitaires ou pas), qui accroissent leur connaissance sur ces thèmes par la découverte de documents issus de nos fonds, inconnus pour eux. Ces collaborations sont très importantes » [chargé de numérisation, BM]

Il n'est pas rare de remarquer, même chez les bibliothécaires, que l'apport intellectuel des professionnels envers les collections doit être subordonné à l'avis des chercheurs, capables de poser « un regard réellement critique ». L'expertise scientifique revenant aux chercheurs, il n'est pas étonnant que les bibliothécaires se sentent faiblement impliqués dans les projets de collaboration. Bien que cette situation ne soit pas la règle, signalons qu'un sentiment d'infériorité vis-à-vis des chercheurs semble affecter négativement les collaborations et conditionner le pouvoir de négociation des professionnels au moment de l'établissement d'une convention.

La prise en compte, depuis le début, des apports et des ambitions des différents acteurs, ainsi que l'attention portée aux aspects relationnels restent des éléments fondamentaux pour trouver un bon équilibre :

« Bilan très positif. Les différents projets ont permis de nouer des liens solides avec des chercheurs de diverses institutions et ils ont abouti à des réalisations sur lesquelles la bibliothèque peut s'appuyer pour les projets suivants. Peut-être un enseignement important tiré est de bien déterminer, avant et pendant les projets, ce que la bibliothèque peut apporter et ce qu'elle ne peut pas faire, et de ne pas hésiter à le dire, diplomatiquement » [responsable valorisation et collections patrimoniales, BIU]

« Les bonnes pratiques, c'est bien ; mais ce qui est le plus important, c'est de s'écouter les uns les autres et de cultiver de bonnes relations, en prenant soin que chaque partenaire trouve son compte dans l'entreprise commune. Dans notre cas, cela était facilité par la clarté de l'objectif commun : redonner l'accès à un fonds gravement endommagé mais d'une grande importance scientifique » [directeur de recherche]

283. « Quels enseignements tirer de ces années riches en événements ? D'abord, que le patrimoine a plus que jamais besoin des chercheurs. Seul un enseignant expérimenté pouvait orienter notre programme de numérisation dans un sens qui soit vraiment utile à des étudiants novices. Si le regard du bibliothécaire est irremplaçable pour acquérir une vision synthétique des collections ou résumer un ouvrage en quelques mots clés, la connaissance détaillée des contenus appartient aux chercheurs, qui connaissent nos documents de l'intérieur, puisqu'ils prennent le temps de les dépouiller, et sont à même de poser sur eux un regard réellement critique. Et force est d'admettre que l'écart entre le socle de connaissances du grand public (et des bibliothécaires...) et ces documents écrits dans des langues anciennes dont nous commençons à perdre les codes d'accès, nécessitera de plus en plus la médiation de spécialistes : l'appropriation du patrimoine écrit par le grand public devra, demain plus que jamais, s'appuyer sur le décodage offert aux bibliothécaires par le travail, de la recherche ». [Claerr, 2012, p. 148-149].

Les institutions patrimoniales sont-elles un guichet de données ?

Si les chercheurs, les ingénieurs et les professionnels de l'information s'accordent sur le principe que les partenariats et les collaborations sont « indispensables », les raisons qui les amènent à entreprendre des collaborations ne sont pas les mêmes :

« Même si chercheurs et bibliothécaires ont pour objet commun l'information et son traitement, leurs pratiques sont dépendantes d'objectifs différents » [IGE]

Bilan, question n° 30 - « mixte - les priorités ne sont pas les mêmes » [PU]

En effet, comme nous l'avons vu précédemment, les besoins et les logiques qui actionnent la numérisation et l'étude de corpus ne sont pas forcément partagés ; lorsque la bibliothèque est sollicitée tardivement ou peu impliquée dans le projet, des frictions peuvent apparaître, voire aboutir à un véritable mécontentement :

« La collaboration est généralement insuffisamment préparée, en particulier de la part des chercheurs. Ceux-ci construisent leur projet, obtiennent des financements, et sollicitent simplement ensuite la B*** pour leur fournir les documents numérisés dont ils ont besoin. Nous souhaitons changer cette situation, et être partie intégrante du projet dès sa conception. Afin également que des crédits du projet soient véritablement dédiés à la numérisation et à la description des corpus, car ce modèle où la B*** est sollicitée après coup ne peut être soutenable »

« J'ai observé dans mon précédent poste en BM, et ici à l'U***, que les projets initiés par les chercheurs portant sur des documents conservés dans les bibliothèques ne tiennent souvent pas assez compte de l'intérêt pour les bibliothèques à diffuser les documents numériques sur leur plateforme (en plus de la plateforme du labo de recherche) ou à être force de proposition sur les contours documentaires. La responsabilité de ce déséquilibre est à coup sûr partagée, il peut être difficile de négocier face à un directeur de labo. Mais la bibliothèque en est réduite à un rôle de prestataire de service alors qu'elle devrait être reconnue comme un interlocuteur scientifique à part entière, dont la mission est aussi de desservir un public que le labo ne touchera pas (grand public dans le cas d'une BM, ou public étudiant) » [BU]

Comme le souligne Jean-Luc Minel, le chercheur voit souvent les institutions patrimoniales comme un guichet où trouver de la donnée : il n'est pas dans une démarche selon laquelle les données qu'il produit pourraient alimenter les bases de données des institutions patrimoniales [ADEMEC, 2017]. En effet, le chercheur est le plus souvent dans une logique impliquant que les données ne servent qu'à appuyer sa production scientifique, et ainsi à avancer dans sa carrière ; comme sa promotion ne tient pas à la mise à disposition de données, il utilise les données comme « terrain de test », mais sans avoir intérêt à ce que les résultats de son travail soient exploités dans un cadre professionnel selon des formats conséquents, ou que l'outil auquel il travaille soit effectivement mis en production.

De leur côté, les institutions patrimoniales cherchent des solutions à long terme pour la gestion et la mise en valeur des données. Leur investissement en temps et en ressources humaines vise à obtenir des résultats aussi concrets et cohérents que possible.

« La collaboration avec les chercheurs est très riche, en particulier sur les projets d'exposition. Sur la conception de la collection numérique, deux approches se confrontent et doivent se concilier : les bibliothécaires recherchent une cohérence et une unité dans la collection numérique, tandis que les chercheurs s'intéressent plus à la valeur de chaque document pris séparément » [BU]

En dehors de quelque exemple positif,

« La bibliothèque théâtrale de la comtesse de Verrue a fait l'objet d'un travail de description et d'identification par un chercheur ; ces descriptions très complètes ont été reprises dans les métadonnées des documents numérisés » [BM]

faut-il se résigner à l'idée que toutes les données et les recherches qui sont produites dans le cadre de ces partenariats ne puissent être réintégrées dans l'institution patrimoniale ? Faut-il que les chercheurs s'engagent pour garantir un retour aux institutions qui sont à l'origine des données ?

Chercheurs et professionnels : une temporalité différente ?

Un élément d'inquiétude revient incessamment, à la fois dans les réponses des chercheurs et dans celles des professionnels : le temps.

Si, pour certains chercheurs, la collaboration avec les bibliothécaires est à l'origine d'une meilleure gestion du temps de travail :

Meilleure répartition des tâches en fonction des compétences. Optimisation du temps de travail et amélioration de la conception d'ensemble du projet. Expérience hautement positive que nous renouvelons à chaque fois que cela est possible [IUF]

ils relèvent souvent que le temps de l'institution, notamment pour la gestion administrative et technique de la collaboration, a une influence négative sur le calendrier des travaux, ce qui peut décourager la mise en place d'une entreprise conjointe ou amener à des désistements en cours de route :

« calendrier très long » [chercheur postdoctoral]

« une collaboration soutenue exige beaucoup de temps » [PU]

« après des longs échanges avec le DSI de la bibliothèque [***], nous nous sommes repliés sur une solution interne au laboratoire » [MCF]

« Les conservateurs ont fait leur possible pour accélérer les choses, mais le temps (long) de la B*** n'est pas toujours celui des projets. . . » [PU-IUF]

Plusieurs chercheurs soulignent aussi qu'il n'est pas toujours simple d'établir le contact avec les institutions, notamment avec les institutions parisiennes, du fait du nombre de sollicitations qu'elles reçoivent :

« That depends. we met several times with colleagues from the British Library. Making contact with colleagues in other libraries (for example the BnF or the Marciana in Venice) was much harder » [PU]

« En province on se croise lors de mon passage dans la ville et on discute. A Paris c'est impossible de croiser les gens plus que 2mn (ce qui ne veut pas dire qu'ils sont de mauvaise volonté) » [chercheur postdoctoral]

Signalons que la plupart des projets de financement est fondée sur un calendrier bien précis, qu'il sera nécessaire de prendre en compte dès le début des travaux :

« Les aspects négatifs consistent en : les temps de réaction de la part de la B*** (trop lents par rapport au calendrier imposé par un programme [***]); la nécessité de reformuler, en cours de réalisation, les ambitions initiales (trop hautes) proposées par la B*** vers des objectifs plus réalisables » [chercheur/chargé de collection]

Du côté des institutions patrimoniales, on retrouve la même conscience et les mêmes préoccupations à l'égard des possibilités de débordement du projet dans le temps :

« Les projets mettent du temps à se concrétiser même quand il y a la volonté de les mener à bien »

mais les inquiétudes semblent porter surtout sur l'implication des chercheurs sur la longue durée :

« Bonne implication au départ, lassitude des partenaires devant les temps de réalisation (identification, nommage, contrôle-qualité), retour une fois ces opérations "techniques" réalisées »

« Bilan positif : enrichissement mutuel, mais très gourmand en temps ; ne peut fonctionner qu'avec des chercheurs vraiment impliqués ; comme il a été difficile pour nous d'en trouver, un grand nombre de corpus n'ont été élaborés que par notre service, ou alors avec des collègues d'autres bibliothèques »

Bilan : « Mitigé à cause du manque de disponibilité des chercheurs (les interlocuteurs ont souvent changé) »

Bilan : « Le bilan est relativement mitigé pour le moment : - rien de systématique : il s'agit surtout de chercheurs avec lesquels nous avons déjà des liens interpersonnels au sein de la BU. La collaboration menée avec ces trois chercheurs est donc difficile à reproduire avec d'autres, ou à plus grande échelle. Le consortium I*** est toutefois un moyen de rendre ces collaborations plus tangibles. - il est souvent très difficile de solliciter les chercheurs sur le long terme. Hormis des événements ou des demandes ponctuelles, une implication pour un projet de diffusion sur plusieurs mois, avec indexation par exemple, est difficile, et oblige à beaucoup d'adaptations et d'insistance. Mais hormis ces réticences qui interrogent la durabilité de telles collaborations, les échanges en eux-mêmes sont riches et intéressants. . . Ne pas abandonner, donc ! »

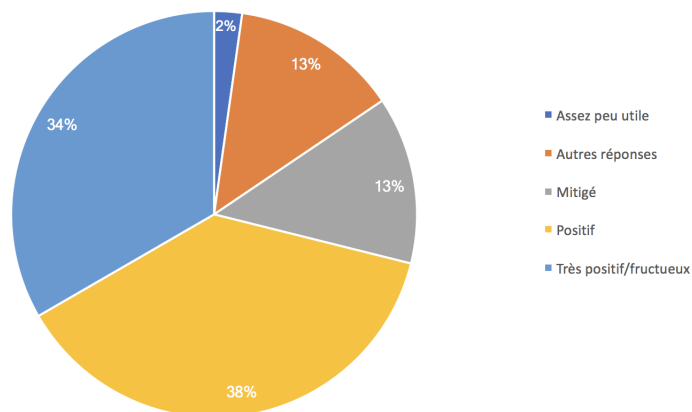


FIGURE 23 – Questionnaire aux bibliothécaires - question n° 20 - Bilan des collaborations

BIBLIOTHÈQUES TERRITORIALES ET BIBLIOTHÈQUES DE L'ESRI : DES SERVICES, DES RESSOURCES ET DES PRATIQUES DIFFÉRENTES ?

Parmi les 82 établissements qui ont répondu à notre enquête, et qui nous semblent assez représentatifs de l'ensemble des bibliothèques, nous pouvons distinguer au moins deux grands groupes : les bibliothèques appartenant à un établissement relevant de la tutelle du ministère en charge de l'Enseignement supérieur, de la Recherche et de l'Innovation (56 %), et les bibliothèques relevant des collectivités territoriales (30 %) ²⁸⁴. En dehors des éléments que nous avons déjà traités auparavant, y-a-t-il des différences de pratique entre ces deux types d'institutions ? Quels sont les services et les ressources de ces différentes institutions face à la numérisation de corpus documentaires ?

Rejoindre les chercheurs par des services adaptés

Destinées à accompagner des publics différents, les bibliothèques territoriales et les bibliothèques de l'enseignement supérieur affichent des services différents. Malheureusement, dans l'échantillon de chercheurs qui a répondu à notre enquête, il n'est pas aisé de faire ressortir des différences marquantes dans le public qui s'adresse à l'une ou à l'autre ; d'autre part, ni les quelques enquêtes récentes sur les besoins des chercheurs en bibliothèque universitaire ²⁸⁵, ni l'enquête de public sur les bibliothèques municipales [MCC - DGMIC, 2016] ²⁸⁶ ne nous renseignent suffisamment sur le public des chercheurs qui s'intéressent aux corpus documentaires. Nous pouvons observer que, si les bibliothèques territoriales et les bibliothèques de l'enseignement supérieur sont toutes les deux sollicitées par les chercheurs, dans ces dernières, le dialogue avec eux est perçu comme plus aisé, parfois « structurel », et souvent facilité par les échanges en salle de lecture :

« Compte-tenu du périmètre de mon institution, le dialogue avec les chercheurs est permanent »

« Ça dépend de la collaboration, il y en a eu beaucoup. La plupart du temps, la collaboration est informelle au début et se fait à la faveur des chercheurs avec qui l'on discute en salle de lecture »

« [...] Oui, je travaille au sein d'un laboratoire : les échanges avec les chercheurs sont quotidiens. Le fonds numérisé est à l'origine un fonds constitué par et pour la recherche. La numérisation a lieu en partenariat [...] »

Quant aux services destinés aux chercheurs, à la question n° 10,

« Votre institution dispose-t-elle d'un guichet ou d'un service consacré aux chercheurs, et notamment aux demandes de numérisation sollicitées par les chercheurs ? »

284. Pour le moment, nous laissons de côté les bibliothèques relevant de la tutelle du ministère de la Culture et de la Communication ; nous analyserons le cas de la BnF dans la section suivante.

285. Nous avons consulté, entre autres, [Paris 3 - Direction des Bibliothèques Universitaires, 2016] et [SDC Paris 8, 2015] ; signalons aussi les résultats de l'enquête menée à Lille 3, où la notion de « bibliothécaires référent » a suscité un fort intérêt chez les étudiants comme chez les enseignants chercheurs, <https://insula.univ-lille3.fr/2013/04/enquete-publics-bibliotheque-universitaire-lille-3/>.

286. Voir aussi l'enquête sur la BML, [Bougéard-Pierron, 2016].

même si les bibliothécaires répondent pour la plupart par la négative (57,7 %), on notera que ce pourcentage est également reparti entre bibliothèques territoriales et bibliothèques de l'enseignement supérieur. Néanmoins, lorsqu'il s'agit d'identifier quels sont les services destinés à la recherche, l'écart entre les deux types d'institutions devient plus important :

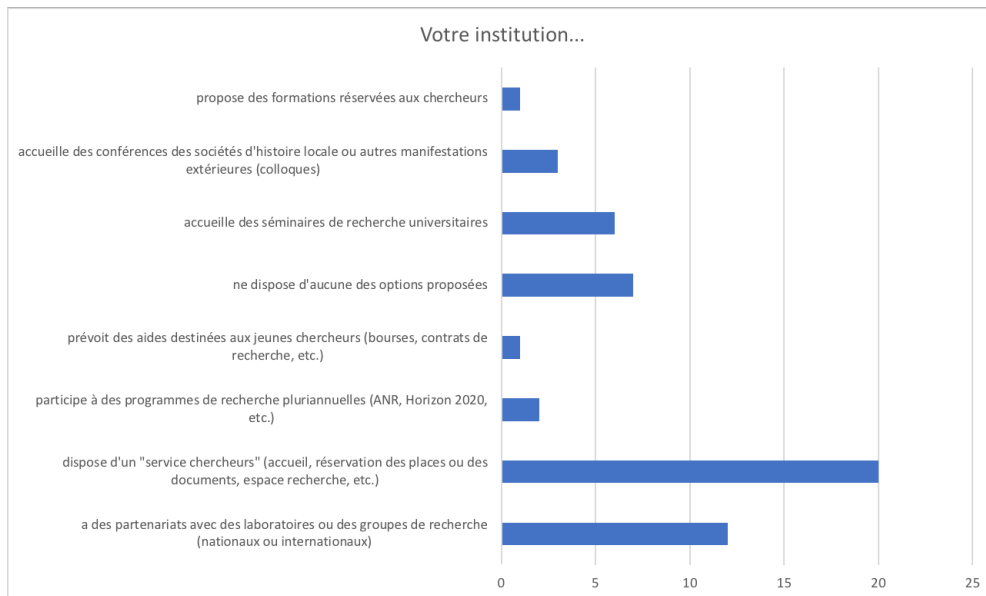


FIGURE 24 – Questionnaire aux bibliothécaires - question n° 23 - Bibliothèques des collectivités territoriales

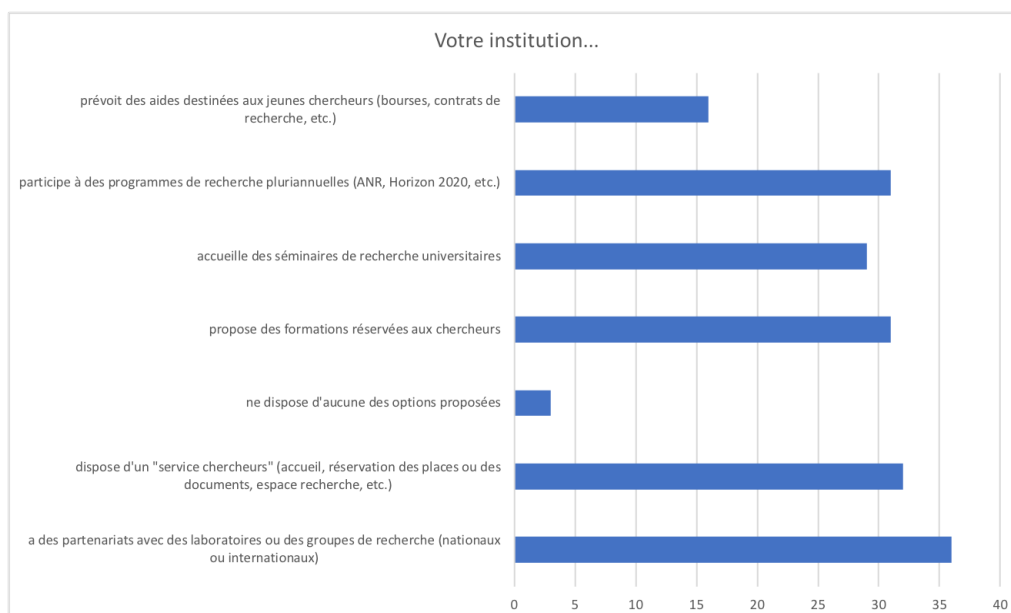


FIGURE 25 – Questionnaire aux bibliothécaires - question n° 23 - Bibliothèques de l'ESRI

Même si certaines bibliothèques de l'ESRI ont répondu au niveau de l'université, et non pas du service dont elles font partie, les deux groupes se distinguent surtout par rapport aux initiatives de formation aux chercheurs, au nombre de participations à des

programmes de recherche et au nombre de partenariats avec des laboratoires nationaux et internationaux, ces liens étant essentiels pour envisager des solides collaborations :

« Pour moi, un axe essentiel de la collaboration BU/chercheurs passe par deux leviers :

- tout ce qui se passe autour d'HumaNum, des consortiums, et d'une façon générale du CNRS. Nouer des liens avec le CNRS (via les documentalistes CNRS par exemple) est crucial pour comprendre comment fonctionnent les chercheurs et échanger avec des communautés de chercheurs intéressés
- le cadre des appels à projets (Région, ANR, H2020) est souvent un excellent moyen d'entrer en contact avec les chercheurs qui ne savent pas qu'ils ont besoin d'une expertise documentaire, mais sont très contents quand on propose spontanément nos services. De bons rapports avec la Direction de la Recherche permettent de faire ce lien » [Université Bordeaux Montaigne]

Autre élément qui revient fréquemment dans les réponses relatives aux bibliothèques relevant de l'ESRI et que nous retrouvons plus rarement dans le cadre des bibliothèques territoriales est celui d'une collaboration orienté à la didactique, permettant de faire connaître le corpus en question aux étudiants ou au doctorants :

« groupe de travail pour identification des documents / journées d'études / présentation à des étudiants dans le cadre de masters / etc. » [Collège de France]

« Sollicitation du SCD par les équipes de recherche / ou l'inverse en fonction des finalités des projets / Le cadre pédagogique est aussi essentiel pour des masters spécifiques ; la numérisation et surtout la valorisation des corpus numérique est aussi un enjeu pédagogique et d'initiation à la recherche » [Université de Caen Normandie]

« TD donné par un E[nseignant] C[hercheur] à des étudiants de L3 puis stages M1 encadrés par cet EC » [Bibliothèques de Nancy]

Bilan : « Très efficace si les chercheurs ont un besoin immédiat. Dans le cas des doctorants, crée une dépendance très forte avec la réussite et la poursuite de leurs études" » [BU]

Les dispositifs de financement

Du point de vue des financements, en dehors d'actions ponctuelles, d'initiatives nationales comme les programmes de numérisation concertée mis en place par le département de la Coopération de la BnF, ainsi que des initiatives rattachées au « Programme Investissements d'avenir » (PIA) que nous avons déjà évoquées, les bibliothèques des collectivités territoriales et celles de l'ESRI font appel à des dispositifs et des interlocuteurs différents ²⁸⁷.

Si plusieurs expérimentations existent pour mettre en place des modèles économiques participatifs de numérisation patrimoniale (institution de redevances, partenariats public-privé), ces derniers ne jouent qu'un rôle d'appoint par rapport aux financements publics [Maurel, 2017a]. Pour les bibliothèques en région, les financements peuvent être issus des collectivités (municipalité, conseil général, région, etc.), des contrats État-Région (CPER) ou des fonds régionaux d'acquisition des bibliothèques (FRAB) pour certaines régions,

287. Pour comprendre comment inscrire son projet dans les politiques nationales et régionales, voir le chapitre [Claerr, 2012] et les dispositifs de soutien cités par [Claerr et Westeel, 2013b, p. 32-38], avec les nécessaires mises à jour.

des crédits déconcentrés en région auprès de la DRAC, des plans de numérisation du ministère de la Culture et de la Communication (PAPE), ou encore, du concours particulier au sein de la dotation générale de décentralisation (DGD), notamment pour les bibliothèques numériques de référence. Pour les institutions relevant de l'ESRI, les contrats avec les organismes de recherche et l'attribution de crédits ponctuels sont la voie privilégiée pour la numérisation et la constitution de corpus scientifiques. Les appels à projets au titre des très grandes infrastructures de recherche (TGIR), de l'Agence nationale de la recherche (ANR), de la Bibliothèque scientifique numérique (BSN), ainsi que les conventions de pôle associé de la BnF, font également partie des aides à la numérisation.

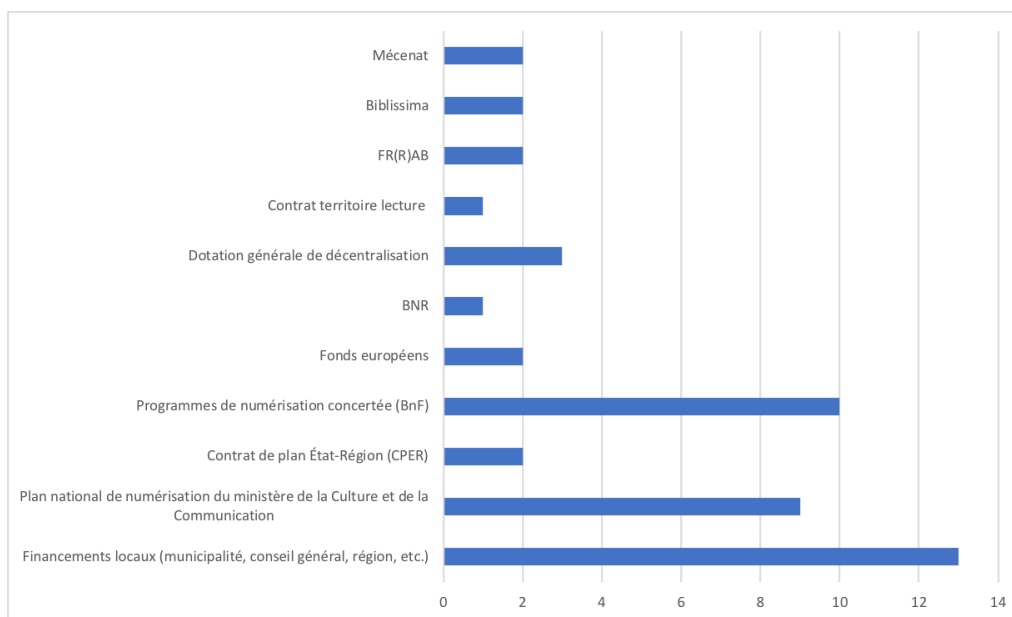


FIGURE 26 – Questionnaire aux bibliothécaires - question n° 22 - Bibliothèques des collectivités territoriales

À la question n° 22,

« Depuis 2012, quelles ont été les sources des financements accordées à votre institution pour la numérisation de corpus documentaires ? »

le dispositif le plus souvent évoqué par les bibliothécaires est celui des financements locaux.

Du côté des bibliothèques relevant de l'ESRI, la plupart des initiatives qui nous ont été rapportées s'inscrivaient dans le cadre d'une collaboration avec la BnF, en tant que pôle associé ou dans un programme de numérisation concertée ; les initiatives qui avaient reçu le soutien de la Bibliothèque scientifique numérique et des organismes de recherche étaient également bien représentés.

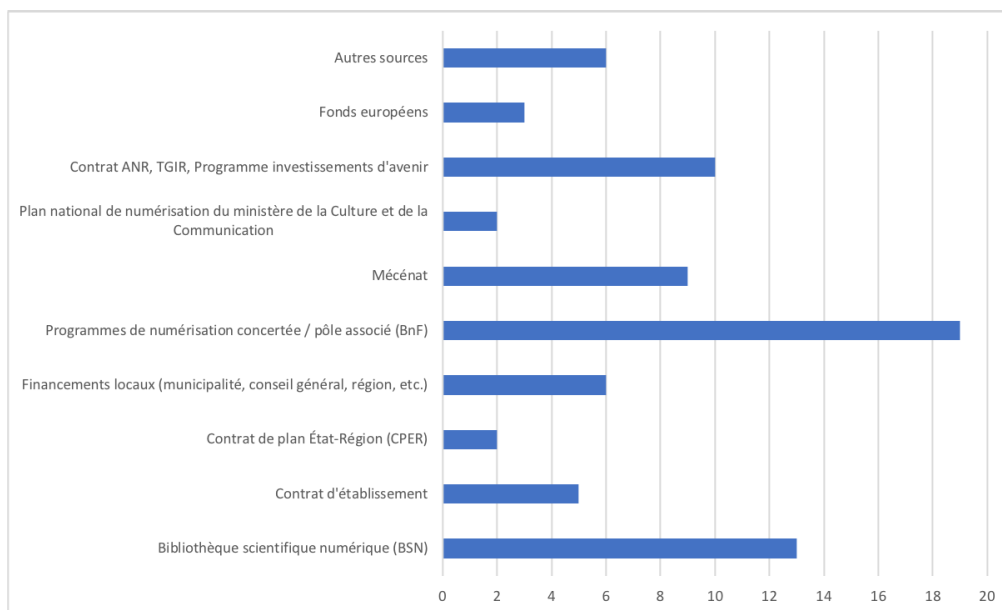


FIGURE 27 – Questionnaire aux bibliothécaires - question n° 22 - Bibliothèques de l'ESRI

Quoiqu'il mette en évidence deux situations de financement bien différentes, ce récapitulatif montre la forte implication de la BnF dans ce type de projets, à la fois pour les bibliothèques territoriales et pour celles de l'ESRI. Il montre également que, sauf quelques exceptions intéressantes, certains dispositifs comme les contrats avec les organismes de recherche, restent presque exclusivement possibles dans le cadre des bibliothèques de l'enseignement supérieur ; faut-il en déduire que cela est en rapport avec le manque de moyens et de services à la recherche dans les bibliothèques territoriales ?

« Ce serait bien en effet qu'il y ait plus de coopération. Nous avons déposé [***] un projet à l'Equipex Biblissima avec la BM de [***]. Le projet n'est pas passé parce que le rôle de la bibliothèque n'était pas assez clair/important. La BM est pourtant intéressée mais elle n'a pas les moyens de s'investir dans ce qui est le coeur de nos activités (édition, textes et outils) et l'appel Biblissima me paraît encore très centré sur les catalogues de bibliothèques, etc., c'est à dire sur ce qui intéresse traditionnellement les bibliothèques mais moins les chercheurs (ou moins de chercheurs) » [MCF]

Exemple de collaboration fructueuse : « La numérisation des manuscrits de Saint-Bertin, conservés principalement à la bibl. mun. de Saint-Omer dans le cadre d'une description par l'IRHT (soutenue par Biblissima) me semble être une belle opération [PU]

LES RÉPONSES DE LA BNF

Dans le cadre de la politique de coopération nationale, la Bibliothèque nationale de France est active depuis 2008 dans plusieurs programmes de numérisation et de valorisation du patrimoine documentaire des bibliothèques françaises. Les dispositifs mis en place comportent un soutien des programmes de numérisation concertée thématiques par le biais de l'attribution de subventions sur crédits du département de la Coopération²⁸⁸, ainsi qu'un soutien indirect, financé par le Centre national du livre, à travers l'intégration d'imprimés patrimoniaux des collections des bibliothèques françaises dans le marché de numérisation de la BnF. Ce soutien comprend aussi les projets bilatéraux de numérisation avec les pôles associés régionaux BnF; dans le cas de ces derniers, la BnF est aussi intervenue de manière occasionnelle pour quelques projets de numérisation de titres de presse locale dans le cadre du *Plan national de numérisation* du MCC; elle a également réalisé un état des lieux en 2013-2014 [Girard, 2014]. Ces programmes ont permis le traitement des documents sur les chaînes de numérisation de la BnF ainsi que la constitution de corpus numérisés qui ont intégré les collections numériques de Gallica et des établissements partenaires.

Au-delà de la numérisation, deux dispositifs récents et en cours d'évolution, permettant la mise en ligne de collections et l'exploration de corpus, ont été mis en place afin de répondre d'un côté aux exigences des institutions, de l'autre aux demandes des chercheurs.

Gallica Marque Blanche

Gallica Marque Blanche est un dispositif de coopération qui met actuellement à disposition cinq bibliothèques numériques : *Numistral*, réalisée en 2013 pour la Bibliothèque Nationale et Universitaire de Strasbourg²⁸⁹, le site de la *Grande Collecte* réalisé avec le SIAF (Service Interministériel des Archives de France) sur la période de la Grande Guerre²⁹⁰, *Rotomagus* pour la bibliothèque municipale de la ville de Rouen²⁹¹, le RFN *Réseau Francophone Numérique*²⁹² et la *Bibliothèque diplomatique numérique* pour le Ministère de l'Europe et des Affaires étrangères²⁹³. Plusieurs autres projets sont en cours, dont une Marque Blanche consacrée au projet Polonsky, visant à mettre en ligne tous les manuscrits numérisés dans le cadre de ce projet de mécénat²⁹⁴, intégrant aussi un visualiseur IIF (*International Image Interoperability Framework*) et la gestion du multilinguisme.

288. Par exemple, avec la Bibliothèque Interuniversitaire Cujas pour les sciences juridiques ou en histoire de l'art avec la Bibliothèque de l'Institut national d'Histoire de l'art (INHA). Voir aussi les résultats de l'appel à initiatives 2015-2017 : <https://lc.cx/Ahpe>; pour un état des lieux des numérisation concertées, voir [Girard, 2015].

289. Voir : <http://www.numistral.fr>.

290. Voir : <http://www.lagrandecollecte.fr>.

291. Voir : <http://www.rotomagus.fr>.

292. Voir : <http://rfnum-bibliotheque.org>.

293. Voir : <http://bibliotheque-numerique.diplomatie.gouv.fr>, (lancement février 2018).

294. Pour les actualités du projet « France et Angleterre, 700-1200 : manuscrits de la British Library et de la Bibliothèque nationale de France », voir le carnet de recherche <https://manuscripta.hypotheses.org/category/programmes-de-recherche/polonsky>.

Comme il s'agit d'un dispositif de coopération, les collections proposées doivent être en lien avec celles de Gallica, tant sur le plan de la ligne éditoriale que des formats :

« L'offre reste néanmoins soumise à deux conditions d'accès essentielles : l'appréciation scientifique des collections numérisées proposées par les partenaires et la capacité d'intégration de ces documents dans la bibliothèque numérique » [BnF, 2016, p. 1]

En cours de développement, ce dispositif reçoit de nombreuses demandes ; actuellement, l'objectif de la BnF est de réaliser environ quatre Marques Blanches par an. Cette coopération numérique s'adresse avant tout à des établissements qui ont numérisé ou qui sont en train de numériser leurs collections, et qui souhaitent créer ou faire évoluer leur bibliothèque numérique ; une attention particulière est consacrée aux BNR (bibliothèques numériques de référence). La BnF propose aux partenaires une infrastructure solide, celle de Gallica, et s'occupe de la maintenance et de l'évolution de la plateforme, ainsi que de l'archivage des fichiers numériques dans SPAR (Système de Préservation et d'Archivage Réparti)²⁹⁵. Cette solution s'avère intéressante dans la mesure où elle permet d'avoir une bibliothèque numérique techniquement à jour, mise aux couleurs du partenaire et intégrant le système de recherche, de visualisation, d'affichage des résultats ainsi que les outils de recherche avancée de Gallica. L'océrisation de tous les documents fera bientôt partie des services intégrés. Signalons que ce dispositif permet aux partenaires de gagner en visibilité : les notices des documents numérisées sont intégrés et moissonnés dans les catalogues BnF (BnF Catalogue Général ou BnF Archives et Manuscrits). De plus, dans un but de cohérence des collections proposées, le dispositif peut intégrer des collections issues d'autres bibliothèques (pour le moment, seulement depuis les bibliothèques qui sont déjà moissonnées dans Gallica).

Pour garantir l'intégration et la pérennité des données, une phase initiale de coordination est importante. La concertation porte d'abord sur l'organisation du site : des créations complémentaires ou des développements spécifiques peuvent être intégrés en passant aussi par des prestataires ; les notices sont ensuite intégrées dans les catalogues BnF, ce qui peut demander une mise à niveau et la mise en place de notices d'autorité ; enfin, les documents doivent être intégrés dans l'entrepôt, nommés, organisés et structurés selon les conventions de la BnF. Comme cet engagement de ressources est important et que le temps de démarrage se révèle relativement long, ce dispositif se prête surtout à des projets de collaboration de longue durée (au moins 4/5 ans) et non, par exemple, à des projets liés à des expositions ou à des événements. Concernant les coûts de mise en œuvre, la convention Marque Blanche prévoit un principe de co-financement à 50/50, proportionnel à la volumétrie de documents numériques à intégrer, aussi bien pour la réalisation de la bibliothèque numérique que pour sa maintenance et son hébergement.

D'autre part, cette solution peut également comporter des limites et des inconvénients. Si le partenaire n'a pas à gérer les questions techniques, il n'a pas non plus la main sur l'ensemble des contenus du site et sur leur éventuelle éditorialisation ; pour le moment, tous les enrichissements et les mises à jour sont effectués par la BnF ; néanmoins, il est prévu que le système d'enrichissement Drupal qui est utilisé pour gérer le site puisse partiellement être ouvert aux partenaires qui ont une politique éditoriale particulièrement

295. Sur le système SPAR, voir : http://www.bnf.fr/fr/professionnels/spar_systeme_preservation_numerique.html ; à propos des formats de données pour la préservation numérique à la BnF, voir : http://www.bnf.fr/fr/professionnels/preservation_numerique_boite_outils/a.pres_num_formats.html.

active. La mise à niveau des données de la bibliothèque partenaire selon le schéma Dublin Core utilisé par Gallica peut également impliquer une perte de richesse des données d'entrée (par exemple lors de la conversion de métadonnées MODS vers le Dublin Core).

Tout en présentant un intérêt réel pour les bibliothèques partenaires, signalons aussi que du point de vue de la BnF, Marque Blanche constitue une sorte de « terrain de test » : les solutions qui sont testées dans ce cadre peuvent ensuite être mises en production dans Gallica. De plus, cette typologie de mise en ligne des contenus peut également être profitable aux chercheurs ; face à l'hétérogénéité des bibliothèques numériques existantes, affichant des formats et des stratégies de recherche multiples, Marque Blanche repose sur un modèle simple, connu et entretenu.

Le projet CORPUS

Mené sous la direction d'Emmanuelle Bermès et inscrit au plan quadriennal de recherche de la BnF (2016-2019), le projet CORPUS vise à préfigurer un service de fourniture de corpus numériques à destination de la recherche, et à fournir aux chercheurs des données et des outils d'analyse dans le respect du droit d'auteur et de la vie privée²⁹⁶. Trois grands corpus sont concernés dans le cadre de ce projet : Gallica et Gallica *intra-muros* (4,5 millions de documents numériques) ; les Archives de l'internet (793 To) ; les métadonnées BnF (plus de 20 millions de notices des catalogues), exposées dans *data.bnf.fr* et accessibles à travers le portail « BnF API et jeux de données ».

Les projets de partenariat menés dans le cadre du projet CORPUS témoignent d'une stricte collaboration entre bibliothécaires, équipes techniques et chercheurs, comme dans le cas du partenariat avec l'équipe ANR Web90 (CNRS/ISCC)²⁹⁷. Comme l'explique Emmanuelle Bermès dans un article faisant suite au *IFLA Satellite Meeting* d'août 2017 [Bermès, 2017b], ainsi que dans sa conférence lors du dernier congrès de l'ADBU [Bermès, 2017a], ce projet est né pour répondre aux nouvelles demandes que les chercheurs ont commencé à adresser à la BnF autour de ses collections numériques : « au lieu de s'intéresser aux collections numériques "à l'unité", ils s'y intéressaient de façon de plus en plus massive ». Selon Emmanuelle Bermès, on peut distinguer trois cas d'usage différents dans ce nouveau type de questionnement. Le premier, le plus fréquent, est « le corpus comme source », c'est-à-dire le corpus dont le contenu est tellement intéressant qu'il mérite d'être fouillé pour faire émerger de nouvelles problématiques de recherche ; c'est le cas, par exemple, de la presse numérisée et du projet de fouille *The Europeana Newspapers Project*, sur lequel est intervenu Jean-Philippe Moreux, chercheur et employé de la BnF [Moreux, 2016]. Le deuxième cas, est celui du « corpus comme bac à sable », dans lequel une équipe de recherche demande un corpus, non pas pour son contenu, mais pour sa nature, par exemple pour effectuer des tests d'algorithmes ; citons le cas de l'équipe ETIS de l'université Cergy Pontoise, dont le but était de tester un algorithme d'indexation automatique sur un ensemble hétérogène d'images²⁹⁸ ; cette initiative rencontrait parallèlement les intérêts de la BnF quant à la conception d'un outil

296. Le projet est organisé sur quatre ans : trois ans d'expérimentation transversale sur l'ensemble des collections de la BnF et une quatrième année consacrée au bilan et l'idéation du nouveau service, mais figurant déjà dans le Contrat d'objectifs et de performances 2017-2021 [BnF, 2017, p. 46-47]. Pour le descriptif du projet, voir le site de la BnF : <http://c.bnf.fr/fom>, ainsi que le billet de blog [Jacquot, 2017]. Mise à jour, mai 2018 : voir le rapport de [Moiraghi, 2018].

297. Voir : <http://c.bnf.fr/fse> et <http://webcorpora.hypotheses.org/>.

298. Pour plus de détails techniques, voir [Picard *et al.*, 2015].

pour l'indexation semi-automatique des images. Le troisième cas d'usage est celui du « corpus comme interface », quand, dans une démarche autre que l'étude de corpus, on se rend compte que ce dernier est le meilleur outil pour répondre aux enjeux de la recherche scientifique en question ; c'est le cas du projet « Le devenir en ligne du patrimoine numérisé : l'exemple de la Grande Guerre », mené en collaboration entre la BnF, la BDIC et Télécom ParisTech dans le cadre du LabEx *Les passés dans le présent* ; ce projet avait pour vocation « d'étudier les pratiques sociales en ligne qui visent à construire une représentation du passé et à entretenir la mémoire de la guerre » [Beaudouin et Pehlivan, 2017, p. 3] : il s'est avéré que la solution la plus efficace était de construire conjointement un corpus web, qui a ainsi permis de cartographier le réseau des relations entre les sites liés à la Grande Guerre.

Comme l'explique Emmanuelle Bermès [Bermès, 2017a], les projets menés dans le cadre du programme CORPUS affichent une grande variété ; l'hétérogénéité des compétences informatiques des chercheurs semble à l'origine de la diversité des réponses qu'ils s'attendent à trouver de la part des bibliothèques : elle doit être prise en compte dans la modulation de l'offre de services. Si certains chercheurs ont besoin d'être accompagnés dans la connaissance des données, des outils et des formats, s'ils demandent des interfaces et des données pré-traitées, d'autres, beaucoup plus à l'aise avec le numérique, sont surtout intéressés par les données brutes et par la possibilité de faire tourner leurs outils avec les données mises à disposition par la bibliothèque. « Il y a un besoin autour des corpus numériques et, *a priori*, ce besoin ne fera qu'augmenter ». Dans le cadre de la réflexion menée autour de la création de ce service, la BnF s'interroge sur plusieurs éléments. L'espace : est-il important de disposer d'un lieu pour se retrouver ? Les chercheurs semblent manifester le souhait d'échanger avec les membres de la bibliothèque. Le volet juridique se révèle également important : certaines données sont accessibles *intra-muros* uniquement, et pas en accès distant. En matière d'infrastructure informatique : les chercheurs veulent un poste sur lequel ils peuvent avoir la main, mais aussi des outils pré-installés et la possibilité d'être guidés ; est-il possible de mettre en place une plateforme, un environnement de travail capable de répondre à tout chercheur « un peu sur le modèle du Fab Lab » ? En termes de ressources humaines : faut-il des équipes dédiées pour accompagner les projets ? Se posent enfin des questions de compétences et de savoir-faire : il faudrait faire en sorte que les chercheurs partagent leurs connaissances et leur expertise, sans que la bibliothèque ait à gérer l'intégralité de la formation.

CONCLUSION

Nous nous sommes intéressés dans ce mémoire aux collaborations entre chercheurs et bibliothécaires dans le cadre de projets de numérisation de corpus documentaires. En choisissant ce sujet, portant sur un cas de collaboration bien précis, nous espérions avoir affaire à un nombre relativement restreint de problématiques et de cas de figure. Toutefois, au fil de notre travail et de nos échanges avec les acteurs de ces collaborations, nous avons été confrontés aux multiples et vastes horizons ouverts par ce sujet d'actualité.

Les corpus intéressent aussi bien les chercheurs et les bibliothécaires, que les ingénieurs et les professionnels de l'information des institutions universitaires, patrimoniales et culturelles. Les chercheurs ont conscience que les données massives permettent d'exposer de nouvelles possibilités de recherche ; ils sont attirés par l'exploitation et le partage de grands corpus documentaires, constituant en eux-mêmes de nouveaux objets de recherche ; les compétences, les ressources ou les droits pour accéder ou exploiter ces gisements de documents et de données leur font parfois défaut. Les institutions, quant à elles, sont conscientes que la numérisation de corpus participe du rayonnement des collections et de la constitution d'une nouvelle offre de services ; elles sont confrontées à la mise à disposition des données selon des logiques d'ouverture et d'interopérabilité, ainsi qu'aux questions que cela soulève en termes techniques, déontologiques et juridiques. Les profils d'ingénieurs et de spécialistes en Humanités numériques, susceptibles de faire le lien entre les équipes de recherche et les institutions patrimoniales et culturelles, se développent pour leur part de plus en plus, malgré une situation de précarité contractuelle et de positionnement professionnel encore incertain. Un équilibre selon lequel tous les acteurs trouveraient des bénéfices à leurs démarches n'est pas simple à établir.

S'inscrivant dans un contexte de pluridisciplinarité et sollicitant différents métiers, le traitement de corpus documentaires implique d'adopter une posture transversale et de mettre en œuvre une capacité d'écoute face aux besoins des différents acteurs impliqués. Si cet objet d'étude est fécond, tant en termes de réalisations techniques que d'enjeux scientifiques, le dialogue entre équipes de recherche et professionnels autour de « corpus intelligents » reste encore à construire ; en effet, issus de pratiques différentes, ces interlocuteurs ne partagent souvent pas les mêmes codes et le même vocabulaire, et peinent encore à se rencontrer et à établir des collaborations. La gestion du temps et les résultats attendus ne sont pas les mêmes de part et d'autre. Néanmoins, si la prise de contact et la mise en relation entre ces différents acteurs ne sont pas toujours faciles, les retours d'expérience montrent que les efforts de rapprochement mutuel amènent à des résultats encourageants et à des collaborations réitérées. Une bonne connaissance des modalités actuelles de la recherche scientifique, appuyée par des instruments et des infrastructures, permet de mieux répondre aux besoins et aux intérêts des chercheurs, ainsi que d'articuler politique de numérisation et politique de recherche. De même, la connaissance des pratiques professionnelles et des modes de classement et d'accès à l'information à l'intérieur des institutions patrimoniales et culturelles permet aux chercheurs d'identifier des gisements de données, des leviers de valorisation scientifique, et d'établir un dialogue fructueux en amont du lancement du projet.

La typologie des projets et des collaborations observés montre leur grande variété, qui est la conséquence de la variété des besoins, des moyens et des compétences des chercheurs, des professionnels et des institutions. Si le défi des professionnels des bibliothèques ne consiste pas simplement à préserver les fonds, mais encore à solliciter

un nouveau regard sur eux, les techniques d'investigation des données ouvrent de nouvelles perspectives, tout en impliquant un véritable questionnement à propos de l'évolution des bibliothèques numériques, des compétences scientifiques et informatiques des professionnels des bibliothèques, des services offerts aux chercheurs pour accéder de manière profitable aux jeux de données et de métadonnées. La numérisation de corpus n'est effectivement que l'un des services pour lesquels les bibliothèques peuvent être sollicitées, les demandes pouvant inclure l'aide à l'interprétation des données des catalogues, la valorisation d'une collection de documents ou d'un travail de recherche, l'expertise sur les formats, sur les solutions de pérennisation des projets, sur la gestion des droits, l'aide à la rédaction d'un plan de gestion de données, etc. Touchant, à différents titres, à l'ensemble de ces pratiques, les corpus constituent une occasion d'échange, un terrain de rencontre et de collaboration stimulant et prometteur entre chercheurs, bibliothécaires, « humanistes numériques » et professionnels de l'information.

BIBLIOGRAPHIE

Toutes les adresses des sites Internet ont été vérifiées le 1^{er} mars 2018.

- [ADEMEC, 2017] ADEMEC (2017). Table ronde : Quel renouvellement des formes de collaboration entre chercheurs et institutions patrimoniales ? Journée d'études. En ligne : <<https://lc.cx/QapS>>.
- [Alberani, 2008] ALBERANI, V., éditeur (2008). *Manuale/ dizionario della biblioteconomia e delle scienze dell'informazione*, volume 1 : Indicizzazione e recupero semantico dell'informazione. Associazione italiane biblioteche, Roma.
- [Alix, 2013] ALIX, Y., éditeur (2013). *Le métier de bibliothécaire*. Éditions du Cercle de la Librairie, Paris, 2e édition.
- [Ambrosino *et al.*, 2015] AMBROSINO, M., BUREAU, B., CASTALDI, G., FASCIONE, S., NICOLARDI, F. et NICOLAS, C. (2015). Hyperdonat de Lyon à Naples. In PETY, D., éditeur : *Patrimoine littéraire en ligne : la renaissance du lecteur ?*, Corpus, pages 25–45. Université de Savoie Mont Blanc, Chambéry.
- [Arènes, 2015] ARÈNES, C. (2015). Les modes de communication de la recherche aujourd'hui : quel rôle pour les bibliothécaires ? Mémoire pour le diplôme de conservateur des bibliothèques, Enssib, Villeurbanne. En ligne : <<http://www.enssib.fr/bibliotheque-numerique/documents/65046-les-modes-de-communication-de-la-recherche-aujourd-hui-quel-role-pour-les-bibliothecaires.pdf>>.
- [Artières, 2016] ARTIÈRES, P. (2016). Du dépouillement au dévoilement. L'historien et les archives numérisées. In *La recherche dans les institutions patrimoniales. Sources matérielles et ressources numériques*, Papiers, pages 18–27. Presses de l'Enssib, Villeurbanne.
- [Baillot *et al.*, 2017] BAILLOT, A., PUREN, M., RIONDET, C. et ROMARY, L. (2017). Access to Cultural Heritage Data : A Challenge for the Digital Humanities. Montréal, Canada. En ligne : <<https://dh2017.adho.org/abstracts/345/345.pdf>>.
- [Balle, 2006] BALLE, F., éditeur (2006). *Lexique d'information communication*. Dalloz, Paris.
- [Barthélemy *et al.*, 2017] BARTHÉLEMY, C., KOSKAS, C. et FLEPP, P. (2017). L'exploitation scientifique d'un corpus numérique épistolaire. Projet HyperPaulhan, projet Ponge. In *Des humanités numériques littéraires ?*, Cerisy. LabEx Obvil. En ligne : <<http://obvil.sorbonne-universite.site/des-humanites-numeriques-litteraires-actes-du-colloque-de-cerisy>>.
- [Baudry, 2016] BAUDRY, J. (2016). Exposition : Madagascar – Cartographie d'hier et cartographie d'aujourd'hui. En ligne : <<http://imageo.hypotheses.org/49>>.
- [Baudry, 2017] BAUDRY, J. (2017). Le rôles des bibliothèques universitaires dans l'utilisation des données patrimoniales par les chercheurs. Journée ADEMEC, Paris. En ligne : <<https://lc.cx/QaJj>>.
- [Bauer et Kaltenböck, 2012] BAUER, F. et KALTENBÖCK, M. (2012). *Linked Open Data : The Essentials. A Quick Start Guide for Decision Makers*. DGS – Druck-u. Graphikservice GmbH, Wien.

- [Beaudouin et Pehlivan, 2017] BEAUDOUIN, V. et PEHLIVAN, Z. (2017). Cartographie de la Grande Guerre sur le Web : Rapport final de la phase 2 du projet "Le devenir en ligne du patrimoine numérisé : l'exemple de la Grande Guerre". Bibliothèque nationale de France, Bibliothèque de documentation internationale contemporaine, Télécom ParisTech. En ligne : <<https://hal.archives-ouvertes.fr/hal-01425600>>.
- [Bensamoun et Latil, 2015] BENSAMOUN, A. et LATIL, A., éditeurs (2015). *Propriété littéraire et artistique et humanités numériques*, Collection des Presses Universitaires de Sceaux, Paris. Mare & Martin.
- [Béquet, 2014] BÉQUET, G. (2014). *Trois bibliothèques européennes face à Google : aux origines de la Bibliothèque numérique (1990-2010)*. Mémoires et documents de l'École des chartes : 97. École nationale des chartes, Paris.
- [Bermès, 2017a] BERMÈS, E. (2017a). Fournir des corpus à destination de la recherche à des fins de TDM : le projet CORPUS, BnF. 47ème congrès de l'ADBU, Lille. En ligne : <<https://lc.cx/AJVK>>.
- [Bermès, 2017b] BERMÈS, E. (2017b). Text, data and link-mining in digital libraries : looking for the heritage gold. IFLA Satellite Meeting - Digital Humanities – Opportunities and Risks : Connecting Libraries and Research, Berlin. En ligne : <<https://hal.inria.fr/hal-01643293/>>.
- [Berra, 2012] BERRA, A. (2012). Faire des humanités numériques. In MOUNIER, P., éditeur : *Read/Write Book 2 : Une introduction aux humanités numériques*, pages 25–43. OpenEdition Press. En ligne : <<http://books.openedition.org/oepe/238>>.
- [Berra, 2015] BERRA, A. (2015). Pour une histoire des humanités numériques. *Critique*, 819-820(8):613–626. En ligne : <<https://www.cairn.info/revue-critique-2015-8-page-613.htm>>.
- [Berra et Mounier, 2016] BERRA, A. et MOUNIER, P. (2016). Humanités numériques : quelle(s) critique(s) ? In *Humanités numériques : théories, débats, approches critiques*, Lille. En ligne : <<https://lc.cx/ATnQ>>.
- [BnF, 2016] BNF (2016). Gallica Marque Blanche. Rapport technique, Bibliothèque nationale de France - DSR/DCP Service Coopération numérique et Gallica. En ligne : <http://www.bnf.fr/documents/gallica_marque_blanche.pdf>.
- [BnF, 2017] BNF (2017). Contrat d'objectifs et de performance 2017-2021. Rapport technique, Bibliothèque nationale de France. En ligne : <http://www.bnf.fr/documents/contrat_performance.pdf>.
- [Bougeard-Pierron, 2016] BOUGEARD-PIERRON, S. (2016). Enquête qualitative 2016. Rapport technique, Bibliothèque municipale de Lyon. En ligne : <<http://www.enssib.fr/bibliotheque-numerique/documents/66730-enquete-qualitative-2016-bibliotheque-municipale-de-lyon.pdf>>.
- [Boullier, 2015] BOULLIER, D. (2015). Les sciences sociales face aux traces du big data ? Société, opinion, répliques. *Revue française de science politique*, 65(5):805–828. En ligne : <<https://www.cairn.info/revue-francaise-de-science-politique-2015-5-page-805.htm>>.
- [Bouqui, 2015] BOUQUI, C. (2015). Du corpus à l'outil herméneutique : la base de données intertextuelle Molière 21. In PETY, D., éditeur : *Patrimoine littéraire en ligne : la renaissance du lecteur ?*, Corpus, pages 57–68. Université de Savoie Mont Blanc, Chambéry.

- [Bouvier-Ajam *et al.*, 2011] BOUVIER-AJAM, L., CHEVALLIER, P. et RIOUST, L. (2011). La consultation de manuscrits en ligne : Une étude qualitative auprès de trois catégories d'utilisateurs potentiels. *Bulletin des bibliothèques de France*, 56(5):17–23. En ligne : <<https://hal.archives-ouvertes.fr/hal-00732154>>.
- [Brault, 2016] BRAULT, J. (2016). Renseigner, orienter le chercheur. Expériences de service public. In ROUSTAN, M., MONJARET, A. et CHEVALLIER, P., éditeurs : *La recherche dans les institutions patrimoniales. Sources matérielles et ressources numériques*, Papiers, pages 28–44. Presses de l'Enssib, Villeurbanne.
- [Briatte, 2014] BRIATTE, K. (2014). HADOC : Penser la production des données culturelles dans un contexte d'interopérabilité sémantique et organisationnelle. En ligne : <<https://lc.cx/QaUM>>.
- [Burdick, 2012] BURDICK, A. (2012). *Digital Humanities*. The Massachusetts Institute of Technology Press, Cambridge, Massachusetts and London, England.
- [Cacaly, 2008] CACALY, S., éditeur (2008). *Dictionnaire de l'information*. A. Colin, Paris.
- [Calderan *et al.*, 2014] CALDERAN, L., LAURENT, P. et LOWINGER, H., éditeurs (2014). *Big Data. Nouvelles partitions de l'information. Actes du séminaire INRIA - IST (octobre 2014)*. Information et stratégie. De Boeck Supérieur, Louvain-la-Neuve. En ligne : <<https://www.cairn.info/big-data-9782804189150.htm>>.
- [Calimaq, 2017] CALIMAQ (2017). L'Open Data culturel est possible (et les bibliothèques le prouvent !). *S.I.Lex - Carnet de veille et de réflexion d'un juriste et bibliothécaire*. En ligne : <<https://scinfolex.com/2017/04/13/lopen-data-culturel-est-possible-et-les-bibliotheques-le-prouvent/>>.
- [Capitani, 2005] CAPITANI, P. (2005). Terminologia controllata : lavori in corso. L'importanza del thesaurus al tempo di Internet. *Biblioteche oggi*, 23(5):41–49.
- [Chartier, 2012] CHARTIER, R. (2012). Qu'est-ce qu'un livre ? Métaphores anciennes, concepts des lumières et réalités numériques. *Le français aujourd'hui*, 3(178):11–26. En ligne : <<https://lc.cx/AJjm>>.
- [Chevallier, 2016] CHEVALLIER, P. (2016). Le patrimoine numérisé : nouveaux usages, nouveaux usagers ? Réflexions méthodologiques autour de Gallica. In ROUSTAN, M., MONJARET, A. et CHEVALLIER, P., éditeurs : *La recherche dans les institutions patrimoniales. Sources matérielles et ressources numériques*, Papiers, pages 45–57. Presses de l'Enssib, Villeurbanne.
- [Chevallier et Amar, 2014] CHEVALLIER, P. et AMAR, M. (2014). Les usages des corpus numérisés de Gallica sur la Grande Guerre. Rapport technique, Bibliothèque nationale de France. En ligne : <<https://halshs.archives-ouvertes.fr/hal-01056704/>>.
- [Chevry Pébayle, 2012] CHEVRY PÉBAYLE, E. (2012). Valorisation des corpus numérisés. *Les Cahiers du numérique*, 8(3). En ligne : <<https://www.cairn.info/revue-les-cahiers-du-numerique-2012-3.htm>>.
- [Claerr, 2012] CLAERR, T. (2012). Inscrire son projet dans les politiques patrimoniales nationales et régionales. In COQ, D., éditeur : *Apprendre à gérer des collections patrimoniales en bibliothèque*, volume 26 de *Boîte à outils*, pages 152–164. Presses de l'Enssib, Villeurbanne. En ligne : <<http://books.openedition.org/pressesenssib/673>>.

- [Claerr et Westeel, 2013a] CLAERR, T. et WESTEEL, I. (2013a). Concevoir la bibliothèque numérique en prenant en compte les usages numériques et les évolutions technologiques. In *Manuel de constitution de bibliothèques numériques*, pages 39–89. Éditions du Cercle de la Librairie, Paris.
- [Claerr et Westeel, 2013b] CLAERR, T. et WESTEEL, I. (2013b). *Manuel de constitution de bibliothèques numériques*. Collection Bibliothèques. Éditions du Cercle de la Librairie, Paris.
- [Clavert et al., 2012] CLAVERT, F., BERRA, A. et HEIMBURGER, F. (2012). L'historien programmeur? In *THATCamp Paris 2012. Non-actes de la non-conférence des humanités numériques*. Édition de la Maison des sciences de l'homme, Paris. En ligne : <<http://books.openedition.org/editionsmsh/305>>.
- [Clivaz et al., 2012] CLIVAZ, C., MEIZOZ, J., VALLOTTON, F. et VERHEYDEN, J., éditeurs (2012). *Reading Tomorrow. From Ancient Manuscripts to the Digital Era / Lire Demain. Des manuscrits antiques à l'ère digitale*. Presses Polytechniques et Universitaires Romandes, Lausanne.
- [Collart, 2011] COLLART, C. (2011). L'offre numérique. In COLLART, C. et MELOT, M., éditeurs : *Images et bibliothèques*, Bibliothèques, pages 87–99. Éditions du Cercle de la Librairie, Paris.
- [Comby et al., 2016] COMBY, É., YANNICK, M. et de CARRARA, S., éditeurs (2016). *Corpus de textes : composer, mesurer, interpréter*. Sociétés, espaces, temps. ENS Éditions, Lyon.
- [Courtin, 2014] COURTIN, A. (2014). Brève introduction au Linked Open Data [appliqué aux institutions culturelles]. En ligne : <<https://fr.slideshare.net/antoinecourtin/brve-introduction-au-linked-open-data-appliqu-aux-institutions-culturelles>>.
- [Dacos, 2011] DACOS, M. (2011). Manifeste des Digital Humanities. *THATCamp*. En ligne : <<http://tcp.hypotheses.org/318>>.
- [Daire, 2016] DAIRE, J. (2016). De la politique d'accès à la coopération scientifique. chercheurs et collections patrimoniales à la Cinémathèque française. In *La recherche dans les institutions patrimoniales. Sources matérielles et ressources numériques*, Papiers, pages 125–136. Presses de l'Enssib, Villeurbanne.
- [Dégez-Selves et Buxtorf, 2017] DÉGEZ-SELVES, C. et BUXTORF, A.-É. (2017). Riche lieu : les portes s'ouvrent sur l'art et l'histoire. *Ar(abes)ques*, 85:22–23.
- [DISIC, 2015] DISIC (2015). Référentiel Général d'Interopérabilité. Standardiser, s'aligner et se focaliser pour échanger efficacement. Rapport technique Version 1.9.7, Direction Interministérielle des Systèmes d'Information et de Communication. En ligne : <<https://lc.cx/AJvb>>.
- [DLWeb INA, 2017] DLWEB INA (2017). La collecte et consultation des tweets à l'INA. En ligne : <<https://lc.cx/QaNk>>.
- [Domange, 2013] DOMANGE, C. (2013). Ouverture et partage des données publiques culturelles pour une (r)évolution numérique dans le secteur culturel. Rapport technique, Ministère de la Culture et de la Communication.
- [Doueïhi, 2010] DOUEIHI, M. (2010). Le livre à l'heure du numérique : objet fétiche, objet de résistance. In DACOS, M., éditeur : *Read/Write Book. Le livre inscriptible*, pages 95–103. OpenEdition Press, Marseille. En ligne : <<http://books.openedition.org/oep/155>>.

- [Doueïhi, 2011] DOUEIHI, M. (2011). *Pour un humanisme numérique*. Éditions du Seuil, Paris.
- [Dupuis, 2015] DUPUIS, M. (2015). L'application du droit d'auteur aux corpus numériques. In BENSAMOUN, A. et LATIL, A., éditeurs : *Propriété littéraire et artistique et humanités numériques*, pages 197–211. Mare & Martin, Paris.
- [Duteil-Mougel et Foulquié, 2006] DUTEIL-MOUGEL, C. et FOULQUIÉ, B., éditeurs (2006). *Corpus en Lettres et Sciences sociales. Des documents numériques à l'interprétation. Colloque international et école d'été*, Albi. Colloques d'Albi Langages et Signification (CALS).
- [Duval, 2006] DUVAL, F. (2006). La philologie française, pragmatique avant tout ? In DUVAL, F., éditeur : *Pratiques philologiques en Europe*, pages 115–150. École nationale des chartes, Paris.
- [Duval, 2017] DUVAL, F. (2017). Pour des éditions critiques numériques. L'exemple des textes français. *Médiévales*, 73:13–30.
- [Ego, 2015] EGO, A. (2015). Exception pédagogique et corpus numérique. In BENSAMOUN, A. et LATIL, A., éditeurs : *Propriété littéraire et artistique et humanités numériques*, pages 73–79. Mare & Martin, Paris.
- [Évrard, 2002] ÉVRARD, É. (2002). Réflexions sur les méthodes quantitatives en domaine littéraire. *Médiévales*, 42:89–100. En ligne : <www.persee.fr/doc/medi_0751-2708_2002_num_21_42_1541>.
- [Feather et Sturges, 2003] FEATHER, J. et STURGES, P., éditeurs (2003). *International encyclopedia of information and library science*. Routledge, London - New York, 2e édition.
- [Ferjoux et Dupond, 2017] FERJOUX, C. et DUPOND, M. (2017). Usages des patrimoines numérisés. Un réseau interdisciplinaire et interinstitutionnel. Mise en réseau d'objectifs, d'expériences, de pratiques et de méthodes 2014-2016 : postures, études et perspectives. Rapport technique, UDPN - Université Sorbonne Paris Cité, Paris.
- [Fiormonte, 2016] FIORMONTE, D. (2016). Toward a cultural critique of digital humanities. In GOLD, M. K., éditeur : *Debates in the Digital Humanities*. University of Minnesota Press. En ligne : <<http://dhdebates.gc.cuny.edu/debates/text/86>>.
- [Gaillard, 2014] GAILLARD, R. (2014). De l'open data à l'open research data : quelle(s) politique(s) pour les données de la recherche ? Mémoire pour le diplôme de conservateur des bibliothèques, Enssib, Villeurbanne.
- [Gasc, 2010] GASC, M. (2010). Concevoir et gérer des bibliothèques en milieu universitaire. *Bulletin des bibliothèques de France*, 2:69–74. En ligne : <<http://bbf.enssib.fr/consulter/bbf-2010-02-0069-001>>.
- [Gaudet, 2011] GAUDET, F. (2011). L'enquête de public en 10 leçons et 1/2. In EVANS, C., éditeur : *Mener l'enquête. Guide des études de publics en bibliothèque*, numéro 22 de La boîte à outils. Presses de l'Enssib, Villeurbanne.
- [Gillium, 2016] GILLIUM, J. (2016). Big data et bibliothèques : traitement et analyse informatique des collections numériques. Mémoire pour le diplôme de conservateur des bibliothèques, Enssib, Villeurbanne. En ligne : <<https://lc.cx/QaqZ>>.
- [Girard, 2014] GIRARD, A. (2014). Enquête sur la numérisation de la presse ancienne locale et régionale : rapport de synthèse. Rapport technique

- BnF-ADM-2014-023933-01, Bibliothèque nationale de France. En ligne : <http://www.bnf.fr/documents/enquete_num_presse2014.pdf>.
- [Girard, 2015] GIRARD, A. (2015). La numérisation concertée de corpus d'imprimés. Etat des lieux des programmes et des partenaires. Rapport technique BnF-ADM-2015-044253-01, Bibliothèque nationale de France. En ligne : <http://www.bnf.fr/documents/num_concertee_imprimes.pdf>.
- [Glorieux, 2015] GLORIEUX, F. (2015). Bibliothèques informatisées et lectures distantes. *Critique*, Des chiffres et des lettres : les humanités numériques 819-820(8):653–666. En ligne : <<https://www.cairn.info/revue-critique-2015-8-page-653.htm>>.
- [Granjon et Magis, 2016] GRANJON, F. et MAGIS, C. (2016). Critique et humanités numériques : pour une approche matérialiste de l'«immatériel». *Variations*, 19. En ligne : <<http://journals.openedition.org/variations/748>>.
- [Grenoble, 2008] GRENOBLE (2008). Projet de convention entre l'Université Stendhal - Grenoble 3 et la ville de Grenoble relative à la consultation de la base documentaire des manuscrits de Stendhal du fonds de la bibliothèque municipale. Ville de Grenoble et Université Stendhal.
- [Guerreau, 2011] GUERREAU, A. (2011). Pour un corpus de textes latins en ligne. *BU-CEMA - Bulletin du centre d'études médiévales d'Auxerre*. En ligne : <<http://journals.openedition.org/cem/11787>>.
- [Guerreau, 2012] GUERREAU, A. (2012). Textes anciens en série : outils informatiques d'organisation et de manipulation de bases de données textuelles. *BU-CEMA - Bulletin du centre d'études médiévales d'Auxerre*. En ligne : <<http://journals.openedition.org/cem/12177>>.
- [Guillot et al., 2013] GUILLOT, C., LAVRENTIEV, A., RAINSFORD, T., MARCHELLO-NIZIA, C. et HEIDEN, S. (2013). La "philologie numérique" : tentative de définition d'un nouvel objet éditorial. In TRACHSLER, R., DUVAL, F. et LEONARDI, L., éditeurs : *Actes du XXVIIe Congrès international de linguistique et de philologie romanes (Nancy, 15-20 juillet 2013)*. En ligne : <<https://halshs.archives-ouvertes.fr/halshs-00846767>>.
- [Haquet, 2012] HAQUET, C. (2012). Patrimoine et politique documentaire : « la bibliothèque dans son entier ». In COQ, D., éditeur : *Apprendre à gérer des collections patrimoniales en bibliothèque*, numéro 26 de Boîte à outils, pages 26–36. Presses de l'Enssib, Villeurbanne. En ligne : <<http://books.openedition.org/pressesenssib/652>>.
- [Hartsell-Gundy et al., 2015] HARTSELL-GUNDY, A., BRAUNSTEIN, L. et GOLOMB, L., éditeurs (2015). *Digital humanities in the library : challenges and opportunities for subject specialists*. The Association of College and Research Libraries.
- [Higgins, 2011] HIGGINS, S. (2011). Digital Curation : The Emergence of a New Discipline. *The International Journal of Digital Curation*, 6(2):79–88. En ligne : <<http://www.ijdc.net/article/view/184/251>>.
- [Hourcade et al., 2010] HOURCADE, J.-C., LALOË, F. et SPITZ, E. (2010). *Longévité de l'information numérique : les données que nous voulons garder vont-elles s'effacer ? Groupe PSN (Pérennité des supports numériques) commun à l'Académie des sciences et à l'Académie des technologies*. EDP Sciences, Paris.
- [IHA, 2013] IHA (2013). Jeunes chercheurs et humanités numériques : un manifeste. En ligne : <<http://dhiha.hypotheses.org/1108>>.

- [Ingarao et Saïdi, 2010] INGARAO, M. et SAÏDI, S. (2010). Guide pratique pour la production de corpus numérique. Rapport technique, Institut des Sciences de l'Homme, MutEC, ENS. En ligne : <<https://lc.cx/AJjN>>.
- [Jacquet, 2012] JACQUET, C. (2012). Mettre à niveau une bibliothèque. In COQ, D., éditeur : *Apprendre à gérer des collections patrimoniales en bibliothèque*, numéro 26 de Boîte à outils, pages 132–141. Presses de l'Essib, Villeurbanne. En ligne : <<http://books.openedition.org/pressesenssib/669>>.
- [Jacquot, 2017] JACQUOT, O. (2017). Décrire, transcrire et diffuser un corpus documentaire hétérogène : méthodes, formats, outils. En ligne : <<http://bnf.hypotheses.org/2214>>.
- [Juanals et Minel, 2017] JUANALS, B. et MINEL, J.-L., éditeurs (2017). *Enjeux numériques pour les médiations scientifiques et culturelles du passé*. Les passés dans le présent. Presses Universitaires de Paris Nanterre, Paris.
- [Keenan et Johnston, 2000] KEENAN, S. et JOHNSTON, C. (2000). *Concise dictionary of library and information science*. Saur, München, 2e édition.
- [Kunstmann et al., 2013] KUNSTMANN, P., BOZZI, A., PARODI, G., PIERREL, J.-M. et STEIN, A. (2013). Quels corpus et quels outils d'exploitation de corpus pour les études de linguistique et philologie romanes : l'unité de la romanistique. Nancy. En ligne : <<https://hal.archives-ouvertes.fr/hal-01113965>>.
- [Langlais, 2017] LANGLAIS, P.-C. (2017). Les bibliothèques numériques sont-elles représentatives ? *Sciences Communes*. En ligne : <<https://scoms.hypotheses.org/799>>.
- [Law, 2004] LAW, D. (2004). Bibliographes spécialisés et bibliothécaires de référence : questions actuelles au royaume-uni. In *Bibliothécaire, quel métier*, pages 53–60. Cercle de la Librairie. En ligne : <<http://strathprints.strath.ac.uk/2944/>>.
- [Le Béhec et Alloing, 2016] LE BÉHEC, M. et ALLOING, C. (2016). Les humanités numériques pour repenser les catégories d'analyse. Le cas du «territoire numérique de marques». *Revue française des sciences de l'information et de la communication*, 8. En ligne : <<http://journals.openedition.org/rfsic/1804>>.
- [Le Coz, 2017] LE COZ, L. (2017). Patrimoine numérisé et Open Content. Quelle place pour le domaine public dans les bibliothèques numériques patrimoniales ? Mémoire pour le diplôme de conservateur des bibliothèques, Essib, Villeurbanne. En ligne : <<https://lc.cx/ATZ6>>.
- [Le Deuff, 2014] LE DEUFF, O., éditeur (2014). *Le temps des humanités digitales. La mutation des sciences humaines et sociales*. Société de la connaissance. FYP Éditions, Limoges.
- [Le Deuff, 2015] LE DEUFF, O. (2015). HUBIC : humanités digitales et bibliothèques, des compétences à identifier. Rapport technique, Laboratoire MICA - M2HDLab. En ligne : <<http://f.hypotheses.org/wp-content/blogs.dir/1057/files/2015/08/R>>.
- [Le Deuff et Berra, 2012] LE DEUFF, O. et BERRA, A. (2012). Quelles compétences et littératies pour les humanités numériques ? In *THATCamp Paris 2012. Non-actes de la non-conférence des humanités numériques*. Édition de la Maison des sciences de l'homme, Paris. En ligne : <<http://books.openedition.org/editionsmsh/334>>.
- [Le Masne de Chermont, 2016] LE MASNE DE CHERMONT, I. (2016). Les enjeux de la reconstitution en ligne de corpus disséminés. L'exemple des manuscrits koutchéens.

- In La recherche dans les institutions patrimoniales. Sources matérielles et ressources numériques*, Papiers, pages 116–124. Presses de l'Enssib.
- [Léger, 2015] LÉGER, P. (2015). La liberté de création au service des humanités numériques. In BENSAMOUN, A. et LATIL, A., éditeurs : *Propriété littéraire et artistique et humanités numériques*, pages 43–59. Mare & Martin, Paris.
- [Longhi, 2017] LONGHI, J. (2017). Humanités, numérique : des corpus au sens, du sens aux corpus. *Questions de communication*, 31(1):7–17. En ligne : <<https://www.cairn.info/revue-questions-de-communication-2017-1-page-7.htm>>.
- [Macke, 2015] MACKE, J.-S. (2015). Les archives zoliennes : création et évolution du site ArchiZ en fonction de ses lecteurs et usages. In PETY, D., éditeur : *Patrimoine littéraire en ligne : la renaissance du lecteur ?*, Corpus, pages 87–95. Université de Savoie Mont Blanc, Chambéry.
- [Magnani, 2017] MAGNANI, E. (2017). Qu'est-ce qu'un corpus ? Compte-rendu de la journée d'études. En ligne : <<https://halshs.archives-ouvertes.fr/halshs-01610087>>.
- [Maillet, 2011] MAILLET, D. (2011). Numérisation d'un fond d'images fixes. In COLLART, C. et MELOT, M., éditeurs : *Images et bibliothèques*, Collection Bibliothèques, pages 123–144. Éditions du Cercle de la Librairie, Villeurbanne.
- [Marino, 2015] MARINO, L. (2015). Des exceptions au droit d'auteur au secours des humanités numériques ? In *Propriété littéraire et artistique et humanités numériques*, pages 61–79. Mare & Martin, Paris.
- [Maurel, 2017a] MAUREL, L. (2017a). Quel modèle économique pour une numérisation patrimoniale respectueuse du domaine public ? In DUJOL, L., éditeur : *Communs du savoir en bibliothèque*. Cercle de la Librairie, Paris. En ligne : <<https://hal-univ-paris10.archives-ouvertes.fr/hal-01528096/document>>.
- [Maurel, 2017b] MAUREL, L. (2017b). Quels choix juridiques pour la médiation culturelle et scientifique dans l'environnement numérique ? In JUANAL, B. et MINEL, J.-L., éditeurs : *Enjeux numériques pour les médiations scientifiques et culturelles du passé*, Les passés dans le présent, pages 179–198. Presses Universitaires de Paris Nanterre, Paris.
- [Mazens, 2017] MAZENS, S. (2017). La politique nationale de l'IST au service de la recherche. *Ar(abes)ques*, 84:4–5.
- [MCC, 2017] MCC (2017). Recommandations pour la rédaction d'un projet culturel, scientifique, éducatif et social (PCSES). Rapport technique, Ministère de la Culture et de la Communication - Service du livre et de la lecture. En ligne : <<https://lc.cx/AJRH>>.
- [MCC - DGMIC, 2016] MCC - DGMIC (2016). Publics et usages des bibliothèques municipales en 2016. Rapport technique, Ministère de la Culture et de la Communication - Direction générale des médias et des industries culturelles. En ligne : <<https://lc.cx/AYTP>>.
- [MCC Archives, 2014] MCC ARCHIVES (2014). Archives et enjeux de société. Ministère de la Culture et de la Communication Culture et recherche.
- [Melzac, 2017] MELZAC, C. (2017). À sept mois du clap de fin, ISTEEX approche les 20 millions d'articles. *Ar(abes)ques*, 84:19.

- [MENESR, 2016] MENESR (2016). Stratégie nationale des infrastructures de recherche. En ligne : <<http://multimedia.enseignementsup-recherche.gouv.fr/SNIR/files/assets/basic-html/page-1.html>>.
- [Mesguich, 2017] MESGUICH, V. (2017). *Bibliothèques : le Web est à vous*. Collection Bibliothèques. Éditions du Cercle de la Librairie, Paris.
- [Mesguich *et al.*, 2012] MESGUICH, V., PIERRE, J., ALLOING, C., GALLETZOT, G., SERRES, A., PEIRANO, R., FROSSARD, F., DESCHAMPS, C., BATTISTI, M. et MARTINET, F. (2012). Enjeux et dimensions. *Documentaliste-Sciences de l'Information*, 49(1):24–45. En ligne : <<https://www.cairn.info/revue-documentaliste-sciences-de-l-information-2012-1-page-24.htm>>.
- [Moiraghi, 2018] MOIRAGHI, E. (2018). Le projet Corpus et ses publics potentiels. Une étude prospective sur les besoins et les attentes des futurs usagers. Rapport technique, Bibliothèque nationale de France.
- [Moirez *et al.*, 2013] MOIREZ, P., MOREUX, J. P. et JOSSE, I. (2013). État de l'art en matière de crowdsourcing dans les bibliothèques numériques. Rapport technique Livrable L-4.3.1, Bibliothèque nationale de France - Projet de R&D du FUI 12. En ligne : <http://www.bnf.fr/documents/crowdsourcing_rapport.pdf>.
- [Morel, 2017] MOREL, A. (2017). PSC / PCSES : pour une diffusion de la démarche projet en bibliothèque. Mémoire de D.E.A., Enssib, Villeurbanne. En ligne : <<http://www.enssib.fr/bibliotheque-numerique/documents/65765-psc-pces-pour-une-diffusion-de-la-demarche-projet-en-bibliotheque.pdf>>.
- [Moretti, 2000] MORETTI, F. (2000). Conjectures on world literature. *New Left Review*, 1:54–68. En ligne : <<https://newleftreview.org/II/1/franco-moretti-conjectures-on-world-literature>>.
- [Moretti, 2005] MORETTI, F. (2005). *La letteratura vista da lontano*. Saggi. Einaudi, Torino.
- [Moretti, 2016] MORETTI, F., éditeur (2016). *La littérature au laboratoire*. Ithaque, Paris.
- [Moreux, 2016] MOREUX, J.-P. (2016). Approches innovantes pour la presse ancienne numérisée : fouille et visualisation de données. *Carnet de la recherche à la Bibliothèque nationale de France*. En ligne : <<https://bnf.hypotheses.org/208>>.
- [Mounier, 2012] MOUNIER, P., éditeur (2012). *Read/Write Book 2 : Une introduction aux humanités numériques*. OpenEdition Press, Marseille. En ligne : <<http://books.openedition.org/oep/226>>.
- [Muller, 2017] MULLER, C. (2017). TDM : Fouille de textes et de données dans le contexte de la loi pour une république numérique - journée d'étude adbu du 13/12/16. En ligne : <<https://lc.cx/QaUN>>.
- [Neuroth, 2015] NEUROTH, H. (2015). Gérer les données : nouveau défi pour les bibliothèques. *I2D – Information, données et documents*, 53(4):66. En ligne : <<https://lc.cx/A4uN>>.
- [Oudenhoven, 2017] OUDENHOVEN, M. (2017). On the role of a university library in the TDM landscape. *FutureTDM*. En ligne : <<http://www.futuretdm.eu/blog/role-university-library-tdmlandscape/>>.

- [Paret, 2012] PARET, P. (2012). Les enseignants et la BU. Diplôme de conservateur de bibliothèques, Enssib, Villeurbanne. En ligne : <<http://www.enssib.fr/bibliotheque-numerique/documents/60405-les-enseignants-et-la-bu.pdf>>.
- [Paris 3 - Direction des Bibliothèques Universitaires, 2016] PARIS 3 - DIRECTION DES BIBLIOTHÈQUES UNIVERSITAIRES (2016). Compte-rendu d'enquête : les pratiques des Enseignants - Chercheurs de l'Université Sorbonne Nouvelle - Paris 3. Rapport technique, Université Sorbonne Nouvelle Paris 3. En ligne : <<http://www.dbu.univ-paris3.fr/images/docs/12573/enquete-enseignants-chercheurs.pdf>>.
- [Pérésan-Roudil, 2014] PÉRÉSAN-ROUDIL, A. (2014). Dans la forêt touffue des bibliothèques numériques patrimoniales françaises : quels choix, quelles stratégies, quelles perspectives ?. Mémoire pour le diplôme de conservateur des bibliothèques, Enssib, Villeurbanne. En ligne : <<https://lc.cx/QaJs>>.
- [Pety, 2015] PETY, D., éditeur (2015). *Patrimoine littéraire en ligne : la renaissance du lecteur ?*, Chambéry. Laboratoire LLSETI, Université de Savoie Mont Blanc.
- [Picard et al., 2015] PICARD, D., GOSSELIN, P.-H. et GASPARD, M.-C. (2015). Challenges in Content-Based Image Indexing of Cultural Heritage Collections. *IEEE Signal Processing Magazine - Institute of Electrical and Electronics Engineers*, 32(4):95–102. En ligne : <<https://hal.archives-ouvertes.fr/hal-01164409>>.
- [Pinon, 2016] PINON, C. (2016). Critères de sélection des données pour un corpus d'arabe contemporain. intégrer les variations génériques et diatopiques. In COMBY, É., MOSSET, Y. et de CARRARA, S., éditeurs : *Corpus de textes : composer, mesurer, interpréter*, pages 25–39. ENS Éditions. En ligne : <<http://books.openedition.org/enseditions/7346>>.
- [Pouyllau, 2012] POUYLLAU, S. (2012). Utilisons RDFa pour nos corpus. In *THATCamp Paris 2012. Non-actes de la non-conférence des humanités numériques*, Paris. Édition de la Maison des sciences de l'homme. En ligne : <<http://books.openedition.org/editionsmsh/379>>.
- [Prévost, 2015] PRÉVOST, S. (2015). Diachronie du français et linguistique de corpus : une approche quantitative renouvelée. *Langages*, 197(1):23–45. En ligne : <<https://www.cairn.info/revue-langages-2015-1-page-23.htm>>.
- [Puren, 2017] PUREN, M. (2017). Créer un plan de gestion des données (DMP). INRIA - ANF – INIST, Vandoeuvre-Les-Nancy. En ligne : <https://anfdonnees2017.sciencesconf.org/data/pages/20170706_dmp_puren.pdf>.
- [Research Consulting, 2017] RESEARCH CONSULTING (2017). La fouille de textes et de données dans l'enseignement supérieur et la recherche publique. Une analyse d'études de cas menées au Royaume-Uni et en France. préparé au nom de l'ADBU. En ligne : <<https://lc.cx/AxuC>>.
- [Romary et al., 2016] ROMARY, L., MERTENS, M. et BAILLOT, A. (2016). Data fluidity in DARIAH - pushing the agenda forward. *BIBLIOTHEK Forschung und Praxis*, 39(3):350–357. En ligne : <<https://hal.archives-ouvertes.fr/DARIAH/hal-01285917v1>>.
- [Rouen, 2003] ROUEN (2003). Convention de coopération de recherche. Programme d'étude Madame Bovary. Transcription des manuscrits. Ville de Rouen et Université de Rouen.

- [Roustan *et al.*, 2016] ROUSTAN, M., MONJARET, A. et CHEVALLIER, P. (2016). *La recherche dans les institutions patrimoniales. Sources matérielles et ressources numériques*. Papiers. Presses de l'Enssib, Villeurbanne.
- [Schaffner et Erway, 2014] SCHAFFNER, J. et ERWAY, R. (2014). Does Every Research Library Need a Digital Humanities Center? Rapport technique, OCLC Research, Dublin, Ohio. En ligne : <<https://eric.ed.gov/?id=ED564811>>.
- [Schuwey, 2015] SCHUWEY, C. (2015). Faire lire les Nouvelles Nouvelles de Donneau de Visé : défis et opportunités d'une édition en ligne. In PETY, D., éditeur : *Patrimoine littéraire en ligne : la renaissance du lecteur ?*, Corpus, pages 69–84. Université de Savoie Mont Blanc, Chambéry.
- [SDC Paris 8, 2015] SDC PARIS 8 (2015). Enquête de besoins menée auprès des chercheurs de l'université de Paris 8. Rapport technique, Université Paris 8.
- [Strauch et Rehm, 2007] STRAUCH, D. et REHM, M. (2007). *Lexikon Buch, Bibliothek, Neue Medien*. Saur, München.
- [Swiatek, 2017] SWIATEK, C. (2017). Quelles compétences pour les bibliothèques de recherche? *Bulletin des bibliothèques de France (BBF)*, 13:22–35. En ligne : <<http://bbf.enssib.fr/consulter/bbf-2017-13-0022-003>>.
- [Tardy et Dobedei, 2015] TARDY, C. et DOBEDEI, V. (2015). Introduction. In *Mémoire et nouveaux patrimoines*. OpenEdition Press, Marseille. En ligne : <<http://books.openedition.org/oep/411>>.
- [Tessier, 2010] TESSIER, M. (2010). Rapport sur la numérisation du patrimoine écrit. Rapport technique, Ministère de la Culture et de la Communication. En ligne : <<https://lc.cx/QRk6>>.
- [Thompson *et al.*, 2011] THOMPSON, M., MELOT, M. et COLLART, C. (2011). Traitement documentaire de l'image fixe. In COLLART, C. et MELOT, M., éditeurs : *Images et bibliothèques*, Collection Bibliothèques, pages 145–186. Éditions du Cercle de la Librairie, Paris.
- [Unesco, 2003] UNESCO (2003). Charte sur la conservation du patrimoine numérique. En ligne : <<https://lc.cx/Qaam>>.
- [Urfist Nice, 2011] URFIST NICE (2011). Enquête sur les pratiques informationnelles des chercheurs : quelques résultats. En ligne : <<http://urfistinfo.hypotheses.org/1901>>.
- [Vial, 2016] VIAL, S. (2016). Le tournant design des humanités numériques. *Revue française des sciences de l'information et de la communication*, 8. En ligne : <<http://journals.openedition.org/rfsic/1828>>.
- [Vinck et Clivaz, 2014] VINCK, D. et CLIVAZ, C. (2014). Les humanités délivrées. Savoir et culture réinventés hors du livre. *Revue d'anthropologie des connaissances*, 8(4):681–704. En ligne : <<https://lc.cx/Qat8>>.
- [Vuillermoz, 2015] VUILLERMOZ, M. (2015). L'invention du lecteur : un enjeu stratégique des préfaces de pièces de théâtre aux XVIe et XVIIe siècles (France, Italie, Espagne). In PETY, D., éditeur : *Patrimoine littéraire en ligne : la renaissance du lecteur ?*, Corpus, pages 47–55. Université de Savoie Mont Blanc, Chambéry.

GLOSSAIRE

Corpus : « le corpus est un ensemble à priori homogène de documents (textuels, mais aussi iconographiques, sonores, etc.) réunis dans un but scientifique » [Comby *et al.*, 2016, p. 178].

Bibliothèque numérique : « Les bibliothèques numériques sont des organisations qui offrent des ressources, y compris en personnel, pour sélectionner, structurer, offrir un accès intellectuel, interpréter, distribuer et conserver l'intégralité de documents sous forme numérique. Une bibliothèque numérique garantit également un accès sur la durée aux oeuvres électroniques dans le but d'être aisément et à un moindre coût disponibles à un ou plusieurs publics spécifiques » [Claerr et Westeel, 2013b, Introduction, p. 26, d'après Gary Cleveland].

Lemme : « dans beaucoup de langues, dont le français, un même mot peut apparaître sous différentes formes fléchies (féminin, pluriel, formes conjuguées des verbes, etc.). Le lemme d'un mot est la forme unique servant de référence ; c'est celle qui est utilisée par convention dans les dictionnaires (infinitif pour le verbe, singulier pour le nom, masculin singulier pour l'adjectif, etc.). La lemmatisation est donc l'opération qui consiste à identifier et étiqueter pour les mots d'un texte le lemme correspondant » [Comby *et al.*, 2016, p. 179].

Lexicométrie : « la lexicométrie se situe à l'interface entre lexicologie, mathématiques et informatique, en proposant un ensemble de méthodes qui réorganise les données (comme les index ou les concordanciers) et qui permet la mise en oeuvre de calculs statistiques (telles les fréquences ou les spécificités) » [Comby *et al.*, 2016, p. 179].

Textométrie : « la textométrie, parfois appelée logométrie, combine des analyses statistiques héritées de la statistique textuelle ainsi que de la lexicométrie et une volonté de retour facilité au texte favorisant la contextualisation des résultats par une édition numérique des textes » [Comby *et al.*, 2016, p. 180].

ANNEXES

LES QUESTIONNAIRES	114
Le questionnaire aux chercheurs	114
Le texte	114
Sélection de résultats	121
Le questionnaire aux bibliothécaires	132
Le texte	132
Sélection de résultats	137

LES QUESTIONNAIRES

Le questionnaire aux chercheurs

Le texte

S'adressant aux chercheurs de toutes disciplines confondues, ce bref questionnaire a pour but de récolter des informations au sujet des différentes possibilités de collaborations, existantes ou souhaitables, entre les acteurs de la recherche et les bibliothécaires concernant la mise en place de projets de numérisation de corpus documentaires. Il s'agit notamment de corpus patrimoniaux, mais aussi de toute sélection de plusieurs documents pouvant répondre à une problématique précise (scientifique, pédagogique, culturelle, etc.).

L'enquête est diffusée par Graziella Pastore (graziella.pastore@enssib.fr), élève conservateur à l'Esssib, dans le cadre de la préparation de son mémoire en vue de l'obtention du diplôme de conservateur d'État des bibliothèques (Esssib, DCB 26).

Ce questionnaire est anonyme et il ne vous prendra pas plus de 15 minutes. Si vous souhaitez apporter plus de précisions, vous pouvez indiquer votre adresse de courriel à la fin du questionnaire.

Les réponses seront collectées jusqu'au 15 décembre 2017.

Merci beaucoup de votre collaboration !

* *Obligatoire*

Initiative (1)

1. Dans le cadre de vos activités de recherche, avez-vous déjà sollicité une bibliothèque pour numériser un corpus de documents ? Si oui, combien de fois ? *

Merci de prendre en compte seulement les demandes de numérisation de corpus effectuées auprès d'une bibliothèque à compter de 2012. Notez que les demandes effectuées auprès des archives, musées, collections privées, etc. ne rentrent pas dans ce calcul ; dans ce cas, merci de sélectionner "Non, jamais" et de poursuivre le questionnaire.

Une seule réponse possible.

- Oui, 1 fois
- Oui, de 2 à 3 fois
- Oui, de 4 à 5 fois
- Oui, plus de 5 fois
- Non, jamais *Passez à la question 22.*

Collaboration (1)

Merci de prendre en compte seulement les projets s'appuyant sur la numérisation d'un corpus menés depuis 2012.

2. Lors de vos projets de numérisation de corpus, avez-vous demandé l'aide d'un professionnel des bibliothèques ? Par exemple, pour le choix, l'indexation, la valorisation, la numérisation, l'hébergement de données produites ou toute autre expertise concernant les documents du corpus.*

Une seule réponse possible.

- Oui

— Non

Après avoir répondu à la dernière question de cette section, passez à la question 8.

3. Si non, pouvez-vous nous expliquer pourquoi ?

Collaboration (2)

Si depuis 2012 vous avez mené plusieurs projets impliquant la numérisation d'un corpus et une collaboration avec les professionnels des bibliothèques, merci de répondre aux questions suivantes en prenant en compte seulement le projet que vous jugiez le plus représentatif de votre expérience.

4. À quel moment de votre projet avez-vous sollicité la collaboration d'un professionnel des bibliothèques ? *

Une seule réponse possible.

- Dès la conception du projet (avant l'approbation, l'obtention de financements, etc.)
- Pendant les phases d'avancement du projet (après l'approbation, l'obtention de financements, etc.)
- À la fin du projet (par exemple pour la diffusion des résultats, pour la mise en ligne ou l'hébergement des données, etc.)

5. Sous quelle forme cette collaboration a-t-elle été menée ? *

Par exemple : comment les tâches se sont-elles réparties ? Quelle était la fréquence des réunions ? Avez-vous travaillé selon une méthode projet (traditionnelle, Agile, etc.) ?

6. Quel bilan tirez-vous de cette collaboration ? *

Une seule réponse possible.

- Plutôt positif
- Ni positif ni négatif
- Plutôt négatif

7. Quels ont été les aspects les plus positifs/négatifs de cette collaboration ?

Nature de votre demande de numérisation

8. Lors de votre demande, de quel type de corpus avez-vous demandé la numérisation ? *

Une seule réponse possible.

- Documents d'archive
- Documents pédagogiques servant à l'histoire d'une discipline
- Manifestations ou expressions différentes d'une même oeuvre
- Presse / Périodiques
- Production d'un même auteur
- Production de plusieurs auteurs sur une période cible
- Production de plusieurs auteurs sur un thème cible
- Production de plusieurs auteurs sur un thème et une période cible
- Autre : ...

9. Quelle était la nature principale des documents faisant partie du corpus dont vous avez demandé la numérisation ? *

Plusieurs réponses possibles.

- Cartes et plans
 - Cartes postales
 - Dessins
 - Documents d'archive
 - Estampes
 - Incunables
 - Livres imprimés avant 1810
 - Livres imprimés après 1810
 - Manuscrits
 - Photographies
 - Presse / Périodiques
 - Thèses
 - Autre : ...
10. Quelle était la finalité principale de votre corpus ? (Historique, linguistique, philologique, littéraire, etc.) *
11. Pouvez-vous nous indiquer le titre de votre projet et éventuellement un lien vers de la documentation à son sujet ? *

Données et métadonnées

12. Concernant les documents du corpus à la base de votre projet, est-ce que vous avez utilisé des métadonnées associées produites par la bibliothèque ? *

Une seule réponse possible.

- Oui
- Non

13. Si oui, dans quel format ?

Plusieurs réponses possibles.

- Dublin Core
- EAD
- METS
- SKOS/RDF
- TEI
- Autre : ...

14. Si vous n'avez pas utilisé les métadonnées produites par la bibliothèque, pouvez-vous nous expliquer pourquoi ?

Une seule réponse possible.

- L'utilisation des métadonnées n'était pas prioritaire dans le cadre de mon projet
- Les formats disponibles ne correspondaient pas aux nécessités de mon projet
- Je ne disposais pas d'un logiciel adapté ou des compétences nécessaires pour traiter les métadonnées
- Autre : ...

15. Concernant les documents numérisés, avez-vous disposé d'autres fichiers textuels produits par la bibliothèque qui conserve ces documents ? Si oui, dans quel format ?*

Plusieurs réponses possibles.

- Oui : fichiers texte brut (rtf, txt)

- Oui : fichiers ALTO
- Oui : fichiers ePub
- Oui : fichiers TEI
- Non
- Autre : ...

16. Dans le cadre de votre recherche sur le corpus en question, avez-vous produit des données ou des métadonnées susceptibles d'intégrer celles produites par la bibliothèque qui conserve les documents ? Si oui, dans quel format ? *

Projet scientifique et financements

17. Dans quel contexte votre demande de numérisation s'inscrivait-elle ? *

Une seule réponse possible.

- Projet scientifique individuel
- Projet scientifique collectif (regroupant une ou plusieurs équipes)

18. Votre projet faisait partie d'un consortium ou d'un partenariat ? Si oui, lequel ?

19. Disposiez-vous d'un financement pour des travaux de numérisation ? *

Une seule réponse possible.

- Oui
- Non

20. Si oui, pouvez-vous nous préciser de quel type de financement il s'agissait ? *

Plusieurs réponses possibles.

- Agence nationale de la recherche (ANR)
- Financement privé (mécénat)
- Financement d'une collectivité locale
- Financement d'un organisme étranger
- Financement d'une association ou d'une fondation
- Horizon 2020
- Programme Investissements d'Avenir (Idex, Labex, Equipex, etc.)
- Aucun financement
- Autre : ...

21. Commentaires

Espace pour d'éventuelles précisions concernant cette section.

Passez à la question 28.

Initiative (2)

22. Vous n'avez jamais sollicité la numérisation d'un corpus de documents auprès d'une bibliothèque : pourquoi ? *

Une seule réponse possible.

- Je n'en ai pas besoin pour mes travaux/recherches
- J'ai besoin de travailler directement sur les documents originaux
- Les documents sur lesquels je travaille ont déjà été numérisés par la bibliothèque
- La numérisation des documents qui m'intéressent a déjà été sollicitée par d'autres intervenants (fondations, mécènes, autres chercheurs ou équipes de recherche, etc.)

- Les documents sur lesquels je travaille sont conservés ailleurs (archives nationales/départementales, musées, collections privées, etc.)
- Je ne sais pas à qui m'adresser
- Autre : ...

23. Participez-vous (ou avez-vous déjà participé) à un projet qui s'appuie sur un corpus de documents numérisés ? *

Une seule réponse possible.

- Oui
- Non

Passez à la question 28.

Collaboration à un projet de numérisation de corpus

Si depuis 2012 vous avez participé à plusieurs projets s'appuyant sur la numérisation de corpus, merci de prendre en compte seulement le projet que vous jugez le plus représentatif de votre expérience

24. Vous participez (ou vous avez participé) à un projet qui s'appuie sur la numérisation d'un corpus de documents. Pouvez-vous nous en dire plus sur ce projet et sur votre rôle ? *

Merci d'indiquer un éventuel lien vers de la documentation en ligne ou le contact d'un référent de projet.

25. Lors de ce projet, avez-vous demandé la collaboration d'un professionnel des bibliothèques ? Par exemple pour le choix, l'indexation, la valorisation, la numérisation, l'hébergement de données produites ou toute autre expertise concernant les documents du corpus ? *

Une seule réponse possible.

- Oui
- Non

26. Si oui, pouvez-vous nous préciser la nature de cette collaboration ?

27. Si oui, quel bilan tirez-vous de cette collaboration ?

Expertise scientifique ou technique

28. Avez-vous déjà été sollicité(e) par une bibliothèque pour une expertise scientifique ou technique à propos d'un corpus de documents numérisé ou à numériser ? *

Une seule réponse possible.

- Oui, pour une expertise scientifique
- Oui, pour une expertise technique
- Oui, pour une expertise scientifique et technique
- Non

29. Si oui, à quel propos ?

Plusieurs réponses possibles.

- Identification des documents originaux et/ou création des métadonnées correspondantes
- Sélection des documents à numériser
- Restauration

- Préparation d'une exposition
- Préparation d'une exposition virtuelle ou d'un produit numérique
- Valorisation d'un fonds à travers un travail de recherche (article, thèse, conférence, etc.)
- Choix ou conversion du modèle des métadonnées ou de mise en consultation des documents numériques
- Choix ou conversion des outils de stockage des documents numériques
- Autre : ...

30. Si oui, quel bilan tirez-vous de cette expérience / ces expériences ?

Suggestions

31. Dans le cadre de vos recherches, en France ou ailleurs, avez-vous rencontré de bonnes pratiques de collaboration entre chercheurs et bibliothécaires que vous souhaitez nous signaler ? Avez-vous des remarques ou des suggestions notamment concernant les projets de numérisation de corpus ?
32. Souhaitez-vous nous signaler un projet de numérisation de corpus documentaires susceptible de témoigner d'une collaboration entre chercheurs et bibliothécaires ?

Contexte de travail du chercheur

33. Aujourd'hui, quel est votre poste principal ? *

Une seule réponse possible.

- Attaché temporaire d'enseignement et de recherche (ATER)
- Chargé de recherche
- Chercheur postdoctoral
- Directeur de recherche
- Doctorant
- Enseignant du secondaire
- Ingénieur d'études
- Ingénieur de recherche
- Maître de conférence
- Professeur des universités
- Autre : ...

34. Quel est votre principal domaine de recherche ? *

D'après la grille Unesco <http://skos.um.es/unesco6/00/html>. Si vous le souhaitez, vous pouvez préciser votre domaine dans la question suivante.

Une seule réponse possible.

- Anthropologie
- Astronomie et astrophysique
- Chimie
- Démographie
- Droit et sciences juridiques
- Éthique
- Géographie
- Histoire (et Sciences liées à l'histoire : Archéologie, Iconographie, Paléographie, Philologie, etc.)

- Linguistique (Linguistique appliquée, Linguistique diachronique/synchronique, Géographie linguistique, etc.)
- Logique
- Mathématiques (Algèbre, Informatique, Statistiques, etc.)
- Pédagogie
- Philosophie
- Physique
- Psychologie
- Science politique
- Sciences agricoles
- Sciences de l'ingénieur
- Sciences de la Terre et de l'espace
- Sciences de la vie
- Sciences économiques
- Sciences médicales
- Sciences relatives aux arts et lettres (Architecture, Théorie littéraire, Théorie des beaux-arts, analyse et critique, etc.)
- Sociologie

35. Souhaitez-vous préciser votre domaine de recherche ?

36. Quelle est votre tranche d'âge ? *

Une seule réponse possible.

- Entre 20 et 30 ans
- Entre 31 et 40 ans
- Entre 41 et 50 ans
- Entre 51 et 60 ans
- Plus de 61 ans

37. Seriez-vous disponible pour être contacté(e) par courriel pour préciser certains points de ce questionnaire ? *

Une seule réponse possible.

- Oui *Passez à la question 38.*
- Non *Arrêtez de remplir ce formulaire.*

Contact

38. Merci d'indiquer votre adresse de courriel *

Fin du questionnaire.

Sélection de résultats

Initiative (1)

1. Dans le cadre de vos activités de recherche, avez-vous déjà sollicité une bibliothèque pour numériser un corpus de documents ? Si oui, combien de fois ? *

Nombre de réponses : 57.

- Oui, 1 fois : 4 (7 %)
- Oui, de 2 à 3 fois : 10 (15,8 %)
- Oui, de 4 à 5 fois : 2 (3,5 %)
- Oui, plus de 5 fois : 11 (19,3 %)
- Non, jamais : 31 (54,4 %)

Collaboration (1)

2. Lors de vos projets de numérisation de corpus, avez-vous demandé l'aide d'un professionnel des bibliothèques ? Par exemple, pour le choix, l'indexation, la valorisation, la numérisation, l'hébergement de données produites ou toute autre expertise concernant les documents du corpus.*

Nombre de réponses : 26.

- Oui : 9 (34,6 %)
- Non : 17 (65,4 %)

3. Si non, pouvez-vous nous expliquer pourquoi ?

Nombre de réponses : 7.

La plupart des répondants menant un projet en collaboration avec la BnF, l'appel à un professionnel de la bibliothèque était courant, sans démarche explicite ; ils soulignent toutefois que cela ne concerne que les aspects de la numérisation : la sélection des documents incombe aux chercheurs, le traitement des données aux spécialistes des Humanités Numériques, etc. Un des répondants nous signale que le problème le plus important concerne la gestion des droits.

Collaboration (2)

4. À quel moment de votre projet avez-vous sollicité la collaboration d'un professionnel des bibliothèques ? *

Nombre de réponses : 17.

- Dès la conception du projet (avant l'approbation, l'obtention de financements, etc.) : 10 (58,8 %)
- Pendant les phases d'avancement du projet (après l'approbation, l'obtention de financements, etc.) : 7 (41,2 %)
- À la fin du projet (par exemple pour la diffusion des résultats, pour la mise en ligne ou l'hébergement des données, etc.) : 0

5. Sous quelle forme cette collaboration a-t-elle été menée ? *

Nombre de réponses : 17.

Faisant suite à des projets de visée et taille différente (individuels, collectifs, régionaux, internationaux, etc.) les méthodes de travail qui ressortent de cette question sont assez hétérogènes (entretiens fixés au cas par cas, élaboration conjointe d'un cahier de charges, mise en place de plusieurs comités de pilotage, etc.). La plupart

des répondants met l'accent sur la fréquence importante de réunions en amont du projet et dans la phase d'élaboration ; le suivi peut ensuite être assuré à travers des réunions ponctuelles et des échanges par courriel.

6. Quel bilan tirez-vous de cette collaboration ? *

Nombre de réponses : 17.

- Plutôt positif : 14 (82,4 %)
- Ni positif ni négatif : 2 (11,8 %)
- Plutôt négatif : 1 (5,9 %)

7. Quels ont été les aspects les plus positifs/négatifs de cette collaboration ?

Nombre de réponses : 15.

Aspects positifs : "disponibilité des conservateurs ; intérêt partagé pour la valorisation et diffusion (selon les cas : mise à disposition des portails de diffusion des bibliothèques)"; "acquisition de nouvelles compétences liées au numérique pour les personnels des bibliothèques impliquées"; "meilleure répartition des tâches en fonction des compétences. Optimisation du temps de travail et amélioration de la conception d'ensemble du projet"; "beaucoup d'écoute et de volonté de rendre vraiment service conformément à notre demande"; "nous avons toujours reçu le meilleur soutien au projet scientifique et le meilleur accueil sur place de la part des collègues bibliothécaires"; "être sûre d'être dans les clous pour les questions d'ingénierie documentaire"; "la richesse de la documentation conservée par la BnF ; les multiples compétences mises à disposition par la BnF pour le programme ; le rôle pivot de la BnF dans le réseau international".

Aspect négatifs : "manque des instrumentations sur place (d'où la nécessité de faire appel à un prestataire, avec augmentation des coûts); difficultés logistiques ; calendrier très long"; "un certain manque de confiance dans la compétence des chercheurs"; "les conservateurs ont fait leur possible pour accélérer les choses, mais le temps (long) de la BnF n'est pas toujours celui des projets ..."; "la difficulté demeure la recherche de fond, laquelle me revient entièrement"; "le service de bibliothèque ne se pose pas trop la question des usages numériques (autre que la consultation de documents) et service surchargé de boulot ..."; "les temps de réaction de la part de la BnF (trop lents par rapport au calendrier imposé par un programme financé par mécénat); la nécessité de reformuler, en cours de réalisation, les ambitions initiales [...] vers des objectifs plus réalisables ; quelques difficultés dans l'identification des interlocuteurs pour des volets ponctuels".

Nature de votre demande de numérisation

8. Lors de votre demande, de quel type de corpus avez-vous demandé la numérisation ? *

Nombre de réponses : 26.

- Documents d'archive : 10
- Presse / Périodiques : 1
- Production d'un même auteur : 3
- Production de plusieurs auteurs sur une période cible : 3
- Production de plusieurs auteurs sur un thème cible : 1
- Production de plusieurs auteurs sur un thème et une période cible : 8

9. Quelle était la nature principale des documents faisant partie du corpus dont vous avez demandé la numérisation ? *

Nombre de réponses : 26.

- Cartes et plans : 2
- Dessins : 1
- Documents d'archive : 10
- Estampes : 2
- Incunables : 3
- Livres imprimés avant 1810 : 13
- Livres imprimés après 1810 : 3
- Manuscrits : 18
- Photographies : 1
- Presse / Périodiques : 3
- Thèses : 1

10. Quelle était la finalité principale de votre corpus ? (Historique, linguistique, philologique, littéraire, etc.) *

Nombre de réponses : 26.

- Catalogage de manuscrits / Reconstruction de catalogues anciens / Catalogage des collections spécialisées : 3
- Édition scientifique numérique : 5
- Humanités numériques : 2
- Littérature : 8
- Philosophie : 2
- Histoire des sciences : 2
- Étude génétique : 1
- Histoire de l'art : 1
- Paléographie : 1
- Histoire : 13
- Histoire du théâtre : 1
- Linguistique : 4
- Philologie : 7

11. Pouvez-vous nous indiquer le titre de votre projet et éventuellement un lien vers de la documentation à son sujet ? *

Nombre de réponses : 26.

Données et métadonnées

12. Concernant les documents du corpus à la base de votre projet, est-ce que vous avez utilisé des métadonnées associées produites par la bibliothèque ? *

Nombre de réponses : 26.

- Oui : 9 (34,6 %)
- Non : 17 (65,4 %)

13. Si oui, dans quel format ?

Nombre de réponses : 8.

- Dublin Core : 5
- EAD : 2
- METS : 1

- SKOS/RDF : 1
 - TEI : 2
 - Extraction des notices concernées dans le catalogue : 1
 - Textes numérisés sur Gallica : 1
14. Si vous n'avez pas utilisé les métadonnées produites par la bibliothèque, pouvez-vous nous expliquer pourquoi ?
- Nombre de réponses* : 16.
- Sélection* :
- "C'est nous qui avons créé les métadonnées (DC) et non la bibliothèque" : 1
 - "Dans la plupart des cas, les bibliothèques ne disposaient pas de métadonnées (fiables). L'un des enjeux du projet était précisément l'enrichissement des catalogues des bibliothèques et la création des métadonnées" : 1
 - "Je pense que la détermination des métadonnées en question relève du travail du chercheur (i.e. quelqu'un qui connaît parfaitement l'ensemble du corpus de l'intérieur, car des éléments des métadonnées sont souvent implicites)" : 1
 - L'utilisation des métadonnées n'était pas prioritaire dans le cadre de mon projet : 7
 - "Le document ne possède pas de métadonnées susceptibles d'être utilisées pour la numérisation" : 1
 - Les formats disponibles ne correspondaient pas aux nécessités de mon projet : 2
15. Concernant les documents numérisés, avez-vous disposé d'autres fichiers textuels produits par la bibliothèque qui conserve ces documents ? Si oui, dans quel format ?*
- Nombre de réponses* : 26.
- Oui : fichiers texte brut (rtf, txt) : 1
 - Oui : fichiers ALTO : 2
 - Oui : fichiers ePub : 0
 - Oui : fichiers TEI : 0
 - Non : 20
 - Fichiers textes bruts produits par nos soins : 1
 - production des conservateurs (descriptions, études des fonds, etc.) : 1
 - la question n'est pas très claire : 1
 - PDF : 1
16. Dans le cadre de votre recherche sur le corpus en question, avez-vous produit des données ou des métadonnées susceptibles d'intégrer celles produites par la bibliothèque qui conserve les documents ? Si oui, dans quel format ? *
- Nombre de réponses* : 26.
- Adobe Illustrator : 1
 - ALTO (corrigé) : 1
 - Catalogage des manuscrits en EAD : 1
 - Dublin Core : 5
 - HTML : 1
 - Livres : 1
 - ODT : 1
 - Photographies numériques des manuscrits : 1

- Proposition de nouvelles fiches autorités BnF PEP, ORG, TIT, ou des mises à jour / corrections : 1
- SPARQL : 1
- XML : 2
- XML-TEI : 8
- Il serait intéressant que les bibliothèques diffusent des éditions de leurs manuscrits à côté des photos, mais pour l’instant elles semblent peu intéressées : 1
- Non : 5
- Pas encore : 2

Projet scientifique et financements

17. Dans quel contexte votre demande de numérisation s’inscrivait-elle ? *

Nombre de réponses : 26.

- Projet scientifique individuel : 10 (38,5 %)
- Projet scientifique collectif (regroupant une ou plusieurs équipes) : 16 (61,5 %)

18. Votre projet faisait partie d’un consortium ou d’un partenariat ? Si oui, lequel ?

Nombre de réponses : 19.

19. Disposiez-vous d’un financement pour des travaux de numérisation ? *

Nombre de réponses : 26.

- Oui : 18
- Non : 8

20. Si oui, pouvez-vous nous préciser de quel type de financement il s’agissait ? *

Nombre de réponses : 26.

- Agence nationale de la recherche (ANR) : 6
- CNRS + Ministère de l’Education : 1
- CRSHC : 1
- Financement d’une association ou d’une fondation : 1
- Financement d’une collectivité locale : 8
- Financement d’un organisme étranger : 1
- Financement privé (mécénat) : 2
- Fonds recherche de l’ENS de Lyon : 1
- Horizon 2020 : 2
- La bibliothèque a financé elle-même la numérisation : 1
- Programme Investissements d’Avenir (Idex, Labex, Equipex, etc.) : 5
- Région, MSH, FEDER, CPER, IUF... : 1
- Aucun financement : 7

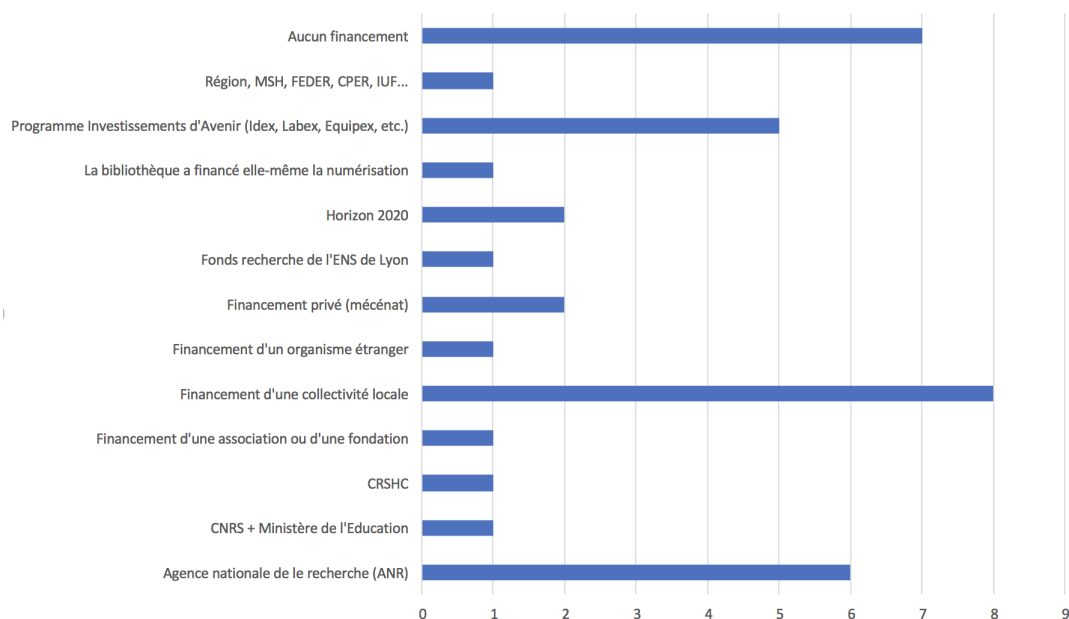


FIGURE 28 – Questionnaire aux chercheurs - question n° 20

21. Commentaires

Nombre de réponses : 4.

Sélection :

- "Le coût de la numérisation a été assumé par la bibliothèque"
- "Nous avons demandé peu de numérisation parce que nous travaillons dans le cadre de documents patrimoniaux déjà numérisés. Seuls quelques exemplaires à refaire en raison de la qualité des images ou des manuscrits qui entraînent dans la politique de numérisation des établissements"

Initiative (2)

22. Vous n'avez jamais sollicité la numérisation d'un corpus de documents auprès d'une bibliothèque : pourquoi ? *

Nombre de réponses : 31.

23. Participez-vous (ou avez-vous déjà participé) à un projet qui s'appuie sur un corpus de documents numérisés ? *

Nombre de réponses : 31.

— Oui : 20 (64,5 %)

— Non : 11 (35,5 %)

Collaboration à un projet de numérisation de corpus

24. Vous participez (ou vous avez participé) à un projet qui s'appuie sur la numérisation d'un corpus de documents. Pouvez-vous nous en dire plus sur ce projet et sur votre rôle ? *

Nombre de réponses : 20 (dont 18 exploitables)

25. Lors de ce projet, avez-vous demandé la collaboration d'un professionnel des bibliothèques ? Par exemple pour le choix, l'indexation, la valorisation, la numérisation, l'hébergement de données produites ou toute autre expertise concernant les documents du corpus. *

Nombre de réponses : 20.

- Oui : 5 (25 %)
- Non : 15 (75 %)

26. Si oui, pouvez-vous nous préciser la nature de cette collaboration ?

Nombre de réponses : 6.

Sélection :

- "Collaboration qui n'a jamais pu être mise en place"
- "La collaboration ne s'est pas concrétisée pour l'instant"
- "Le travail initial a été fait par des documentalistes et des archivistes. Il m'est arrivé d'intervenir pour corriger certaines données"
- "Renseignements sur l'histoire des manuscrits du corpus"

27. Si oui, quel bilan tirez-vous de cette collaboration ?

Nombre de réponses : 6.

- "bilan négatif"
- "Excellent"
- "Le projet devrait être relancé dans le prochains mois"
- "positif, retour prompt et précis de la part d'un conservateur"
- "Très positif"
- "Une collaboration fructueuse et une pérennisation qui est assurée (sauf erreur de ma part) conjointement par la bibliothèque et par l'université porteuse du projet"

Expertise scientifique ou technique

28. Avez-vous déjà été sollicité(e) par une bibliothèque pour une expertise scientifique ou technique à propos d'un corpus de documents numérisé ou à numériser ? *

Nombre de réponses : 57.

- Oui, pour une expertise scientifique : 10 (17,5 %)
- Oui, pour une expertise technique : 1 (1,8 %)
- Oui, pour une expertise scientifique et technique : 5 (8,8 %)
- Non : 41 (71,9 %)

29. Si oui, à quel propos ?

Nombre de réponses : 17.

- Identification des documents originaux et/ou création des métadonnées correspondantes : 9 (52,9 %)
- Sélection des documents à numériser : 7 (41,2 %)
- Restauration : 2 (11,8 %)
- Préparation d'une exposition : 3 (17,6 %)
- Préparation d'une exposition virtuelle ou d'un produit numérique : 2 (11,8 %)
- Valorisation d'un fonds à travers un travail de recherche (article, thèse, conférence, etc.) : 6 (35,3 %)
- Choix ou conversion du modèle des métadonnées ou de mise en consultation des documents numériques : 2 (11,8 %)
- Choix ou conversion des outils de stockage des documents numériques : 3 (17,6 %)
- Non, mais j'ai fourni spontanément une identification de livre mal répertoriée : 1 (5,9 %)

— Stratégie de numérisation patrimoniale dans les bibliothèques argentines : 1 (5,9 %)

30. Si oui, quel bilan tirez-vous de cette expérience / ces expériences ?

Nombre de réponses : 16.

— Excellent : 4

"Excellent enrichissant et motivant même si le manque de moyens rend parfois la tâche du difficile"

"Excellent. Collaboration efficace et fructueuse qui a permis de nouer des liens durables avec le conservateur et les collections dont il avait la charge"

— Très enrichissant : 2

— Très positif : 2

— Positif : 5

— "Mixte : les priorités ne sont pas les mêmes"

— Autres réponses : 2

"L'échange entre métiers est toujours riche d'enseignement"

"Il s'agissait de demandes informelles de conservateurs me demandant si je voyais des manuscrits à faire numériser dans leur fonds"

Suggestions

31. Dans le cadre de vos recherches, en France ou ailleurs, avez-vous rencontré de bonnes pratiques de collaboration entre chercheurs et bibliothécaires que vous souhaitez nous signaler ? Avez-vous des remarques ou des suggestions notamment concernant les projets de numérisation de corpus ?

Nombre de réponses : 27.

32. Souhaitez-vous nous signaler un projet de numérisation de corpus documentaires susceptible de témoigner d'une collaboration entre chercheurs et bibliothécaires ?

Nombre de réponses : 21 (dont 20 exploitables).

a) Archives de la Bibliothèque centrale du MNHN

b) Bibliothèque virtuelle de Clairvaux <https://www.bibliotheque-virtuelle-clairvaux.com/>, Bibliothèque virtuelle du Mont-Saint-Michel <http://www.unicaen.fr/bvmsm/>

c) BIU Santé @medica ouvert en 2000

d) Deux projets en cours : <http://philosophie-clandestine.huma-num.fr/> + <http://rey.huma-num.fr>

e) En cours, par ex. : le projet partenariat Biblissima "Bibliothèques humanistes ligériennes" avec les bibliothèques du Mans, Angers et La Flèche

f) galerie.verdun.fr

g) <http://www.handschriftenzentrum.de/> et <http://www.manuscripta-mediaevalia.de/#|5>

h) Roman de la Rose digital library

i) Projet Fonte Gaia

j) Le travail de numérisation de la Herzog August Bibliothek Wolfenbüttel

k) MHARS. Numapress

- l) BAnQ et le Bibliothèque de l'université McGill
- m) Plusieurs projets financés par Biblissima ; ManusOnLine et ses bibliothèques partenaires (mss de la Lombardie ; mss de la Sardaigne ; Censur des mss de Dante ; Censur des mss franciscains) ; Fragmentarium (<http://www.fragmentarium.unifr.ch/>) ou e-codices (programmes suisses), ...
- n) Non : 5

Contexte de travail du chercheur

33. Aujourd'hui, quel est votre poste principal ? *

Nombre de réponses : 57.

- Bibliothécaire : 1
- Chargé de recherche : 5
- Chargé de collection : 1
- Chercheur postdoctoral : 6
- Conservateur : 1
- Directeur de recherche : 3 (+ 1 retraité)
- Doctorant : 1
- Enseignant du secondaire : 1
- Ingénieur d'études : 2
- Ingénieur de recherche : 5
- Maître de conférence : 15 (+ 2 IUF)
- Professeur des universités : 13 (+ 1 professeur émérite)

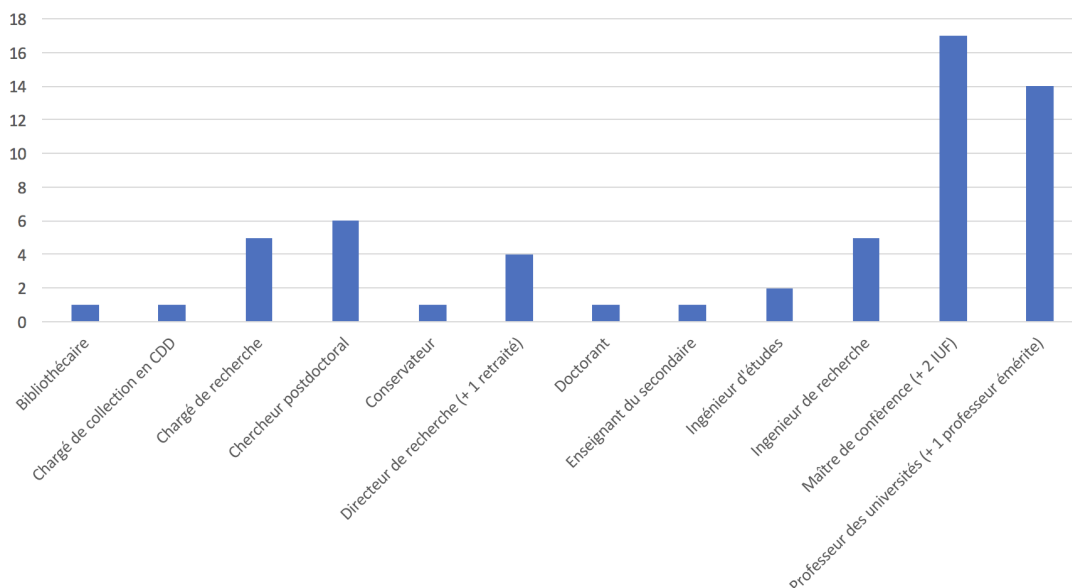


FIGURE 29 – Questionnaire aux chercheurs - question n° 33

34. Quel est votre principal domaine de recherche ? *

Nombre de réponses : 57.

- Géographie : 1
- Histoire (et Sciences liées à l'histoire : Archéologie, Iconographie, Paléographie, Philologie, etc.) : 23 (40,4 %)

- Linguistique (Linguistique appliquée, Linguistique diachronique/synchronique, Géographie linguistique, etc.) : 8 (14 %)
- Mathématiques (Algèbre, Informatique, Statistiques, etc.) : 5 (8,8 %)
- Pédagogie : 2
- Philosophie : 2
- Physique : 1
- Sciences de l'ingénieur : 3 (5,3 %)
- Sciences économiques : 1
- Sciences médicales : 1
- Sciences relatives aux arts et lettres (Architecture, Théorie littéraire, Théorie des beaux-arts, analyse et critique, etc.) : 11 (19,3 %)

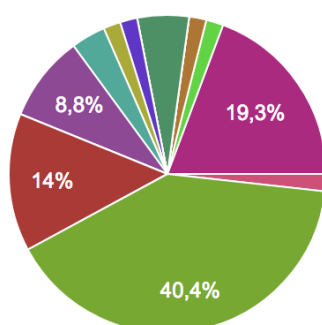


FIGURE 30 – Questionnaire aux chercheurs - question n° 34

35. Souhaitez-vous préciser votre domaine de recherche ?

Nombre de réponses : 38.

36. Quelle est votre tranche d'âge ? *

Nombre de réponses : 57.

- Entre 20 et 30 ans : 3
- Entre 31 et 40 ans : 24
- Entre 41 et 50 ans : 15
- Entre 51 et 60 ans : 8
- Plus de 61 ans : 7

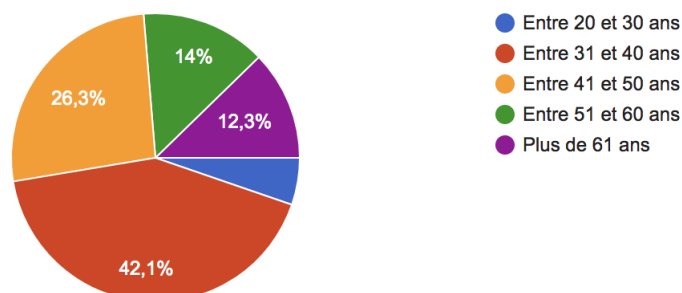


FIGURE 31 – Questionnaire aux chercheurs - question n° 36

37. Seriez-vous disponible pour être contacté(e) par courriel pour préciser certains points de ce questionnaire ? *

Nombre de réponses : 57.

— Oui : 42

— Non : 15

Le questionnaire aux bibliothécaires

Le texte

S'adressant aux professionnels des bibliothèques (principalement aux responsables des services de numérisation, des bibliothèques numériques et des services aux chercheurs), ce questionnaire a pour but de récolter des informations au sujet des différentes possibilités de collaborations, existantes ou souhaitables, entre les acteurs de la recherche et les bibliothécaires concernant la mise en place de projets de numérisation de corpus documentaires. Il s'agit notamment de corpus patrimoniaux, mais pas uniquement (fonds anciens et modernes, presse, photographies, thèses, etc.). L'enquête prend en compte les projets de numérisation de corpus documentaires menés depuis 2012. Par projet de numérisation de corpus l'on entend la numérisation d'une sélection de plusieurs documents répondant à une problématique précise (scientifique, pédagogique, culturelle, patrimoniale, etc.); l'on exclut la numérisation courante d'ouvrages, ainsi que la numérisation à de strictes fins de conservation. L'enquête est diffusée par Graziella Pastore (graziella.pastore@enssib.fr), élève conservateur à l'Esssib, dans le cadre de la préparation de son mémoire en vue de l'obtention du diplôme de conservateur d'État des bibliothèques (Esssib, DCB 26). Ce questionnaire ne vous prendra pas plus de 15 minutes. Merci beaucoup de votre collaboration !

Les réponses seront collectées jusqu'au 15 décembre 2017.

**Obligatoire*

Institution et mission de l'agent

1. Indiquez l'institution dans laquelle vous travaillez *

Merci d'indiquer votre affectation principale

2. Indiquez votre mission principale *

Numérisation de corpus documentaires

Cette enquête prend en compte les projets de numérisation de corpus documentaires menés depuis 2012.

Par projet de numérisation de corpus l'on entend la numérisation d'une sélection de plusieurs documents répondant à une problématique précise (scientifique, pédagogique, culturelle, patrimoniale, etc.); l'on exclut la numérisation courante d'ouvrages, ainsi que la numérisation à de strictes fins de conservation.

3. Depuis 2012, combien de projets de numérisation de corpus documentaires ont été menés dans votre service ? *

Une seule réponse possible.

- 1 projet
- 2 projets
- 3 projets
- 4 projets
- 5 projets
- Entre 5 et 10 projets
- Plus de 10 projets
- Estimation impossible
- Aucun projet

Passez à la question 17.

Typologie de corpus documentaires

4. Quelle est la typologie des corpus numérisés depuis 2012 ? *

Plusieurs réponses possibles.

- Production d'un même auteur
- Production de plusieurs auteurs sur un thème cible
- Production de plusieurs auteurs sur une période cible
- Manifestations ou expressions différentes d'une même œuvre
- Documents pédagogiques servants à l'histoire d'une discipline
- Documents d'archive
- Autre : ...

5. Quelle est la nature principale des documents numérisés ? *

Plusieurs réponses possibles.

- Cartes et plans
- Dessins
- Documents d'archive
- Estampes
- Incunables
- Livres imprimés avant 1810
- Livres imprimés après 1810
- Manuscrits
- Photographies
- Autre : ...

6. Pouvez-vous indiquer le sujet principal des derniers corpus numérisés dans votre service, ainsi qu'un lien vers la documentation éventuellement en ligne à leur sujet ?

Initiative

7. Parmi les numérisations de corpus effectuées depuis 2012, savez-vous combien ont été sollicitées par un chercheur ou par une équipe de chercheurs ? *

Merci de prendre en compte seulement les demandes effectuées à partir de 2012.

Une seule réponse possible.

- De 1 à 4
- Entre 5 et 10
- Plus de 10
- Aucune
- Estimation impossible

8. Est-ce que votre service catégorise la provenance des demandes de numérisation ?
Si oui, selon quelle typologie (particuliers, chercheurs, professionnels, etc.) ?

Gestion de la demande de numérisation

9. Parmi les demandes reçues, quelle est la plus fréquente ? *

Une seule réponse possible.

- Production d'un même auteur
- Production de plusieurs auteurs sur un thème cible

- Production de plusieurs auteurs sur une période cible
 - Manifestations ou expressions différentes d'une même œuvre
 - Documents pédagogiques servants à l'histoire d'une discipline
 - Documents d'archive
 - Estimation impossible
 - Autre : ...
10. Votre institution dispose-t-elle d'un guichet ou d'un service consacré aux chercheurs, et notamment aux demandes de numérisation sollicitées par les chercheurs ? *

Une seule réponse possible.

- Oui
- Non

11. Si oui, pouvez-vous nous dire quelques mots sur ce service et nous indiquer un référent à contacter ?

Travaux et outils de la numérisation

12. Depuis 2012, les travaux de numérisation de corpus ont-ils été réalisés en interne directement par votre institution ? *

Une seule réponse possible.

- Oui, tous les travaux ont été menés en interne dans votre institution
- Non, votre institution a eu recours à un (ou plusieurs) prestataire(s) externe(s)
- Non, votre institution s'est appuyée sur un partenariat avec d'autres institutions
- Autre : ...

13. Quelles sont les techniques de numérisation principalement proposées (ou directement utilisées) dans votre institution depuis 2012 ? *

Plusieurs réponses possibles.

- Numérisation par prise de vue photographique
- Numérisation par balayage
- Numérisation des microformes Autre : ...

14. Dans le cadre des numérisations de corpus effectuées depuis 2012, votre institution a-t-elle proposé (ou effectué directement) la conversion de certains documents en mode texte ? Si oui, dans quel format ? *

Plusieurs réponses possibles.

- Oui, simple reconnaissance des caractères (OCR)
- Oui, au format ALTO
- Oui, au format ePub
- Oui, au format TEI
- Non
- Autre : ...

15. Votre institution sollicite-t-elle (ou effectue directement) la création de métadonnées pour les documents numérisés ? Si oui, dans quels formats ? *

Plusieurs réponses possibles.

- Oui, au format METS
- Oui, au format TEI
- Oui, au format EAD

- Non
- Autre : ...

16. Votre institution dispose-t-elle d'un outil pour la mise en ligne des corpus numérisés ? Si oui, lequel (CMS, système de gestion documentaire, etc.) ? *

Expertise

17. Depuis 2012, votre institution a-t-elle sollicité des chercheurs pour une expertise technique ou scientifique dans le cadre de nouvelles numérisations de corpus documentaires ou pour des corpus documentaires précédemment numérisés ? *

Une seule réponse possible.

- Oui, pour des nouvelles numérisations
- Oui, concernant des corpus déjà numérisés
- Oui, concernant des corpus déjà numérisés ainsi que pour des nouvelles numérisations
- Non

18. Si oui, à quel propos ?

Plusieurs réponses possibles.

- Identification des documents originaux et/ou création des métadonnées correspondantes
- Sélection des documents à numériser
- Restauration
- Préparation d'une exposition
- Préparation d'une exposition virtuelle ou d'un produit numérique
- Valorisation d'un fonds à travers un travail de recherche (article, thèse, conférence, etc.)
- Choix ou conversion du modèle des métadonnées ou de mise en consultation des documents
- Choix ou conversion des outils de stockage des documents numériques
- Autre : ...

19. Si oui, pouvez-vous nous expliquer sous quelle forme cette collaboration a été menée ?

20. Si oui, quel bilan tirez-vous de cette collaboration ?

Financement des projets (1)

21. Depuis 2012, votre institution a-t-elle reçu des financements pour des projets de numérisation de corpus documentaires ? *

Une seule réponse possible.

- Oui *Passez à la question 22.*
- Non *Passez à la question 23.*

Financement des projets (2)

22. Depuis 2012, quelles ont été les sources des financements accordées à votre institution pour la numérisation de corpus documentaires ? *

Plusieurs réponses possibles.

- Bibliothèque scientifique numérique (BSN)

- Concours particulier (mécénat)
- Contrat d'établissement
- Contrat de plan État Région (CPER)
- Financements locaux (municipalité, conseil général, région, etc.)
- Fonds européens
- FR(R)AB
- Programmes de numérisation concertée (BnF)
- Plan national de numérisation du ministère de la Culture et de la Communication
- Autre : ...

Dispositifs de valorisation et d'aide à la recherche

Cette section concerne les dispositifs permettant le développement des échanges de la communauté scientifique autour des collections numérisées ou susceptibles d'être numérisées.

23. Votre institution ... *

Plusieurs réponses possibles.

- dispose d'un "service aux chercheurs" (accueil, réservation des places ou des documents, espace recherche, etc.)
- propose des formations réservées aux chercheurs
- prévoit des aides destinées aux jeunes chercheurs (bourses, contrats de recherche, etc.)
- accueille des séminaires de recherche universitaires
- participe à des programmes de recherche pluriannuelles (ANR, Horizon 2020, etc.)
- a des partenariats avec des laboratoires ou des groupes de recherche (nationaux ou internationaux)
- ne dispose d'aucune des options proposées
- autre : ...

Passez à la question 24.

Suggestions

24. Dans votre expérience professionnelle ou scientifique, en France ou ailleurs, avez-vous rencontré de bonnes pratiques de collaboration entre chercheurs et bibliothécaires que vous souhaitez nous signaler ? Avez-vous de remarques ou des suggestions notamment concernant les projets de numérisation de corpus documentaires ?
25. Seriez-vous disponible pour être contacté(e) par courriel pour préciser certains points de ce questionnaire ? *

Une seule réponse possible.

- Oui *Passez à la question 26.*
- Non *Arrêtez de remplir ce formulaire.*

Contact

26. Merci d'indiquer votre adresse de courriel *

Sélection de résultats

Institution et mission de l'agent

1. Indiquez l'institution dans laquelle vous travaillez *

Nombre de réponses : 87.

- Aix-Marseille Université
- Archives départementales de l'Hérault
- Bibliothèque nationale de France
- Bibliothèque de documentation internationale contemporaine
- Bibliothèque de l'Institut de France : 2
- Bibliothèque de l'Institut national d'Histoire de l'Art
- Bibliothèque de l'Université Clermont Auvergne
- Bibliothèque du Grand Verdun
- Bibliothèque du musée Condé, Chantilly
- Bibliothèque du Patrimoine de Clermont Auvergne Métropole
- Bibliothèque francophone multimédia de Limoges
- Bibliothèque interuniversitaire Cujas
- Bibliothèque Mazarine
- Bibliothèque multimédia intercommunale Epinal
- Bibliothèque municipale d'Angers
- Bibliothèque municipale d'Autun
- Bibliothèque municipale de Aix en Provence
- Bibliothèque municipale de Besançon
- Bibliothèque municipale de Bordeaux
- Bibliothèque municipale de Bourges
- Bibliothèque municipale de Brest
- Bibliothèque municipale de Chalon sur Saône
- Bibliothèque municipale de Dijon
- Bibliothèque municipale de Lyon
- Bibliothèque municipale de Mulhouse
- Bibliothèque municipale de Nantes
- Bibliothèque municipale de Nîmes
- Bibliothèque municipale de Tours
- Bibliothèque municipale du Havre
- Bibliothèque nationale et universitaire
- Bibliothèque patrimoniale Pau - Réseau des médiathèques de la Communauté d'agglomération Pau Béarn Pyrénées
- Bibliothèque Sainte-Geneviève
- Bibliothèque universitaire historique de médecine de Montpellier
- Bibliothèques de la Ville de Compiègne
- Bibliothèques de Nancy
- BIS
- BIU Santé : 1 + 1 Pôle pharmacie
- BMVR de Nice
- BUPMC Université Pierre et Marie Curie
- Centre Pompidou - Bibliothèque Kandinsky
- CIRDÒC (Centre InterRégional de développement de l'Occitan)

- Collège de France
- Conseil départemental du Val-d'Oise
- Conservatoire national des arts et métiers, Service commun de la documentation
- École française d'Extrême-Orient
- École nationale des chartes
- École nationale vétérinaire d'Alfort
- ENSAM
- INRA
- Institut catholique de Paris, Bibliothèque de Fels
- Institut français d'archéologie orientale du Caire
- Institut national de l'information géographique et forestière
- Médiathèque de Vincennes
- Médiathèque François-Mitterrand, Poitiers
- Médiathèque Jean-Jacques Rousseau
- Médiathèque Valéry-Larbaud - Vichy
- MMSH Phonothèque (AMU - CNRS)
- Observatoire de Paris
- Prytanée national militaire
- Rouen Nouvelles Bibliothèques
- Sciences Po : 2
- Université Bordeaux Montaigne
- Université de Bordeaux : 2
- Université de Caen Normandie
- Université de Franche-Comté, SCD
- Université de Haute-Alsace
- Université de la Polynésie française
- Université de Lille 1, Sciences et technologies
- Université de Lorraine
- Université de Perpignan, SCD
- Université de Poitiers, SCD
- Université de Reims Champagne-Ardenne
- Université de Valenciennes et du Hainaut-Cambrésis
- Université fédérale Toulouse Midi-Pyrénées, SICD
- Université Lille 2, SCD
- Université Lyon 1, SCD
- Université Nice Sophia Antipolis, SCD
- Université Paris II Panthéon Assas, SCD
- Université Paris Sorbonne
- Université Paris-Sud
- Université Rennes 2
- Université Sorbonne Paris 4, SCD²⁹⁹
- Ville de Reims

2. Indiquez votre mission principale *

Nombre de réponses : 87.

- Administrateur de la bibliothèque numérique : 2

299. Signalons qu'en janvier 2018, l'université Paris-Sorbonne a depuis fusionné avec l'UPMC pour devenir Sorbonne-Université.

- Archiviste : 1
- Bibliothécaire responsable de la bibliothèque patrimoniale : 1
- Chargé de Conservation de documents patrimoniaux : 1
- Chargé de mission Valorisation et soutien à la recherche / Services aux chercheurs : 3
- Chargé de numérisation : 1
- Chef de projet BNR : 1
- Chef du projet Bibliothèque numérique et coordinatrice des services à la recherche pour le Pôle Lettres Arts Sciences Humaines et Sociales : 1
- Conservateur chargé de la Programmation culturelle et des Fonds patrimoniaux : 1
- Conservateur chargé des collections patrimoniales : 6
- Conservateur chargé des collections patrimoniales et de la numérisation : 1
- Coordinateur des contenus documentaires (poldoc) : 1
- Directeur adjoint / Responsable du pôle Collections & Patrimoine : 1
- Directeur de la conservation et du patrimoine : 1
- Directeur scientifique : 1
- Directeur/Directrice adjoint-e de la bibliothèque, Conservateur des collections patrimoniales : 1
- Directeur/Directrice adjoint-e de la bibliothèque/SCD : 5
- Directeur/Directrice de la bibliothèque/SCD : 6
- Ingénieur d'études : 1
- Professionnel chargé de la gestion des étudiants : 1
- Professionnel chargé de la Valorisation des fonds patrimoniaux : 1
- Professionnel chargé du pôle Jeunesse : 1
- Professionnel chargé du *Text mining* : 1
- Responsable archives / bibliothèque patrimoniale numérique : 1
- Responsable BNR et innovation numérique : 1
- Responsable de fonds et collections : 1
- Responsable de l'informatique documentaire : 1
- Responsable de la bibliothèque : 2
- Responsable de la bibliothèque de la recherche : 1
- Responsable de la bibliothèque numérique : 4
- Responsable de la bibliothèque patrimoniale : 1
- Responsable de la valorisation et de la conservation des collections patrimoniales : 1
- Responsable des bibliothèques de Droit et d'Histoire et de la bibliothèque numérique patrimoniale : 1
- Responsable des collections de conservation : 1
- Responsable des collections iconographiques et de la bibliothèque numérique patrimoniale : 1
- Responsable des services aux chercheurs et des productions numériques : 1
- Responsable du centre de documentation : 3
- Responsable du département Recherche/ d'appui à la recherche : 2
- Responsable du département transversal de l'ingénierie documentaire : 1
- Responsable du patrimoine et de la numérisation : 1
- Responsable du pôle Bibliothèque documentation numérisation : 1
- Responsable du pôle/service Numérisation : 6

- Responsable du service Bibliothèque numérique et Conservation : 1
- Responsable du service de la Bibliothèque numérique et des catalogues : 1
- Responsable du service/pôle Patrimoine/du fonds ancien : 14

Numérisation de corpus documentaires

3. Depuis 2012, combien de projets de numérisation de corpus documentaires ont été menés dans votre service ? *

Nombre de réponses : 87.

- 1 projet : 7 (8 %)
- 2 projets : 10 (11,5 %)
- 3 projets : 10 (11,5 %)
- 4 projets : 6 (6,9 %)
- 5 projets : 8 (9,2 %)
- Entre 5 et 10 projets : 17 (19,5 %)
- Plus de 10 projets : 18 (20,7 %)
- Estimation impossible : 3 (3,4 %)
- Aucun projet : 8 (9,2 %)

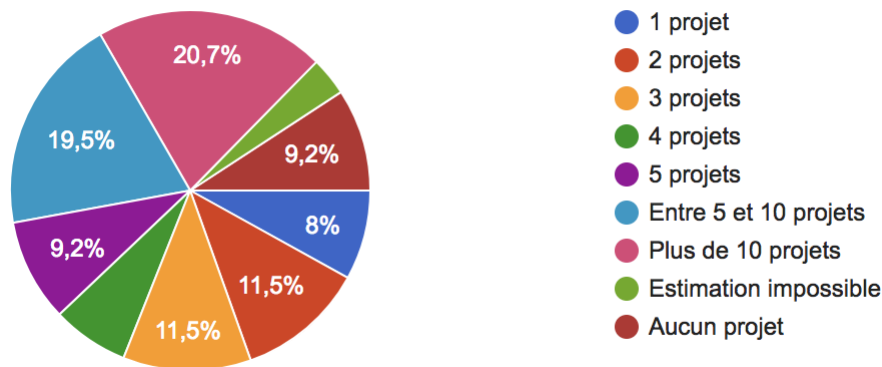


FIGURE 32 – Questionnaire aux bibliothécaires - question n° 3

Typologie de corpus documentaires

4. Quelle est la typologie des corpus numérisés depuis 2012 ? *

Nombre de réponses : 79.

- Production d'un même auteur : 23 (16 %)
- Production de plusieurs auteurs sur un thème cible : 40 (28 %)
- Production de plusieurs auteurs sur une période cible : 16 (11 %)
- Production de plusieurs auteurs sur un thème et une période cible : 3 (2 %)
- Manifestations ou expressions différentes d'une même œuvre : 6 (4 %)
- Documents pédagogiques servant à l'histoire d'une discipline : 8 (6 %)
- Documents d'archive : 28 (20 %)
- Presse (locale, ancienne, etc.) et périodiques : 10 (7 %)
- Photographies, cartes postales : 2 (2 %)
- Objets antiques : 2 (1 %)

— Typologies non exploitables : 4 (3 %)

5. Quelle est la nature principale des documents numérisés ? *

Nombre de réponses : 79.

— Cartes et plans : 18 (7 %)

— Dessins : 23 (9 %)

— Documents d'archive : 31 (12 %)

— Estampes : 18 (7 %)

— Incunables : 14 (5 %)

— Livres imprimés avant 1810 : 33 (12 %)

— Livres imprimés après 1810 : 37 (14 %)

— Manuscrits : 36 (14 %)

— Presse et périodiques : 12 (4 %)

— Photographies : 30 (11 %)

— Cartes postales : 4 (1 %)

— Thèses anciennes : 2 (1 %)

— Objets divers (médailles, objets, etc.) : 3 (1 %)

— Autre (partitions, publicités, enquêtes enregistrées, menus, films) : 5 (2 %)

6. Pouvez-vous indiquer le sujet principal des derniers corpus numérisés dans votre service, ainsi qu'un lien vers la documentation éventuellement en ligne à leur sujet ?

Nombre de réponses : 74.

Université de Lorraine - "Corpus juridique lorrain, mené en lien avec un laboratoire, la BM de Nancy et la BnF (programme numérisation des sources juridiques); histoire de l'archéologie antique, à partir de publications de la fin du XIX^e-début XX^e (livres conservés au Centre Edouard Will de l'Université de Lorraine) et de certains documents iconographiques du fonds Burnouf; documents divers représentatifs de différentes disciplines (médecine, littératures et langues, histoire locale, sciences...); Fonds sur le jansénisme et son impact politique, à partir de publications XVII^e ayant appartenu à l'historien René Taveneaux - À venir : numérisation de la totalité de la revue Annales de l'Est, à la demande du CRULH à l'été 2017 (Centre de Recherches Universitaire Lorrain d'Histoire). Plusieurs autres corpus ont été numérisés à la demande de chercheurs ou de labos, mais n'ont pas été mis en ligne (numérisation "à la demande" par les BU lorraines). Ils ne sont pas comptés ici, mais il s'agit des corpus comptés à la question suivante (incluant donc le corpus des Annales de l'Est)"

Bibliothèque municipale de Chalon sur Saône - "1- Roman de la Rose (participation au projet international conduit par la BnF pour la France); 2- pas de sujet : les manuscrits médiévaux (programme IRHT)"

École française d'Extrême-Orient - "Archéologie"

Médiathèque Jean-Jacques Rousseau - "Archives d'un cartographe de l'Afrique Lannoy de Bissy / fonds d'estampe sur la Savoie / cartes postales Savoie : <http://bibliotheque-numerique.chambery.fr/>"

Bibliothèque municipale d'Angers - "Archives du statuaire David d'Angers ; Correspondance de François Grille ; Projet BHLi : production humaniste en Anjou et Maine (1450-1630); Fonds photographique Jacques Mallet ; Premiers récits imprimés des Guerres de Vendée ; Manuscrits musicaux du Moyen Âge ; etc."

INRA - "Bioagresseurs des cultures agricoles : <https://archive.org/details/inra>"

Bibliothèque de l'INHA - "Collections et collectionneurs / ornements / architecture / dessins"

Université Nice Sophia Antipolis, SCD - "Corpus d'écrivains ; corpus sur l'histoire coloniale ; corpus sur l'œuvre d'un chercheur ; corpus sur l'histoire du droit ; corpus sur l'histoire de l'Université"

Bibliothèque francophone multimédia de Limoges - "Corpus de partitions d'un compositeur local"

BIU Santé - "Corpus des dessins originaux de la BIU Santé (<http://www.biusante.parisdescartes.fr/histoire/medica/dessins.php>) ; Numérisation d'un corpus pour le laboratoire LittéPub (pas encore en ligne - <http://littepub.net/>). Numérisation des thèses d'Ancien Régime (en cours de réalisation), ..."

Université de Lille 1, Sciences et technologies - "Corpus en histoire des sciences et des techniques (<http://iris.univ-lille1.fr>)"

BUPMC Université Pierre et Marie Curie - "Correspondance du mathématicien Montessus de Ballore (mise en ligne en cours), corpus "Histoire des mathématiques", thèses de la Faculté des sciences (XIX^e siècle)

ENSAM - "Cours des Écoles nationales d'Arts et Métiers, dessins, revues et cartes postales <http://patrimoine.ensam.eu/>"

Université Bordeaux Montaigne - "Dernier corpus mis en ligne : productions institutionnelles de l'université, qui comprend notamment les annuaires de la faculté, les programmes des cours, les rentrées solennelles et, pour le XX^e siècle, les supports de communication (revue de l'université, bulletin mensuel, etc.). - Dernier corpus numérisé mais non encore mis en ligne : revues de sociétés savantes, numérisées en concertation avec lesdites et sociétés et avec co-financement BnF"

MMSH Phonothèque (AMU - CNRS) - "Disciplines : anthropologie, sociologie, littérature - Aire géographique : Iran, Yémen, Liban, Syrie, Provence"

Université de Caen Normandie - "Documents d'archives liés à l'histoire de l'établissement / Documents d'archives scientifiques : pour ces derniers corpus, mise en ligne prévue en 2018. Les corpus précédents sont déjà consultables sur Gallica"

Université de Poitiers, SCD - "Droit colonial (pas encore mis en ligne)"

Université de Bordeaux - "Droit de la famille"

Université fédérale Toulouse Midi-Pyrénées, SICD - "Droit et sciences juridiques à Toulouse , Archives des préhistoriens en Midi-Pyrénées, Catholicisme dans le Midi toulousain, Protestantisme dans le Sud de la France, Sciences et techniques en pays toulousain, Enseignement en pays toulousain, Langue occitane <http://tolosana.univ-toulouse.fr/fr/corpus>"

Institut catholique de Paris, bibliothèque de Fels - "École française de spiritualité ; exégèse biblique ; relations Église et Etat ; photographies du Proche-Orient cf. <https://bibliotheques.icp.fr/rechercher/collections-numerisees-icp/>"

- Médiathèque François-Mitterrand, Poitiers - "Éditions poitevines du 16^e siècle; archives relatives à l'histoire du Poitou"
- Sciences Po - "Elections françaises V^e république <https://archive.org/details/archiveselectoralesducevipof>"
- Bibliothèque municipale du Havre - "Essentiellement des photographies de Bains de mer et de la ville du Havre avant destruction <http://ged.lireauhavre.fr/>"
- École nationale des chartes - "Estampes / Revue numismatique / Tirés à part, <http://bibnum.enc.sorbonne.fr/collections/browse>"
- Conservatoire national des arts et métiers, Service commun de la documentation - "Expositions universelles, catalogues de constructeurs, Description des arts et métiers (collection d'ouvrages)"
- Centre Pompidou - Bibliothèque Kandinsky - "Fonds d'artistes, architecture et design, archivesetdocumentation.centrepompidou.fr"
- Bibliothèque municipale de Bordeaux - "Fonds Delpit (estampes sur Bordeaux et le Sud-Ouest)"
- Bibliothèque municipale de Dijon - "Fonds Jules Legras (archives, photographies) dans le cadre de la commémoration des révolutions russes de 1917, http://patrimoine.bm-dijon.fr/pleade/ead.html?id=FR212316101_etat_fonds&c=FR212316101_etat_fonds_BMDIJON6812L, https://twitter.com/jules_legras"
- Bibliothèque Sainte-Geneviève - "Fonds Tessin, histoire de l'art, [https://archive.org/details/bibliothequesaintegenevieve?and\[\]=Fonds%20Tessin](https://archive.org/details/bibliothequesaintegenevieve?and[]=Fonds%20Tessin)"
- Bibliothèque de documentation internationale contemporaine - "Guerre 1914-1918 (photographies), Immigration d'Amérique latine (périodiques), Guerre 1939-1945 : Stalags et Oflags (périodiques), Guerre d'Algérie (Tracts, documents d'archives). Ligue des droits de l'homme (périodiques), <http://argonnaute.u-paris10.fr/>"
- BIU Santé, pôle Pharmacie - "Histoire de la pharmacie, cosmétologie, botanique, chimie"
- Bibliothèques de la Ville de Compiègne - "Histoire de la Picardie"
- Université Lyon 1, SCD - "Histoire des Mathématique et thèses de santé lyonnaises"
- Université Paris-Sud - "Histoire du droit"
- Université de Haute-Alsace - "Histoire industrielle et histoire du travail en Alsace au XIX^e siècle, voir sur <http://www.phototheque-scd.uha.fr>"
- Bibliothèque municipale de Nantes - "Histoire littéraire et artistique de Nantes (surréalisme : fonds Sarment et manuscrits de Jacques Vaché; correspondance de Julien Lanoë), manuscrits de Marcel Schwob, cartes postales locales, fonds d'autographes (fonds Labouchère en cours de numérisation), manuscrits de Jules Verne, périodique Le Phare de la Loire (en cours de numérisation), <https://bm.nantes.fr/home/espaces-dedies/patrimoine/patrimoine-numerise.html>"

- Bibliothèque municipale de Aix en Provence - "<http://bibliotheque-numerique.citedulivre-aix.com/>"
- Bibliothèque Mazarine - "<http://mazarinum.bibliotheque-mazarine.fr/a-propos/>"
- Bibliothèque municipale de Nîmes - "Incunables ; hebraica ; impressions nîmoises anciennes ; bibliothèque théâtrale de la comtesse de Verrue ; histoire locale ; protestantisme à Nîmes ; manuscrits de Séguier ; presse locale
- Bibliothèque de l'Institut de France - "Inventaire des dessins de Charles Percier, en partenariat avec la RMN-GP et l'INHA (http://agorha.inha.fr/inhaprod/jsp/reference.jsp?reference=INHA__METADONNEES__51); Numérisation du journal "La Vie parisienne" de 1863 à 1914 en partenariat avec l'INHA (<https://www.inha.fr/fr/ressources/actualites/en-2017/mise-en-ligne-de-la-base-la-vie-parisienne.html?search-keywords=vieparisienne>); Correspondance de Flaubert, en partenariat avec le Centre Flaubert de l'Université de Rouen (<http://flaubert.univ-rouen.fr/correspondance/>)"
- Bibliothèque municipale de Lyon - "Journaux des canuts pour alimenter L'Echo de la fabrique (<http://echo-fabrique.ens-lyon.fr/document.php?id=1090>) et Numelyo, bibliothèque numérique de la Ville de Lyon; Corpus iconographique dans le cadre du projet Biblissima "Le livre illustré à Lyon : 1480-1600. Il faudrait pouvoir restituer également de la numérisation sur demande de la part des chercheurs qui prend une part importante des flux de numérisation notamment dans le cadre du marché de numérisation de masse des imprimés libres de droits (2008-2017) Pour exemple : en lien avec l'ENS Lyon dans le cadre d'un séminaire sur les "Voyageuses britanniques" : collaboration autour de corpus numérisés ou à numériser et la production de contenus par les acteurs du séminaire"
- Aix-Marseille Université - "Jurisprudence Provençale"
- Bibliothèque du Patrimoine de Clermont Auvergne Métropole - "L'Auvergne Pittoresque (pas encore en ligne), À l'épreuve de la Grande Guerre (<https://www.bibliotheques-clermontmetropole.eu/overnia/>), Un provincial nommé Blaise Pascal (<https://www.bibliotheques-clermontmetropole.eu/overnia/>)"
- Bibliothèque municipale de Brest - "La Dépêche de Brest"
- Université de Valenciennes et du Hainaut-Cambrésis - "Thèses. La numérisation est confiée à un prestataire (Atelier National de Reproduction des Thèses). Les corpus sont actuellement en cours de numérisation. Nous prévoyons de les mettre en ligne via notre archive institutionnelle dédiée aux thèses et aux mémoires Theoreme (<http://theoreme.univ-valenciennes.fr/index.html>)"
- Université de Perpignan, SCD - "Langue Catalane / Mexique / Histoire de l'université de Perpignan, <https://estudi.univ-perp.fr/collections/browse>, <https://estudi-hypotheses-org.ezproxy.univ-perp.fr/20>"

- Bibliothèque municipale de Bourges - "Le Berry savant de la Renaissance ; Corpus d'archives Alain-Fournier ; cartes postales de Bourges - cf. <http://www.mediatheque-bourges.fr/EXPLOITATION/PATRIMOINE/bibliotheque-patrimoine.aspx>"
- École nationale vétérinaire d'Alfort - "Le cheval"
- Rouen Nouvelles Bibliothèques - "Le corpus des manuscrits et oeuvres de Flaubert, visibles sur Bovary.fr : <http://www.bovary.fr/>"
- Bibliothèque interuniversitaire Cujas - "Les fondamentaux du droit ; les publications officielles des anciennes colonies françaises ; les crises constitutionnelles françaises"
- Université Sorbonne Paris 4, SCD - "Les projets de numérisation ne sont pas menés par le SCD mais par les laboratoires (organisation liée au fonctionnement spécifique de Paris-Sorbonne où les UFR sont très indépendants). Il s'agit d'une forme de mécénat pour numérisation (qui n'a pas encore été faite). Le SCD a financé partiellement l'achat du scanner. Un laboratoire travaille actuellement sur Tolstoï : <http://eurorbem.paris-sorbonne.fr/spip.php?article636>"
- Bibliothèque municipale de Mulhouse - "Lithographie (Godefroy Engelmann) , Histoire industrielle (Bulletin de la Société industrielle de Mulhouse)"
- Université Rennes 2 - "Littérature celtique (<http://bibnum.univ-rennes2.fr/>)"
- Prytanée national militaire - "Livres d'architecture du XVI^e siècle et Humanités ligériennes"
- Bibliothèque municipale d'Autun - "Manuscrits médiévaux (IRHT)"
- Bibliothèque universitaire historique de médecine de Montpellier - "Médecins et pharmaciens montpelliérains, <http://www.biu-montpellier.fr/florabium/jsp/redir.jsp?redir=folia>; Manuscrits de la BU médecine et notamment les 72 manuscrits du fonds Clairvaux en lien avec la Médiathèque du Grand Troyes pour le projet : <https://www.bibliotheque-virtuelle-clairvaux.com/>"
- Bibliothèque du Grand Verdun - Numérisation d'une partie des manuscrits médiévaux de la Bibliothèque; de documents iconographiques issus du fonds "14/18"; de plans réalisés par le chanoine Marie-André de Chaligny de Plaine (XVIII^e siècle)"
- Université de la Polynésie française - "Numérisation d'une revue le BSEO grâce à une convention avec la société savante, numérisation de cartes postales et de revues en langue tahitienne : <http://anaite.upf.pf>"
- Ville de Reims - "Numérisation des titres de presse quotidienne régionale"
- Bibliothèque du musée Condé, Chantilly - "Orangerie"
- Sciences Po - "Ouvrages de littérature, <https://archive.org/details/sciencespo>"
- Université de Bordeaux - "Patrimoine médical : <http://www.babordnum.fr/collections/show/1>"

Bibliothèque patrimoniale Pau - Réseau des médiathèques de la Communauté d'agglomération Pau Béarn Pyrénées - "Pau Béarn Pyrénées : manuscrits et iconographie + presse"

Archives départementales de l'Hérault - "Périodiques locaux traitant de l'Hérault, <http://archives-pierresvives.herault.fr/archive/recherche/bibliotheque/n:27>"

CIRDÒC (Centre InterRégional de développement de l'Occitan) - "Périodiques sur ou en langue d'oc (dont presse régionale)"

Bibliothèque multimédia intercommunale Epinal - "Presse / manuscrits médiévaux / Vosges"

Bibliothèque nationale et universitaire - "Presse alsacienne; Théâtre alsacien; Livres alsaciens du 16e siècle; Livres d'amitié; etc. <http://www.bnu.fr/numistral/projets-en-cours>. Numérisation à la demande de documents particuliers"

Conseil départemental du Val-d'Oise - "Presse ancienne locale (www.archives.valdoise.fr puis je recherche (en mode texte pour la presse ancienne))"

Bibliothèques de Nancy - "Presse quotidienne régionale (<http://www.kiosque-lorrain.fr>)"

Bibliothèque nationale de France - "Projet Deloynes (<http://c.bnf.fr/gcM>); projet Collecta (<http://c.bnf.fr/gcP>); projet Verspera (<http://c.bnf.fr/gcV>)"

Observatoire de Paris - "Publications de l'Observatoire de Paris et des observatoires français de province"

Institut français d'archéologie orientale du Caire - "Relations entre la France et l'Égypte, <http://heritage.bnf.fr/bibliothequesorient/fr>"

Médiathèque Valéry-Larbaud - Vichy - "Thermalisme : <http://bibliotheque.clermont-universite.fr/collections-numerisees>"

Bibliothèque de l'Université Clermont Auvergne - "Thermalisme, littérature populaire début du 20^e siècle, herbiers, Factums"

Initiative

7. Parmi les numérisations de corpus effectuées depuis 2012, savez-vous combien ont été sollicitées par un chercheur ou par une équipe de chercheurs ? *

Nombre de réponses : 79.

- De 1 à 4 : 20 (25,3 %)
- Entre 5 et 10 : 16 (20,3 %)
- Plus de 10 : 15 (19 %)
- Aucune : 11 (13,9 %)
- Estimation impossible : 17 (21,5 %)

8. Est-ce que votre service catégorise la provenance des demandes de numérisation ? Si oui, selon quelle typologie (particuliers, chercheurs, professionnels, etc.) ?

Nombre de réponses : 63.

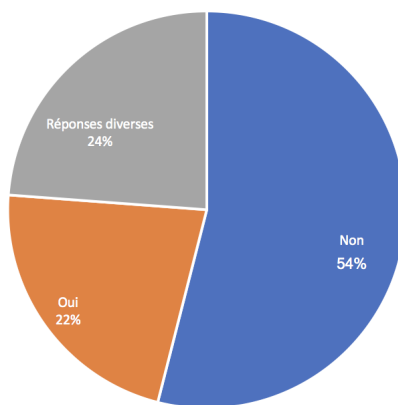


FIGURE 33 – Questionnaire aux bibliothécaires - question n° 8

Gestion de la demande de numérisation

9. Parmi les demandes reçues, quelle est la plus fréquente ? *

Nombre de réponses : 79.

- Production d'un même auteur : 4 (5 %)
- Production de plusieurs auteurs sur un thème cible : 15 (19 %)
- Production de plusieurs auteurs sur une période cible : 3 (4 %)
- Manifestations ou expressions différentes d'une même œuvre : 1 (1 %)
- Documents d'archive : 15 (19 %)
- Documents pédagogiques servant à l'histoire d'une discipline : 3 (4 %)
- Presse et périodiques locaux anciens : 2 (3 %)
- Corpus contenant des données à extraire par *text mining* : 1 (1 %)
- Estimation impossible : 27 (34 %)
- Demandes ponctuelles et variées : 8 (10 %)

"documents isolées faisant l'objet d'une recherche personnelle", "Les chercheurs demandent rarement des corpus mais des ouvrages isolés", "les demandes reçues ne portent en général que pour des documents ponctuelles, en dehors du corpus Droit et sciences juridiques", "Les demandes varient en fonction de leurs recherches. Difficile à estimer. Les deux premières réponses je dirais", "ne correspond pas à ce que nous faisons", "on ne traite pas la numérisation à la demande", "Pas de demande de numérisation de corpus, mais demande de numérisation d'un document en particulier", "qu'ils soient libres de droit ou non, et tous supports confondus"

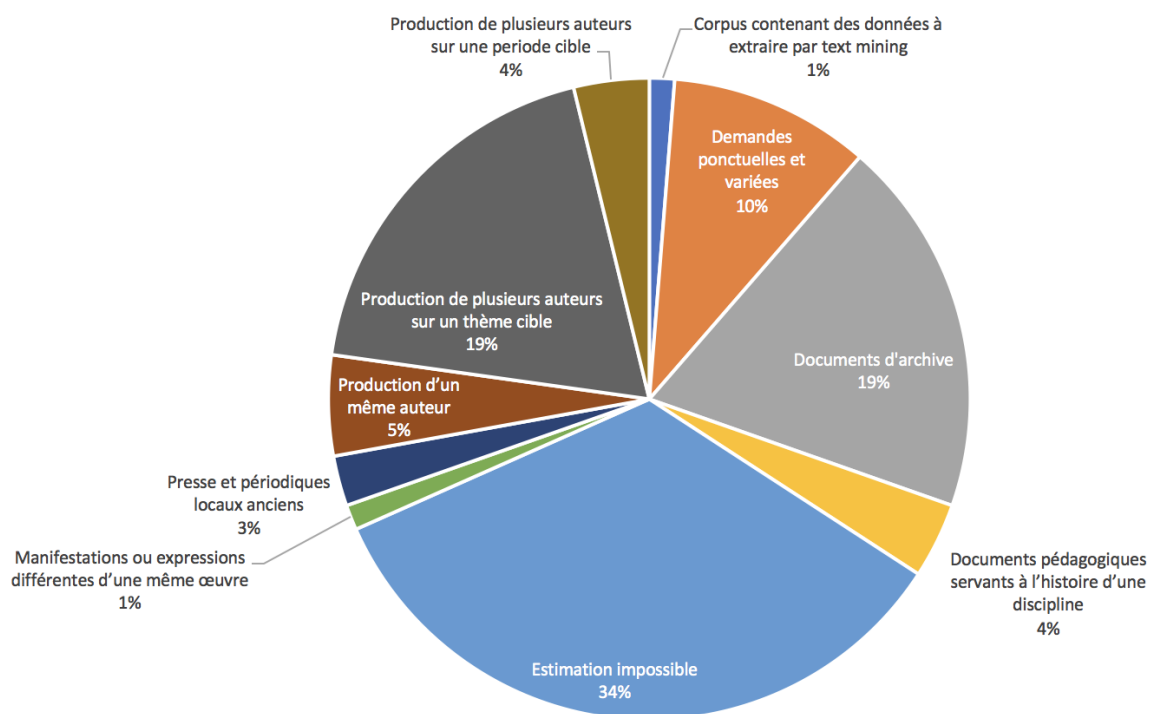


FIGURE 34 – Questionnaire aux bibliothécaires - question n° 9

10. Votre institution dispose-t-elle d'un guichet ou d'un service consacré aux chercheurs, et notamment aux demandes de numérisation sollicitées par les chercheurs ?*

Nombre de réponses : 79.

— Oui : 35 (44,3 %)

— Non : 44 (57,7 %)

11. Si oui, pouvez-vous nous dire quelques mots sur ce service et nous indiquer un référent à contacter ?

Nombre de réponses : 41.

Travaux et outils de la numérisation

12. Depuis 2012, les travaux de numérisation de corpus ont-ils été réalisés en interne directement par votre institution ? *

Nombre de réponses : 79.

— Oui, tous les travaux ont été menés en interne dans votre institution : 10 (13 %)

— Non, votre institution a eu recours à un (ou plusieurs) prestataire(s) externe(s) : 46 (58 %)

— Non, votre institution s'est appuyée sur un partenariat avec d'autres institutions (BnF, BSN 5, etc.) : 16 (20 %)

— Non, votre institution s'est appuyée sur des prestataires et sur un partenariat avec d'autres institutions (BnF, BSN 5, etc.) : 7 (9 %)

13. Quelles sont les techniques de numérisation principalement proposées (ou directement utilisées) dans votre institution depuis 2012 ? *

Nombre de réponses : 79.

- Numérisation par prise de vue photographique : 56
- Numérisation par balayage : 38
- Numérisation des microformes : 8
- Aucune sur les documents textuels. Travail de numérisation interne de photographies anciennes : 1
- Numérisation droite format waveave : 1
- Production d'une image numérique grâce à la station de num Cobalt : 1
- Numérisation par scanner de numérisation patrimoniale (ne fonctionne pas par balayage) : caméra linéaire, vitre patrimoniale amovible, pas de contact avec le document : 1
- Réponses non exploitables : 1

14. Dans le cadre des numérisations de corpus effectuées depuis 2012, votre institution a-t-elle proposé (ou effectué directement) la conversion de certains documents en mode texte ? Si oui, dans quel format ? *

Nombre de réponses : 79.

- Oui, simple reconnaissance des caractères (OCR) : 34
 Université fédérale Toulouse Midi-Pyrénées, SICD - OCR très très ponctuelle (brochures annotées pour les Archives de préhistoriens)
- Oui, au format ALTO : 21
- Oui, au format ePub : 4
 Bibliothèque municipale de Lyon ; Sciences Po ; Inra ; Bibliothèque Sainte-Geneviève
- Oui, au format TEI : 5
 Bibliothèque municipale de Lyon ; CIRDOC (Centre InterRégional de développement de l'Occitan) ; Bibliothèque Sainte-Geneviève ; Université Nice Sophia Antipolis, SCD ; Bibliothèque de documentation internationale contemporaine (de manière très marginale)
- Autres formats (TIFF, PDF) : 3
- Non : 23
 Médiathèque François-Mitterrand, Poitiers - Projet 2017 : OCR + Alto prévu ; Université de Caen Normandie - Nous sommes en train de structurer la chaîne interne pour OCR et TEI

15. Votre institution sollicite-t-elle (ou effectue directement) la création de métadonnées pour les documents numérisés ? Si oui, dans quels formats ? *

Nombre de réponses : 79.

- Dublin Core : 13
- Oui, au format EAD : 18
- Oui, au format METS : 25
- Oui, au format MODS : 1
- Oui, au format TEI : 6
- OLAC, EDM : 1
- IPTC : 1
- Non, en projet : 3
- Non : 27
- Format maison : 2
- UNIMARC (pour les périodiques) : 1
- "Métadonnées encapsulées au format XMP avec des balises Dublin Core" : 1

- "Conversion UNIMARC en XML développée en interne" 1
 - "PDF (métadonnées, indexation autour de la table des matières)" 2
 - Réponses non exploitables : 2
16. Votre institution dispose-t-elle d'un outil pour la mise en ligne des corpus numérisés? Si oui, lequel (CMS, système de gestion documentaire, etc.)? *
- Nombre de réponses : 79.*
- Non : 8
 - CMS de gestion de bibliothèque numérique : 4
 - Omeka : 15
 - Limb Gallery™ : 7
 - Bibliothèque numérique en cours de développement, en évolution : 11
 - Ligeo Archives : 2
 - Base de données avec interface développée en interne : 2
 - Serveur de stream : 1
 - Système de gestion électronique des documents (GED) : Keepeek, Inmédia, etc. : 5
 - Syracuse SIGB Digital Ready : 1
 - EZpublish : 1
 - Flora SIGB : 2
 - Base FileMaker Pro : 1
 - Drupal : 1
 - CMS Armadillo : 1
 - CMS Mnesys : 2
 - CMS développé en interne : 1
 - Gallica Marque Blanche : 1
 - Pleade outil de gestion d'instruments de recherche en EAD : 2
 - Syracuse : 1
 - Visionneuse associée aux inventaires en ligne : 1
 - Internet Archive : 3
 - FlippingBook : 1
 - ORI-OAI : 1
 - Numelyo (FEDORA pour le stockage, FEZ et E-Babel pour la visualisation) : 1
 - Osiros : 1
 - D-Space : 1
 - Inmedia : 1

Expertise

17. Depuis 2012, votre institution a-t-elle sollicité des chercheurs pour une expertise technique ou scientifique dans le cadre de nouvelles numérisations de corpus documentaires ou pour des corpus documentaires précédemment numérisés? *
- Nombre de réponses : 87.*
18. Si oui, à quel propos?
- Nombre de réponses : 60.*
- Identification des documents originaux et/ou création des métadonnées correspondantes : 29

- Sélection des documents à numériser : 41
 - Restauration : 2
 - Préparation d'une exposition : 15
 - Préparation d'une exposition virtuelle ou d'un produit numérique : 17
 - Valorisation d'un fonds à travers un travail de recherche (article, thèse, conférence, etc.) : 30
 - Choix ou conversion du modèle des métadonnées ou de mise en consultation des documents : 4
 - Constitution de l'arborescence des corpus thématiques : 1
 - Introduction à des instruments de recherche : 1
19. Si oui, pouvez-vous nous expliquer sous quelle forme cette collaboration a été menée ?
Nombre de réponses : 53.
20. Si oui, quel bilan tirez-vous de cette collaboration ?
Nombre de réponses : 45.
- Assez peu utile : 1 (2 %)
 - Autres réponses : 6 (13 %)
 - a) "Appui pour montrer l'intérêt scientifique d'un projet"
 - b) "Collaboration nécessaire pour l'expertise scientifique mais parfois complexe sur le plan relationnel"
 - c) "Collaborations indispensables"
 - d) "Encore trop tôt pour un bilan"
 - Mitigé : 6 (13 %)
 - a) "Mitigé à cause du manque de disponibilité des chercheurs (les interlocuteurs ont souvent changé)"
 - b) "Mitigé. Bonne implication au départ, lassitude des partenaires devant les temps de réalisation (identification, nommage, contrôle-qualité), retour une fois ces opérations "techniques" réalisées"
 - c) "Participation limitée des chercheurs pour la rédaction, implication forte pour la sélection des titres du corpus"
 - d) "Pas de vrai retour (= liste) de la part des chercheurs"
 - Positif : 17 (38 %)
 - a) "Bilan positif : enrichissement mutuel, mais très gourmand en temps ; ne peut fonctionner qu'avec des chercheurs vraiment impliqués ; comme il a été difficile pour nous d'en trouver, un grand nombre de corpus n'ont été élaborés que par notre service, ou alors avec des collègues d'autres bibliothèques"
 - b) "Cette collaboration est indispensable pour répondre aux besoins des chercheurs. Triptyque gagnant pour les humanités numériques : chercheur / bibliothécaire ou documentaliste / informaticien"
 - c) "La mise en œuvre de ces missions nécessite la prise en compte des contraintes intellectuelles, techniques, juridiques liées au statut du document numérique"

- d) "La valeur ajoutée de la mise en ligne de ce corpus en histoire provient de l'implication des chercheurs dans le catalogage des documents et dans la création des tables de matières servant à naviguer dans ces documents non océrisables"
 - e) "Positif : toutes les parties y gagnent (d'un côté, accès aux prises de vue, de l'autre, création de métadonnées), ainsi que le public et l'établissement (visibilité et accès aux oeuvres)"
 - f) "Positif, mais sous-exploitation actuelle du résultat"
 - g) "Positive. Ils sont le public cible et les meilleurs connaisseurs des fonds"
 - h) "Tout projet mené en collaboration avec les chercheurs a plus de chance d'aboutir"
 - i) "Toutes ces collaborations sont constructives. Notre politique de numérisation s'oriente de plus en plus vers les demandes des chercheurs"
 - j) "Très efficace si les chercheurs ont un besoin immédiat. Dans le cas des doctorants, crée une dépendance très forte avec la réussite et la poursuite de leurs études"
 - k) "Une mauvais expérience sur un projet sinon les autres projets se sont bien déroulés"
- Très positif/fructueux : 15 (34 %)
- a) "Bilan très positif, les enseignants se sont emparés du dossier et ont grandement enrichi la bibliothèque numérique par leurs notices. Mais cela nécessite de trouver des chercheurs ayant un peu de temps et de l'intérêt pour le patrimoine, ce qui est parfois difficile"
 - b) "Bilan très positif. Les différents projets ont permis de nouer des liens solides avec des chercheurs de diverses institutions et ils ont abouti à des réalisations sur lesquelles la bibliothèque peut s'appuyer pour les projets suivants. Peut-être un enseignement important tiré est de bien déterminer, avant et pendant les projets, ce que la bibliothèque peut apporter et ce qu'elle ne peut pas faire, et de ne pas hésiter à le dire, diplomatiquement"
 - c) "La collaboration avec les chercheurs est très riche, en particulier sur les projets d'exposition. Sur la conception de la collection numérique, deux approches se confrontent et doivent se concilier : les bibliothécaires recherchent une cohérence et une unité dans la collection numérique, tandis que les chercheurs s'intéressent plus à la valeur de chaque document pris séparément"
 - d) "La plupart du temps, très positif. C'est un des publics privilégiés de la numérisation patrimoniale, il est très important que leurs travaux s'appuient sur nos collections et les mettent en valeur"
 - e) "Très positif : une vrai plus value sur les fonds. un travail complexe quand on est pas formé aux métiers de l'édition"

Financement des projets (1)

21. Depuis 2012, votre institution a-t-elle reçu des financements pour des projets de numérisation de corpus documentaires ? *

Nombre de réponses : 87.

- Oui : 67 (77 %)
- Non : 20 (23 %)

Financement des projets (2)

22. Depuis 2012, quelles ont été les sources des financements accordées à votre institution pour la numérisation de corpus documentaires ? *

Nombre de réponses : 67.

- Bibliothèque scientifique numérique (BSN) : 16
- Contrat d'établissement : 6
- Contrat de plan État-Région (CPER) : 4
- Mécénat privé : 12
- Fonds européens : 7
- COMUE : 1
- Plan national de numérisation du ministère de la Culture et de la Communication : 11
- Crédits Investissements d'avenir (Biblissima, LabEx, IdEx, etc.) : 5
- Financement TGIR Huma-Num : 3
- Financements locaux (municipalité, conseil général, région, etc.) : 20
- FR(R)AB : 2
- Dotation générale de décentralisation (BNR, DRAC, etc.) : 4
- Programmes de numérisation concertée (BnF) : 26
- Contrat territoire lecture : 1
- Financement d'un laboratoire universitaire (Bonus Qualité Recherche, fonds ANR, etc.) : 4
- Autres sources : 2
- Ne sais pas : 1

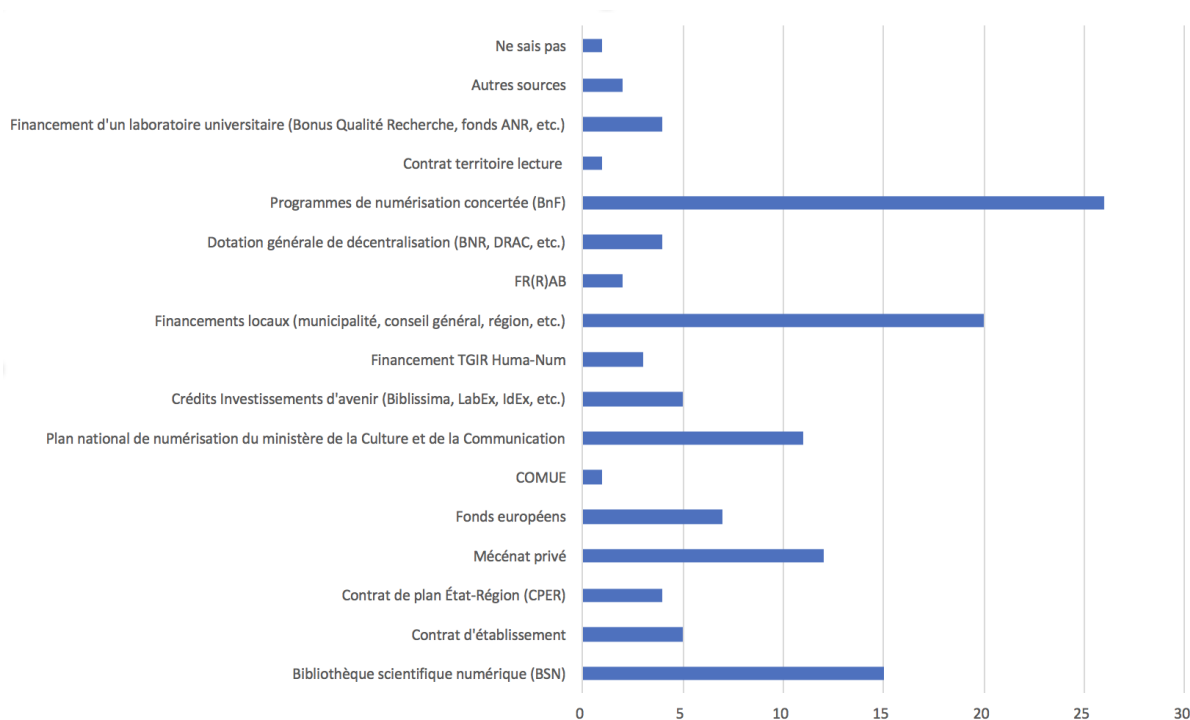


FIGURE 35 – Questionnaire aux bibliothécaires - question n° 22

Dispositifs de valorisation et d'aide à la recherche

23. Votre institution ... *

Nombre de réponses : 87.

- dispose d'un "service chercheurs" (accueil, réservation des places ou des documents, espace recherche, etc.) : 56
- propose des formations réservées aux chercheurs : 36
- prévoit des aides destinées aux jeunes chercheurs (bourses, contrats de recherche, etc.) : 21
- a des partenariats avec des laboratoires ou des groupes de recherche (nationaux ou internationaux) : 53
- participe à des programmes de recherche pluriannuelles (ANR, Horizon 2020, etc.) : 33
- accueille des séminaires de recherche universitaires : 41
- ne dispose d'aucune des options proposées : 10

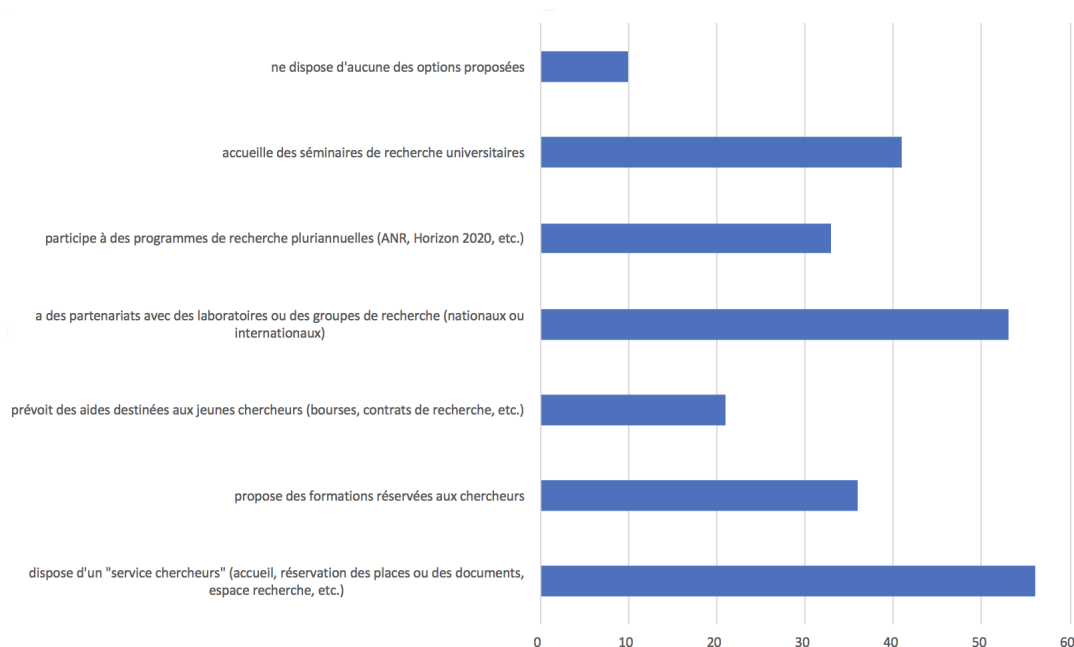


FIGURE 36 – Questionnaire aux bibliothécaires - question n° 23

Suggestions

24. Dans votre expérience professionnelle ou scientifique, en France ou ailleurs, avez-vous rencontré de bonnes pratiques de collaboration entre chercheurs et bibliothécaires que vous souhaitez nous signaler? Avez-vous de remarques ou des suggestions notamment concernant les projets de numérisation de corpus documentaires?

Nombre de réponses : 18.

Selection :

25. Seriez-vous disponible pour être contacté(e) par courriel pour préciser certains points de ce questionnaire? *

Nombre de réponses : 87.

- Oui : 79 (90,8 %)

— Non : 8 (9,2 %)

TABLE DES ILLUSTRATIONS

1	Questionnaire aux chercheurs - question n° 10	24
2	Questionnaire aux chercheurs - question n° 8	25
3	Questionnaire aux chercheurs - question n° 9	26
4	Questionnaire aux bibliothécaires - question n° 4	29
5	Questionnaire aux bibliothécaires - question n° 5	30
6	Questionnaire aux chercheurs - question n° 22	36
7	Questionnaire aux chercheurs - question n° 1	53
8	Questionnaire aux chercheurs - Synthèse de la participation aux projets de numérisation de corpus	54
9	Questionnaire aux bibliothécaires - question n° 7	54
10	Questionnaire aux chercheurs - question n° 4	56
11	Questionnaire aux bibliothécaires - question n° 17	61
12	Questionnaire aux chercheurs - question n° 28	62
13	Questionnaire aux bibliothécaires - question n° 18	62
14	Questionnaire aux chercheurs - question n° 29	62
15	Questionnaire aux bibliothécaires - question n° 15	65
16	Questionnaire aux bibliothécaires - question n° 14	65
17	Questionnaire aux chercheurs - question n° 13	66
18	Questionnaire aux chercheurs - question n° 16	67
19	Questionnaire aux bibliothécaires - question n° 13	68
20	Questionnaire aux bibliothécaires - question n° 12	68
21	Questionnaire aux bibliothécaires - question n° 16	71
22	Questionnaire aux chercheurs - question n° 6	81
23	Questionnaire aux bibliothécaires - question n° 20 - Bilan des collaborations	85
24	Questionnaire aux bibliothécaires - question n° 23 - Bibliothèques des collectivités territoriales	88
25	Questionnaire aux bibliothécaires - question n° 23 - Bibliothèques de l'ESRI	88
26	Questionnaire aux bibliothécaires - question n° 22 - Bibliothèques des collectivités territoriales	90
27	Questionnaire aux bibliothécaires - question n° 22 - Bibliothèques de l'ESRI	91
28	Questionnaire aux chercheurs - question n° 20	126
29	Questionnaire aux chercheurs - question n° 33	129
30	Questionnaire aux chercheurs - question n° 34	130
31	Questionnaire aux chercheurs - question n° 36	130
32	Questionnaire aux bibliothécaires - question n° 3	140
33	Questionnaire aux bibliothécaires - question n° 8	147
34	Questionnaire aux bibliothécaires - question n° 9	148
35	Questionnaire aux bibliothécaires - question n° 22	153
36	Questionnaire aux bibliothécaires - question n° 23	154

TABLE DES MATIÈRES

SIGLES ET ACRONYMES	9
INTRODUCTION	11
Présentation du sujet	11
Périmètre et méthodologie de l'enquête	13
Intérêt et enjeux de l'étude de corpus	17
QU'EST CE QU'UN CORPUS ?	19
Du latin au français	19
Le point de vue des chercheurs	19
Le point de vue des professionnels des bibliothèques	23
Quelles typologies de corpus ?	24
CONSTITUTION ET ACCÈS AUX CORPUS : LES ENJEUX DE LA NUMÉRISA- TION ET DU TRAITEMENT DES DONNÉES	33
Les enjeux didactiques et scientifiques	34
Du travail en salle de lecture à l'approche distante	34
Les Humanités numériques et la création de corpus intelligents	36
Les bibliothèques numériques vers de nouveaux usages ?	42
La gestion et la valorisation des fonds patrimoniaux	42
Une offre numérique en cours d'évolution	44
Les services d'accompagnement aux chercheurs et les infrastructures de recherche	46
Des services d'accompagnement <i>ad hoc</i> ?	46
Le développement des infrastructures de recherche	47
Les infrastructures internationales	48
Les infrastructures nationales	49
Autour de la numérisation de corpus : la feuille de route de la coopération	51
DE L'INITIATIVE À LA MISE EN PLACE DE LA COOPÉRATION	53
D'où vient la demande de numérisation ? À quel moment ?	53
Quel mode projet ?	56
La question des droits	57
L'établissement de la convention	59
DU DÉROULEMENT À LA VALORISATION DU PROJET	61
Les choix documentaires et scientifiques	61
La sélection des documents, le catalogage et l'enrichissement des notices	63
Quel set de métadonnées ? Quel modèle conceptuel ?	64
Les choix techniques de numérisation et d'archivage	67
Qui numérise et comment ?	68

Le stockage et la pérennisation : méthodes et outils	70
La diffusion et la mise en ligne de corpus	70
Les dispositifs de valorisation et de soutien	72
La numérisation et l'étude de corpus : un terrain de rencontre stimulant	75
LES JALONS D'UNE COLLABORATION INTER-MÉTIER	77
Les bonnes pratiques nationales et internationales	77
Quel équilibre entre médiation aux données et formation numérique des chercheurs ?	79
Expertise scientifique et technique : une question de légitimité ou de complémentarité ?	81
Les institutions patrimoniales sont-elles un guichet de données ?	83
Chercheurs et professionnels : une temporalité différente ?	84
BIBLIOTHÈQUES TERRITORIALES ET BIBLIOTHÈQUES DE L'ESRI : DES SERVICES, DES RESSOURCES ET DES PRATIQUES DIFFÉRENTES ?	87
Rejoindre les chercheurs par des services adaptés	87
Les dispositifs de financement	89
LES RÉPONSES DE LA BNF	93
Gallica Marque Blanche	93
Le projet CORPUS	95
CONCLUSION	97
GLOSSAIRE	111
ANNEXES	113
TABLE DES ILLUSTRATIONS	157
TABLE DES MATIÈRES	159