

Diplôme de conservateur de bibliothèque

Mémoire d'étude / mars 2023

## **Les modèles de valorisation numérique de la presse ancienne en France**

**Eugénie Martin**

Sous la direction de Frédéric Clavert  
Professeur assistant en histoire contemporaine européenne au *Luxembourg*  
*Centre for Contemporary and Digital History* - Université du Luxembourg



## **Remerciements**

*Je tiens avant tout à remercier mon directeur de mémoire, Frédéric Clavert, pour l'aide apportée dans la délimitation du sujet et ses conseils méthodologiques à divers stades de mon travail.*

*Un grand merci aux collègues qui ont eu la gentillesse de me recevoir en entretien et à toutes les personnes qui ont répondu à mon enquête. Leurs éclairages m'ont été particulièrement précieux.*

*Jacques Deville, Arnaud Laborderie, Sophie Robert, Sophie Bros et Morgane Avellaneda m'ont permis de mieux cerner, au commencement de cette étude, les enjeux passionnants autour de la presse ancienne en France ; je les en remercie vivement. Merci aussi à Irène Bastard, Noémie Couillard, Sophie Boudarel, François Godin et Pierre-Yves Martin pour leur aide dans la réalisation de l'enquête, ainsi qu'à Lionel Dujol pour ses cours très stimulants sur la médiation numérique des savoirs.*

*Enfin, mes derniers remerciements vont à l'ensemble de la promotion Christine de Pizan, tout particulièrement aux Jeanjean et à ma colocataire pour les bons moments – et repas – partagés.*

### **Résumé :**

*À l'heure où une part importante des collections de presse ancienne en France se trouve dans un état de conservation préoccupant, la sauvegarde de ce patrimoine écrit, notamment grâce à la numérisation, constitue une urgence au niveau national comme à l'échelle locale. Ces deux dernières décennies, la mise en ligne de corpus massifs a eu pour effet d'accroître considérablement l'offre de ressources à disposition du chercheur spécialiste de l'étude de la presse, comme de l'amateur d'histoire ou de généalogie amené à consulter ce type d'archives. Alors que de nombreuses initiatives sont mises en œuvre pour promouvoir ce patrimoine et que les progrès technologiques démultiplient les possibilités d'exploitation des corpus, ce travail s'attache à mieux cerner les usages des lecteurs pour réfléchir à des pistes de valorisation en adéquation avec leurs besoins.*

### **Descripteurs :**

*Presse -- France -- Histoire ; Journaux -- Conservation et restauration ; Numérisation ; Bibliothèques numériques ; Humanités numériques*

### **Abstract:**

*At a time when a significant portion of historic French newspapers are in a worrying state of conservation, it is urgent to preserve these collections at a national and local level, particularly through digitization programs. Over the past two decades, the amount of online resources available to the public – researchers, historians, but also history enthusiasts and amateur genealogists – has significantly increased. While various initiatives are being implemented to promote this heritage, and while technological progress makes it easier to study huge corpora, this work aims to better understand the readers' research habits, in order to suggest the most adequate ways to answer their needs.*

### **Keywords:**

*Press -- France -- History; Newspapers -- Conservation and restoration; Digitization; Digital libraries; Digital humanities*

### **Droits d'auteur**

Droits d'auteur réservés.

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

# Sommaire

<b>SIGLES ET ABREVIATIONS .....</b>	<b>7</b>
<b>INTRODUCTION.....</b>	<b>9</b>
<i>Cadre général et problématique d'étude.....</i>	<i>9</i>
<i>Périmètre de l'étude et de l'enquête de publics.....</i>	<i>11</i>
<b>I. PRESERVER, VALORISER : LE PAYSAGE DE LA PRESSE ANCIENNE EN FRANCE.....</b>	<b>15</b>
<b>1.1. Numériser pour sauvegarder un patrimoine en danger : état des lieux, constats et premières réponses .....</b>	<b>15</b>
<i>1.1.1. La presse ancienne, un objet complexe .....</i>	<i>15</i>
<i>1.1.2. Un patrimoine en péril .....</i>	<i>19</i>
<i>1.1.3. Les solutions mises en œuvre pour améliorer le signalement et la conservation des collections .....</i>	<i>21</i>
<b>1.2. Numériser pour mieux valoriser : les axes de valorisation de la presse ancienne .....</b>	<b>23</b>
<i>1.2.1. Programmes de numérisation à l'échelle nationale et régionale : les grands jalons .....</i>	<i>23</i>
<i>1.2.2. Une coopération renforcée pour mieux valoriser la presse ancienne régionale .....</i>	<i>25</i>
<i>1.2.3. Gallica et RetroNews, deux plateformes nationales pour diffuser les collections numérisées .....</i>	<i>27</i>
<b>1.3. Différentes stratégies de valorisation selon les territoires.....</b>	<b>31</b>
<i>1.3.1. Des modes de diffusion pluriels : de l'indépendance à l'intégration dans Gallica, différents positionnements à l'égard de l'opérateur national .....</i>	<i>31</i>
<i>1.3.2. Exemples de valorisation numérique en région .....</i>	<i>34</i>
<i>1.3.3. Un écosystème en pleine mutation : vers une mise en commun des ressources ?.....</i>	<i>38</i>
<b>II. DE LA VALORISATION A L'EXPLOITATION DES CORPUS : DE NOUVEAUX BESOINS LIES A LA NUMERISATION DE LA PRESSE ANCIENNE .....</b>	<b>41</b>
<b>2.1. Le lecteur de presse ancienne, usager aux mille visages.....</b>	<b>41</b>
<i>2.1.1. L'exploitation des collections de presse, enjeu-clé de leur valorisation.....</i>	<i>41</i>
<i>2.1.2. Diversité et porosité des usages.....</i>	<i>42</i>
<i>2.1.3. Le « sacre de l'amateur » : vers un public amateur de plus en plus éclairé ?.....</i>	<i>43</i>
<b>2.2. Ce que le numérique fait au journal : un rapport nouveau à la presse ancienne en tant que source .....</b>	<b>45</b>
<i>2.2.1. Presse numérisée et « remédiation » : de nouvelles approches de recherche.....</i>	<i>45</i>

2.2.2. « Feuilleter la presse par Giga Octets » : les enjeux de la fouille et de la visualisation de données.....	47
<b>2.3. Étude de publics : mieux connaître les usages et les besoins de deux catégories de lecteurs .....</b>	<b>50</b>
2.3.1. Cadrage de l'étude.....	50
2.3.2. Synthèse des résultats.....	51
<b>III. ADAPTER LES SERVICES AUX BESOINS POUR MIEUX VALORISER LA PRESSE ANCIENNE .....</b>	<b>69</b>
<b>3.1. Vers plus de médiation numérique et de formation .....</b>	<b>69</b>
3.1.1. La prise en compte du parcours utilisateur : une réflexion à poursuivre sur l'ergonomie des interfaces .....	69
3.1.2. L'importance de la formation .....	72
<b>3.2. Vers une amélioration des outils de recherche et d'exploitation des corpus.....</b>	<b>73</b>
3.2.1. La qualité de l'OCR, préalable indispensable à l'évolution des moteurs et fonctionnalités de recherche .....	73
3.2.2. Un rôle accru à jouer pour les bibliothèques dans l'accompagnement à la recherche : les chantiers actuels .....	75
3.2.3. Des approches innovantes pour améliorer la qualité de la médiation.....	78
<b>3.3. Vers une coopération renforcée : désenclaver pour mieux valoriser .....</b>	<b>80</b>
3.3.1. Faciliter la recherche fédérée dans les corpus de presse : un horizon réaliste aujourd'hui .....	80
3.3.2. Intensifier les collaborations avec les partenaires académiques et les bibliothèques étrangères .....	83
3.3.3. S'appuyer sur les communautés de pratique.....	86
<b>CONCLUSION .....</b>	<b>89</b>
<b>SOURCES.....</b>	<b>91</b>
<b>BIBLIOGRAPHIE.....</b>	<b>95</b>
<b>ANNEXES.....</b>	<b>105</b>
<b>TABLE DES ILLUSTRATIONS .....</b>	<b>135</b>
<b>TABLE DES MATIERES.....</b>	<b>137</b>

## *Sigles et abréviations*

- AAF : Association des archivistes français
- ACRPP : Association pour la conservation et la reproduction photographique de la presse
- AD : Archives départementales
- ANR : Agence nationale de la recherche
- API : *Application Programming Interface*
- ATER : Attaché temporaire d'enseignement et de recherche
- BDIC : Bibliothèque de documentation internationale contemporaine
- BDLI : Bibliothèque de dépôt légal imprimeur
- BIPFIG : Bibliographie de la presse française politique et d'information générale
- BM : Bibliothèque municipale
- BnF : Bibliothèque nationale de France
- BnF-P : BnF-Partenariats
- Bn-R : Bibliothèque numérique de Roubaix
- BNU : Bibliothèque nationale et universitaire
- CARISM : Centre d'analyse et de recherche interdisciplinaires sur les médias
- CÉMTI : Centre d'études sur les médias, les technologies et l'internationalisation
- CHCSC : Centre d'histoire culturelle des sociétés contemporaines
- CLARIN : *Common Language Resources and Technology Infrastructure*
- CLEMI : Centre pour l'éducation aux médias et à l'information
- CNRS : Centre national de la recherche scientifique
- CollEx : Collections d'excellence
- CRD : Chargé de recherche documentaire
- CTHS : Comité des travaux historiques et scientifiques
- DARIAH : *Digital Research Infrastructure for the Arts and Humanities*
- DGD : Dotation générale de décentralisation
- DNA : *Dernières Nouvelles d'Alsace*
- DRAC : Direction régionale des Affaires culturelles
- EAC : Éducation artistique et culturelle
- EMI : Éducation aux médias et à l'information
- ENS : École normale supérieure
- ESR : Enseignement supérieur et recherche
- GRIPIC : Groupe de recherches interdisciplinaires sur les processus d'information et de communication
- HDR : Habilitation à diriger des recherches

IHRIM : Institut d'Histoire des Représentations et des Idées dans les Modernités

IIIF : *International Image Interoperability Framework*

INSPE : Institut national supérieur du professorat et de l'éducation

IRMÉCCEN : Institut de Recherche Médias, Cultures, Communication et Numérique

LabSIC : Laboratoire des Sciences de l'Information et de la Communication

OAI-PMH : *Open Archives Initiative - Protocol for Metadata Harvesting*

OCR : *Optical character recognition*

OLR : *Optical layout recognition*

PCPP : Plans de conservation partagée des périodiques

PHisTeM : Poétique historique des textes modernes

PNV : Programme de numérisation et de valorisation des contenus culturels

R&D : Recherche et développement

REN : Reconnaissance d'entités nommées

RGPD : Règlement général sur la protection des données

RIRRA 21 : Représenter, inventer la réalité, du Romantisme au XXIe siècle

SHS : Sciences humaines et sociales

SIAF : Service interministériel des Archives de France

SPAR : Système de Préservation de l'Archivage Réparti

SPHM : Société pour l'histoire des médias

Sudoc-PS : Système universitaire de documentation pour les publications en série

UMR : Unité mixte de recherche

URL : *Uniform Resource Locator*

UX : *User eXperience*

# INTRODUCTION

---

## Cadre général et problématique d'étude

Pour leur 16<sup>e</sup> édition, intitulée « À la une ! Numériser, valoriser et conserver la presse ancienne », les Journées du Patrimoine Écrit ont mis à l'honneur la presse ancienne et souligné à la fois l'importance et l'urgence de la réflexion autour de ces collections et de leur conservation. En raison de leur nature particulièrement fragile, mais aussi bien souvent du fait de mauvaises conditions de conservation, ces collections sont aujourd'hui en grand danger. Or, les collections de presse sont aussi parmi les plus demandées et consultées par différents types de publics, qu'il s'agisse de chercheurs professionnels, biographes, généalogistes ou simples amateurs d'histoire locale, pour qui la presse ancienne représente une inépuisable mine d'informations. Les chiffres d'un rapport de 2018 sur la collection de presse de la Bibliothèque nationale de France<sup>1</sup> indiquent que 11% des demandes de communication physique sur le site François-Mitterrand portent sur des fascicules de presse. En ligne, la presse ancienne représente 30% des consultations sur la bibliothèque numérique Gallica, soit 5 millions de visites par an<sup>2</sup>. Véritable « lieu de mémoire » au sens de Pierre Nora<sup>3</sup>, la presse nous informe sur l'évolution des mentalités au fil de l'histoire, le paysage politique propre à une époque, les activités syndicales, la vie culturelle et artistique du pays, etc. À ce titre, elle constitue un patrimoine d'une exceptionnelle richesse.

Pourtant, cette prise de conscience est relativement récente : la presse, média de masse populaire, a longtemps fait figure de « sous-patrimoine écrit », mal-aimé des chercheurs et des institutions, et dont les autorités ne faisaient que peu de cas dans les discours comme dans les budgets alloués à sa conservation. Mais désormais, dans le paysage du patrimoine écrit comme de la recherche, le changement de perception sur les collections de presse est patent. Il suffit, pour s'en convaincre, de voir les nombreux colloques, journées d'étude et projets de recherche en études médiatiques ces dernières années. Voici quelques exemples de thématiques explorées : « À la recherche des femmes dans la presse ancienne numérisée » (mars 2021)<sup>4</sup>, « Les feuilletons dans la presse » (avril 2021)<sup>5</sup>, « Fictions criminelles européennes et discours médiatique du crime » (octobre 2021)<sup>6</sup>, « Presses anciennes et modernes à l'ère du numérique » (juin 2022)<sup>7</sup>,

---

<sup>1</sup> Aline Girard et Catherine Aurélin, *La presse physique et numérique à la BnF : état des lieux et perspectives*, rapport de la Bibliothèque nationale de France, mars 2018. Nous l'abrégerons « Rapport Presse 2018 », selon la dénomination qui lui a été donnée lors des Journées du Patrimoine Écrit 2022.

<sup>2</sup> *Journées du Patrimoine Écrit 2022 : Projet centre de conservation de la BnF à Amiens*, intervention de Julie Ladant, directrice du département Droit, économie, politique (Bibliothèque nationale de France), 23-24 juin 2022.

<sup>3</sup> Pierre Nora, *Les lieux de mémoire*, Paris, France, Gallimard, 1997, vol. 3/.

<sup>4</sup> *À la recherche des femmes dans la presse ancienne numérisée (1850-1950)*, <https://www.bnf.fr/fr/agenda/la-recherche-des-femmes-dans-la-presse-ancienne-numerisee-1850-1950>, 15 mars 2021 (consulté le 13 février 2023).

<sup>5</sup> *Les feuilletons dans la presse*, <https://www.bnf.fr/fr/agenda/les-feuilletons-dans-la-presse>, 10 avril 2021 (consulté le 13 février 2023).

<sup>6</sup> *Crimes fictionnels/crimes factuels. Fictions criminelles européennes et discours médiatiques du crime*, <https://www.medias19.org/actualites/crimes-fictionnelscrimes-factuels-fictions-criminelles-europeennes-et-discours-mediatiques-du-crime-7-8-octobre-2021>, 7-8 octobre 2021 (consulté le 13 février 2023).

<sup>7</sup> *Congrès Médias 19 - Numapresse : Presses anciennes et modernes à l'ère du numérique*, <https://www.bnf.fr/fr/agenda/congres-medias-19-numapresse-presses-anciennes-et-modernes-lere-du-numerique>, 2-3 juin 2022 (consulté le 13 février 2023).

« Couper/coller. Le journalisme aux ciseaux du XIXe siècle à Internet » (février 2023)<sup>8</sup>, ou encore « La civilisation du journal dans les espaces méditerranéens de la fin du XVIIIe siècle à la Première Guerre mondiale » (à venir)<sup>9</sup>. La presse ancienne concentre autour d'elle des enjeux de recherche très forts, son étude se révélant essentielle dans de nombreux champs des sciences humaines et sociales (histoire politique, culturelle, économique, histoire des mentalités, sociologie, littérature, etc.), mais aussi pour toute personne souhaitant mener des recherches sur sa propre famille, une personnalité locale, sa ville ou un événement particulier.

Aussi la nécessité de mieux conserver et valoriser ce patrimoine est-elle davantage prise en considération ces dernières années. D'une part, le sauvetage des collections de presse ancienne est aujourd'hui devenu une priorité du ministère de la Culture. En témoigne le projet d'ouverture, à Amiens, d'un Conservatoire national de la presse destiné à accueillir à horizon 2030 l'ensemble des collections de la BnF publiées sur le territoire français ou en français<sup>10</sup>. D'autre part, à l'heure où se cristallisent de très nombreux enjeux autour de la qualité de l'information et de ses sources, faciliter l'accès à la presse d'hier et d'aujourd'hui devient plus que jamais une question politique. D'où l'importance des programmes de numérisation pour améliorer quantitativement et qualitativement la conservation et les possibilités d'exploitation de ces collections. Depuis une quinzaine d'années, l'intensification de la numérisation, par le biais de divers programmes concertés entre la Bibliothèque nationale de France et les établissements détenteurs de collections de presse en région (bibliothèques territoriales, services d'archives, centres de documentation, musées, etc.), va de pair avec une intensification de la recherche sur ces questions : à mesure que la numérisation progresse, les possibilités de recherche s'accroissent et s'affinent dans un corpus sans cesse élargi, accessible en seulement quelques clics. Le journal devient un objet patrimonial d'un genre nouveau.

De ces remarques préliminaires, il ressort un constat essentiel : la numérisation a changé la nature de la presse en tant que source primaire, transformant par là même les pratiques mais aussi les questions de recherche. Le corpus de l'historien ne se résume plus, désormais, à quelques titres épluchés un par un en bibliothèque, mais peut potentiellement englober les données de plusieurs centaines de titres. Il devient alors possible d'entreprendre des recherches sur l'« objet journal », ou d'étudier la diffusion et la circulation de l'information à une très large échelle. Pour les publics non-experts aussi, la numérisation modifie le geste de consultation : à mesure que les établissements mettent en ligne leurs archives de presse, le déplacement en salle de lecture est de moins en moins nécessaire – et ce d'autant plus qu'une fois le titre numérisé, l'original n'est souvent plus disponible à la consultation. Ces dernières années, la multiplication des portails en région a considérablement accru l'offre documentaire en matière de presse ancienne locale. Or, les transformations du paysage de la

---

<sup>8</sup> *Couper/coller. Le journalisme aux ciseaux du XIXe siècle à Internet* (ANR Numapresse, ENS Lyon), <https://www.fabula.org/actualites/109625/colloque--coupercoller-le-journalisme-aux-ciseaux-du-xixe-siecle.html>, 9-10 février 2022 (consulté le 13 février 2023).

<sup>9</sup> *La « civilisation du journal » dans les espaces méditerranéens de la fin du XVIIIe siècle à la Première Guerre mondiale*, <https://calenda.org/1015668>, 23-24 mars 2023 (consulté le 13 février 2023).

<sup>10</sup> *BnF - Rapport d'activité 2021*, <https://www.bnf.fr/fr/bnf-rapport-dactivite-2021>, p. 131-136 (consulté le 13 février 2023).

presse ancienne en France ne sont pas sans conséquences sur les pratiques de recherche dans ces corpus.

Le contexte de la création du futur Conservatoire national de la presse est l'occasion pour nous de faire un point d'étape. Sans omettre les problématiques spécifiques liées à la conservation des collections physiques, nous nous demanderons surtout comment le tournant du numérique infléchit les usages et les besoins des lecteurs de presse ancienne, experts ou non. Une étude de publics menée auprès de deux panels de chercheurs – professionnels et amateurs – constituera un préalable à l'identification de pistes d'amélioration pour la mise en valeur de ce patrimoine. Comment valoriser au mieux la presse ancienne numérisée, tout en répondant efficacement aux besoins spécifiques des utilisateurs de ces collections ? À l'heure de la démultiplication des ressources de presse patrimoniale en ligne, comment organiser une médiation efficace et adaptée autour de ces contenus ? Enfin, comment le numérique peut-il apporter au lecteur de presse ancienne, usager aux mille visages, des outils pertinents pour explorer et exploiter ces corpus dématérialisés ?

Pour répondre à ces questions, il nous faudra tout d'abord revenir aux fondements de la politique de numérisation relative à la presse ancienne en France : nous montrerons comment celle-ci répond à un double objectif de sauvegarde et de valorisation de collections extrêmement menacées, qui posent des difficultés multiples en raison de leur ampleur, de leur dispersion et de la relative imprécision de leur périmètre. Or, toute pensée de la valorisation s'accompagne nécessairement d'une réflexion sur l'usager et ses besoins : aussi nous efforcerons-nous, dans un deuxième temps, de mieux cerner les profils d'utilisateurs et leurs pratiques de recherche dans ces corpus de presse. Nous verrons ainsi comment le numérique induit de nouveaux usages auxquels il convient de s'adapter : nous prendrons appui sur les besoins exprimés par les utilisateurs pour proposer des pistes concrètes d'amélioration en matière de valorisation et de services à la recherche dans les collections en ligne.

## **Périmètre de l'étude et de l'enquête de publics**

### *De la valorisation numérique à l'exploitation*

S'il existe de nombreuses manières de valoriser des collections de presse ancienne (expositions, ateliers d'écriture, etc.), notre étude se concentrera sur le versant numérique de cette valorisation. À l'ère numérique, la notion de valorisation suppose avant tout une logique d'accès : mettre à disposition des collections sur le Web. Cela implique tout d'abord de numériser et de rendre accessible, mais aussi de favoriser la diffusion, la circulation et l'appropriation des contenus, notamment à l'aide de formats ouverts permettant l'interopérabilité. Une nouvelle compétence métier est de plus sollicitée : l'éditorialisation des contenus, c'est-à-dire la sélection, le conseil, la recommandation. Conséquence de l'abondance informationnelle qui caractérise la « culture numérique »<sup>11</sup>, l'économie de la recommandation est devenue, de fait, l'un des piliers du Web aujourd'hui.

Qui dit valorisation numérique dit aussi médiation numérique : non seulement l'accompagnement dans l'appropriation des interfaces et des outils, mais

---

<sup>11</sup> Milad Doueïhi, *Digital cultures*, Cambridge, Mass., Harvard University Press, 2011.

aussi médiation culturelle à travers des dispositifs numériques mettant souvent en avant des formes narratives : billets de blog, vidéos, podcasts, etc. Penser une médiation suppose nécessairement l'intervention d'un tiers construisant une relation : en médiation culturelle, la fonction du médiateur consiste à rapprocher les publics et le patrimoine dans une perspective à la fois éducative, récréative et citoyenne<sup>12</sup>. Comme l'écrit Bertrand Calenge au sujet des bibliothèques, le tropisme de la collection s'efface derrière un partage actif, organisé, des savoirs et des connaissances<sup>13</sup>. En effet, toute démarche de médiation a pour horizon un public cible : le dispositif de médiation numérique est à la croisée d'un besoin perçu, d'un usage identifié et d'un outil mis à disposition pour répondre au besoin<sup>14</sup>. Nous garderons en tête ces trois composantes dans notre étude et nos préconisations.

Mais en matière de presse ancienne, on ne peut s'en tenir à une définition de la valorisation qui se limiterait à la mise à disposition, à l'éditorialisation et à la médiation autour de ces collections. La valorisation numérique passe aussi, de plus en plus, par le fait de favoriser au maximum l'exploitation des contenus. En effet, un fascicule de presse n'est pas une archive qui « parle » immédiatement, contrairement à une photographie ancienne par exemple. Bien que la composante visuelle des journaux ne soit pas à négliger (caricatures, Unes illustrées, photographies de presse, etc.), c'est généralement le contenu textuel du journal qui intéresse le lecteur désireux d'y retrouver une information. Permettre la recherche plein texte dans ces archives est une première étape indispensable, mais les outils nous permettent aujourd'hui d'aller plus loin dans l'extraction et l'analyse des données du journal, à une échelle de plus en plus massive. Il convient alors de sortir d'une vision strictement académique de l'exploitation, puisque ces nouveaux outils, en se démocratisant, concernent tout aussi bien aujourd'hui les chercheurs amateurs.

### *Délimitation du périmètre de l'étude*

Notre étude se limite à la presse française, bien que cela n'exclue pas l'évocation ponctuelle de projets étrangers ou internationaux. Dans le panorama général que nous proposons, nous nous concentrons avant tout sur les initiatives menées à l'échelon national et en région, en insistant sur les stratégies de coopération entre la Bibliothèque nationale de France et les bibliothèques territoriales autour de la numérisation et de la diffusion de la presse ancienne. La BnF étant pilote sur de nombreux projets, une part importante de notre étude sera consacrée à présenter son action à plusieurs niveaux, et certaines de nos préconisations lui seront applicables en priorité. Face à l'ampleur des collections de presse ancienne sur le territoire et au nombre très important d'établissements

---

<sup>12</sup> Serge Chaumier et François Mairesse, « Chapitre 2 - Les fonctions du médiateur » dans *La médiation culturelle*, Paris, Armand Colin, 2013, p. 61-114.

<sup>13</sup> « Arrêtons de parler de “la” bibliothèque : la réalité de l'institution réside dans les capacités actives de médiation des bibliothécaires ! Cette dynamique-là ne se construit pas *ex nihilo*, par le seul miracle de ressources accumulées et proposées à une population. Elle exige une mobilisation des acteurs. C'est sans doute, à mon sens, la principale émergence de ce dernier demi-siècle : bien plus que l'informatique ou le numérique, c'est l'urgence de la médiation humaine des connaissances qui est réclamée pour construire du sens. » Bertrand Calenge, *Les bibliothèques et la médiation des connaissances*, Paris, France, Éditions du Cercle de la librairie, 2015.

<sup>14</sup> « Un dispositif n'existe qu'au travers de trois dimensions étroitement imbriquées : un faisceau de besoins d'informations perçus, des usages constatés et des moyens techniques ». Lionel Dujol et Silvère Mercier, *Médiation numérique des savoirs : des enjeux aux dispositifs*, Montréal, Canada, Les éditions Asted, 2017.

qui en conservent, nous ne sommes pas en mesure de détailler les actions de tous les partenaires tels que les services d'archives, musées ou structures régionales pour le livre, bien que nous ayons conscience qu'ils constituent des relais essentiels en matière de signalement, diffusion et valorisation de ces fonds. Nous nous concentrerons donc sur quelques initiatives significatives en région.

De même, nous n'intégrons pas les bibliothèques universitaires<sup>15</sup> dans notre périmètre, bien que nombre d'établissements disposent également de collections de presse, parfois sur des thématiques spécialisées : presse de spectacles<sup>16</sup>, presse sportive, etc. La place de la presse dans les collections universitaires – notamment celles disposant du label CollEx visant à identifier et valoriser les collections d'excellence – mériterait une étude à part entière.

### *Délimitation des publics*

Afin de mener notre enquête de publics, il nous a fallu au préalable cerner les différents panels de lecteurs ayant recours à de la presse ancienne pour un usage personnel ou de recherche. Nous avons commencé par identifier les catégories de lecteurs suivantes :

**1. Les chercheurs de l'Enseignement supérieur (ESR)**, titulaires et non-titulaires. Il ne s'agit pas là, bien sûr, d'un panel uniforme : le fait d'être un doctorant en début ou en fin de thèse, par exemple, n'implique pas le même degré de familiarité avec les corpus et ressources numériques. De même, la maîtrise d'outils d'extraction et d'analyse des corpus n'est évidemment pas la même selon que l'on est ou non familier des humanités numériques. Il conviendra d'apprécier plus finement les différences entre diverses catégories de chercheurs, dont les besoins peuvent varier fortement en fonction de leurs domaines et pratiques de recherche.

**2. Les généalogistes et amateurs<sup>17</sup> d'histoire.** Leurs recherches peuvent être mues par la simple curiosité ou servir un objectif précis, qu'il s'agisse de membres de sociétés d'histoire locale ou de particuliers retraçant la généalogie de leur famille. Pour ce type de lectorat, le degré d'expertise – et plus largement de familiarité numérique – est variable selon les individus ; d'où l'importance de cerner au mieux leurs habitudes de recherche, celles-ci induisant des besoins potentiellement très différents d'une personne à l'autre.

**3. Les enseignants du secondaire** (en particulier les professeurs d'histoire), pouvant être amenés à consulter de la presse ancienne numérisée à des fins pédagogiques. S'il s'agit bien là d'une modalité de recherche à part entière, nous choisissons cependant de ne pas inclure dans notre analyse cette utilisation particulière des ressources de presse qui, selon nous, pourrait donner lieu à un sujet d'étude à part entière sur les usages pédagogiques de la presse ancienne. Nous avons donc choisi de ne pas prévoir de questionnaire d'enquête à destination

---

<sup>15</sup> À l'exception de la Bibliothèque nationale et universitaire de Strasbourg et de La contemporaine (anciennement la BDIC), en vertu de leur statut particulier : bibliothèque nationale et universitaire dans le premier cas ; bibliothèque inter-universitaire, musée et centre d'archives dans le second.

<sup>16</sup> On peut prendre pour exemple la Théâtrothèque Gaston Baty de l'Université Sorbonne Nouvelle, dont les collections théâtrales sont labellisées CollEx.

<sup>17</sup> Le terme d'« amateurs » ne renvoie ici aucunement à une absence supposée d'expertise, mais désigne une catégorie d'usagers partageant un goût prononcé pour l'histoire et pratiquant à titre individuel des recherches sur divers sujets en lien avec leurs intérêts.

de ce public, qu'il nous aurait fallu approcher de manière tout à fait différente des deux autres, via les établissements scolaires et les professeurs documentalistes en particulier. En revanche, nous n'excluons pas la possibilité que notre second questionnaire, à destination des amateurs d'histoire, puisse toucher également ce public, bien qu'il ne soit pas *a priori* le cœur de cible de ce sondage.

**4. Le « grand public »** : il s'agit là aussi d'une catégorie que nous choisissons de ne pas inclure dans notre étude, du fait de la trop grande variété de profils et modes d'approche des collections de presse ancienne numérisée, qu'il s'agisse d'une recherche scolaire, d'une exploration par simple curiosité, d'un besoin d'information ponctuel, etc. La notion de « grand public » ne nous semble pas véritablement opérante pour appréhender les usages poreux et multiples de la presse ancienne numérisée – sans compter la difficulté pour toucher ces usagers ponctuels dans le cadre d'une enquête de publics.

Nous avons donc fait le choix de concentrer notre enquête sur les deux premiers panels qui, bien que n'étant pas totalement homogènes, se distinguent par des besoins de recherche clairement identifiés, délimités par un ou plusieurs axes majeurs de recherche. Nous n'écarterons pas de notre réflexion des modes de valorisation à vocation plus « grand public », mais ceux-ci ne constitueront pas le cœur de notre questionnement et de nos préconisations.

Un pont entre les deux axes de notre étude – valorisation et exploitation – s'opère naturellement dans l'enquête menée auprès des deux panels interrogés, puisqu'il s'agit non seulement de les questionner sur leur connaissance globale des ressources de presse ancienne en ligne, mais aussi de détailler leurs usages. L'objectif de l'étude est donc de réfléchir à la meilleure adéquation possible entre les besoins réels des usagers et l'offre de services proposée en ligne pour la recherche dans ces corpus.

Le détail de notre méthodologie d'enquête (connaissance des publics ciblés, élaboration et diffusion du questionnaire) est donné en annexe.

# I. PRESERVER, VALORISER : LE PAYSAGE DE LA PRESSE ANCIENNE EN FRANCE

---

## 1.1. NUMERISER POUR SAUVEGARDER UN PATRIMOINE EN DANGER : ETAT DES LIEUX, CONSTATS ET PREMIERES REPONSES

### 1.1.1. La presse ancienne, un objet complexe

#### *De quelle presse parle-t-on ?*

La difficulté du corpus étudié tient tout d'abord à la définition de celui-ci : derrière son apparente simplicité, l'expression « presse ancienne » est en réalité loin de faire consensus. On fait généralement remonter la naissance de la presse écrite à la *Gazette* de Théophraste Renaudot en 1631<sup>18</sup>, bien que le journal papier ait connu son véritable essor au cours du XIXe siècle, en particulier après la loi du 29 juillet 1881 sur la liberté de la presse marquant la fin de la censure. Pour le *terminus ad quem*, c'est le plus souvent la coupure historique de 1944 qui est retenue : les ordonnances prises à la Libération, visant à garantir le pluralisme de la presse et la liberté d'expression après la Seconde Guerre mondiale, ont profondément réorganisé le système médiatique en interdisant les journaux qui avaient collaboré avec l'Allemagne – en l'occurrence, la quasi-totalité des titres hormis les journaux clandestins de la Résistance et ceux qui avaient interrompu leur parution durant le conflit<sup>19</sup>.

Si la date de 1944 est souvent retenue dans les plans de numérisation en France<sup>20</sup>, ce n'est pas toujours le choix fait par les diverses institutions qui numérisent de la presse ancienne – bibliothèques territoriales, archives départementales et municipales, structures régionales pour le livre, etc. Une autre borne possible consiste à prendre en compte un délai de 70 ans suivant la publication : le journal tombe alors dans le domaine public et n'est plus protégé par le droit d'auteur, ce qui permet alors sa reproduction et sa diffusion. Ainsi, dans le cadre du récent marché de numérisation « Presse 2021 » coordonné par la BnF, les partenaires en région sont-ils incités à numériser leurs titres jusqu'en 1958, ce qui correspond à l'application de ce délai de 70 ans au terme du programme. Toutefois, comme cela fut souligné lors des Journées du Patrimoine Écrit 2022, cette notion de domaine public est particulièrement complexe pour ce qui concerne la presse, car elle s'applique au droit des éditeurs uniquement. Pour ce qui est du droit des contributeurs individuels, en théorie, le délai s'apprécie à compter du décès de chacun d'entre eux. Or, si l'on peut aisément faire des extrapolations pour un journal très ancien – on estimera, pour un journal datant de

---

<sup>18</sup> « La fondation, en 1631, par Renaudot (1586-1653) de *La Gazette* est considérée comme l'acte de naissance de la presse périodique française, même si un autre titre peut revendiquer une antériorité de quelques mois. » Encyclopædia Universalis, « La Gazette (T. Renaudot) », <https://www.universalis.fr/encyclopedie/la-gazette/> (consulté le 31 janvier 2023).

<sup>19</sup> Christian Delporte, Claire Blandin et François Robinet, *Histoire de la presse en France : XXe-XXIe siècles*, Malakoff, France, Armand Colin, 2016, p. 144-149.

<sup>20</sup> C'est notamment la borne chronologique que retient le site Presse locale ancienne, qui recense les journaux d'intérêt local parus en France, des origines à 1944.

1900, qu'il est vraisemblable que tous les contributeurs soient décédés depuis plus de 70 ans –, cela s'avère plus complexe à mesure que l'on progresse dans le temps. La mise en ligne de titres plus récents ne va donc pas sans poser un certain nombre de difficultés d'ordre juridique<sup>21</sup>. Nous faisons toutefois le choix de ne pas revenir ultérieurement sur ces questions, qui mériteraient de faire l'objet d'une étude distincte de celle-ci<sup>22</sup>.

Les difficultés posées par la circonscription de ce corpus ne sont pas seulement chronologiques : ce qu'on inclut sous le terme de « presse ancienne » est lui-même sujet à discussion, et les choix quant à la délimitation de ce périmètre peuvent varier d'un établissement à un autre. En témoignent notamment les différentes publications pouvant être mises en valeur dans les portails patrimoniaux : il n'est pas rare d'y voir les journaux regroupés avec d'autres types de périodiques – bulletins paroissiaux, revues savantes, etc. Cela se justifie, le plus souvent, par une volonté de mettre en avant des ressources périodiques jugées particulièrement intéressantes au regard de l'histoire locale, même si elles sortent du champ de la presse *stricto sensu*. C'est le cas par exemple de la bibliothèque numérique Pireneas : aux côtés de quotidiens et hebdomadaires documentant l'histoire de Pau, du Béarn et des Pyrénées, on trouve diverses publications telles que le Bulletin de la Société académique des Hautes-Pyrénées ou le Bulletin de l'Association Francis Jammes<sup>23</sup>. De la même façon, le groupe de travail coordonné par la structure régionale Mobilis (Pays de la Loire) autour de la presse ancienne a fait le choix « d'axer ses travaux de coopération autour des titres périodiques d'actualité locale et immédiate imprimés en région Pays de la Loire avant 1945 » : cela inclut, en plus de la presse d'information générale, les bulletins paroissiaux, municipaux et syndicaux numérisés par les services d'archives départementales et très consultés par les lecteurs<sup>24</sup>. La bibliothèque numérique nationale Gallica rassemble elle aussi presse et revues, entretenant une certaine porosité entre les deux.

Il est à noter que la périodicité ne fait pas non plus consensus : si de nos jours, d'un point de vue législatif, on regroupe sous le terme de « presse d'information générale » des publications ayant une périodicité au maximum hebdomadaire, cela ne s'applique par nécessairement pour la presse ancienne. De fait, la Bibliographie de la presse française politique et d'information générale des origines à 1944 (BIPFPIG) répertorie une centaine de titres ayant une périodicité annuelle. Le corpus délimité dans le cadre du rapport Presse 2018, quant à lui, inclut les parutions allant du quotidien au trimestriel<sup>25</sup>.

---

<sup>21</sup> *Journées du Patrimoine Écrit 2022 : Les enjeux juridiques de la presse ancienne*, intervention de David Pouchard, adjoint à la cheffe du bureau de la propriété intellectuelle (Service des affaires juridiques et internationales, ministère de la Culture), 23-24 juin 2022. Voir aussi David Pouchard, « La valorisation des fonds photographiques, ou comment concilier le droit d'auteur et l'accès au patrimoine culturel », *In Situ. Revue des patrimoines*, 15 octobre 2018, n° 36.

<sup>22</sup> Un colloque a d'ailleurs été consacré à ces questions juridiques à la BnF en novembre 2022. *Droit(s) et presse*, <https://www.bnf.fr/fr/agenda/droits-et-presse>, 18 novembre 2022 (consulté le 13 février 2023).

<sup>23</sup> *Pireneas : la presse locale en ligne*, <https://www.pireneas.fr/pireneas/fr/content/pireneas-la-presse-locale-en-ligne> (consulté le 1 février 2023).

<sup>24</sup> Entretien avec Émilie Chevalme, chargée de mission Patrimoine Écrit à Mobilis (Pôle régional de coopération des acteurs du livre et de la lecture en Pays de la Loire), le 12/10/2022.

<sup>25</sup> A. Girard et C. Aurérin, *La presse physique et numérique à la BnF*, *op. cit.*, p. 24.

Cependant, pour la clarté de notre étude, nous nous en tiendrons aux critères fixés par les historiens spécialistes de la presse<sup>26</sup>. Le premier critère est celui de l'intérêt général : par presse ancienne, on désigne la presse d'information générale, couvrant les actualités – politiques notamment – à l'échelon national et local. Si cette définition n'inclut pas les revues savantes, réservées à un public trop restreint, elle n'exclut pas pour autant la presse thématique voire spécialisée, tant que sa portée relève du caractère d'intérêt général. Le deuxième critère est celui d'une périodicité dense, le plus souvent quotidienne ou hebdomadaire, voire bimensuelle ou mensuelle – en particulier avec l'émergence du reportage au cours du XXe siècle. Enfin, ce qui distingue la presse d'autres types de périodiques, c'est sa dimension commerciale, puisqu'un journal est une entreprise dirigée par un éditeur.

Quant au périmètre géographique, il prend en compte ce qui a été publié sur l'ensemble du territoire français, dans toutes les langues. À ce titre, la presse allophone publiée en France relève bien du patrimoine national. De fait, un journal comme *Haratch*, principal journal en arménien occidental de France, reflète bien l'histoire de la communauté arménienne installée en France après le génocide<sup>27</sup>. Établir une cartographie documentaire s'avère donc particulièrement complexe pour la presse ancienne : selon les critères mis en avant, les frontières sont poreuses et difficiles à établir de manière définitive.

### *Des collections immenses et disséminées, difficiles à signaler*

La difficulté à circonscrire le périmètre de la presse ancienne est accentuée par l'immense masse documentaire qu'elle représente sur l'ensemble du territoire. Les chiffres de la collection de presse à la BnF suffisent à mesurer l'ampleur de sa volumétrie : la collection, qui s'accroît depuis 1631 par dépôt légal et par acquisitions, représente actuellement près de 44 km linéaires dans les magasins des différents sites et compte environ, en incluant les doublons, 273 000 titres publiés en France<sup>28</sup>. L'accroissement permanent des collections rend particulièrement complexe la gestion des flux : pour la seule année 2017, à la BnF, 15 948 nouveaux titres sont entrés dans la collection<sup>29</sup> – soit un accroissement annuel d'environ 500 mètres linéaires selon les projections actuelles. En région, les bibliothèques disposant du dépôt légal imprimeur reçoivent chaque jour les titres de presse quotidienne régionale, y compris les éditions locales. Évoquant les deux grands titres de presse quotidienne alsacienne, les *Dernières Nouvelles d'Alsace* et *L'Alsace*, Jérôme Schweitzer indique :

Les versions PDF de chacune des éditions locales de ces derniers titres sont sauvegardées chaque jour depuis 2014. Ce mode de dépôt complète les

---

<sup>26</sup> Ces critères, définis par des historiens tels que Pierre Albert, Gilles Feyel ou Claude Bellanger, ont été rappelés par Philippe Mezzasalma, chef du service Presse à la BnF, lors des Journées du Patrimoine Écrit à Rouen. *Journées du Patrimoine Écrit 2022 : Numériser les collections de presse*, 23-24 juin 2022.

<sup>27</sup> Entretien avec Philippe Mezzasalma, chef du service Presse à la BnF, et Alexia Bauville, adjointe au chef du service Presse, le 07/11/2022. Voir *Ces journaux des diasporas qui ont fait la presse parisienne*, colloque international Transfopress à la BULAC, <https://www.bulac.fr/ces-journaux-des-diasporas-qui-ont-fait-la-presse-parisienne>, 9-11 mars 2022 (consulté le 23 février 2023).

<sup>28</sup> *Congrès Médias 19 - Numapresse : Presses anciennes et modernes à l'ère du numérique*, BnF, 2-3 juin 2022, intervention d'Arnaud Laborderie, chef de projet au service de la Coopération numérique et de Gallica. Ces chiffres s'appuient sur le rapport Presse 2018.

<sup>29</sup> *La collection de presse conservée à la BnF*, <https://www.bnf.fr/fr/la-collection-de-presse-conservée-la-bnf>, (consulté le 16 septembre 2022).

collections de presse locale de la BNU, constituées depuis la fin du XIXe siècle. Il remplace désormais les microfilms et complète la conservation de l'exemplaire papier qui se poursuit d'autre part<sup>30</sup>.

La presse pose donc des difficultés récurrentes « de gestion dynamique des collections, de gestion de l'accroissement, avec des problématiques de saturation »<sup>31</sup>. Si ces défis liés à la masse documentaire se sont posés dès le début du XIXe siècle, le phénomène n'a fait que s'amplifier avec l'explosion du nombre de titres de presse au cours de ce même siècle – en particulier à partir de la Monarchie de Juillet – et ce jusqu'à aujourd'hui. Retracer l'histoire de ces collections, c'est avant tout faire le constat de la complexité à en maîtriser les flux.

En outre, les collections physiques peuvent être conservées par de multiples établissements ainsi que par les éditeurs de presse eux-mêmes. Elles se trouvent donc extrêmement dispersées sur le territoire, ce qui complexifie leur signalement. Il n'est pas rare que les bibliothèques et centres d'archives des collectivités territoriales conservent des exemplaires uniques (*unica*), d'où l'importance des plans de conservation partagée des périodiques (PCPP) pour améliorer la visibilité et le signalement de ces titres et assurer la conservation des exemplaires les plus précieux<sup>32</sup>. À la BnF, la collection de presse a également pour particularité d'être très disséminée : concernant toutes les disciplines, elle se répartit entre tous les départements<sup>33</sup>. Si environ 70% des publications sont conservées au département Droit, économie, politique, des fonds significatifs se trouvent également à la bibliothèque de l' Arsenal, au département Littérature et art, ou encore au département des Arts du spectacle. Nombre d'exemplaires conservés à la BnF sont vraisemblablement des *unica*<sup>34</sup>.

Le problème du signalement de la presse n'est pas nouveau. Dès 1866, l'historien de la presse Eugène Hatin regrettait, dans sa préface à la *Bibliographie historique et critique de la presse périodique française*, la dispersion de la collection et l'impossibilité d'en établir une cotation satisfaisante ainsi qu'une bibliographie exhaustive ; or, tout titre non signalé est invisibilisé<sup>35</sup>. Signaler

---

<sup>30</sup> Jérôme Schweitzer, « Autre chose que des livres... : les entrées du dépôt légal dans les collections de la BNU », *La Revue de la BNU*, 1 novembre 2018, n° 18, p. 26-33.

<sup>31</sup> *Journées du Patrimoine Écrit 2022 : Projet centre de conservation de la BnF à Amiens*, intervention de Julie Ladant, directrice du département Droit, économie, politique (Bibliothèque nationale de France), 23-24 juin 2022.

<sup>32</sup> Une étude de 2019 sur les PCPP indique : « Les PCPP offrent l'assurance qu'un désherbage ciblé ne se fait qu'avec la certitude qu'au moins un exemplaire complet du titre soit conservé et disponible, dans au moins un établissement participant au Plan, pour répondre au besoin essentiel des chercheurs, enseignants, étudiants, et citoyens. Bientôt, la notion de sauvegarde s'est teintée d'une coloration de valorisation des titres rares, précieux ou témoignant d'une caractéristique locale unique, ce qui a permis, dans le cadre des PCPP régionaux, de dégager des sous-thématiques régionales fortes et dans le cadre des PCPP thématiques, de définir les contours intellectuels d'une discipline et de mettre en lumière des corpus scientifiques d'excellence ou d'éventuelles ressources "de niche", valorisables dans le cadre du programme CollEx-Persée ». Julie Mistral et Yves Desrichard, *Les plans de conservation partagée des périodiques : un état de l'art quantitatif et qualitatif*, <https://abes.fr/wp-content/uploads/2020/01/Etude-PCPP-2019.pdf>, décembre 2019 (consulté le 14 février 2023).

<sup>33</sup> Lise Devreux et Philippe Mezzasalma, *Des sources pour l'histoire de la presse : guide*, Paris, France, Bibliothèque nationale de France, 2011.

<sup>34</sup> A. Girard et C. Aurérin, *La presse physique et numérique à la BnF*, op. cit., p. 24.

<sup>35</sup> « Et en attendant, il n'est pas possible, et il ne le sera pas de longtemps, de savoir ce que la Bibliothèque possède de journaux de ces différentes spécialités, pas possible, par conséquent, d'en faire une bibliographie quelque peu complète, à moins que l'on ne soit de la maison et cela me semble fâcheux. Ce qui n'est pas catalogué est, en effet, lettre morte pour les profanes, même pour les plus favorisés. » Eugène Hatin, *Bibliographie historique et critique de la presse périodique française ou Catalogue systématique et raisonné de tous les écrits périodiques de quelque valeur publiés ou ayant circulé en France depuis l'origine du journal jusqu'à nos jours* (1866), Torino, Italie, Bottega d'Erasmus, 1960. Préface, p. IX.

correctement la presse ancienne est donc un préalable indispensable pour la porter à la connaissance du public et ainsi permettre sa valorisation.

### *Des collections hétérogènes, posant des difficultés d'indexation*

L'indexation matière des titres de presse représente également un véritable défi pour les établissements qui en conservent et cherchent à les valoriser : le rattachement de mots-clés (ou vedettes-matière) est en effet beaucoup plus complexe à réaliser que pour d'autres types de collections. Pour la presse ancienne, du fait de la variété des sujets couverts, cette indexation fine des contenus est quasi impossible à réaliser manuellement (à l'inverse de collections iconographiques par exemple, pour lesquelles il est possible, à l'aide de mots-clés ou tags, de générer automatiquement un parcours thématique à travers le portail patrimonial). Actuellement, la presse requiert donc de la part du professionnel une éditorialisation très particulière, impliquant un important travail d'écriture et de production. Si la presse ancienne spécialisée est relativement facile à valoriser à l'aide de modules éditoriaux thématiques (la presse sportive, les journaux de mode, les magazines pour enfants, etc.), on ne peut en dire autant de la presse d'information générale, beaucoup plus complexe à thématiser. L'hybridité de ses contenus, à la fois textuels et illustrés, n'est pas sans poser des difficultés – nous y reviendrons.

## **1.1.2. Un patrimoine en péril**

### *Un état de conservation des collections alarmant*

Mais surtout, ce qui caractérise la presse ancienne en France, c'est l'extrême fragilité d'une grande partie de ces collections. Comme cela a été rappelé à l'occasion des Journées du Patrimoine Écrit 2022, le rapport Presse 2018 de la BnF a fait le constat d'un état de conservation particulièrement alarmant des collections physiques. Le rapport note entre autres l'inadéquation du conditionnement d'une grande partie des fascicules, parfois encore conservés dans des enveloppes en papier Kraft, pliés et ficelés. En l'état, la conservation de ces collections n'est pas assurée de manière satisfaisante et le constat d'un manque d'espace s'impose : le rapport estime qu'un conditionnement optimal reviendrait à multiplier par quatre les kilomètres linéaires actuels<sup>36</sup>.

Au fil du temps, la qualité du papier des journaux a pu fortement varier, ce qui engendre des problématiques de conservation spécifiques à chaque période de l'histoire. Paradoxalement, les journaux les plus anciens (antérieurs à la Troisième République), en raison de leur papier non acide et de bonne qualité, se préservent bien mieux que ceux de la fin du XIXe siècle et de la première moitié du XXe siècle, journaux de masse au papier acide et aux encres de basse qualité. Le problème de la conservation des documents papier se pose de manière accrue pour la presse des années 1870 et suivantes, avec le passage au papier de pâte de bois encollé en milieu acide. Le processus d'acidification de ces papiers entraîne, en quelques décennies seulement, « un jaunissement, puis une perte de souplesse et de résistance mécanique des feuillets, jusqu'à les rendre incommunicables aux

---

<sup>36</sup> A. Girard et C. Aurérin, *La presse physique et numérique à la BnF*, op. cit., p. 31.

lecteurs »<sup>37</sup>. À titre d'exemple, en 2009, deux tiers des journaux sur l'affaire Dreyfus n'étaient déjà plus utilisables, ce qui pose d'évidents problèmes pour la recherche<sup>38</sup>. Si la désacidification des documents est un moyen de les préserver, il arrive cependant que la fragilisation des feuillets soit telle que les traitements pour les désacidifier ne soient plus envisageables.

La numérisation apparaît alors comme l'unique recours pour sauvegarder des collections entières vouées à une disparition certaine. Cela requiert toutefois une restauration minimale et un coût que les structures ne sont pas toujours en mesure d'assumer. Les états de conservation parfois critiques et les nombreuses lacunes dans les collections rendent souvent difficile, pour un établissement documentaire, la prise en charge de la numérisation d'un titre de presse, même local, dès lors que celui-ci a une certaine périodicité. Pourtant, la question ne se limite pas qu'à la sphère des bibliothèques. C'est aussi, plus largement, une question politique ; car à travers ces collections physiques en grand danger, c'est toute une mémoire collective qui se trouve menacée.

***Numériser pour sauvegarder : une volonté politique réaffirmée,  
mais des moyens à renforcer***

La sauvegarde des collections de presse est le but premier de la numérisation de masse engagée à l'échelle nationale depuis une vingtaine d'années. Si celle-ci a été impulsée à la BnF par le mandat de l'historien des médias Jean-Noël Jeanneney<sup>39</sup> au début des années 2000, la décennie suivante a vu se renforcer cette prise de conscience : fin 2009, le Schéma numérique des bibliothèques de Bruno Racine<sup>40</sup> a mis en avant la place prépondérante que devait occuper la presse ancienne dans les programmes de numérisation nationaux. Le rapport insistait alors sur les moyens considérables et le fort engagement dont l'État devait faire preuve pour assurer cette numérisation de conservation<sup>41</sup>. Au-delà de l'impératif de « constituer des corpus documentaires numériques les plus exhaustifs possibles dans des domaines de référence répondant à la demande du public et à des nécessités de préservation du patrimoine », la perspective d'une valorisation des collections était également envisagée. Dès le début de la décennie 2010, une action nationale est ainsi appelée pour « rassembler et valoriser les collections de presse nationale et régionale déjà numérisées »<sup>42</sup>.

Par la suite, plusieurs rapports d'évaluation de l'action publique sont venus corroborer les observations et préconisations du Schéma numérique des bibliothèques, tant pour appuyer l'idée d'un plan d'urgence de numérisation que pour déplorer l'insuffisance des moyens déployés. Prenant appui sur les conclusions du rapport Presse 2018, ce constat a été renouvelé lors des Journées du

---

<sup>37</sup> Thi-Phuong Nguyen et Philippe Vallas, « La conservation des documents papier : point sur l'évolution des techniques et des stratégies », *Bulletin des bibliothèques de France* (BBF), 2006, n° 4, p. 11-21.

<sup>38</sup> Bruno Racine, *Schéma numérique des bibliothèques*, rapport élaboré dans le cadre du Conseil du Livre, décembre 2009, p. 13.

<sup>39</sup> Président de la BnF de 2002 à 2007.

<sup>40</sup> Président de la BnF de 2007 à 2016.

<sup>41</sup> Le financement sollicité par la BnF (20 M d'euros), bien que devant constituer une impulsion décisive à ce programme de numérisation, y était alors jugé insuffisant au regard des enjeux et des coûts. B. Racine, *Schéma numérique des bibliothèques*, *op. cit.*, p. 65.

<sup>42</sup> *Ibid.*, p. 7.

Patrimoine Écrit 2022, qui ont été l'occasion de rappeler l'implication du ministère de la Culture sur ces questions :

Ce sont des collections extrêmement importantes et précieuses, et qui sont consultées très fortement par nos concitoyens. Et si [...] nous ne le disons pas avec force dans un discours incontestablement politique et du niveau national, ces collections disparaîtront à bas bruit, et dans un certain nombre d'années, le pays se réveillera collectivement avec le constat que ces collections n'existent plus<sup>43</sup>.

Malgré tout, en dépit d'une volonté réaffirmée depuis plusieurs décennies, les budgets dégagés jusqu'à présent n'ont pas été à la hauteur des enjeux et la numérisation de masse des collections est encore loin d'être achevée. Il est donc nécessaire de renforcer encore les moyens mis en œuvre pour sauver ce patrimoine, en particulier les titres les plus vulnérables. Conscient de l'urgence de la situation, le ministère de la Culture fait désormais de la presse ancienne une priorité nationale. C'est dans ce contexte qu'a été décidée la création du futur centre de conservation à Amiens, et qu'un soutien supplémentaire va être accordé à la BnF à compter de 2024-2025 pour accélérer ses deux grands chantiers : la numérisation et le signalement des collections.

### **1.1.3. Les solutions mises en œuvre pour améliorer le signalement et la conservation des collections**

#### *Le rôle de Presse locale ancienne, héritier de la BIPFPIG*

Dès les années 1950-1960, la Bibliothèque nationale a lancé plusieurs programmes destinés à remédier aux problèmes de signalement et de conservation des collections de presse : d'une part, par le biais de l'ACRPP (Association pour la conservation et la reproduction photographique de la presse, créée en 1959), elle a entrepris de microfilmer 6000 titres de presse et de périodiques parmi les plus consultés en France<sup>44</sup>. Puis la Bibliographie de la presse française politique et d'information générale (BIPFPIG) a été créée pour améliorer le signalement de la presse ancienne régionale et locale. À l'heure actuelle, le signalement des collections apparaît comme un enjeu majeur de leur traitement et de leur valorisation, et l'achèvement de cette bibliographie constitue un axe fort de la politique du ministère en faveur de la presse ancienne.

Prolongeant la BIPFPIG, le site Presse locale ancienne<sup>45</sup>, lancé en 2016, est l'outil développé par la BnF pour repérer et décrire les collections de presse ancienne locale, qu'il s'agisse de titres conservés à la BnF ou en région. Cette version en ligne de la bibliographie s'appuie sur les données du Sudoc-PS, catalogue collectif français pour les périodiques, et des catalogues de la BnF. Le corpus couvert par Presse locale ancienne a été élargi par rapport à celui de la BIPFPIG : les critères définis dans les premiers volumes de la bibliographie papier<sup>46</sup> ont évolué au fil du temps, à mesure que le renouvellement des sciences humaines amenait à prendre en compte de nouveaux aspects. Le site recense ainsi

---

<sup>43</sup> Journées du Patrimoine Écrit 2022, discours d'ouverture de Nicolas Georges, Directeur chargé du livre et de la lecture au ministère de la Culture, 23-24 juin 2022.

<sup>44</sup> Lise Devreux et P. Mezzasalma, *Des sources pour l'histoire de la presse*, op. cit., p. 45.

<sup>45</sup> *Presse locale ancienne*, <https://presselocaleancienne.bnf.fr/accueil> (consulté le 17 janvier 2023).

<sup>46</sup> Celle-ci a commencé à paraître en 1964 et continue à paraître à l'heure actuelle.

des revues ou bulletins qui ne sont pas des journaux à proprement parler, mais qui complètent avantageusement le corpus de presse en incluant des publications importantes pour l'histoire d'un département : publications de sociétés d'agriculture, de chambres de commerce, etc. Le signalement a également été élargi à des journaux antérieurs à 1865 qui n'étaient pas recensés au départ, ainsi qu'à des titres des départements d'Outre-mer<sup>47</sup>.

Parallèlement, le site Presse locale ancienne a évolué dans ses missions : conçu au départ pour donner plus de visibilité à la BIPFPIG, il est apparu comme un outil bien plus commode que le volume papier pour signaler de manière fine les très nombreux projets de numérisation menés dans les territoires. En ligne, il est en effet possible d'indiquer le lien correspondant à une numérisation et de mettre à jour la bibliographie en temps réel : toute modification opérée sur une notice apparaît sur le site en seulement quelques jours. Cela implique toutefois un travail de veille permanent et des relations soutenues avec les nombreux établissements en région, à qui il revient de signaler les dernières numérisations effectuées. L'actualisation constante du signalement des titres numérisés est donc complexe à organiser ; pourtant, ce travail est essentiel pour la valorisation de ces titres.

### ***Le futur Conservatoire national de la presse, une réponse apportée aux problèmes de conservation et de stockage***

Malgré la multiplication des plans de conservation préventive curative depuis les années 1950 (reliure pour protéger les fascicules, restauration et doublage par thermocollage, programmes de désacidification, reproduction sur microfilms puis numérisation de masse), une partie significative des collections nationales n'est pas conservée dans des conditions optimales aujourd'hui, faute de place. La décision de bâtir un centre de conservation dédié en particulier à la presse<sup>48</sup> est d'abord venue du constat d'une saturation quasi complète des magasins de la BnF, en particulier sur le site François-Mitterrand.

Le repérage des titres de presse dans le cadre du rapport Presse 2018 a constitué un travail préalable à l'établissement d'une « collection de référence » correspondant aux titres les plus emblématiques, sélectionnés à la fois sur des critères d'usage (titres les plus demandés par les chercheurs) et de vulnérabilité physique. Ce premier ensemble d'environ 3000 titres<sup>49</sup> constitue le chantier prioritaire de l'implantation des collections dans le futur Conservatoire : la collection de référence sera en effet le premier ensemble à déménager. Elle fera également l'objet d'un traitement prioritaire, puisqu'il est prévu que l'intégralité

---

<sup>47</sup> À ses débuts, la BIPFPIG ne recensait que les journaux de 1865 à 1944. Mais ce choix initial a été réévalué à partir de 1980 : il est en effet apparu plus pertinent d'inventorier toute la presse ancienne, qui a souvent, dans les départements, démarré à la Révolution française voire sous l'Ancien Régime. Source : Entretien avec Nathalie Fabry, chargée du traitement rétrospectif des collections de périodiques et responsable du site Presse locale ancienne, le 04/10/2022.

<sup>48</sup> Celui-ci comprendra en réalité deux ensembles : un Conservatoire national de la presse et un centre de conservation destiné à accueillir tous types de collections. Mais ces deux ensembles seront complètement associés, et les structures (qu'il s'agisse des magasins ou des ateliers) leur seront communes. *Journées du Patrimoine Écrit 2022 : Projet centre de conservation de la BnF à Amiens, 2022.*

<sup>49</sup> Le rapport d'activité 2021 de la BnF indique : « Au total, ce sont 2 867 titres qui ont été retenus, des origines à 1958, permettant de couvrir le premier titre de presse paru en 1631, *La Gazette* de Renaudot, à ceux de la fin de la IV<sup>e</sup> République. Cette sélection sera soumise à l'avis d'un conseil scientifique constitué de chercheurs et de journalistes. 50% de ces titres correspondent à des journaux créés sous la III<sup>e</sup> République. » *BnF - Rapport d'activité 2021*, <https://www.bnf.fr/fr/bnf-rapport-dactivite-2021> (consulté le 13 février 2023).

des titres soit numérisée et ocrisée<sup>50</sup>. À terme, toute la collection de presse de la BnF (soit 70% de l'ensemble des collections déménagées) sera délocalisée dans le futur Conservatoire à Amiens.

Prévu pour être un centre de stockage des collections à très long terme, le Conservatoire entend offrir des conditions optimales pour la conservation des collections de presse : celles-ci seront conservées à plat, non pliées, dans des magasins robotisés à oxygène raréfié (sur le modèle de Boston Spa au Royaume-Uni et de Büron en Suisse, les deux seuls exemples de ce type en Europe). Leur traitement sera assuré par des ateliers de restauration et de numérisation, la restauration des fascicules étant le plus souvent un préalable nécessaire à toute opération de numérisation. En amont de l'ouverture du site, d'importants travaux de reconditionnement, de récolement et d'identification des documents devront être menés en interne.

Traduisant l'accélération des préoccupations autour de la presse ancienne, ce projet de centre de conservation et de collection de référence manifeste aujourd'hui une vraie volonté politique et la prise de conscience que, pour protéger ce qui relève tout autant d'un patrimoine matériel que d'une mémoire immatérielle, il y a urgence à numériser les titres les plus vulnérables. Mais au-delà de la préservation de cet héritage, il en va aussi du rayonnement de la collection : c'est là le second objectif de la numérisation. Comprendre la manière dont s'organisent ces programmes de numérisation à l'échelle du territoire est donc un préalable à toute réflexion sur la valorisation de la presse ancienne en France.

## 1.2. NUMÉRISER POUR MIEUX VALORISER : LES AXES DE VALORISATION DE LA PRESSE ANCIENNE

### 1.2.1. Programmes de numérisation à l'échelle nationale et régionale : les grands jalons

Si la numérisation est identifiée comme solution de sauvegarde de la presse ancienne, elle sert aussi à la valorisation de ce patrimoine écrit. Les premiers programmes d'envergure remontent au début des années 2000. Un premier plan quinquennal de numérisation de la presse a été lancé en 2005 sous l'impulsion de Jean-Noël Jeanneney, afin notamment de faire face au géant américain Google<sup>51</sup> : l'objectif était la mise en ligne sur Gallica de 22 titres de quotidiens nationaux parmi les plus représentatifs du XIXe siècle et du début du XXe siècle (jusqu'aux années 1940), correspondant à 3,2 millions de pages numérisées<sup>52</sup>. Ce plan, qui concernait de célèbres titres tels que *Le Temps*, *La Presse*, *Le Figaro*, *L'Humanité*, *Le Journal des débats*, *La Croix*, etc., a également servi de laboratoire pour impulser la modernisation de Gallica. De leur côté, les collectivités ont adopté, dès le tournant des années 1990-2000, des stratégies différentes pour numériser une

---

<sup>50</sup> L'ocrisation désigne le recours à la technologie OCR (*Optical character recognition*), ou reconnaissance optique de caractères, qui permet la recherche en texte intégral.

<sup>51</sup> Le président de la BnF de l'époque appelait les institutions européennes à une réaction concertée face à la force de frappe de Google, qui se proposait alors de numériser gratuitement – mais non sans contreparties – les documents des plus grandes bibliothèques mondiales. L'enjeu était alors, selon Jeanneney, celui de la préservation de la souveraineté de l'Europe sur son patrimoine écrit. Jean-Noël Jeanneney, *Quand Google défie l'Europe : plaidoyer pour un sursaut*, Paris, France, Mille et une nuits, 2005.

<sup>52</sup> A. Girard et C. Aurérin, *La presse physique et numérique à la BnF*, op. cit., p. 36.

partie de leurs collections de presse locale. Parmi les programmes les plus anciens, on peut citer celui de la région Rhône-Alpes, lancé dès 1996, celui de la Ville de Lyon, ou encore celui de la région Languedoc-Roussillon<sup>53</sup>.

La logique de ces programmes de numérisation fut d'abord géographique. Les premières campagnes menées par la BnF ont commencé par les journaux publiés à Paris, d'abord à diffusion nationale – grands quotidiens et quelques grands hebdomadaires – avant d'élargir à la presse locale parisienne. La question s'est ensuite posée pour tous les autres titres, à savoir la grande majorité de la production sur l'ensemble du territoire français. Dans les années 2005-2010, priorité a été donnée par le ministère aux établissements qui souhaitaient se lancer dans la numérisation de leur fonds local patrimonial, selon une logique de rayonnement local épousant la diffusion historique de ces titres de presse<sup>54</sup>.

Par la suite, c'est une logique davantage thématique qui a été adoptée : d'une part, dans une optique plus grand public, les programmes de numérisation ont été pensés en fonction des grandes thématiques couvertes par la presse du XIXe siècle, afin de refléter l'ensemble des tendances politiques et courants de pensée de l'époque. D'autre part, des sections thématiques ont vu le jour sur les différents portails numériques, correspondant à des catégories plus spécialisées telles que la presse satirique, la presse résistante, les journaux de mode, etc. La création de Numelyo, plateforme patrimoniale de la bibliothèque municipale de Lyon née en 2012, s'est ainsi appuyée sur plusieurs corpus thématiques préexistants, comprenant les revues savantes, un vaste ensemble de 300 titres de presse locale, ainsi qu'un important corpus Presse 1914-1918, dont une petite partie est aujourd'hui en ligne<sup>55</sup>.

D'importants programmes de numérisation ont également vu le jour à la faveur de grandes manifestations telles que les commémorations nationales ou certaines expositions, qui ont impulsé des collaborations fructueuses entre établissements pour numériser des corpus spécifiques. C'est le cas par exemple des journaux de tranchées, publications réalisées par des soldats au front et diffusées à l'échelle d'un régiment. Dès 2003, ces titres ont fait l'objet d'un programme de numérisation spécifique et concertée entre la BnF, la BDIC<sup>56</sup> et d'autres établissements documentaires – la Bibliothèque historique de la Ville de Paris, la Bibliothèque nationale et universitaire de Strasbourg, ou encore la bibliothèque municipale de Lyon. Ce type de programmes autour de corpus spécifiques, pilotés par la Bibliothèque nationale de France, se poursuit encore aujourd'hui ; en témoigne l'appel à projets de numérisation et de valorisation concertées en sport

---

<sup>53</sup> B. Racine, *Schéma numérique des bibliothèques*, op. cit., p. 13.

<sup>54</sup> La question s'est également posée pour la presse des territoires sous domination française pendant la colonisation : protectorats comme le Maroc ou la Tunisie, ou encore départements français comme l'Algérie. Source : *Journées du Patrimoine Écrit 2022 : Numériser les collections de presse*, intervention de Philippe Mezzasalma, chef du service Presse à la BnF, 23-24 juin 2022.

<sup>55</sup> Il s'agit d'une documentation préalablement acquise par la BM de Lyon sur les principaux belligérants, comprenant notamment de la presse étrangère. Il est à noter que plusieurs titres, bien que numérisés, n'ont pas été mis en ligne sur Numelyo. C'est le cas notamment du *Salut public*, quotidien publié dans le Rhône pendant près d'un siècle (1848-1944). Ce titre est diffusé sur le site de l'Arald (Agence Rhône-Alpes pour le livre et la documentation), Lectora Plus, mais pas encore sur Numelyo. L'objectif est toutefois qu'à terme, tous ces titres soient versés dans la plateforme Numelyo, numérisés en mode texte afin de faciliter la recherche. Source : Entretien avec Pierre Guinard, responsable du fonds ancien de la bibliothèque municipale de Lyon, le 28/11/2022.

<sup>56</sup> Bibliothèque de documentation internationale contemporaine, désormais La contemporaine.

dans la perspective des Jeux olympiques de 2024, dont l'un des axes thématiques porte sur les périodiques sportifs<sup>57</sup>.

## 1.2.2. Une coopération renforcée pour mieux valoriser la presse ancienne régionale

### *L'action territoriale de la Bibliothèque nationale de France*

Depuis cinq ans environ, la politique de coopération menée par la BnF avec ses partenaires en région s'est intensifiée. L'attention particulière prêtée aux collections régionales s'inscrit de fait dans les axes de valorisation jugés prioritaires par le ministère, comme le soulignait déjà, en 2009, le Schéma numérique des bibliothèques :

Dans le domaine de la valorisation du patrimoine culturel, la région est souvent l'échelon retenu par le ministère de la Culture et de la communication pour la mise en place de politiques concertées sous la forme de plans régionaux de numérisation et de portails régionaux de ressources numériques<sup>58</sup>.

Divers financements existent pour valoriser le patrimoine écrit en région. Le Programme de numérisation et de valorisation des contenus culturels (PNV) permet de soutenir les programmes de numérisation des établissements en région, ou encore le transfert des données de leur base régionale vers la BnF. Les bibliothèques territoriales peuvent aussi solliciter la Dotation générale de décentralisation (DGD), moyen privilégié par l'État pour accompagner ces programmes. Une participation financière de la BnF peut en outre être sollicitée dans le cadre de la Coopération : cette aide à la production numérique, d'abord accordée à hauteur de 50%, a été rehaussée à 80% en 2021 afin de maximiser le soutien apporté par l'État à ces chantiers<sup>59</sup>.

À partir de 2018, la BnF a commencé à déployer son action de coopération sur deux plans, d'une part en se faisant le relais en matière de subventions sur des projets déjà existants, d'autre part en étant force de proposition sur un élément de corpus jugé pertinent : la presse d'information générale publiée dans tous les chefs-lieux d'arrondissement de l'époque. Reposant sur une mutualisation documentaire et de financements, cette politique de coopération s'est développée dans le cadre de la politique de pôles associés régionaux<sup>60</sup>. Cette concertation régionale est redoublée par de nombreux partenariats bilatéraux : la mise en réseau repose essentiellement sur le binôme constitué par la bibliothèque municipale et le

---

<sup>57</sup> Le règlement pour la mise en œuvre du programme (2019-2024) précise : « En complément des périodiques déjà numérisés ou en cours de numérisation, cet axe concerne les titres relatifs au sport. [...] Cet axe écarte donc la presse généraliste nationale et régionale pour ne retenir que les revues spécialisées nationales et régionales ainsi que les périodiques des associations sportives. Il vise à offrir à la fois un accès aux titres les plus emblématiques et un focus sur des titres plus éphémères mais souvent rares. ». *Règlement pour la mise en œuvre du Programme de numérisation et de valorisation concertées en Sport*, [https://www.bnf.fr/sites/default/files/2020-05/Sport2020\\_Reglement.pdf](https://www.bnf.fr/sites/default/files/2020-05/Sport2020_Reglement.pdf) (consulté le 24 février 2023).

<sup>58</sup> B. Racine, *Schéma numérique des bibliothèques*, op. cit., p. 11.

<sup>59</sup> *Journées du Patrimoine Écrit 2022 : Numériser les collections de presse*, intervention d'Arnaud Dhermy, chef de la mission de la coopération régionale à la BnF, 23-24 juin 2022.

<sup>60</sup> En 2019, des conventions de « pôle associé » ont été signées avec la BnF dans huit régions : Auvergne-Rhône-Alpes, Provence-Alpes-Côte-d'Azur, Hauts-de-France, Nouvelle-Aquitaine, Grand-Est, Normandie, la Réunion, Antilles-Guyane. *Coopération régionale et action territoriale de la BnF*, <https://www.bnf.fr/fr/cooperation-regionale-et-action-territoriale-de-la-bnf> (consulté le 17 janvier 2023).

service d'archives départementales voisin, avec des possibilités de répartition, en termes à la fois documentaire et de conservation.

La décision de limiter l'action de la Coopération à la presse d'actualité<sup>61</sup> a été prise afin de garantir une homogénéité d'offre documentaire pour toutes les régions, quel que soit l'avancement des chantiers. À elle seule, la presse d'information générale représente un énorme corpus sur l'ensemble du territoire, et la phase de numérisation de masse est encore loin d'être achevée. Toutefois, les résultats de cet effort de numérisation concertée sont encourageants, avec une avancée annuelle de 2 à 3% du corpus total<sup>62</sup>.

***Constituer une collection de référence complète et homogène :  
l'objectif du marché « Presse 2021 »***

Aujourd'hui, le projet de création d'un Conservatoire de la presse accélère encore la cadence de ces programmes, d'une part en prévision du déménagement des collections, d'autre part dans le cadre de la sélection documentaire opérée pour la future collection de référence. En 2021, la BnF s'est lancée dans une vaste campagne de numérisation de titres conservés en région, notamment dans le but d'enrichir cette collection de référence : sur les 3000 titres que celle-ci comporte, 2000 sont déjà présents dans Gallica et le millier restant est en cours de numérisation<sup>63</sup>. Les chantiers initiés dans le cadre du marché Presse 2021 concernent d'importants titres de presse quotidienne régionale, tels que *Le Bien public*, *Le Progrès*, *La Voix du Nord*, les *Dernières Nouvelles d'Alsace*, *Le Mémorial bordelais*, etc. Dans le cadre de ce programme, il est demandé aux établissements partenaires l'envoi de 130 000 pages par an à la BnF pour numérisation. Un cahier des charges a été fixé pour garantir l'homogénéité de traitement des fascicules et l'harmonisation des métadonnées.

Si la coopération est déjà ancienne avec certains partenaires comme la BNU, les chiffres montrent l'accélération constante de cette dynamique de coopération et le passage à une autre échelle de volumétrie : la BnF numérise environ 170 000 pages par mois à l'heure actuelle, contre 35 000 à 40 000 pages mensuelles en 2015<sup>64</sup>. L'objectif visé est la complétude des collections, en s'appuyant notamment sur les bibliothèques de dépôt légal imprimeur (BDLI) qui possèdent les titres de presse quotidienne régionale les plus exhaustifs. C'est le cas par exemple de la bibliothèque municipale de Lyon, qui détient la collection la plus complète du *Progrès* depuis sa création en 1859 : non seulement l'édition de Lyon, mais aussi seize éditions locales qui lui parviennent quotidiennement<sup>65</sup>. Ce chantier de comblement des lacunes s'inscrit également dans une logique partenariale à l'échelle des collectivités territoriales. Au sujet de la numérisation des *Dernières Nouvelles d'Alsace* et de *L'Alsace*, Jérôme Schweitzer indique :

---

<sup>61</sup> Le marché de numérisation « Presse 2021 » exclut, de fait, de nombreux autres pans de la production éditoriale de l'époque : presse satirique, presse des spectacles, presse associative, presse littéraire de création, etc.

<sup>62</sup> Entretien avec Arnaud Dhermy, chef de la mission de la coopération régionale à la BnF, le 29/09/2022.

<sup>63</sup> Entretien avec Julie Ladant, directrice du département Droit, économie, politique à la BnF, le 13/02/2023.

<sup>64</sup> Entretien avec Mathilde Dutertre, adjointe au chef du service Numérisation à la BnF et Marie-Noëlle Gourdel, cheffe de projet numérisation, le 23/11/2022.

<sup>65</sup> Entretien avec Gaétan Carlier, conservateur stagiaire à la bibliothèque municipale de Lyon en charge de la préparation à la numérisation du *Progrès*, le 14/11/2022.

Outre ce partenariat noué dans le cadre du réseau des pôles associés de la BnF, ce projet concernera d'autres partenaires potentiels comme les Médiathèques de Strasbourg ou les archives des DNA qui disposent d'exemplaires qui permettront de compléter les collections de la BNU parfois lacunaires pour certaines années<sup>66</sup>.

Plus que jamais, la coopération apparaît donc essentielle pour compléter les collections et en assurer la diffusion auprès du public. À terme, ce corpus numérisé homogène sera véritablement exploitable pour la recherche : représentatif de la production éditoriale en région, il ne sera pas limité par des choix documentaires variant d'un territoire à un autre, mais permettra d'avoir des taux de comparaison éclairants sur divers sujets d'actualité. Il sera ainsi permis d'analyser la lecture qu'un territoire a pu faire de l'actualité nationale ou internationale, ou encore la manière dont l'information était filtrée au niveau des populations dans telle ou telle région, influençant les mentalités.

### 1.2.3. Gallica et RetroNews, deux plateformes nationales pour diffuser les collections numérisées

#### *La mise en valeur de la presse dans Gallica : différents modes d'accès pour différents publics*

À l'échelle nationale, la presse ancienne a la particularité d'être consultable sur deux plateformes, Gallica et RetroNews – la première étant pilotée par la BnF et la seconde par la filiale, BnF-Partenariats. À la différence de RetroNews, plateforme exclusivement consacrée à la presse, Gallica est la bibliothèque de tout le patrimoine écrit et audiovisuel français. Cette différence de nature entre les deux sites de la BnF explique en grande partie leur positionnement distinct dans leur manière de diffuser et de valoriser les collections de presse ancienne. Dans le cas de Gallica, qui s'appuie sur le catalogue général de la BnF, presse et revues sont indifférenciées ; la recherche par types de documents amène donc à explorer ces deux corpus sans distinction. De plus, la presse étant l'un des multiples supports diffusés sur Gallica, le moteur de recherche ne propose pas des filtres spécifiques à la presse, mais des champs de recherche suffisamment larges pour s'adapter à tout type de document<sup>67</sup>. La presse représente cependant la part majoritaire de la collection de Gallica : 5,8 millions de documents, soit 70% de la collection. 78% des numéros sont numérisés<sup>68</sup>.

À l'échelle d'une bibliothèque numérique aussi large que Gallica, éditorialiser les collections de presse relève donc de la gageure. Aussi les pages « Sélections » de Gallica jouent-elles un rôle déterminant dans la mise en valeur des contenus. Les sélections proposées permettent d'orienter la recherche dans les corpus de presse ancienne par tranches diachroniques ou thématiques ; pour l'utilisateur, elles jouent ainsi le rôle de filtres. Ces sélections sont régulièrement enrichies par les équipes de Gallica ou par des contributions de chercheurs. La

---

<sup>66</sup> Jérôme Schweitzer, « La presse : axe majeur de la numérisation en 2022 », <https://bnu.hypotheses.org/9893> (consulté le 21 septembre 2022).

<sup>67</sup> Quelques ajustements sont toutefois proposés dans le moteur de recherche avancée : ainsi, lorsqu'on sélectionne « Presse et revues », l'année d'édition est remplacée par la date du fascicule, critère prédominant pour la recherche dans ces corpus.

<sup>68</sup> Présentation de Gallica par Arnaud Laborderie et Marie Carlin à l'Enssib dans le cadre du cursus de formation DCB (UE « Données et Science ouvertes »), le 08/02/2023.

sélection sur les feuillets dans la presse, par exemple, est le fruit du travail d'une chargée de recherche spécialiste des romans-feuillets du XIXe siècle, partie intégrante de la presse quotidienne de cette époque<sup>69</sup>. Les axes de valorisation mis en avant épousent souvent des thématiques de recherche prisées des historiens et chercheurs en Lettres : presse allophone, presse des colonies, presse ouvrière et syndicale, presse résistante, presse et littérature, etc.

Il est également possible d'accéder aux sélections de presse par d'autres chemins que la recherche par typologie documentaire : par thématiques ou par aires géographiques notamment. Ainsi par exemple, la thématique « Histoire » pourra-t-elle guider l'utilisateur vers des ressources périodiques relatives à la Première Guerre mondiale (journaux de tranchées, bulletins de presse étrangère, médaillés posthumes de la Grande Guerre, etc.). On peut aussi consulter la presse francophone éditée en Algérie en cherchant dans les ressources disponibles via l'accès par pays. De même, *La Dépêche de Brest* est accessible aussi bien via l'entrée « Presse locale et régionale » que par une recherche géographique, via l'accès par régions ou par départements français.

Ainsi éditorialisés, les contenus sont rendus plus accessibles à un public peu expert de la recherche dans les catalogues ou les moteurs spécialisés. Ces différents modes d'entrée possibles sont particulièrement utiles dans le cadre de recherches généalogiques, par exemple, où l'utilisateur peut être amené à croiser différentes informations relatives à la vie d'un individu. Une recherche géographique pourra ainsi s'avérer complémentaire d'une requête sur le nom de la personne dans le moteur de recherche avancée. Il est à supposer que le chercheur professionnel, davantage accoutumé à rechercher dans les catalogues, aura plus spontanément recours au moteur de recherche avancée. Les sélections peuvent néanmoins s'avérer précieuses pour sa recherche, si celle-ci porte sur un type de corpus bien défini : une chercheuse interrogée nous a ainsi indiqué avoir eu recours de manière quasi exclusive, pour une communication sur la bande dessinée dans la presse du XIXe siècle, aux titres mis en avant par Gallica dans la sélection correspondante<sup>70</sup>. Il est à noter que le moteur de recherche de Gallica a connu une récente évolution, permettant désormais la recherche par proximité de termes et facilitant ainsi l'exploration des titres océrisés.

La presse présente dans Gallica est aussi mise en valeur par différentes actions de médiation embrassant un plus large public, notamment sur les réseaux sociaux et via le blog et la lettre de Gallica<sup>71</sup>. Ces opérations visent moins à faciliter les usages de la presse qu'à mettre en lumière la diversité des corpus numérisés par la BnF. Faisant la part belle à l'anecdote, au fait divers, la presse ancienne se prête particulièrement bien à ce type de valorisations ponctuelles. Celles-ci peuvent être en lien avec un événement d'actualité : l'entrée de Joséphine Baker au Panthéon le 30 novembre 2021 a ainsi donné lieu à un billet de blog consacré à ses portraits dessinés dans la presse<sup>72</sup>. En mai 2022, en écho à la

---

<sup>69</sup> *Les feuillets dans la presse*, intervention de Morgane Avellaneda, chargée de recherche documentaire (BnF/Université de Saint-Etienne), <https://www.bnf.fr/fr/agenda/les-feuillets-dans-la-presse>, 10 avril 2021 (consulté le 4 janvier 2023).

<sup>70</sup> Entretien avec Sophie Bros, doctorante à l'Université Paris Nanterre et attachée temporaire d'enseignement et de recherche à l'Université Sorbonne Nouvelle, le 01/12/2023.

<sup>71</sup> La lettre de Gallica, mensuelle, propose différents focus et met l'accent sur les nouveautés : « Dernières mises en ligne », « À la une », « Les derniers billets du blog », « Dans l'œil de Gallica », etc.

<sup>72</sup> *Joséphine Baker, portraits dessinés dans la presse - Le blog de Gallica*, <https://gallica.bnf.fr/blog/30112021/josephine-baker-portraits-dessines-dans-la-presse?mode=desktop> (consulté le 1 février 2023).

Journée de l'Europe et à la présidence française du Conseil de l'Union européenne, Gallica a proposé un focus sur sa nouvelle sélection consacrée à la presse européenne<sup>73</sup>. Début 2023, ce sont les collections de périodiques des XIXe et XXe siècles publiés en Algérie coloniale (1830-1962), qui ont été mises à l'honneur à travers deux billets sur le blog de Gallica<sup>74</sup>.

### *RetroNews, le « site média d'histoire » de la BnF : l'accent mis sur l'éditorialisation*

Parallèlement à la bibliothèque numérique Gallica, la BnF a lancé en 2016 sa plateforme RetroNews, entièrement dédiée à la presse ancienne. Fruit d'un partenariat entre la filiale privée de la BnF et la société Immanens, ce site éditorialisé a été constitué à partir du corpus de microfilms du catalogue de l'ACRPP. Le corpus initial comprenait 5 millions de pages numérisées à partir de microfilms conservés par la BnF et 3 millions de pages déjà visibles dans Gallica, auxquelles des traitements sémantiques complémentaires ont été appliqués afin d'en améliorer l'accessibilité. Aujourd'hui, plus de 19 millions de pages sont accessibles, correspondant à 2000 titres de presse publiés en France entre 1631 et 1951<sup>75</sup>. Puisqu'elle contient exclusivement des archives de presse, la plateforme RetroNews n'est pas confrontée aux difficultés posées par une typologie documentaire à la fois massive et hétérogène, comme c'est le cas sur Gallica. Elle est donc à même de proposer un moteur de recherche dédié, avec des options de filtrage ciblées : filtrage temporel, par type de périodicité, par événement, par thématique, etc.

Si la partition entre les titres numérisés par la BnF et ceux numérisés par sa filiale s'est faite avant tout sur la base du support – microfilms pour BnF-P, originaux pour la BnF –, les critères de sélection des titres à numériser sont peu ou prou identiques : sur les deux plateformes, on trouve de la presse quotidienne nationale et régionale, de la presse spécialisée, politique, satirique ou provenant des anciens territoires colonisés, avec une périodicité allant principalement du quotidien au mensuel. On compte à l'heure actuelle 1 million de documents en commun entre RetroNews et Gallica<sup>76</sup>. Éviter une forme de redondance dans les ressources que proposent les deux sites est donc un enjeu important aujourd'hui, notamment en termes de clarté pour l'utilisateur<sup>77</sup>.

Le modèle économique constitue toutefois une différence importante entre les deux plateformes. Contrairement à Gallica, où l'accès aux collections est gratuit sans restriction autre que juridique, RetroNews obéit à un modèle Freemium : les fonctionnalités avancées du site (filtres de recherche avancés,

---

<sup>73</sup> Focus sur l'activité européenne de la BnF et publication du corpus Presse européenne sur Gallica, <https://www.bnf.fr/fr/actualites/focus-activite-europeenne-bnf-et-publication-corpus-presse-europeenne-gallica> (consulté le 10 janvier 2023).

<sup>74</sup> Les premiers titres de presse en Algérie : 1830-1880 - Le blog de Gallica, <https://gallica.bnf.fr/blog/16012023/les-premiers-titres-de-presse-en-algerie-1830-1880?mode=desktop> (consulté le 24 janvier 2023) ; L'âge d'or de la presse algérienne : 1880-1914 - Le blog de Gallica, <https://gallica.bnf.fr/blog/23012023/lage-dor-de-la-presse-algerienne-1880-1914?mode=desktop> (consulté le 24 janvier 2023).

<sup>75</sup> Entretien avec Étienne Manchette, responsable Partenariats et Contenus auprès de BnF-Partenariats, le 30/09/2022. Voir aussi *Retronews - Qui sommes-nous ?*, <https://www.retronews.fr/content/qui-sommes-nous>, 19 février 2016 (consulté le 1 avril 2022).

<sup>76</sup> Présentation de Gallica par Arnaud Laborderie et Marie Carlin à l'Enssib dans le cadre du cursus de formation DCB (UE « Données et Science ouvertes »), le 08/02/2023.

<sup>77</sup> A. Girard et C. Aurérin, *La presse physique et numérique à la BnF*, op. cit., p. 43-44.

outils de visualisation et d'analyse sémantique) sont accessibles sur abonnement payant. Le modèle de partenariat public-privé auquel obéit RetroNews prévoit également un embargo de sept ans sur les titres dont la numérisation a été assurée par BnF-P. Au terme de ces sept années, les titres basculent en accès libre sur Gallica<sup>78</sup>. Désormais, la clause d'exclusivité de diffusion touchant à sa fin pour les titres du corpus initial de RetroNews, la filiale est amenée à évoluer dans son positionnement. Dans la lignée du projet RetroNews, BnF-P a annoncé en septembre 2022 un nouveau programme de numérisation, centré cette fois-ci sur la presse moderne et contemporaine encore sous droits<sup>79</sup>.

Au fur et à mesure, RetroNews s'est positionné de plus en plus comme le « site média d'histoire » de la BnF, mettant fortement l'accent sur l'éditorialisation des contenus :

L'usage de ces contenus, en particulier en direction des publics non experts, passe par l'éditorialisation et la diffusion. Élargir le champ des utilisateurs et valoriser ce patrimoine exigeaient de proposer un site qui s'inscrive au-delà de la consultation experte d'archives et sous-entendait de développer à la fois un outil et un site média<sup>80</sup>.

Cette éditorialisation est assurée par une équipe de contributeurs – journalistes, historiens, universitaires – rémunérés pour leurs productions. De célèbres vulgarisateurs tels que Benjamin Brillaud, créateur de la chaîne YouTube *Nota Bene*, sont également mis à contribution pour valoriser les contenus du site. Centrées autour d'une thématique particulière comme par exemple la naissance de la photographie, ces courtes pastilles de quelques minutes mettent en avant des archives de presse variées : extraits du *Figaro*, du *Journal pour Rire*, du *Monde illustré*, de *L'Excelsior*, etc.<sup>81</sup>. Autre mode de vulgarisation développé, le podcast permet d'aborder ces archives par le biais de la lecture, dans un format d'une demi-heure environ. Au sujet du podcast « Séries noires à la Une », qui convoque divers spécialistes de l'histoire de la justice et des médias pour commenter des faits divers relayés dans la presse de l'époque (l'affaire Landru, les crimes de la Terreur, la bête du Gévaudan, etc.), Étienne Manchette explique :

Nous avons interrogé dans cette série les marges de la société : les avorteuses, les révolutionnaires, les criminels et criminelles politiques. Nous montrons comment tout cela se dessine et comment la presse participe d'un ordre moral, ou en tout cas d'une structuration de l'information, jouant un rôle déterminant dans la société<sup>82</sup>.

Au-delà de son rôle de diffusion, le site RetroNews devient donc véritablement producteur ou coproducteur de contenus. Valorisant la dimension

---

<sup>78</sup> Ils sont auparavant disponibles en accès restreint sur Gallica *intra muros*. Cette interface donne accès aux contenus de Gallica protégés au titre de la propriété intellectuelle, qui ont été numérisés par la BnF et ses partenaires.

<sup>79</sup> Un accord avec douze éditeurs de presse, dont *Libération*, *Paris Match*, *L'Express* et *Le Télégramme*, a été passé en ce sens. La future plateforme, développée par la société Contentside, devrait voir le jour fin 2023, avec un volume initial de 8,5 millions de pages numérisées. *ActuaLitté*, « Après la presse ancienne, la BnF numérise les titres modernes », <https://actualitte.com/article/108013/numerisation/apres-la-presse-ancienne-la-bnf-numerise-les-titres-modernes> (consulté le 11 octobre 2022).

<sup>80</sup> Étienne Manchette et Nathalie Thouny, « RetroNews : éditorialiser la presse ancienne », *Bulletin des bibliothèques de France* (BBF), 2018, n° 15, p. 32-35.

<sup>81</sup> *Quand la France découvre la photographie (Léa Bello feat Nota Bene)*, chaîne YouTube RetroNews-BnF, <https://www.youtube.com/watch?v=bs4nT7Fj9NM>, 2020.

<sup>82</sup> Entretien avec Étienne Manchette, responsable Partenariats et contenus pour BnF-Partenariats, le 30/09/2022.

narrative de la presse ancienne, le podcast est un moyen de conquérir de nouveaux publics. Un récent rapport de l'Inspection générale des affaires culturelles souligne ainsi le succès, auprès de publics jeunes en particulier, de ces formats audionumériques renouvelant la fiction, notamment « le documentaire en série : enquêtes, valorisation d'archives, témoignages, création »<sup>83</sup>. Les formats vidéo sont quant à eux un moyen de valoriser la composante illustrée du journal : « en devenant un objet culturel, la photographie de presse va circuler [...] en dehors des pages imprimées des journaux »<sup>84</sup>, ce qui permet d'en offrir une autre lecture. RetroNews a également fait le pari de la rematérialisation du numérique en lançant sa revue papier. Accessible dans des kiosques et points relais en gare, celle-ci vise à toucher un public qui n'irait pas spontanément vers les collections de presse de la BnF. L'archive est alors transformée en un modèle éditorial qui permet d'offrir une épaisseur historique sur divers sujets d'actualité comme le réchauffement climatique<sup>85</sup>. Un effort concret d'accessibilité et de vulgarisation est demandé aux universitaires et autres contributeurs écrivant dans la revue, pensée comme le prolongement direct du « média RetroNews ».

### 1.3. DIFFERENTES STRATEGIES DE VALORISATION SELON LES TERRITOIRES

#### 1.3.1. Des modes de diffusion pluriels : de l'indépendance à l'intégration dans Gallica, différents positionnements à l'égard de l'opérateur national

La dernière décennie a vu se multiplier, en région, les plateformes visant à mettre en valeur le patrimoine local. Qu'il s'agisse du portail d'un service d'archives, d'une agence régionale pour le livre et la lecture ou d'une collectivité (région, département, ville, communauté d'agglomération, etc.), la presse occupe le plus souvent une place de premier ordre parmi les ressources mises en ligne. Toutefois, les choix opérés en termes de diffusion de ces collections varient selon les cas, reflétant des différences de positionnement parfois significatives d'un établissement ou d'un territoire à l'autre.

Pour la presse, actuellement, il existe trois modes d'existence de plateformes numériques. Pour certaines collectivités, la volonté de garder une indépendance, à la fois en matière de diffusion et d'archivage, implique de maintenir une plateforme indépendante à l'égard de Gallica. C'est le cas notamment de grandes villes telles que Lyon avec sa bibliothèque patrimoniale Numelyo<sup>86</sup>, Bordeaux avec Séléne<sup>87</sup>, Marseille avec MARIus<sup>88</sup>, Orléans avec Aurelia<sup>89</sup>, ou encore divers

---

<sup>83</sup> François Hurard et Nicole Phoyu-Yedid, *L'écosystème de l'audio à la demande (« podcasts ») : enjeux de souveraineté, de régulation et de soutien à la création audionumérique*, rapport de l'Inspection générale des affaires culturelles, n°2020-11, <https://www.culture.gouv.fr/Espace-documentation/Rapports/L-ecosysteme-de-l-audio-a-la-demande-podcasts-enjeux-de-souverainete-de-regulation-et-de-soutien-a-la-creation-audionumerique>, 19 novembre 2020 (consulté le 2 février 2023).

<sup>84</sup> Stéphanie Le Maître, « La valorisation numérique du patrimoine photographique de presse », *Médias19*, <https://www.medias19.org/publications/presse-et-patrimoine/la-valorisation-numerique-du-patrimoine-photographique-de-presse>, 20 juin 2022 (consulté le 4 avril 2022).

<sup>85</sup> *RetroNews - la Revue*, n°4, « Le temps des crises écologiques », automne 2022.

<sup>86</sup> *Numelyo - Bibliothèque numérique de Lyon*, <https://numelyo.bm-lyon.fr/> (consulté le 15 février 2023).

<sup>87</sup> *Séléne - Bibliothèque numérique de Bordeaux*, <https://selene.bordeaux.fr/> (consulté le 15 février 2023).

portails de services d'archives départementales. À l'origine de ce choix, l'argument de l'identité territoriale est le plus souvent invoqué : les documents étant conservés grâce à des fonds publics territoriaux, c'est l'échelon local, et non l'opérateur national, qui est privilégié pour la diffusion de ces ressources. Toutefois, ce modèle tend à devenir minoritaire aujourd'hui, notamment pour des questions de conservation pérenne. En effet, l'archivage pérenne des données numérisées représente un coût significatif et s'avère difficilement soutenable, d'un point de vue financier et technique, pour bon nombre de collectivités. Or, la BnF, avec son Système de Préservation de l'Archivage Réparti (SPAR)<sup>90</sup>, est à même d'offrir une solution de préservation numérique répondant à ce besoin.

C'est pourquoi certaines plateformes ont opté pour la double diffusion – dans Gallica et dans leur propre interface – de leurs ressources de presse. L'exemple le plus connu est la plateforme de presse du Sillon lorrain, Limédia Kiosque. Dans ce type de cas de figure, la solution privilégiée est hybride : la collaboration avec Gallica ne signifie pas la perte de la diffusion mais porte sur l'archivage pérenne des ressources, offrant une garantie de conservation dans le cas d'une défaillance technique de la plateforme externe<sup>91</sup>. Le choix de cette double diffusion permet à l'établissement d'obtenir la gratuité de l'archivage pérenne dans SPAR. En retour, la BnF obtient la copropriété sur les ressources numériques produites dans le cadre de cette coopération. Cette solution, également proposée aux services d'archives départementales, tend à se développer de plus en plus. À l'instar de Limédia, plusieurs portails régionaux vont bientôt passer sous ce modèle, qui a amené la BnF à améliorer son offre en matière d'intégration sans monopole de diffusion<sup>92</sup>.

La troisième modalité possible est l'intégration dans Gallica. Cette modalité inclut notamment le dispositif « Gallica marque blanche ». Inauguré en 2013 avec Numistral, la bibliothèque numérique de la Bibliothèque nationale et universitaire de Strasbourg<sup>93</sup>, ce projet propose le développement d'une bibliothèque numérique patrimoniale aux couleurs d'un partenaire en région mais avec le socle technique, le moteur de recherche et le visualiseur de Gallica. L'URL et l'appareil éditorial sont, quant à eux, totalement autonomes. Une quinzaine de projets de la sorte ont vu le jour depuis la naissance du dispositif<sup>94</sup>. En font partie, notamment, les bibliothèques numériques de la Ville de Rouen (Rotomagus)<sup>95</sup>, de Toulouse (Rosalis)<sup>96</sup>, de la Communauté d'agglomération Pau Béarn Pyrénées (Pireneas)<sup>97</sup>,

---

<sup>88</sup> Marius, *MARseille Images Universelles et Singulières*, <http://marius.marseille.fr/marius/jsp/site/Portal.jsp> (consulté le 15 février 2023).

<sup>89</sup> Aurelia - *Collections patrimoniales numérisées*, <https://mediatheques.orleans-metropole.fr/patrimoine/collections-patrimoniales-numerisees> (consulté le 15 février 2023).

<sup>90</sup> SPAR (*Système de Préservation et d'Archivage Réparti*), <https://www.bnf.fr/fr/spar-systeme-de-preservation-et-d-archivage-reparti> (consulté le 9 janvier 2023).

<sup>91</sup> *Journées du Patrimoine Écrit 2022 : Numériser les collections de presse*, intervention d'Arnaud Dhermy, chef de la mission de la coopération régionale à la BnF, 23-24 juin 2022.

<sup>92</sup> Entretien avec Arnaud Dhermy, chef de la mission de la coopération régionale à la BnF, le 29/09/2022.

<sup>93</sup> Numistral, <https://www.numistral.fr/fr> (consulté le 14 février 2022).

<sup>94</sup> « Gallica marque blanche : qu'est ce que c'est ? », Auvergne-Rhône-Alpes Livre et Lecture, <https://auvergnerrhonealpes-livre-lecture.org/gallica-marque-blanche-qu-est-ce-que-c-est/pdf> (article du 15 juin 2021, dernière édition du 13 février 2023).

<sup>95</sup> Rotomagus, <https://www.rotomagus.fr/rotomagus/?mode=desktop> (consulté le 14 février 2022).

<sup>96</sup> Rosalis, <https://rosalis.bibliotheque.toulouse.fr/rosalis/fr/content/accueil-fr?mode=desktop> (consulté le 15 février 2023).

de Brest (Yroise)<sup>98</sup>, du Havre (Nutrisco)<sup>99</sup>, ou encore de l'Établissement public territorial de Plaine Commune (regroupement de neuf villes du nord de Paris)<sup>100</sup>. En 2022, deux nouvelles bibliothèques ont été créées en marque blanche : PaGella pour la Ville de Grenoble<sup>101</sup> et Mémomum pour Montpellier Méditerranée Métropole<sup>102</sup>. D'autres types d'établissements peuvent également intégrer le projet, à l'instar de l'École des Ponts Paristech, qui a opté pour ce mode de diffusion de ses archives – de presse notamment<sup>103</sup>. Si l'apport documentaire provient du partenaire et fait l'objet d'une double diffusion (dans le portail en marque blanche et dans Gallica), le partenaire a également la possibilité de verser dans son interface des titres provenant de Gallica. Ainsi peut-il, en retour de sa contribution à l'enrichissement de la bibliothèque numérique nationale, combler les lacunes de ses propres collections. Il n'est donc pas rare de trouver, dans une bibliothèque en marque blanche, des titres d'origines multiples. La page « Presse ancienne » de la bibliothèque patrimoniale de Brest mentionne explicitement cette diversité de provenances : « Remontez le temps à travers deux grands quotidiens régionaux, la *Dépêche de Brest* (numérisée en partenariat avec le *Télégramme*, les Archives départementales et les Archives municipales) et *Ouest Eclair* (provenant de Gallica) »<sup>104</sup>.

L'intégration dans Gallica peut aussi se faire en dehors du cadre d'une marque blanche – ce qui n'empêche en aucun cas la bibliothèque partenaire de rejoindre ultérieurement le dispositif. Une filière possible d'intégration est le référencement par interopérabilité OAI-PMH, qui a gagné en importance à partir de la décennie 2010 avec la multiplication des bibliothèques numériques patrimoniales. Cela répondait aux préconisations du Schéma numérique des bibliothèques, qui recommandait d'en « faire un critère majeur de la contractualisation en cas d'attribution de subventions publiques pour la numérisation »<sup>105</sup> :

Les documents référencés dans Gallica par moissonnage des bibliothèques numériques sont issus d'opérations de numérisation largement soutenues par la BnF par attribution de subventions. Les opérations subventionnées s'inscrivent pour la plupart dans les programmes nationaux thématiques et les programmes de complétude de corpus. Le soutien à la numérisation est conditionné par l'obligation pour les partenaires de mettre

---

<sup>97</sup> *Pireneas - Bibliothèque numérique de Pau*, <https://www.pireneas.fr/pireneas/fr/content/accueil> (consulté le 14 février 2022).

<sup>98</sup> *Yroise*, <https://yroise.biblio.brest.fr/yroise/fr/content/accueil-fr?mode=desktop> (consulté le 15 février 2023).

<sup>99</sup> *Nutrisco - Le Havre*, <https://nutrisco-patrimoine.lehavre.fr/nutrisco/?mode=desktop> (consulté le 15 février 2023).

<sup>100</sup> *Commun Patrimoine*, <https://www.communpatrimoine.fr/communpatrimoine/> (consulté le 15 février 2023).

<sup>101</sup> *Pagella*, <https://pagella.bm-grenoble.fr/pagella/fr/content/accueil-fr> (consulté le 15 février 2023).

<sup>102</sup> *Mémomum*, <https://memonum-mediatheques.montpellier3m.fr/memonum/fr/content/accueil-fr> (consulté le 15 février 2023).

<sup>103</sup> *L'Héritage des ponts et chaussées - Presse et revues*, <https://heritage.ecoledesponts.fr/enpc/fr/content/les-collections-numerisees-types-de-documents-presse> (consulté le 15 février 2023).

<sup>104</sup> *Presse ancienne - Yroise*, <https://yroise.biblio.brest.fr/yroise/fr/content/presse-ancienne> (consulté le 15 février 2023).

<sup>105</sup> B. Racine, *Schéma numérique des bibliothèques*, op. cit., p. 7.

en œuvre l'interopérabilité des bibliothèques numériques, permettant le moissonnage<sup>106</sup>.

Les bibliothèques peuvent alors, de la même manière, enrichir leurs collections de documents présents dans Gallica. C'est le cas du portail basque Bilketa, qui depuis sa création en 2015 n'a eu de cesse d'intégrer de nouveaux contenus de différents partenaires, dont la BnF : au total, 10 000 documents proviennent des collections de la Bibliothèque nationale. L'interface a récemment évolué pour faciliter la consultation de ses 25 000 documents, dont 17 000 articles de presse issus de 60 titres de journaux<sup>107</sup>. Un accès direct aux journaux basques présents sur Gallica, comme le journal des Basques de Paris *Elgar*, est permis grâce à l'intégration du visualiseur de Gallica à la plateforme. Le réseau Bilketa joue ainsi un rôle moteur, à l'échelle de la France, pour la valorisation du patrimoine écrit basque et bayonnais.

### 1.3.2. Exemples de valorisation numérique en région

Ces différences de positionnements s'expliquent aussi, bien souvent, par des raisons historiques, mémorielles ou encore linguistiques. Marquées par une histoire singulière, certaines régions ont en effet tout intérêt à privilégier une porte d'entrée locale pour mettre en valeur leur patrimoine. Au-delà de la diffusion de ces collections, l'existence d'un portail local se justifie alors par un réel enjeu de valorisation : la plateforme régionale, par son appréhension fine des données propres à son territoire, offre en effet une granularité que l'opérateur national n'est pas en mesure de fournir. Si notre étude n'a pas pour vocation de proposer un panorama exhaustif de la valorisation de la presse ancienne en région, nous nous proposons toutefois de revenir sur quelques exemples qui, à nos yeux, illustrent bien ces logiques.

#### *Limédia Kiosque : une expérimentation innovante de mutualisation des ressources d'un territoire*

En matière de valorisation de la presse ancienne régionale, le modèle proposé par la plateforme du Sillon lorrain Limédia Kiosque (anciennement Le Kiosque lorrain) est souvent cité comme une réalisation exemplaire<sup>108</sup>. D'une part en raison de la particularité de ce projet fédérant quatre intercommunalités autour d'un portail numérique mutualisé. D'autre part en raison du choix original de scinder le portail en trois sites, correspondant à trois types d'usages différents<sup>109</sup>. À la manière d'une plateforme digitale de contenus culturels, Limédia met au premier

---

<sup>106</sup> *Les partenaires de Gallica en 2017*, Direction des services et des réseaux, Département de la coopération, [https://www.bnf.fr/sites/default/files/2018-11/partenaires\\_gallica.pdf](https://www.bnf.fr/sites/default/files/2018-11/partenaires_gallica.pdf), février 2018 (consulté le 23 février 2023).

<sup>107</sup> *Bilketa, le portail numérique des fonds documentaires basques fait peau neuve*, <https://www.baskulture.com/article/bilketa-le-portail-numerique-des-fonds-documentaires-basques-fait-peau-neuve-5054> (consulté le 10 janvier 2023).

<sup>108</sup> Nous nous appuyons ici sur divers entretiens menés avec des interlocuteurs en région.

<sup>109</sup> « Galeries » pour les mémoires culturelles du Sillon lorrain, « Mosaique » pour la médiathèque numérique et « Kiosque » pour la presse ancienne. Ces différents usages priment sur les logiques institutionnelles, comme l'explique Malik Diallo : « Les trois sites sont donc conçus comme des bibliothèques de contenus en ligne les plus homogènes possibles, avec leur identité propre. Il s'agit de dissocier les sites de contenus de ceux des bibliothèques, pour mieux les rendre visibles. [...] Celles-ci seront d'autant plus utiles et valorisées que leurs contenus seront vus et utilisés. » Malik Diallo, « Limédia, prix de l'innovation numérique 2019 », *Bulletin des bibliothèques de France* (BBF), 16 janvier 2020.

plan les contenus, clé de voûte de la médiation des savoirs à l'ère numérique selon Lionel Dujol<sup>110</sup>.

Le choix de créer Limédia Kiosque, site exclusivement dédié à la presse ancienne, en parallèle d'une autre plateforme patrimoniale plus généraliste (Limédia Galeries), s'appuyait au départ sur un constat : celui de la difficulté à éditorialiser les contenus de presse, pour lesquels l'information bibliographique s'avère assez pauvre du fait des problèmes d'indexation qu'ils posent. D'où le choix de proposer des modules éditoriaux portant à la fois sur l'histoire du titre, mais aussi sur des sujets issus de la presse. Toutefois, le site a dû récemment réorienter sa ligne éditoriale en basculant dans Galeries, plateforme plus grand public, ces contenus produits au départ pour Kiosque. En effet, les données d'usage ont révélé une consultation très différente des ressources patrimoniales des deux sites : tandis que l'utilisateur de Galeries explore les documents sur le mode du butinage, celui de Kiosque est davantage dans une démarche experte de recherche. Autrement dit, les focus thématiques proposés dans Kiosque – la grippe espagnole, les Jeux olympiques, le Tour de France dans la presse quotidienne régionale du siècle dernier, etc. – ne trouvent pas leur public au sein d'une plateforme utilisée en priorité pour la recherche. En revanche, ils sont bien plus à même de séduire l'utilisateur sensible aux images anciennes et documents anciens, qui n'est autre que le public cible de Galeries<sup>111</sup>.

En outre, l'éditionnalisation des contenus dans Limédia procède d'une logique de valorisation d'un territoire marqué par une histoire particulière. L'intérêt d'une telle plateforme régionale est de proposer un corpus de presse dense – près de 160 000 journaux et 56 titres – sur une période longue et particulièrement riche pour l'histoire de la Lorraine : 1760 à 1955, soit près de 200 ans. Faire de Limédia un projet commun aux différents établissements de conservation du Sillon lorrain prend tout son sens au regard de la spécificité de ce territoire : les collections sont complémentaires d'une ville à l'autre, du fait du déplacement des populations durant cette période. Ainsi, les titres en allemand conservés à la bibliothèque de Metz complètent-ils avantageusement les collections de la Bibliothèque Stanislas à Nancy. En matière d'éditionnalisation, mutualiser les ressources permet donc d'aborder des thématiques transversales telles que la presse d'annexion, par exemple.

### *Impliquer les habitants dans la redécouverte de leur patrimoine*

Ces dernières années, plusieurs projets menés en région autour de la presse ancienne se sont distingués par un réel effort de médiation en direction de leurs publics. La région Occitanie, notamment, est depuis longtemps impliquée dans la réhabilitation de ce patrimoine local. Si l'accent est surtout mis à l'heure actuelle sur l'accélération des numérisations et l'intégration progressive des données dans Gallica, le souhait de se consacrer à davantage d'opérations de valorisation comme celles menées par le passé est clairement exprimé par l'agence Occitanie Livre et Lecture<sup>112</sup>. À partir de fin 2013, l'ancienne région Languedoc-Roussillon avait en

---

<sup>110</sup> Entretien avec Lionel Dujol, responsable prospective, innovation et accompagnement au changement (Valence Romans Agglo, Direction Lecture Publique), le 16/11/2022.

<sup>111</sup> Entretien avec Cindy Hopfner, chargée de valorisation et médiation numérique des collections patrimoniales à la Bibliothèques de Nancy, le 22/11/2022.

<sup>112</sup> Entretien avec Mélanie Marchand, chargée de mission Bibliothèques & Patrimoine (Occitanie Livre & Lecture), le 27/10/2022.

effet porté un projet intitulé « Les Chroniques retrouvées du Midi », vaste opération grand public visant à valoriser les 600 000 pages mises en ligne. En complément des diverses actions de médiation menées dans les établissements partenaires (exposition itinérante, ateliers d'écriture en lycée ou en prison, etc.), le projet comportait un volet numérique : création d'un site dédié avec recréation de feuillets et de réclames inspirés de la presse occitane du XIX<sup>e</sup> siècle, une entrée « C'est arrivé près de chez vous » avec géolocalisation de faits divers locaux, une entrée par dates, ainsi qu'un jeu de questions-réponses amenant le public à se plonger dans la lecture d'articles d'époque<sup>113</sup>.

Parfois, c'est à la faveur du lancement d'une nouvelle plateforme que le public est impliqué. Le développement de L'Empreinte, portail patrimonial de la Drôme et de l'Ardèche, s'est ainsi appuyé sur une démarche UX invitant les habitants à partager leur propre vision du patrimoine local. Cette démarche de co-construction a abouti à un portail résolument tourné vers l'utilisateur : « Nous avons décidé que la démarche sur le portail se ferait à partir de l'habitant », explique à ce sujet Lionel Dujol<sup>114</sup>. Le lecteur est ainsi invité à saisir dans la barre de recherche le nom de son village, d'un lieu de sa jeunesse, ou encore à chercher dans les collections iconographiques « une image pour illustrer un souvenir »<sup>115</sup>. Si c'est avant tout une logique grand public qui a guidé le développement de la plateforme, celle-ci est aussi adaptée à un usage de recherche. Quoi qu'il en soit, que l'utilisateur pose sur les collections un œil d'expert ou de novice, le portail est conçu pour avant tout faire appel en lui à une dimension sensible : son rapport à la mémoire et au patrimoine de son territoire, son « goût de l'archive » selon les termes d'Arlette Farge<sup>116</sup>.

Les grandes commémorations sont également des occasions privilégiées pour mettre à l'honneur la presse ancienne tout en impliquant le public. L'exemple le plus connu est la « Grande Collecte », organisée à l'occasion du centenaire de la Première Guerre mondiale. Ce programme, coordonné par le Service interministériel des Archives de France (SIAF) et par la BnF, a été lancé en direction des particuliers possédant chez eux des documents liés à la Première Guerre mondiale et à leur histoire familiale : il leur était alors proposé de faire numériser leurs archives familiales dans les bibliothèques et centres d'archives afin de constituer, à partir de ces ensembles, une base de données relative au conflit. Faisant entrer en résonance ces documents personnels avec les collections plus académiques des bibliothèques, le projet de la Grande Collecte a su lier l'intime et le collectif, la petite et la grande histoire, autour d'enjeux mémoriels forts<sup>117</sup>. Le projet a donné lieu à la création, en marque blanche, d'une plateforme – aujourd'hui fermée – dédiée à l'événement.

---

<sup>113</sup> *Agir ensemble pour les patrimoines - Les Chroniques retrouvées du Midi : la presse ancienne préservée et révélée*, <https://www.dailymotion.com/video/x2x83uh>, 8 juillet 2015 (consulté le 27 octobre 2022). Le site n'existe plus aujourd'hui.

<sup>114</sup> Entretien avec Lionel Dujol, responsable prospective, innovation et accompagnement au changement (Valence Romans Agglo, Direction Lecture Publique), le 16/11/2022.

<sup>115</sup> *L'Empreinte*, <https://www.lempreinte.valenceromansagglo.fr/> (consulté le 14 février 2022).

<sup>116</sup> Arlette Farge, *Le goût de l'archive*, Paris, France, Éditions du Seuil, 1997.

<sup>117</sup> *Grande Collecte : élargir les sources de notre histoire*, <https://www.culture.gouv.fr/Actualites/Grande-Collecte-élargir-les-sources-de-notre-histoire> (consulté le 1 octobre 2022).

### *Opérations en lien avec l'EMI et l'EAC*

L'implication des habitants peut aussi se faire par le biais d'actions de médiation en lien avec l'Éducation artistique et culturelle (EAC) et l'Éducation aux médias et à l'information (EMI), axes forts de la valorisation de la presse auprès des scolaires. La presse ancienne constitue de fait un matériau très riche pour décortiquer le traitement de l'information (comment on s'informe à l'époque, comment l'événement est perçu et restitué, comment l'information transmise influence les mentalités) et par là même exercer son esprit critique, compétence essentielle pour une citoyenneté éclairée. Cet axe de valorisation est porté au niveau national par la BnF, qui coopère avec le CLEMI pour produire des expositions en ligne et sur panneaux, puis les diffuser auprès des établissements en proposant divers parcours : « Histoires de fausses nouvelles »<sup>118</sup>, « De la Gazette à Internet »<sup>119</sup>, etc. La décision d'ouvrir prochainement à Paris une Maison européenne du dessin de presse et du dessin satirique s'inscrit dans cette même logique.

En région aussi, des médiations de ce type sont proposées pour valoriser la presse locale. L'Exprimante, le distributeur de presse ancienne proposé par Lectura Plus (portail du patrimoine écrit et graphique en Auvergne-Rhône-Alpes), en est un exemple original. Partant sur la rématérialisation des 60 titres numérisés disponibles sur le portail, ce dispositif innovant vise à faire « (re)découvrir la presse ancienne parue entre 1807 et 1945 » dans la région<sup>120</sup>. Présenté comme un outil ludique d'action culturelle et d'éducation aux médias, il est proposé aux établissements scolaires, documentaires et culturels de la région, accompagné de fiches méthodologiques pour aider les enseignants et médiateurs à animer ces ateliers EMI.

Autre opération de rématérialisation d'archives de presse numérisées, la reconstitution grandeur nature du procès du parricide de Leyr, menée à la BM de Nancy en 2019, a été pensée à la fois comme une action culturelle grand public et une forme d'éducation aux médias et à l'information<sup>121</sup>. Faisant revivre un fait divers local par le biais de la presse, le projet a mis en lumière l'influence des médias sur l'opinion publique, amenant chacun des participants à s'immerger dans les débats de l'époque et à tisser des parallèles avec des chroniques judiciaires contemporaines. Désormais, le site Limédia Kiosque intègre pleinement la question de l'éducation aux médias dans sa nouvelle ligne éditoriale. Pensé comme une possible voie d'entrée pour les enseignants ou animateurs d'ateliers EMI, il devrait comporter, à terme, divers dossiers pédagogiques portant sur des thématiques généralistes (la caricature<sup>122</sup>, la représentation homme-femme dans la publicité ancienne, etc.) ou sur des événements historiques pour en mettre en

---

<sup>118</sup> *Histoires de fausses nouvelles*, exposition BnF / CLEMI en prêt en 2022-2023, <https://essentiels.bnf.fr/fr/enseignants/b64967cc-0065-4fe1-9490-9b1c7d51250b-fausses-nouvelles> (consulté le 15 février 2023).

<sup>119</sup> *De la Gazette à Internet*, exposition BnF / CLEMI en prêt en 2022-2023, <https://essentiels.bnf.fr/fr/societe/medias/b301c6f3-73f9-4882-81c8-5b81d897a876-gazette-internet> (consulté le 15 février 2023).

<sup>120</sup> *L'Exprimante, un distributeur de presse ancienne*, <https://www.lectura.plus/5486-lexprimantenbsp-un-distributeur-de-presse-ancienne.html> (consulté le 27 octobre 2022).

<sup>121</sup> *Limédia - Le procès du parricide de Leyr, ou comment parler de la presse ancienne*, <https://epitome.hypotheses.org/4288> (consulté le 28 septembre 2022).

<sup>122</sup> Ce dossier est déjà en ligne actuellement : *Rire, réagir, réfléchir*, <https://kiosque.limedia.fr/histoire-presse/rire-reagir-reflechir/> (consulté le 15 février 2023).

évidence le traitement médiatique local (l'affaire Dreyfus vue au prisme de journaux des deux bords, par exemple). La presse locale ancienne, de fait, se prête aisément à ce type de relecture contemporaine.

### 1.3.3 Un écosystème en pleine mutation : vers une mise en commun des ressources ?

À l'heure actuelle, le paysage de la presse ancienne numérisée en France évolue rapidement. D'une part, le contexte de la création du futur centre de conservation à Amiens donne une impulsion nouvelle aux projets menés en région et aux dynamiques de coopération avec la BnF. D'autre part, de nouveaux portails locaux voient régulièrement le jour, tandis que d'autres plateformes ferment, souvent en raison d'un coût de maintenance trop élevé ou de limites techniques (stockage insuffisant, moteur de recherche jugé trop peu performant, absence d'OCR sur les collections, etc.). La base régionale occitane, par exemple, ne sera pas maintenue une fois réalisée l'intégration de ses titres de presse dans Gallica<sup>123</sup>. D'autres plateformes se créent sous le régime de la marque blanche, qu'il s'agisse de bibliothèques de collectivités territoriales comme Deauville ou Roubaix (souvent labellisées « Bibliothèques numériques de référence »), ou d'autres types d'établissements comme le Musée du sport<sup>124</sup>.

Surtout, la mutualisation des ressources, à la fois documentaires et budgétaires, est encouragée à divers échelons. La fusion des régions a entraîné une reconfiguration territoriale, qui n'est pas sans conséquences sur la valorisation des fonds régionaux. Plus encore que par le passé, les logiques de mise en réseau sont soutenues par les régions : en témoigne par exemple l'élargissement du réseau Numistral, qui ne se limite plus désormais à la BNU, mais est devenu depuis quelques années le portail commun aux bibliothèques numériques de l'ensemble du site universitaire alsacien (BNU, Université de Strasbourg et Université de Haute-Alsace) et pourrait devenir un point d'accès unique au patrimoine numérisé alsacien<sup>125</sup>. Cette logique de mutualisation vaut également pour les services d'archives ; ainsi par exemple, les Archives départementales des Deux-Sèvres et de la Vienne ont-elles opté pour un portail commun<sup>126</sup>. Concernant les plans de subventionnement pour la numérisation de la presse, des expérimentations sont en cours pour mettre en place une mutualisation des aides de l'État. Une expérimentation est actuellement menée en ce sens dans les Hauts-de-France, avec un dossier de candidature soumis à un examen conjoint par la DRAC et la BnF : l'accord des deux administrations sur la validation des listes documentaires rendrait alors possible un cofinancement en vue de maximiser l'aide de l'État<sup>127</sup>.

---

<sup>123</sup> *Ressources Patrimoines en Occitanie*, <https://ressourcespatrimoines.laregion.fr/> (consulté le 27 octobre 2022).

<sup>124</sup> *Gallica marque blanche*, <https://www.bnf.fr/fr/gallica-marque-blanche> (consulté le 15 février 2023).

<sup>125</sup> Le réseau vient d'intégrer la bibliothèque municipale de Mulhouse et travaille avec la médiathèque de Strasbourg qui envisage, de son côté, de développer une bibliothèque en marque blanche interopérable avec Numistral. Désormais, puisque l'Alsace et la Lorraine font partie de la même région Grand-Est, il serait également possible d'envisager une coopération avec le Sillon lorrain, la région encourageant ce type de valorisation. La BNU, qui a numérisé plusieurs titres de presse mosellane qu'elle était la seule à conserver, pourrait ainsi venir compléter les collections de Limédia Kiosque. Source : Entretien avec Jérôme Schweitzer, directeur du pôle Partage, diffusion et réseaux à la Bibliothèque nationale et universitaire de Strasbourg, le 07/12/2022.

<sup>126</sup> *Archives départementales des Deux-Sèvres et de la Vienne*, <https://archives-deux-sevres-vienne.fr/> (consulté le 15 février 2023).

<sup>127</sup> L'objectif serait de pouvoir proposer aux partenaires le dépôt d'un dossier mutualisé (sur le modèle du PAPE, Plan d'Action pour le Patrimoine Écrit), prévoyant dans son plan de financement à la fois une aide PNV (Programme de Numérisation et de Valorisation des contenus culturels) et une aide BnF. *Journées du Patrimoine Écrit 2022 : Numériser*

Ces logiques de mutualisation ne se justifient pas seulement pour des raisons budgétaires. Il en va aussi d'un enjeu de lisibilité pour l'utilisateur, au sein du paysage complexe de la presse ancienne en France. Car en bout de chaîne, c'est toujours l'utilisation des ressources en ligne qui doit guider la pensée de la valorisation. Un glissement s'opère donc de la mise en ligne des collections à leur exploitation : ces deux faces sont en réalité inséparables. La numérisation ayant bouleversé l'accès aux collections, il est désormais nécessaire de réfléchir aux nouveaux enjeux autour des usages de la presse ancienne.

---

*les collections de presse*, intervention d'Arnaud Dhermy, chef de la mission de la coopération régionale à la BnF, 23-24 juin 2022.



## II. DE LA VALORISATION A L'EXPLOITATION DES CORPUS : DE NOUVEAUX BESOINS LIES A LA NUMERISATION DE LA PRESSE ANCIENNE

---

### 2.1. LE LECTEUR DE PRESSE ANCIENNE, USAGER AUX MILLE VISAGES

#### 2.1.1. L'exploitation des collections de presse, enjeu-clé de leur valorisation

Réfléchir à la valorisation en ligne des collections implique aussi de prendre en compte, en aval de cette médiation, l'usager auquel ces stratégies sont destinées. De fait, la valorisation ne se pense jamais en-dehors de l'utilisateur : elle est une triangulation entre l'établissement qui met à disposition ses fonds, la collection et le lecteur qui en fait usage<sup>128</sup>. La question de l'exploitation des corpus mis en ligne n'est donc pas séparable de la problématique de la valorisation. Pour ce qui concerne la presse, elle en constitue même l'un des enjeux majeurs. De fait, ce qui distingue la presse ancienne d'autres types de collections réservées en priorité à des usages experts, c'est la multiplicité des utilisations qui peuvent en être faites. La recherche dans ces collections n'est pas toujours académique, loin de là : les recherches généalogiques, notamment, concentrent une large part des usages. Dans les centres d'archives en particulier, la majeure partie des lecteurs en salle sont des généalogistes ou historiens amateurs : « Le lecteur vient dans un but de recherche historique (50%) ou généalogique (40%) », indique une enquête du SIAF de 2015 sur les publics des services d'archives<sup>129</sup>. Penser la valorisation à l'aune de cette variété de profils de chercheurs nous apparaît donc primordial. Au-delà de la mise à disposition des corpus de presse ancienne numérisée et des stratégies de médiation mises en œuvre, nous nous proposons d'étudier plus avant ce qui relève de leur exploitation pour la recherche – universitaire ou non.

Tout d'abord, il est important de considérer ce que le numérique fait au patrimoine : en rendant possible la consultation de documents à domicile, le numérique a fait reculer la salle de lecture<sup>130</sup>. Bien plus qu'une simple vitrine, le portail patrimonial est devenu le lieu de consultation privilégié pour de très nombreux chercheurs et généalogistes. Tout l'enjeu des bibliothèques numériques patrimoniales est donc de proposer une expérience en ligne qui s'approche au plus près de l'expérience de la salle de lecture, et même en améliore la qualité. Accès aisé et rapide à la ressource, possibilités de manipulation du document que n'offre pas l'archive physique (découpage, annotation, etc.), excellente qualité de visualisation, ou encore recherche plein texte à l'aide de l'OCR sont autant d'exigences de service attendues de la part des bibliothèques numériques. Celles-ci se doivent de réfléchir à la valeur ajoutée qu'elles apportent : mettre à disposition

---

<sup>128</sup> Entretien avec André-Pierre Syren, directeur de la valorisation à l'Enssib, le 31/05/2022.

<sup>129</sup> Brigitte Guigueno et Emmanuel Pénicaut, *Qui sont les publics des archives ? Enquêtes sur les lecteurs, les internautes et le public des activités culturelles dans les services publics d'archives (2013-2014)*, <https://francearchives.gouv.fr/fr/article/38196>, 2015, p. 7 (consulté le 15 février 2023).

<sup>130</sup> Voir le séminaire *Travailler en salle de lecture à l'ère numérique*, organisé par Frédéric Clavert et Caroline Muller à l'Université Rennes 2 le 16/03/2022, <https://gout-numerique.net/1109> (consulté le 4 avril 2022).

non pas simplement des documents, mais aussi et surtout des outils de qualité pour les exploiter. On sort ici du simple cadre de la « médiation numérique des savoirs »<sup>131</sup> pour toucher à une autre question : celle des services à la recherche qu'une offre patrimoniale en ligne est en mesure de fournir.

### 2.1.2. Diversité et porosité des usages

À cet égard, il convient de ne pas avoir une vision trop stéréotypée des usages. Si la prise en compte du public non expert est déterminante concernant la presse ancienne, la frontière entre l'usager expert et l'usager non expert est cependant bien plus poreuse qu'elle ne peut le sembler à première vue. D'une part, le degré d'expertise d'un lecteur est en lui-même difficile à appréhender : est-ce la connaissance approfondie d'un sujet ou la maîtrise des méthodes de recherche documentaire qui fait l'expertise ? Si l'on reconnaît aisément aux chercheurs de l'ESR un haut niveau d'expertise en raison de leur familiarité avec les outils de recherche, qu'en est-il de l'étudiant de master ou du doctorant en début de thèse, par exemple ? Le degré de maîtrise de ces outils varie fortement d'un usager à l'autre, d'autant que les traditions en matière de formation à la recherche documentaire ne sont pas forcément équivalentes d'une discipline à l'autre. De même, on peut être un chercheur expert dans un sujet et débutant dans un autre. Inversement, un passionné d'histoire et de généalogie pourra faire montre d'une grande familiarité avec les plateformes patrimoniales, bien que n'étant pas diplômé de l'enseignement supérieur.

D'autre part, comme l'observe Philippe Mezzasalma, la catégorie du « grand public » est relativement peu opérante pour la presse ancienne<sup>132</sup>. Les usages ne sont jamais figés, et un professionnel de la recherche peut très bien avoir un usage grand public de ces collections, en flânant par exemple dans les sélections éditorialisées du portail patrimonial de sa région. Le public enseignant constitue également un entre-deux intéressant : les enseignants du secondaire qui utilisent la presse ancienne dans le cadre de la préparation d'un cours mobilisent ces collections dans un cadre professionnel, mais ont tendance à privilégier des ressources vulgarisées, directement réutilisables avec des classes. Ce public constitue donc une cible privilégiée de contenus éditorialisés en fonction des thématiques des programmes scolaires<sup>133</sup>. En marge de ces usages, les contenus de presse peuvent également être consultés pour des raisons très concrètes, parfois anecdotiques : recherche de preuves d'un événement donné (décès, accident, naturalisation, etc.), dans le cadre de démarches administratives par exemple<sup>134</sup>. Ces requêtes ponctuelles sont particulièrement complexes à appréhender par les établissements, car elles n'entrent dans aucune catégorie d'usage bien définie.

Quant au public des généalogistes, il constitue un panel représentatif de la diversité des usages qui peuvent être faits de ces contenus. Dans les formations à la recherche qu'elle organise, la généalogiste professionnelle Sophie Boudarel

---

<sup>131</sup> L. Dujol et S. Mercier, *Médiation numérique des savoirs*, op. cit.

<sup>132</sup> Entretien avec Philippe Mezzasalma, chef du service Presse à la BnF, et Alexia Bauville, adjointe au chef du service Presse, le 07/11/2022.

<sup>133</sup> Entretien avec Irène Bastard, cheffe de projet Publics et usages à la BnF (Délégation à la stratégie et à la recherche), le 05/10/2022.

<sup>134</sup> Entretien avec Nathalie Fabry, chargée du traitement rétrospectif des collections de périodiques et responsable du site Presse locale ancienne, le 04/10/2022.

constate que ce public tend à s'élargir de manière croissante : si la discipline reste majoritairement pratiquée par un public de 50 ans et plus, on constate qu'elle intéresse de plus en plus un public jeune, parfois même étudiant. Il n'est pas rare, par exemple, de voir de jeunes internautes ouvrir des blogs de généalogie, ce qui est un phénomène nouveau. Or, ce public-là n'a pas forcément une approche identique à celle d'un public de 50 ans et plus. Cette catégorie d'âge reste toutefois prédominante, ce qui s'explique avant tout par la liberté qu'offre la retraite pour de telles activités. Hormis quelques exceptions, le public de généalogistes est un lectorat qui a pris le virage du numérique, car au fur et à mesure des numérisations et des mises en ligne, de très nombreuses sources ne sont plus consultables en version papier. Aussi la plupart des généalogistes mènent-ils désormais leurs recherches sur le net. La presse représente généralement un pan important de leurs recherches, en complément de l'état civil : en plus des actes de naissance, mariage et décès, la presse peut renseigner sur d'autres aspects de la vie de leur ancêtre (obtention d'un prix, d'une médaille, mention dans un fait divers, etc.)<sup>135</sup>.

Il convient donc, comme nous y invitent Valérie Beaudoin et Jérôme Denis dans une étude sur les usages de Gallica<sup>136</sup>, de « passer de la notion anthropologiquement et sociologiquement connotée d'"usager" à celle d'"usage". Autrement dit, identifier la cohérence interne des usages, sans présumer de celle de l'usager supposé rationnel et régulier dans ses engagements – un même usager pouvant avoir plusieurs usages, à des moments différents et pour des motivations différentes »<sup>137</sup>.

### 2.1.3. Le « sacre de l'amateur »<sup>138</sup> : vers un public amateur de plus en plus éclairé ?

Marquant un véritable tournant dans les usages, le numérique transforme également le profil des utilisateurs. Les récentes études de publics mettent en lumière un phénomène significatif ces dernières années : la présence accrue d'utilisateurs amateurs depuis la décennie 2010 en particulier, en raison de la numérisation des collections patrimoniales des bibliothèques. Cette augmentation est particulièrement frappante à la BnF : un rapport conjoint du SIAF et de la BnF publié en 2022 constate que « ces usagers intensifs, qui consultent des documents pour des recherches personnelles comme la généalogie ou l'histoire locale, deviennent massifs parmi les usagers de la bibliothèque numérique, alors qu'ils étaient restés très occasionnels en salle de lecture »<sup>139</sup>. De son côté, l'Observatoire des publics de la BnF note en 2020 une diversification des profils d'utilisateurs : si le public de la BnF se compose toujours en majorité de chercheurs académiques (35% des enquêtés), 33% disent consulter les ressources de l'établissement dans le cadre

---

<sup>135</sup> Entretien avec Sophie Boudarel, généalogiste professionnelle et créatrice du blog *La Gazette des Ancêtres*, le 11/10/2022.

<sup>136</sup> Valérie Beaudoin et Jérôme Denis, *Observer et évaluer les usages de Gallica. Réflexion épistémologique et stratégique*, rapport de recherche, BnF ; Telecom ParisTech, 2014.

<sup>137</sup> Philippe Chevallier, « Les données au service de la connaissance des usages en ligne : l'exemple de l'analyse des logs de Gallica », *Les Enjeux de l'information et de la communication*, 2018, vol. 19/2, n° 2, p. 57-67.

<sup>138</sup> Patrice Flichy, *Le sacre de l'amateur: sociologie des passions ordinaires à l'ère numérique*, Paris, France, La République des idées : Seuil, 2010.

<sup>139</sup> Noémie Couillard et Irène Bastard, *Qui sont les usagers en ligne communs entre les Archives et la BnF ?*, <https://francearchives.gouv.fr/fr/article/677845362>, septembre 2022, p. 9 (consulté le 16 février 2023).

de recherches en amateur<sup>140</sup>. Ces chercheurs amateurs constituent une proportion beaucoup plus importante que par le passé et apparaissent surreprésentés parmi les répondants en ligne, ce qui témoigne de l'impact direct du numérique sur la diversification des publics de l'établissement.

Si les amateurs étaient déjà des habitués des centres d'archives, ils constituent une relative nouveauté pour la BnF, et ce sur une assez grande diversité de sujets : beaucoup d'amateurs en histoire et en littérature notamment, ainsi qu'une proportion faible, mais néanmoins non négligeable, de généalogistes (6%). Contrairement aux chercheurs de l'ESR, qui travaillent souvent au croisement de plusieurs champs d'étude, les amateurs ont tendance à se concentrer sur les disciplines historiques de la BnF que sont la littérature et l'histoire. Ils sont également plus susceptibles de travailler sur des sujets singuliers qui ne s'inscrivent pas dans les tendances actuelles de la recherche académique. Au contraire, leurs sujets sont souvent très atomisés et diversifiés<sup>141</sup>. Le fait qu'ils ne disposent généralement pas, comme les chercheurs académiques, d'une formation documentaire, explique également leur approche différente des fonds, avec une spécialisation forte sur un sujet précis mais une connaissance plus superficielle des autres ressources documentaires à leur disposition<sup>142</sup>.

Parmi ces pratiques de recherche non-professionnelles, de récentes études ont également montré une forte implication de ces chercheurs dans des communautés d'« amateurs éclairés » en ligne. Comme l'a montré Patrice Flichy dans *Le sacre de l'amateur*<sup>143</sup>, Internet a mis en avant ce type de pratiques en créant – à l'aide des réseaux sociaux et des forums notamment – des communautés de passionnés sur des sujets variés. Les internautes peuvent ainsi échanger des méthodes et des conseils pour mener leur travail documentaire. Le forum Pages 14-18 sur la Grande Guerre, étudié dans le cadre du Labex *Les Passés dans le Présent*<sup>144</sup>, illustre bien ces usages sociaux du patrimoine en ligne. Valérie Beaudoin montre l'interconnexion très forte entre sites institutionnels et sites amateurs :

Ces sites et le forum sont fortement reliés entre eux, car la plupart des auteurs sont aussi des membres actifs du forum, ayant atteint un bon niveau d'expertise à la fois dans la recherche documentaire en général et dans leur domaine de recherche propre. Sur le forum, les participants échangent des informations et construisent collectivement de la connaissance ; sur les sites, ils publient leurs recherches personnelles nourries des échanges dans le forum<sup>145</sup>.

On observe donc aujourd'hui une réelle évolution du profil de l'amateur dans son rapport aux collections patrimoniales et notamment à la presse ancienne,

---

<sup>140</sup> Synthèse de l'Observatoire des publics 2020 (Une « photo de famille »), <https://www.bnf.fr/fr/lactualite-des-etudes> (consulté le 6 décembre 2022).

<sup>141</sup> Entretien avec Irène Bastard, cheffe de projet Publics et usages à la BnF (Délégation à la stratégie et à la recherche), le 05/10/2022.

<sup>142</sup> N. Couillard et I. Bastard, *Qui sont les usagers en ligne communs entre les Archives et la BnF ?*, op. cit., p.24.

<sup>143</sup> P. Flichy, *Le sacre de l'amateur*, op. cit.

<sup>144</sup> Valérie Beaudoin, « Le forum Pages 14-18 et sa communauté de pratique » dans Philippe Chevallier et Lionel Maurel (dir.), *Le web français de la Grande Guerre : Réseaux amateurs et institutionnels*, Nanterre, Presses universitaires de Paris Nanterre, 2022, p. 207-256.

<sup>145</sup> Valérie Beaudoin, « Les usages sociaux du patrimoine en ligne : l'exemple de la Grande Guerre » dans Étienne Anheim et al. (dir.), *Les patrimoines en recherche(s) d'avenir*, Nanterre, Presses universitaires de Paris Nanterre, 2022, p. 265-278.

ressource privilégiée des recherches en sciences humaines. Ce que met en lumière le rapport sur les usagers en ligne communs aux Archives et à la BnF, c'est que la figure de l'amateur ne peut en aucun cas être renvoyée dos à dos avec celle de l'expert, ou vue comme synonyme de dilettante. À l'heure du numérique, la frontière tendrait même à se réduire entre pratiques académiques et non-académiques. Ce qui caractérise ce public de chercheurs non-professionnels, c'est notamment une réelle intensité dans la pratique :

On [...] découvre des usagers en ligne à la fréquentation assidue des sites internet des archives et aux pratiques culturelles et savantes particulièrement développées. L'accès en ligne devient l'accès majoritaire pour de nombreux usagers et en particulier les généalogistes aux pratiques encore plus intensives que les étudiants et les chercheurs<sup>146</sup>.

De même, dans la synthèse de l'Observatoire des publics de la BnF, 37% des généalogistes affirment se connecter plusieurs fois par semaine, ce qui témoigne d'un usage plus assidu que chez les autres amateurs : « Cette intensité montre bien un investissement intermédiaire entre les pratiques des publics académiques [...] et celles des usagers dans une démarche de loisirs »<sup>147</sup>. Cette évolution de la composition des publics et de leurs pratiques constitue un vrai défi pour les bibliothèques numériques : en multipliant leurs publics cibles potentiels, elle ouvre aussi la voie à de nouveaux usages et besoins. L'objet de notre enquête est de mieux cerner ces derniers pour ce qui concerne la presse ancienne numérisée.

## **2.2. CE QUE LE NUMERIQUE FAIT AU JOURNAL : UN RAPPORT NOUVEAU A LA PRESSE ANCIENNE EN TANT QUE SOURCE**

### **2.2.1. Presse numérisée et « remédiation » : de nouvelles approches de recherche**

Pour la recherche académique aussi, le numérique représente un chamboulement. En préambule de son article « Lire la presse du XIXe siècle en contexte numérique : vers une nouvelle historiographie des objets médiatiques ? », Guillaume Pinson émet l'hypothèse selon laquelle « la numérisation et la lecture à l'écran ne sont pas des gestes anodins, mais qu'au contraire ils modifient profondément les regards, les manières de travailler, de collecter les données et d'établir des liens entre les corpus »<sup>148</sup>. Autrement dit, le numérique, support désormais privilégié pour l'étude de la presse ancienne, changerait en profondeur la manière dont le chercheur lit et s'approprie ces contenus. Si le geste de la consultation numérique peut paraître banal aujourd'hui, il n'en demeure pas moins une véritable révolution :

Fort de l'expérience de l'histoire de la lecture, qui montre que de tout temps les supports ont participé de l'élaboration du sens des objets symboliques, on sait que la « remédiation » de la presse papier par l'écran, la

---

<sup>146</sup> N. Couillard et I. Bastard, *Qui sont les usagers en ligne communs entre les Archives et la BnF ?*, op. cit., p.10.

<sup>147</sup> *Ibid.*, p.23.

<sup>148</sup> Guillaume Pinson, « Lire la presse du XIXe siècle en contexte numérique : vers une nouvelle historiographie des objets médiatiques? », *Dix-Neuf*, 2 octobre 2017, vol. 21, n° 4, p. 378-388.

## II. De la valorisation à l'exploitation des corpus : de nouveaux besoins liés à la numérisation de la presse ancienne

conversion de la page imprimée en données numériques, fait en sorte que l'on peut être à peu près assuré que nous ne lisons plus la presse ancienne comme nous le faisons il y a encore à peine cinq ou dix ans<sup>149</sup>.

Cette « grande conversion numérique », selon les termes de Milad Doueïhi<sup>150</sup>, n'est pas sans conséquences sur notre appréciation de la matière médiatique. Celle-ci, devenant un objet de valorisation patrimoniale par le truchement du numérique, s'érige d'une part en nouveau lieu de mémoire porteur d'un discours sur une société, une nation ou un territoire : « Le numérique réinvente la presse du passé, la fait exister une seconde fois »<sup>151</sup>. D'autre part, le numérique nous invite à nous intéresser à l'objet journal en lui-même : la structuration du quotidien, de l'hebdomadaire ou du mensuel, son rubriquage, les interactions qui s'opèrent entre les différentes zones de la page, deviennent en eux-mêmes une matière à étudier pour le chercheur. Les bibliothèques numériques, avec les options de visualisation qu'elles proposent (feuilletage du fascicule, zoom dans le texte et ses illustrations, recherche plein texte grâce à l'OCR, parfois segmentation de la page) offrent un nouveau regard sur la matérialité de l'objet.

En outre, le journal peut s'étudier en tant qu'objet textuel et littéraire : l'étude de la presse mobilise en effet de très nombreux chercheurs en littérature ou histoire des idées, qu'il s'agisse des gazettes de l'Ancien Régime ou de la Révolution<sup>152</sup>, des journaux dans lesquels les écrivains publiaient en feuilleton ou contribuaient en tant que rédacteurs. L'étude des poétiques journalistiques a pris une grande ampleur dans le champ de la recherche ces dernières années, et cela n'est bien sûr pas étranger à la mise en ligne de milliers de corpus numérisés, désormais accessibles en quelques clics et constamment étoffés. Diverses équipes en France se sont spécialisées dans ce champ de recherche, notamment autour de projets internationaux tels que Médias 19<sup>153</sup> ou Numapresse<sup>154</sup>.

On observe de surcroît que les politiques de numérisation menées par les établissements ont un impact direct sur les thématiques de recherche : certains sujets sont conditionnés par ce qui est accessible en ligne<sup>155</sup>. Inversement, l'accès facilité à certains corpus au détriment d'autres, non encore numérisés, engendre un certain nombre de points aveugles dans la recherche. Cela peut donner lieu à des effets de mode, mais aussi à des biais si l'on cède à l'illusion du « tout numérique » : « Le numérique opère autant de nouvelles liaisons que des cloisonnements qui ne sont pas toujours interrogés », note à cet égard Guillaume Pinson<sup>156</sup>.

---

<sup>149</sup> *Ibid.*

<sup>150</sup> Milad Doueïhi, *La grande conversion numérique*, trad. Paul Chelma, Paris, France, Éditions du Seuil, 2008.

<sup>151</sup> G. Pinson, « Lire la presse du XIXe siècle en contexte numérique », art. cit., p. 2.

<sup>152</sup> Voir *Le Gazetteur universel et Le Gazetteur révolutionnaire*, les deux répertoires créés par Denis Reynaud en 2009 et 2015 pour la presse du XVIIIe siècle : *Le Gazetteur universel*, <https://gazetier-universel.gazettes18e.fr/> (consulté le 16 février 2023) ; *Le Gazetteur révolutionnaire*, <https://gazetier-revolutionnaire.gazettes18e.fr/> (consulté le 16 février 2023).

<sup>153</sup> *Médias 19 - Littérature et culture médiatique*, <https://www.medias19.org/> (consulté le 16 février 2023).

<sup>154</sup> *Numapresse - Du papier à l'écran. Mutations culturelles, transferts génériques, poétiques médiatiques de la presse française*, <http://www.numapresse.org/>, 9 février 2023 (consulté le 16 février 2023).

<sup>155</sup> Voir Gabriela Ossenbach, « If it's not online, it doesn't exist », *IJHE*, n° 5, 2015, p. 80-82. Article cité par Frédéric Clavert, « Une histoire par les données ? Le futur très proche de l'histoire des relations internationales », *Bulletin de l'Institut Pierre Renouvin*, 2016, vol. 44, n° 2, p. 119-130.

<sup>156</sup> G. Pinson, « Lire la presse du XIXe siècle en contexte numérique », art. cit., p. 6.

## 2.2.2. « Feuilletter la presse par Giga Octets »<sup>157</sup> : les enjeux de la fouille et de la visualisation de données

### *L'émergence de nouveaux besoins de recherche : text-mining, datamining, datavisualization*

Au-delà de ces nouvelles approches dans la manière de considérer le journal comme objet d'étude, le numérique fait surtout apparaître de nouveaux besoins à l'égard de l'exploitation des données. Le numérique bouleverse en effet le statut de la presse comme ressource pour l'historien : de source primaire impliquant un dépouillement manuel et un traitement humain de l'information, celle-ci bascule vers le statut de données. Frédéric Clavert décrit ainsi ces nouveaux processus de « mise en données » (*datafication*) de l'histoire :

Appliquée aux usages historiens, la mise en données est un processus contenant toutes les étapes allant de la numérisation d'un artefact physique – les archives numérisées – [...] aux possibilités [d'analyse] via des outils informatiques, c'est-à-dire une lecture des sources au moyen d'une médiation informatique. La mise en données du monde, telle qu'elle est ici définie, se traduit par l'émergence de grands ensembles de données à disposition des chercheurs, dont Google Books ou Gallica<sup>158</sup>.

La presse ancienne devient alors exploitable par les outils de fouille massive et de visualisation de données, ce qui ouvre de nouvelles pistes pour la recherche : étude des logiques de circulation médiatique et de diffusion de l'information à une très large échelle (plusieurs centaines de titres de presse sur plusieurs décennies), stratégies de comparaison entre différents pays ou périodes historiques, analyses statistiques et visualisation graphique, etc. On peut en prendre pour exemple la thèse pionnière de Pierre-Carl Langlais sur la formation de la chronique boursière dans la presse quotidienne française des années 1830 au Second Empire<sup>159</sup>, ou encore le projet ISOCO de l'Université Paris Nanterre sur la production de séries statistiques à partir des données textuelles de RetroNews<sup>160</sup>. Ce changement d'échelle – à la fois temporelle et spatiale – bouleverse la manière de faire de la recherche : les résultats, mais aussi les questions de recherche elles-mêmes s'en trouvent transformés. Divers sujets classiques des études de presse, comme le rôle de l'illustration dans le journal, peuvent désormais s'étudier de manière tout à fait inédite, à l'aide d'outils de visualisation mettant en évidence les micro-faits comme les macro-tendances<sup>161</sup>.

C'est là un tournant majeur dans les études sur la presse, dont les bibliothèques prennent désormais la mesure. La BnF est aujourd'hui impliquée

---

<sup>157</sup> Claire-Lise Gaillard, « Feuilletter la presse ancienne par Giga Octets », dans Frédéric Clavert et Caroline Muller (dir.), *Le goût de l'archive à l'ère numérique*, <https://gout-numerique.net/table-of-contents/archives-numerisees-la-salle-de-lecture-virtuelle-et-physique/feuilletter-la-presse-ancienne-par-giga-octets> (consulté le 18 septembre 2022).

<sup>158</sup> F. Clavert, « Une histoire par les données ? », art. cit., p. 120.

<sup>159</sup> Ce travail a été l'occasion, pour l'auteur, de développer l'application Pyllica, un outil de récupération automatisée de données sur Gallica (extracteur Python de corpus de périodiques). Cf. Pierre-Carl Langlais, *La formation de la chronique boursière dans la presse quotidienne française (1801-1870) : métamorphoses textuelles d'un journalisme de données*, Thèse de doctorat, Université Paris-Sorbonne, 1970-2017, France, 2015.

<sup>160</sup> *Congrès Médias 19 - Numapresse : Presses anciennes et modernes à l'ère du numérique*, BnF, 2-3 juin 2022, intervention de Philipp Leu, responsable des opérations auprès de BnF-Partenariats.

<sup>161</sup> Jean-Philippe Moreux, « Approches innovantes pour la presse ancienne numérisée : fouille et visualisation de données », <https://bnf.hypotheses.org/208> (consulté le 30 septembre 2022).

dans de nombreux projets de recherche et de R&D, tels qu'*Europeana Newspapers*<sup>162</sup> et *NewsEye*<sup>163</sup>. En témoigne l'organisation, en juin 2022, d'une journée consacrée à la présentation du travail de la Bibliothèque nationale sur l'amélioration des fonctionnalités permettant d'entrer dans la granularité des documents : optimisation de l'OCR, reconnaissance de la mise en page, segmentation des articles, ou encore description fine des contenus pour catégoriser et classer les différents éléments d'une page de presse<sup>164</sup>. Le développement de l'OLR (*Optical layout recognition*) constitue désormais un axe important pour favoriser la recherche dans ces corpus hétérogènes. Le traitement OLR « consiste en la description de la structure logique de chaque fascicule et de ses articles (emprise spatiale, titre et sous-titre, etc.) [...] et en la classification des types de contenus (information, feuilleton littéraire, programme des spectacles, publicités, etc.) »<sup>165</sup>.

En 2016, Jean-Philippe Moreux, expert scientifique de Gallica à la BnF, témoignait du résultat d'une telle expérimentation appliquée à un corpus numérisé dans le cadre du grand projet européen *Europeana Newspapers* (2012-2015). L'analyse d'un ensemble de six titres à l'aide de techniques d'extraction et de visualisation de données annonçait alors des perspectives prometteuses pour l'exploitation de la presse ancienne, à l'heure de la massification des corpus numérisés. La BnF a aussi eu l'occasion d'expérimenter, sur sa plateforme RetroNews, un certain nombre de fonctionnalités appliquées aux collections de presse<sup>166</sup>. Des outils de fouille sur les collections de Gallica ont également été développés, comme l'outil de lexicométrie Gallicagram pour visualiser l'évolution de l'usage des mots au cours du temps. L'approche lexicométrique des fonds de presse permet d'approcher au plus près les événements, les mentalités et les opinions, ces corpus étant « par nature sensibles aux soubresauts de l'actualité »<sup>167</sup> au fil de l'histoire.

### ***La démocratisation de ces usages experts à l'heure de la massification des données***

Ces besoins de recherche ne sont pas totalement nouveaux mais s'observent déjà depuis une dizaine d'années dans les communautés de chercheurs spécialistes des humanités numériques. Toutefois, la grande évolution de ces dernières années est la démocratisation de ces usages naguère réservés à des communautés expertes. Ceux-ci se sont à la fois renforcés et diversifiés : aujourd'hui, de plus en plus de

---

<sup>162</sup> *Europeana Newspapers - A Gateway to European Newspapers Online*, <http://www.europeana-newspapers.eu/> (consulté le 12 novembre 2022).

<sup>163</sup> *NewsEye*, <https://www.newseye.eu/> (consulté le 5 décembre 2022).

<sup>164</sup> *Congrès Médias 19 - Numapresse : Presses anciennes et modernes à l'ère du numérique*, BnF, 2-3 juin 2022, intervention d'Arnaud Laborderie, chef de projet au service de la Coopération numérique et de Gallica.

<sup>165</sup> J.-P. Moreux, « Approches innovantes pour la presse ancienne numérisée », art. cit.

<sup>166</sup> Différents tests ont été menés sur l'amélioration de l'OCR, l'enrichissement sémantique des collections de RetroNews ou encore l'identification de thématiques et d'événements historiques. L'enrichissement se fait à l'aide du VIAF (*Virtual International Authority File*) pour les organisations et les personnes, et de Geonames pour la géolocalisation. L'identification de thématiques et d'événements historiques se fait à partir des rubriques IPTC-Media topics et d'une base de données de 3000 événements historiques. Source : *Congrès Médias 19 - Numapresse : Presses anciennes et modernes à l'ère du numérique*, BnF, 2-3 juin 2022, intervention de Philipp Leu, responsable des opérations auprès de BnF-Partenariats.

<sup>167</sup> Selon les termes des deux créateurs de l'outil, Benjamin Azoulay (Université Paris-Saclay) et Benoît de Courson (ENS-PSL). Cf. *Utiliser les API de Gallica : l'exemple de Gallicagram*, <https://api.bnf.fr/fr/gallicagram-un-util-de-lexicographie> (consulté le 5 décembre 2022).

chercheurs s'intéressent à ces collections de presse, qui demeurent par ailleurs toujours très sollicitées par les historiens amateurs – généalogistes en particulier – et le grand public. On constate par exemple que le *Journal officiel de la République française* fait l'objet de fréquentes fouilles dans le cadre de recherches de patronymes, questionnements sur les décrets de naturalisation, etc.<sup>168</sup>. L'amélioration des capacités de recherche dans les collections numérisées augmente mathématiquement le nombre de requêtes de ce type, le plus souvent pour des usages scientifiques, mais pas seulement. Parallèlement, l'ouverture des données de très nombreuses institutions, couplée à la puissance de calcul permise par les techniques d'intelligence artificielle, démocratise aussi ce type de recherche : il est désormais bien plus aisé pour un chercheur, une équipe de recherche ou un citoyen de travailler sur des quantités massives de données à l'extérieur des portails des bibliothèques.

Or, l'accroissement et la démocratisation de ces usages impliquent aussi des attentes supérieures à ce qu'elles pouvaient être à la création des bibliothèques numériques. À l'époque, l'OCR était conçu pour faciliter l'accès aux textes, et non pour faire de la fouille de texte. Mais désormais, un OCR simple ne suffit plus toujours à satisfaire les exigences croissantes des chercheurs. D'année en année, les programmes de numérisation ont amené à produire des corpus très vastes, parfois à l'échelle européenne voire mondiale. Or plus la collection est vaste, plus la recherche est complexe et rend nécessaire le recours à des outils de fouille de données massive. Jean-Philippe Moreux résume en ces termes l'évolution parallèle des outils de recherche et des usages vers une sophistication croissante :

L'histoire des méthodes de recherche sur des contenus documentaires ou archivistiques est en partie déterminée par une autre histoire, celle de la numérisation de l'information [...]. Les modalités d'interaction des chercheurs avec les corpus numériques ont évolué simultanément, du feuilletage à la fouille de données, sans oublier l'étape de la recherche en texte intégral. En quelques décennies, l'œil du chercheur se sera posé sur microfilms, images numériques, listes de résultats de recherche plein texte, pour finalement s'éloigner des documents, en déléguant à des algorithmes (analyse d'images et de documents, analyse statistique, traitement automatique de la langue, modélisation des textes, visualisation de données) la tâche de nourrir une nouvelle pratique, la lecture distante<sup>169</sup>.

Bien entendu, la « lecture distante », telle que théorisée par Franco Moretti<sup>170</sup>, n'annule pas les pratiques traditionnelles et le travail sur des textes singuliers (« lecture rapprochée »). Or, parallèlement à la démultiplication des possibles usages de la presse ancienne, les besoins se complexifient. Il est donc nécessaire d'appréhender plus finement ces besoins, afin d'adapter au mieux l'offre de services proposée par les bibliothèques numériques.

---

<sup>168</sup> Entretien avec Jean-Philippe Moreux, expert scientifique de Gallica à la BnF, le 14/11/2022.

<sup>169</sup> J.-P. Moreux, « Approches innovantes pour la presse ancienne numérisée », art. cit.

<sup>170</sup> Franco Moretti, *Graphes, cartes et arbres : modèles abstraits pour une autre histoire de la littérature*, Paris, Les Prairies ordinaires, 2008.

## 2.3. ÉTUDE DE PUBLICS : MIEUX CONNAITRE LES USAGES ET LES BESOINS DE DEUX CATEGORIES DE LECTEURS

### 2.3.1. Cadrage de l'étude

Nous avons adressé notre questionnaire d'enquête à deux panels de lecteurs : chercheurs de l'ESR d'une part, généalogistes et amateurs d'histoire d'autre part. Ces deux catégories ont été ciblées pour leur utilisation relativement proche de la presse ancienne : l'exploitation de ces corpus numérisés dans un but de recherche – dans un contexte professionnel ou non-professionnel. Si les établissements opèrent généralement une distinction nette entre leurs publics professionnels et amateurs, nous avons jugé intéressant, pour notre problématique d'enquête, de nous poser les questions suivantes : au regard de la diversité et de la porosité des usages de la presse ancienne numérisée, cette distinction entre chercheurs amateurs et chercheurs professionnels est-elle véritablement opérante pour ces collections aujourd'hui ? Le changement du geste de consultation à l'ère numérique, en bouleversant les pratiques, amène-t-il celles-ci à se rejoindre, ou constate-t-on toujours deux usages très différenciés de ces collections ? En d'autres termes, observe-t-on une différence nette dans les besoins exprimés par ces deux communautés de chercheurs, ou plutôt un *continuum* de pratiques ?

Hormis la première section de l'enquête (« Profil général »), les deux questionnaires étaient identiques en tout point. Ils ont recueilli un nombre de réponses satisfaisant : 241 réponses du côté des généalogistes et amateurs d'histoire, 147 réponses du côté des chercheurs de l'ESR, avec des répondants aux profils variés dans les deux cas. Si ces échantillons nous semblent suffisamment parlants pour en tirer des conclusions, les interprétations que nous proposons ne se fondent toutefois que sur les réponses obtenues auprès des personnes sondées. L'étude n'a donc pas vocation à couvrir l'ensemble des usages et des besoins de ces deux catégories d'utilisateurs. Pour les deux questionnaires, nous avons fait le choix d'une diffusion ciblée : pour celui destiné aux universitaires, nous nous sommes concentrés sur les laboratoires et unités de recherche spécialisés dans l'étude de la presse et des médias et/ou travaillant en lien avec les humanités numériques. Pour celui destiné aux généalogistes et amateurs d'histoire, nous sommes passés par des réseaux de généalogistes et de sociétés savantes, et par les listes de diffusion Bibliopat et AAF.

Nous proposons ici une synthèse des résultats dépouillés et analysés. Certaines questions ont été recodées après dépouillement pour faciliter la lecture et la compréhension des résultats de l'enquête. Dans notre méthode de calcul, nous avons choisi d'inclure les non-réponses (indiquées « NR ») comme une modalité de réponse à part entière. Les données qualitatives (verbatim des réponses) pourront être ponctuellement enrichies par des références aux entretiens menés avec plusieurs répondants à l'issue du sondage. Le détail de la méthodologie d'élaboration du questionnaire, ainsi que l'ensemble des questions et des réponses, se trouvent en annexe<sup>171</sup>.

---

<sup>171</sup> Cf. Annexe n°1 : « Détail de la méthodologie d'enquête » et Annexe n°3 : « Détail des réponses ».

### 2.3.2. Synthèse des résultats

#### *Profils des répondants*

En termes de genres, le panel de chercheurs de l'ESR interrogé (147 répondants) se répartit de manière très équilibrée : 50% d'hommes pour 48% de femmes (les 2% restants préfèrent ne pas répondre). L'origine géographique des répondants est relativement variée : si 38% habitent à Paris ou en Ile-de-France, 55% résident hors région parisienne. Des chercheurs étrangers ont également participé à l'enquête : 4% ont indiqué vivre à l'étranger et 5% ont déclaré un établissement de rattachement situé hors de France. On note aussi une réelle diversité des structures de rattachement : principalement des universités, laboratoires et UMR, mais également plusieurs grandes écoles et instituts de recherche, ainsi que d'autres types d'établissements (INSPE, école d'ingénieur, école vétérinaire, etc.). Les différents types de personnels de l'enseignement supérieur sont représentés dans ce sondage (fig. 1), avec une majorité de maîtres de conférences ou professeurs d'université (37%), mais aussi, entre autres, une part importante de jeunes chercheurs (27% de doctorants et/ou ATER).

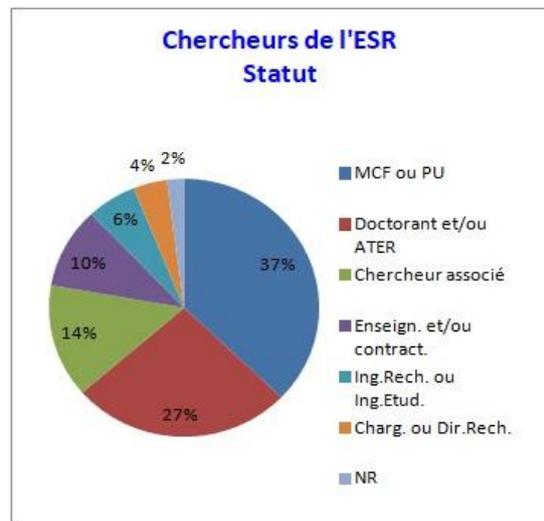


Figure 1 - Statut des chercheurs de l'ESR

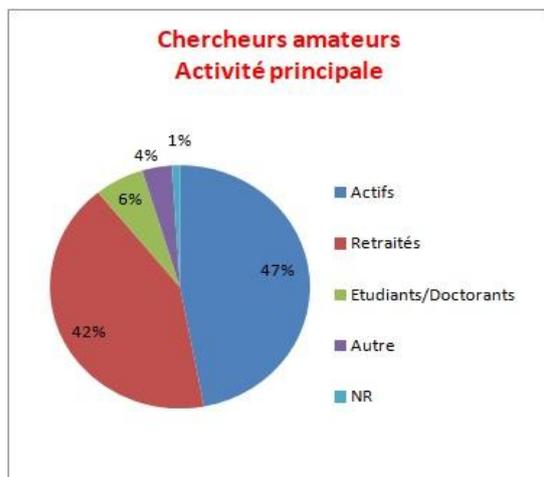


Figure 2 - Activité des chercheurs amateurs

Du côté des généalogistes et amateurs d'histoire (241 répondants), on note qu'une majorité de femmes a répondu à l'enquête : 60% de femmes, contre 38% d'hommes<sup>172</sup>. Contrairement à ce à quoi l'on pourrait s'attendre, ce panel ne comprend pas une majorité de retraités mais une légère majorité d'actifs : 47% (fig. 2). Quinze étudiants ont également pris part à l'enquête, ce qui corrobore le constat de Sophie Boudarel selon lequel la généalogie ou les recherches en amateur tendent à s'ouvrir à des publics plus jeunes que par le passé.

Les principales disciplines d'étude représentées sont les Lettres et Sciences humaines (47%), Sciences politiques et droit (20%), Histoire du livre et archivistique (13%).

<sup>172</sup> Notons cette différence par rapport l'enquête conjointe du SIAF et de la BnF, où le panel comportait à l'inverse une majorité d'hommes.

## II. De la valorisation à l'exploitation des corpus : de nouveaux besoins liés à la numérisation de la presse ancienne

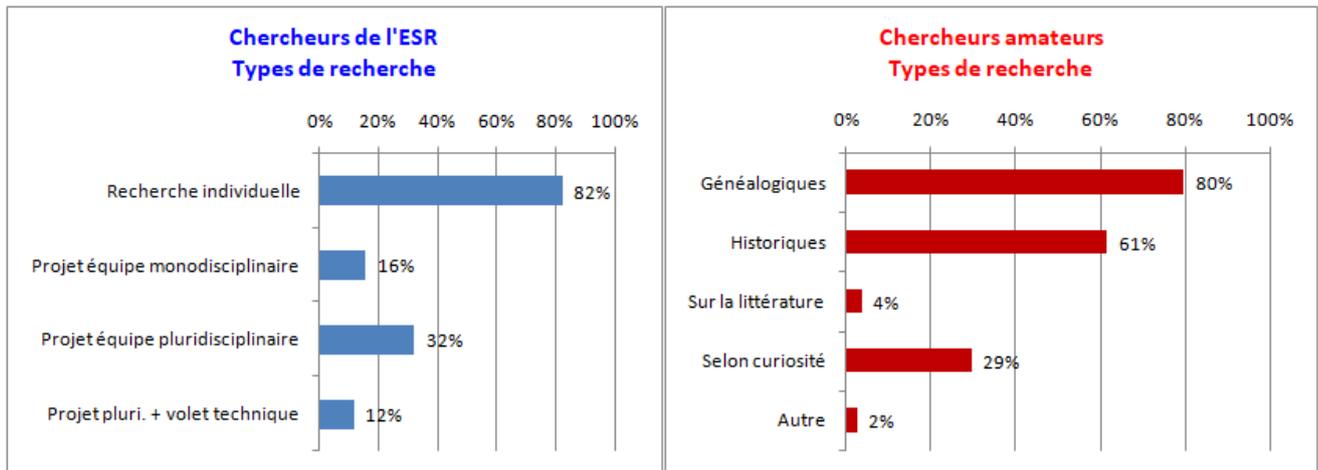
En majorité, ces répondants ont déclaré ne pas avoir une activité ou ancienne activité en lien avec la recherche et/ou l'histoire. Si le panel interrogé est donc bien représentatif d'une pratique de recherche menée en amateur, les professionnels de la recherche ne sont pas pour autant absents de ce sondage. Parmi ceux qui déclarent une activité – actuelle ou ancienne – en lien avec la recherche (19% des enquêtés) et/ou l'histoire (30% des enquêtés), on compte des professionnels des archives ou des bibliothèques, quelques généalogistes professionnels, historiens et chargés de recherches historiques locales, quelques enseignants ou ex-enseignants d'histoire (à l'université et dans le secondaire), ainsi que d'autres professions en lien avec le marché de l'art, la conservation ou la médiation du patrimoine. Actifs et retraités confondus, on compte 30 professionnels des archives et bibliothèques et 9 généalogistes professionnels. Cela s'explique en partie par les canaux de diffusion que nous avons privilégiés : listes de discussion de l'AAF, de Bibliopat et réseaux de généalogistes professionnels. On dénombre donc quelques profils experts dans le second panel interrogé, bien que ceux-ci demeurent largement minoritaires.

### *Types de recherche*

Quand on compare les types de recherche des deux panels (fig. 3 et 4), on constate bien évidemment des différences. Si une grande majorité de chercheurs de l'ESR (82%) déclare mener actuellement des recherches individuelles (poursuite d'un doctorat ou d'une HDR, rédaction d'ouvrages et d'articles, publication d'actes de colloques, parfois contribution à RetroNews, etc.), ils sont également nombreux à mener des projets en équipe – par ordre d'importance : projets menés au sein d'une équipe pluridisciplinaire (32%), projets menés au sein d'une équipe monodisciplinaire (16%) et projets disciplinaires impliquant un ou des ingénieurs de recherche (12%). Plusieurs chercheurs déclarent ainsi s'inscrire dans le cadre de projets ANR mono et pluridisciplinaires (l'ANR Numapresse en particulier) ou participer à des réseaux internationaux. Certains mentionnent également des projets numériques, comme la création d'une base de données.

Chez les chercheurs amateurs, on observe – sans surprise – une nette prédominance de recherches généalogiques (80% des enquêtés). Viennent ensuite les recherches historiques, chez 61% des enquêtés. Les recherches en littérature sont nettement plus rares (4%), alors que l'approche littéraire des fonds de presse est beaucoup plus répandue chez les chercheurs professionnels – nous nous appuyons ici sur le détail de leurs sujets de recherche indiqué en réponse libre du questionnaire, ainsi que sur des entretiens qualitatifs menés avec plusieurs enseignants-chercheurs et doctorants.

## II. De la valorisation à l'exploitation des corpus : de nouveaux besoins liés à la numérisation de la presse ancienne



Figures 3 et 4 - Types de recherche

On observe en outre que près de la moitié des chercheurs amateurs (46%) croisent recherches généalogiques et historiques<sup>173</sup>. La porosité entre les deux disciplines est forte, comme en témoignent un grand nombre de thématiques de recherche listées par les répondants, dont nous sélectionnons ici quelques exemples :

- « Recherches généalogiques et historiques sur la vie personnelle et professionnelle d'un peintre niçois »
- « Un peu de généalogie personnelle et sur l'histoire d'Orsay »
- « Contexte historique autour de certains ancêtres »

La recherche généalogique mène souvent à des recherches historiques plus poussées et vice-versa, comme l'illustrent ces différents témoignages :

- « Dans le cadre de mes recherches historiques, il m'arrive fréquemment de devoir faire des recherches généalogiques. »
- « Ce sont au départ des recherches sur ma famille qui m'ont conduite à m'intéresser à l'un de mes ancêtres, savant et astronome du XIXe siècle. Par la suite, je me suis prise de passion pour son travail scientifique et ai décidé de consacrer un ouvrage à sa mémoire, sous la forme d'une biographie romancée. »
- « Actuellement, je participe au #ChallengeAZ<sup>174</sup> et mes articles sont issus de *La Dépêche de Toulouse* (site Rosalis). J'ai aussi dépouillé *L'Ariégeois* et *L'Ariégeois magazine* [...]. La presse nous informe sur les conditions de vie de nos ancêtres : mercuriales, état des chemins, catastrophes naturelles, politique etc. C'est pourquoi outre les recherches classiques en généalogie, je fais un important usage de la presse. »

De manière générale, les sujets indiqués par les amateurs ont une coloration souvent locale : « Histoire de la Gascogne », « Manufactures de lin et autres à Nantes et en Normandie », « Histoire de la presse en Nord Pas-de-Calais et

<sup>173</sup> Cf. Annexe n°3 : « Détail des réponses ».

<sup>174</sup> Défi mêlant généalogie et écriture initié par Sophie Boudarel, généalogiste et créatrice du blog *La Gazette des Ancêtres*.

commerce des imprimés à Roubaix », « Sucrerie de Saint-Erme », etc. Il n'est cependant pas rare de trouver des thématiques de recherche beaucoup plus larges telles que « Femmes dans l'art », « Armée de l'air dans la presse », « Histoire de l'Algérie en période française », « Histoire de l'électronique », etc. Autrement dit, il n'y pas toujours une différence de nature entre les sujets de prédilection des chercheurs amateurs et ceux des chercheurs de l'enseignement supérieur. Plus que sur les thématiques de recherche, qui peuvent se rejoindre, c'est plutôt sur le type de projet de recherche et le cadre dans lequel il s'inscrit que se fait la différence.

### *Usages de la presse ancienne numérisée*

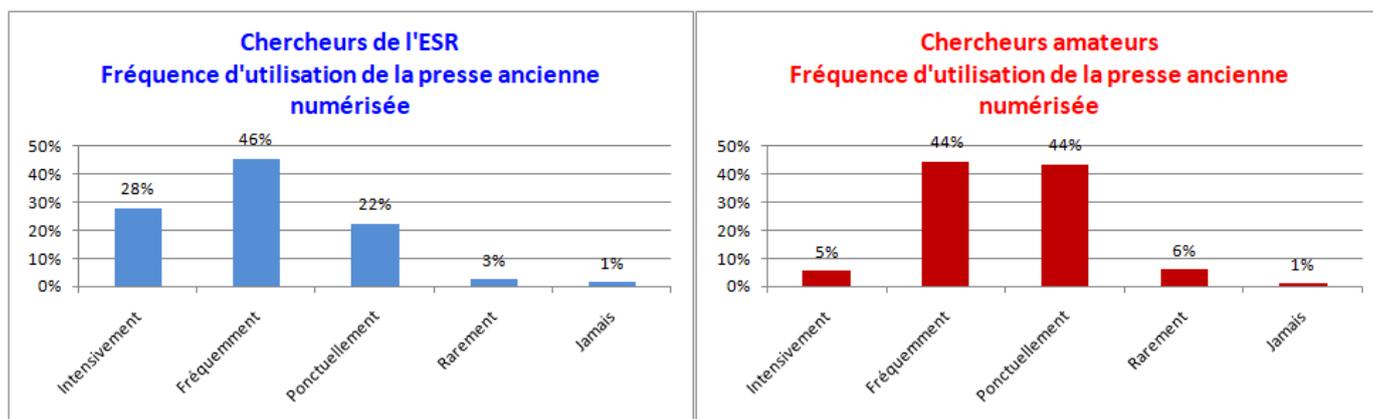
En matière de fréquence de consultation (fig. 5 et 6), les chercheurs professionnels se distinguent par une plus grande intensité de consultation que le public amateur. Tandis que pour les généalogistes et amateurs d'histoire, la presse numérisée constitue majoritairement – à part égale – une ressource utilisée de manière fréquente ou ponctuelle, la plupart des universitaires interrogés disent en avoir une utilisation fréquente ou intensive. Si la majeure partie du public amateur désigne la presse comme une ressource importante<sup>175</sup> (fig. 8), rares sont ceux qui l'utilisent comme corpus d'étude principal (4%). Chez les chercheurs de l'ESR (fig. 7), la presse est plus fréquemment désignée comme corpus principal (14%). C'est là une probable conséquence de la numérisation qui, en mettant à disposition des corpus massifs, généralise les études ayant comme source prioritaire la presse ancienne. Il peut aussi s'agir d'un biais de l'enquête, diffusée notamment auprès de spécialistes des humanités numériques.

On observe néanmoins que les deux panels considèrent en majorité la presse comme une ressource importante, et secondairement comme une ressource parmi d'autres non prioritaire – et ce, dans des proportions très similaires. Quelques universitaires affirment encore privilégier la consultation de collections physiques à celle de collections numérisées. Si cela traduit l'attachement de quelques-uns au papier, cela tient aussi à certains sujets de recherche portant sur l'objet journal. Une répondante souligne ainsi que « Même si la numérisation est très pratique quand on habite loin de Paris, elle ne peut pas tout et ne se substitue pas entièrement à l'étude du support (ex : contraintes liées au format) ». Cette préférence pour l'objet papier ne transparait quasiment pas dans les réponses des amateurs. Contrairement à certaines idées reçues, le public des généalogistes amateurs est donc un public qui a véritablement pris le virage du numérique. Là encore, nous avons conscience qu'il s'agit d'un biais naturel de l'enquête, qui n'a été diffusée que via des canaux numériques. Nous assumons toutefois ce biais, dans la mesure où notre étude s'intéresse aux usages des collections numérisées.

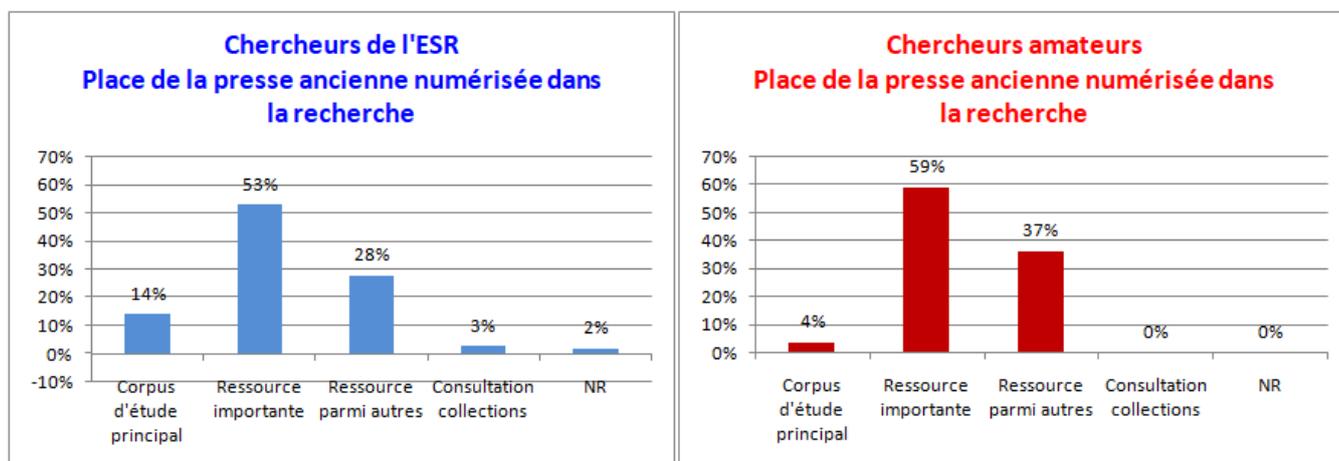
---

<sup>175</sup> Cela corrobore l'avis des généalogistes professionnels, qui présentent la presse comme une ressource importante pour la recherche généalogique. Voir par exemple les formations en ligne proposées par Sophie Boudarel, qui donnent des clefs pour rechercher dans les collections de presse numérisées, au même titre que dans les actes administratifs ou les cartes par exemple. Voir aussi Tony Neulat, *Gallica et RetroNews: deux eldorados généalogiques*, Paris, France, Archives & Culture, 2021.

## II. De la valorisation à l'exploitation des corpus : de nouveaux besoins liés à la numérisation de la presse ancienne



Figures 5 et 6 - Fréquence d'utilisation de la presse ancienne numérisée



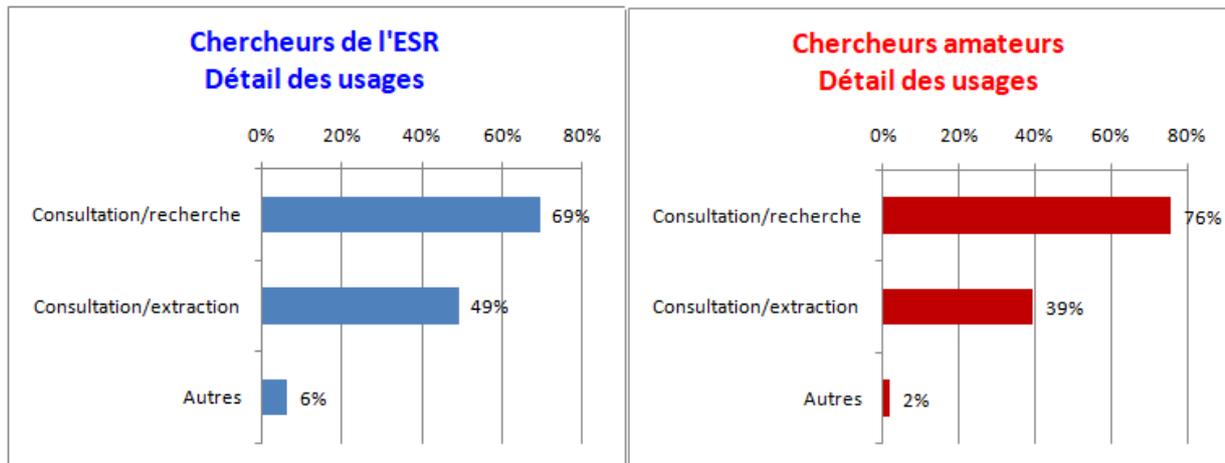
Figures 7 et 8 - Place de la presse ancienne numérisée dans la recherche

Si l'on regarde davantage dans le détail les usages (fig. 9 et 10), on remarque que pour les deux panels, l'usage le plus massif est la consultation avec recherche d'informations « simple » (sans extraction de données). Vient ensuite la consultation avec extraction de données. Sans surprise, cette pratique est plus répandue chez les chercheurs de l'ESR (49%, contre 39% chez les chercheurs amateurs). Toutefois, on note qu'elle constitue également un usage important chez les généalogistes et amateurs d'histoire : ceux-ci peuvent aussi avoir recours à des outils d'extraction parfois très sophistiqués, traduisant un réel niveau d'expertise comme en témoignent certains commentaires libres sur les méthodes de traitement et d'analyse employées :

- « Python, outils Unix, Hugging Face »<sup>176</sup>
- « Scripts Python avec utilisation de bibliothèques spécifiques pour l'extraction, la manipulation et l'analyse des données »

<sup>176</sup> Python est un langage de programmation. UNIX est une famille de systèmes d'exploitation multitâche et multi-utilisateur. Hugging Face est une société qui développe des outils pour créer des applications utilisant l'apprentissage automatique.

## II. De la valorisation à l'exploitation des corpus : de nouveaux besoins liés à la numérisation de la presse ancienne



Figures 9 et 10 - Détail des usages

On constate que deux types d'usages se dégagent chez les généalogistes et historiens amateurs (fig. 12) : usages « traditionnels » avec outils simples (captures d'écran ou extraction d'images au format JPEG, mise en page dans Word, réalisation de son arbre généalogique dans un tableur, etc.<sup>177</sup>) et pratiques plus perfectionnées, témoignant d'une grande familiarité numérique. Parmi elles, on peut relever l'utilisation de divers logiciels de généalogie (*Heredis*, *Gramps*, logiciels personnels), ou le recours à des outils de visualisation de données comme *Timeline* pour la création de frises chronologiques ou *uMap* pour l'édition et le partage de cartes. Il apparaît que ces pratiques sont loin d'être anecdotiques chez eux.

Chez les chercheurs de l'ESR, bien que l'usage du tableur reste dominant pour traiter les données, on note un recours assez important (12%) aux bases de données et logiciels de *data mining* (fig. 11). Cela s'explique par l'augmentation conséquente du nombre de projets de recherche impliquant ce type d'outils, notamment dans les études médiatiques où la fouille de données tend à se pratiquer de plus en plus. Parmi les méthodologies citées par les chercheurs ayant indiqué avoir recours à ces techniques de fouille massive, on retrouve fréquemment des méthodes et outils de *data analysis* (analyse de données ou de texte), *text-mining* (fouille de texte), extraction de données, *machine learning* (apprentissage automatique), ou encore de *datavisualisation* (visualisation de données). En voici quelques exemples :

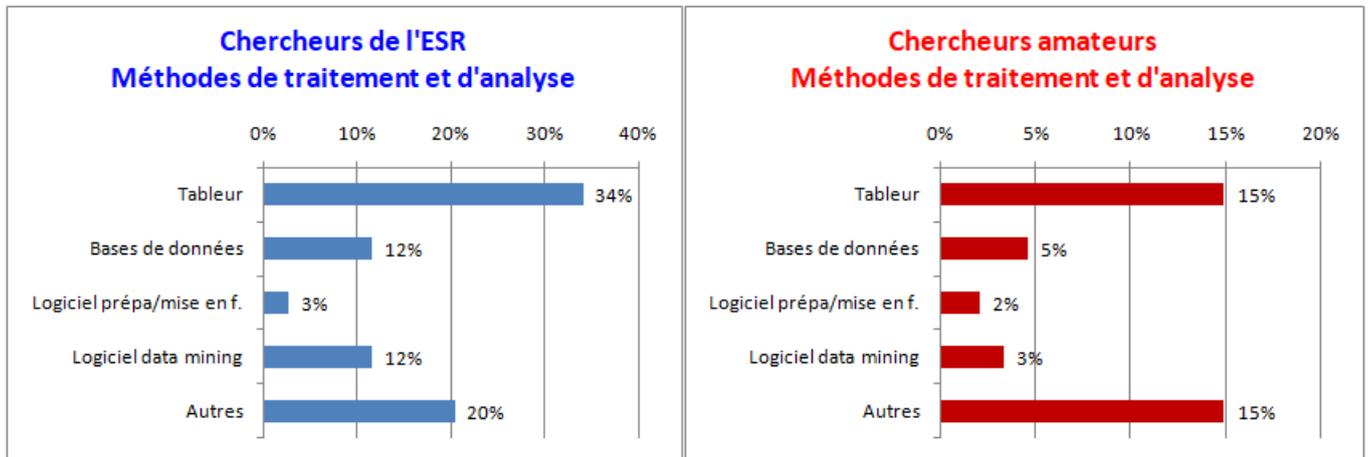
- « Utilisation de Cosma (d'Arthur Perret) »<sup>178</sup>
- « Algorithmes de *machine learning*, traitement de documents en lots, vectorisation de documents »
- « Analyse de texte, extraction d'information, séparation d'articles (Python/Ruby) »
- « Python (pandas, nltk, PyTorch) »
- « *Text-mining* avec R »
- « Gallica extraction text »

<sup>177</sup> De nombreux amateurs qui avaient indiqué « Autre » en réponse à la question les interrogeant sur leurs méthodes de traitement et d'analyse ont détaillé de tels usages.

<sup>178</sup> Logiciel de visualisation de graphe documentaire.

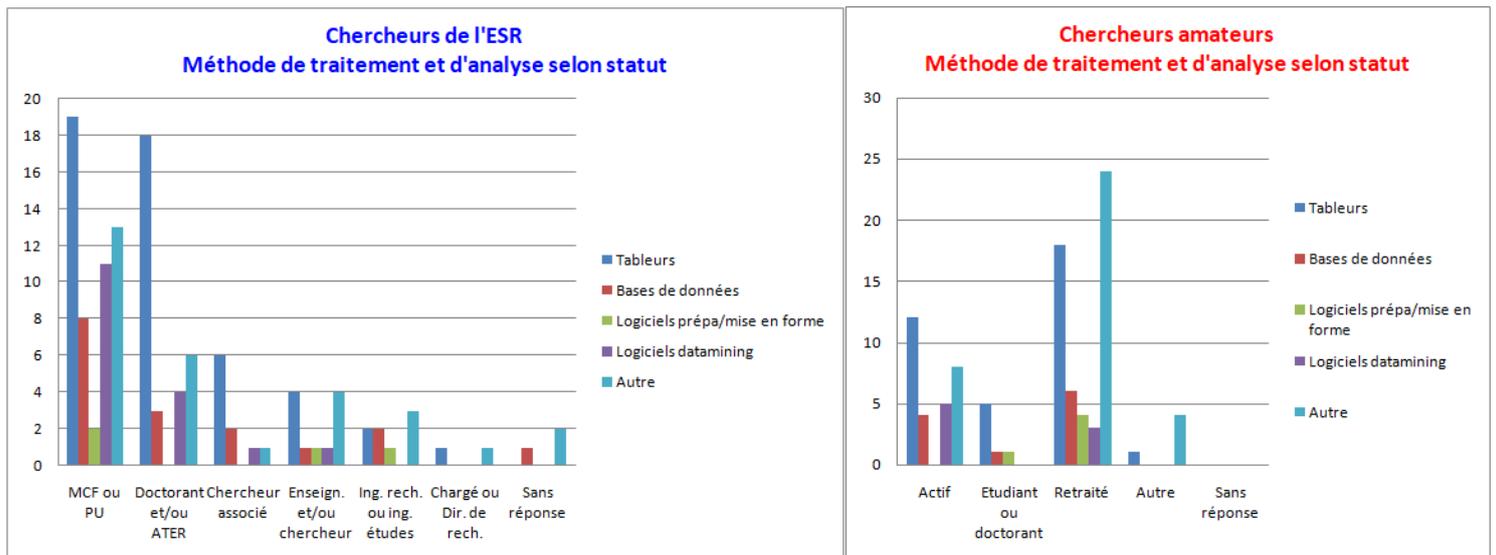
**II. De la valorisation à l'exploitation des corpus : de nouveaux besoins liés à la numérisation de la presse ancienne**

- « Plateformes d'analyse, telles que NewsEye »
- « Utilisation de l'application Gallicagram »



**Figures 11 et 12 - Méthodes de traitement et d'analyse**

Si l'on compare l'usage de telles méthodes avec le statut renseigné par les répondants, on constate que le tableur reste d'usage largement majoritaire, mais que d'autres méthodes sont aussi utilisées (en particulier par les maîtres de conférences et professeurs d'université pour l'ESR, et par les retraités pour les amateurs).



**Figures 13 et 14 - Méthodes de traitement et d'analyse privilégiées selon le statut**

***Plateformes utilisées***

Interrogés sur les différentes plateformes utilisées pour trouver de la presse ancienne, 93% des deux panels ont indiqué avoir recours (entre autres) à la bibliothèque numérique Gallica (fig. 15 et 16). La plateforme nationale apparaît donc largement majoritaire dans les usages. La deuxième plateforme utilisée par les chercheurs de l'ESR est RetroNews (48% des répondants l'utilisent), tandis qu'il s'agit de Geneanet<sup>179</sup> pour les amateurs (59%). Mais l'usage de RetroNews

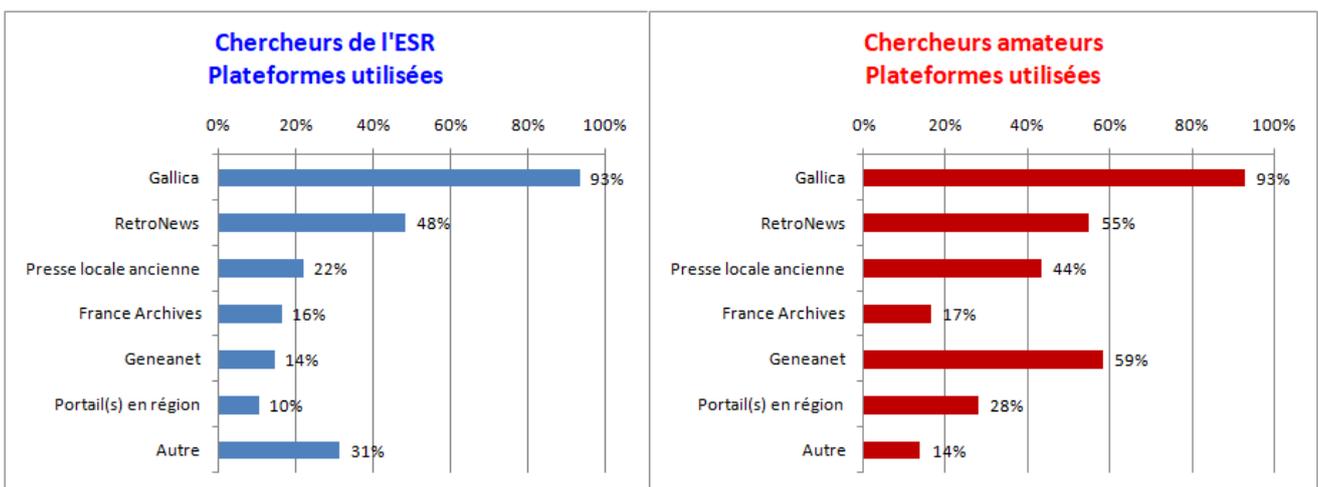
<sup>179</sup> Geneanet est un site généalogique proposant une base de données collaborative, alimentée par ses membres : <https://www.geneanet.org/>. Le site a été racheté en 2021 par le groupe américain Ancestry.

**II. De la valorisation à l'exploitation des corpus : de nouveaux besoins liés à la numérisation de la presse ancienne**

est également très répandu chez les généalogistes (55%). De manière générale, on note dans ce second panel un usage plus diversifié des différentes plateformes numériques existantes, moins concentré autour d'une ressource dominante (Gallica) comme chez les universitaires. Certains répondants interrogés en entretien ont indiqué utiliser Geneanet comme première porte d'entrée vers d'autres collections – celles de Gallica notamment – auxquelles une recherche via ce portail leur donne accès.

Les portails régionaux et sites des archives départementales sont également beaucoup plus consultés par les chercheurs amateurs, ce qui n'est guère surprenant au regard de leurs thèmes de prédilection. Cela confirme à quel point les services d'archives constituent des partenaires privilégiés des bibliothèques pour la valorisation de la presse ancienne, en particulier auprès de ce public de passionnés d'histoire locale et familiale. En plus des nombreux sites individuels des AD nommés par les répondants<sup>180</sup>, divers portails en région sont cités : Limédia (Sillon lorrain), Numelyo (Lyon), Overnia (Clermont Auvergne Métropole), Lectura Plus (Auvergne-Rhône-Alpes), Normannia (Normandie), Bilketa (Pays basque), la bibliothèque numérique Méjanas (Aix-en-Provence), ainsi que les différentes bibliothèques en marque blanche (Yroise, Rosalis, Rotomagus, etc.).

Si les universitaires désignent également certains de ces portails régionaux, ils citent davantage de plateformes de presse étrangère et plateformes académiques spécialisées ou généralistes. Parmi les corpus spécialisés, le *Gazetier universel* est cité comme ressource majeure pour l'étude de la presse du XVIII<sup>e</sup> siècle. Les archives de certains éditeurs de presse sont également mentionnées. 33 répondants ont indiqué recourir à des plateformes étrangères, dont Google Books et Internet Archive, ainsi que divers portails de presse correspondant aux aires géographiques étudiées : Europe du Nord, Chine, Belgique, Allemagne, Israël, etc. Si nous ne les détaillons pas ici puisque notre étude porte sur la presse française, il est à noter que plusieurs historiens amateurs ont également déclaré avoir recours à ces plateformes étrangères dans le cadre de leurs recherches.



**Figures 15 et 16 - Plateformes utilisées**

<sup>180</sup> Nous ne les listons pas ici les réponses, car les portails mis en avant par les différents répondants dépendent généralement de leur sujet de recherche et/ou lieu d'habitation.

### *Utilisation de Gallica et RetroNews*

La suite du questionnaire étudiait plus spécifiquement l'utilisation des deux plateformes proposées par la BnF pour consulter de la presse ancienne : Gallica et RetroNews – la comparaison se justifiant par la dimension nationale des deux sites. Nous souhaitions avant tout mettre en regard la manière dont les utilisateurs abordent la recherche dans ces deux plateformes : sachant que les interfaces ne proposent pas les mêmes fonctionnalités de recherche, il s'agissait de déterminer si les usagers adoptent une méthodologie de recherche similaire et viennent y chercher ou non le même type de contenus<sup>181</sup>. Nous tenions aussi à identifier d'éventuels besoins de formation à la recherche dans les très vastes collections proposées par l'une et par l'autre.

Concernant Gallica, 81% des universitaires et 76% des chercheurs amateurs déclarent l'utiliser régulièrement (fig. 17 et 18). Près de la moitié des amateurs estiment avoir besoin d'une formation à la recherche dans les collections de Gallica (fig. 20). C'est deux fois plus que chez les universitaires, où la demande de formation est tout de même exprimée par un quart des répondants (fig. 19). Ce besoin d'accompagnement est donc réel chez les chercheurs, professionnels ou non, y compris ceux qui se déclarent usagers réguliers de la plateforme (fig. 21 et 22). De fait, un rapport du consortium Couperin de 2020 sur les pratiques de recherche documentaire des chercheurs français souligne que « la question de la maîtrise et de l'appropriation des outils constitue un véritable enjeu ». Dans cette étude, les chercheurs étaient plus de 60% à exprimer un besoin en formation à la recherche documentaire (tous types de plateformes confondus), mettant en avant la nécessité d'une pratique régulière pour s'approprier les outils et parallèlement, le manque de temps pour le faire. Le rapport notait également que « les moteurs de recherche généralistes, Google en tête, constituent le principal mode de recherche d'information du chercheur »<sup>182</sup>.

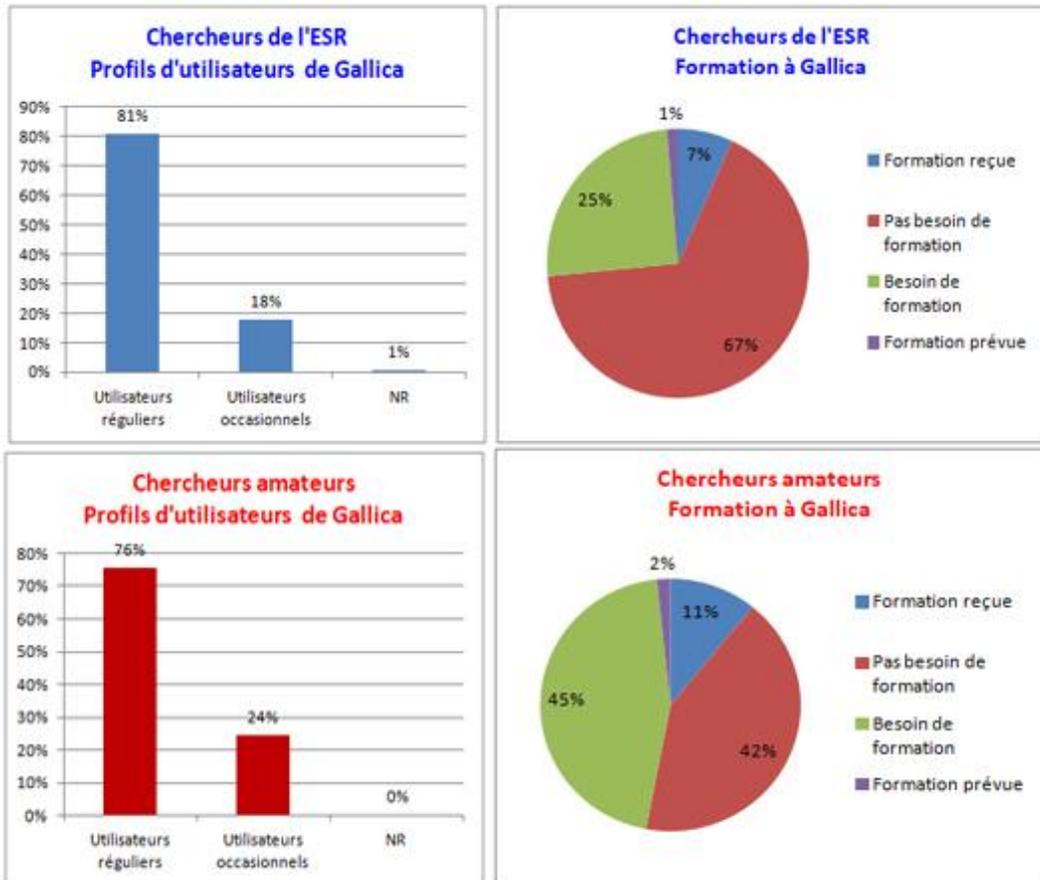
Dans notre enquête, seule une faible partie des enquêtés des deux panels a reçu une formation à Gallica. Chez les généalogistes, cette transmission se fait souvent par des pairs, mais également par d'autres moyens : autoformation (*Revue française de généalogie*, salon de généalogie, YouTube), formation délivrée par un professionnel de la généalogie, ou dans le cadre des études. Seules deux personnes mentionnent les ressources en ligne de la BnF : le blog de Gallica et les pages thématiques de Gallica, notamment celle dédiée à la généalogie. En revanche, les universitaires formés à Gallica l'ont été à moitié par la BnF, en présentiel ou via les tutoriels en ligne.

---

<sup>181</sup> Cf. Annexe n°1 : « Détail de la méthodologie d'enquête ».

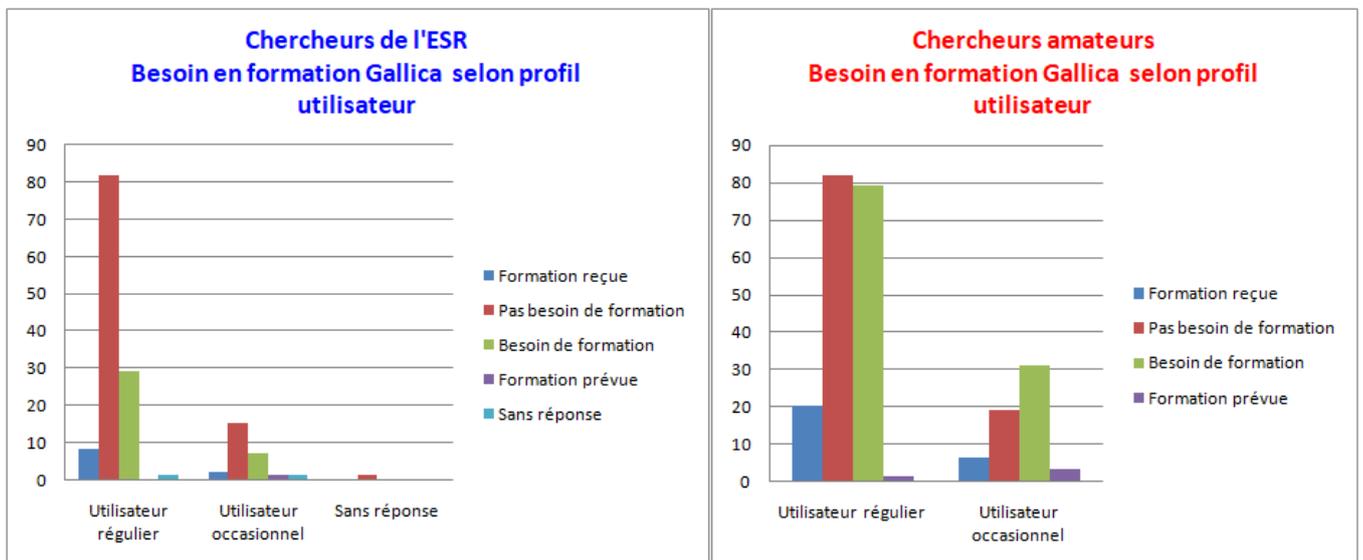
<sup>182</sup> Marie Pascale Baligand et al., *Les pratiques de recherche documentaire des chercheurs français en 2020 : étude du consortium Couperin*, rapport Couperin n°2, Couperin.org (hal-03148285), 2021.

**II. De la valorisation à l'exploitation des corpus : de nouveaux besoins liés à la numérisation de la presse ancienne**



**Figures 17 et 18 (gauche) - Profils d'utilisateurs de Gallica**

**Figures 19 et 20 (droite) - Formation à Gallica**



**Figures 21 et 22 - Besoin en formation à Gallica selon le profil d'utilisateur**

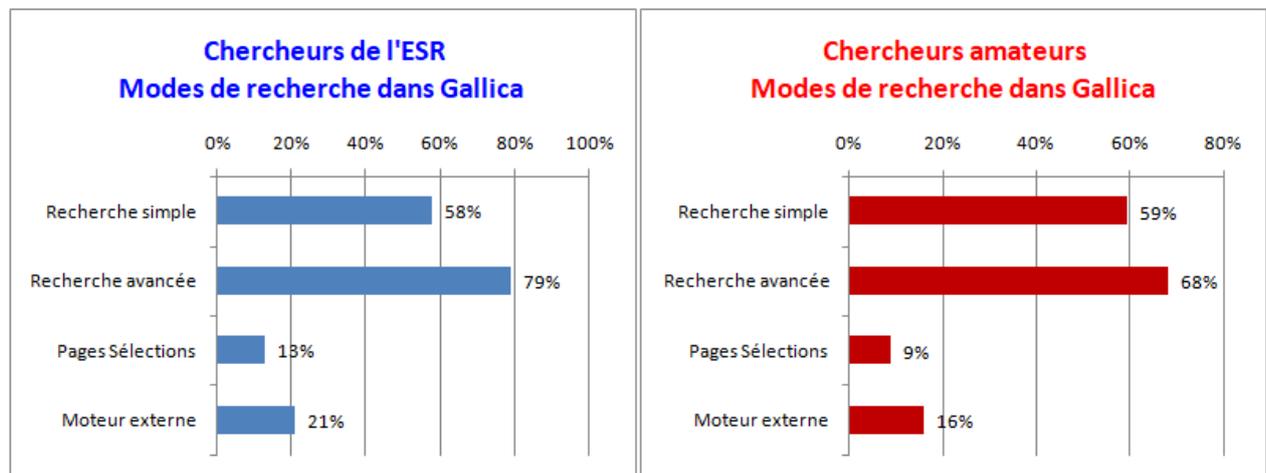
Les modes de recherche privilégiés dans Gallica se révèlent très proches pour nos deux panels (fig. 23 et 24). Les deux modes les plus utilisés sont sans conteste la recherche simple et la recherche avancée (on observe d'ailleurs que plus d'un tiers des participants font usage des deux<sup>183</sup>). Les pages Sélections semblent être

<sup>183</sup> Cf. Annexe n°3 : « Détail des réponses ».

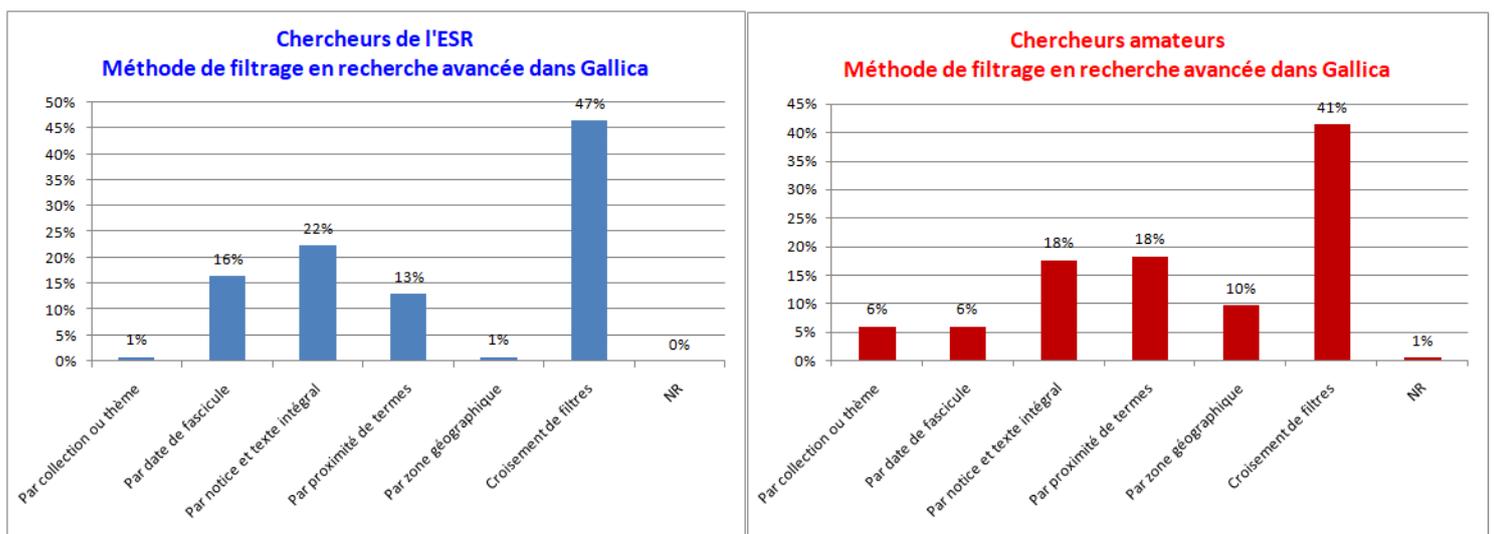
## II. De la valorisation à l'exploitation des corpus : de nouveaux besoins liés à la numérisation de la presse ancienne

peu identifiées par les lecteurs comme manière de filtrer les collections (par thématique, type de presse, zone géographique, etc.). Ceux qui emploient le moteur de recherche avancée croisent le plus souvent plusieurs filtres et utilisent prioritairement la recherche plein texte dans les ressources océrisées : la recherche plein texte à l'aide de l'OCR est la fonctionnalité la plus souvent citée par les usagers en commentaire libre.

Mais dans le détail (fig. 25 et 26.), une différence d'approche peut être observée entre les chercheurs de l'ESR et les chercheurs amateurs : alors que la recherche par zone géographique n'est quasiment pas utilisée par les universitaires, elle l'est davantage – en toute logique – chez les amateurs. De même, la plupart de ceux d'entre eux qui indiquent recourir aux pages Sélections cherchent prioritairement par zone géographique (46%)<sup>184</sup>. Les universitaires ont plus tendance, en revanche, à chercher les journaux par date de fascicule et à privilégier les sélections par type de presse et revues (58%)<sup>185</sup>. Certains indiquent rechercher d'abord dans le texte puis affiner par dates.



Figures 23 et 24 - Modes de recherche dans Gallica



Figures 25 et 26 - Méthode de filtrage en recherche avancée dans Gallica

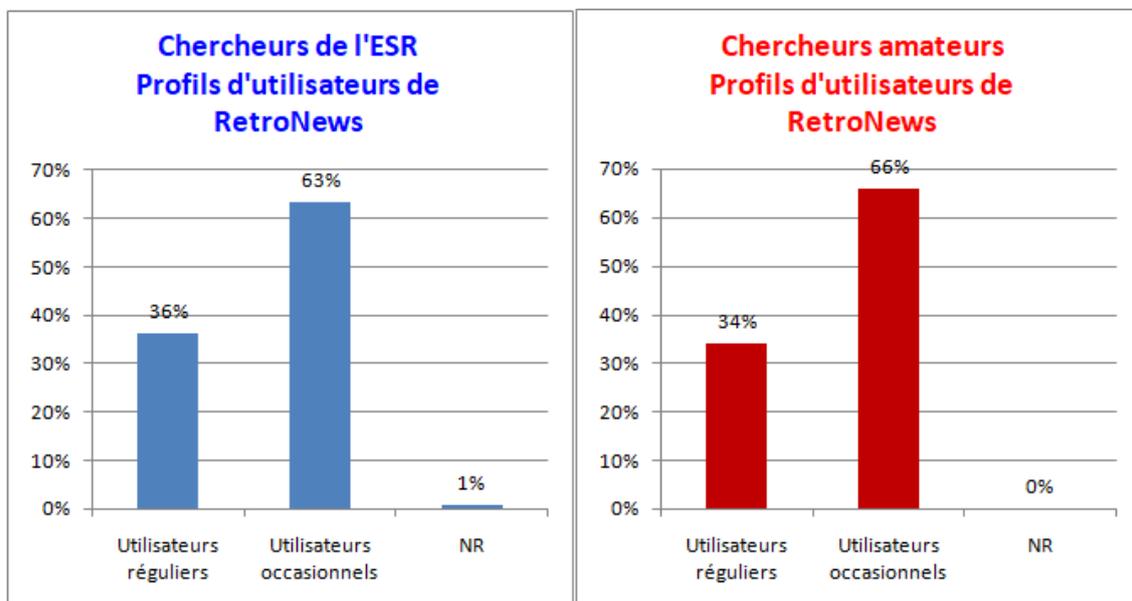
<sup>184</sup> Sur les 22 répondants amateurs indiquant filtrer leurs résultats à l'aide des Sélections de Gallica, ils sont 10 à privilégier la recherche par zone géographique, 7 par thématique, et 5 par type de presse et revues.

<sup>185</sup> Sur les 19 répondants universitaires indiquant filtrer leurs résultats à l'aide des Sélections de Gallica, ils sont 11 à privilégier la recherche par type de presse et revues, 6 par thématique, et 2 par zone géographique.

## II. De la valorisation à l'exploitation des corpus : de nouveaux besoins liés à la numérisation de la presse ancienne

Le niveau d'aisance pour mener une recherche dans Gallica est contrasté selon les répondants, chez les chercheurs amateurs comme ceux de l'ESR : si l'utilisation du moteur de recherche est parfois jugée complexe (difficultés liées au bruit généré par une requête et au tri des résultats, limites de la recherche par titre de journal), certains trouvent son ergonomie supérieure à des plateformes étrangères équivalentes. Plusieurs répondants saluent en particulier la recherche par proximité de termes, jugée très utile. Parmi les fonctionnalités les plus utilisées dans Gallica, les universitaires listent notamment les fonctions de téléchargement et d'exportation des résultats (rapport de recherche, téléchargement du fichier ocrisé, extraction d'images), de visualisation (zoom) et de transcription. Le calendrier est fréquemment utilisé par ces chercheurs, ce qui corrobore la tendance à la recherche par date observée dans ce panel. Parmi les fonctionnalités plébiscitées par les chercheurs amateurs, on trouve le classement par ordre de mise en ligne, l'espace personnel pour mémoriser les résultats et le rapport de recherche. Cette volonté de retrouver facilement ses résultats et de voir les nouveaux ajouts depuis sa dernière recherche traduit, chez nombre de ces amateurs, un usage intensif de la plateforme. Certains mentionnent également IIF<sup>186</sup>, l'identifiant pérenne des fascicules, ou encore les API<sup>187</sup> de Gallica pour récupérer les images ou le texte issu de l'ocrisation, comme des outils précieux pour leurs recherches. Toutefois, plusieurs répondants des deux groupes font remonter des problèmes liés à l'OCR, au zoom ou encore à l'affichage des images.

Concernant RetroNews, un peu plus d'un tiers des participants des deux panels s'en déclarent utilisateurs réguliers (fig. 27 et 28). Chez les chercheurs de l'ESR, les trois quarts des abonnés le sont au titre de leur institution, ce qui explique vraisemblablement la plus faible proportion d'abonnés chez les particuliers : 16% des répondants, contre 38% chez les universitaires (fig. 29 et 30). Les enquêtés des deux groupes estiment en majorité ne pas avoir besoin des tutoriels explicatifs, bien qu'environ un tiers en expriment le besoin (fig. 31 et 32).

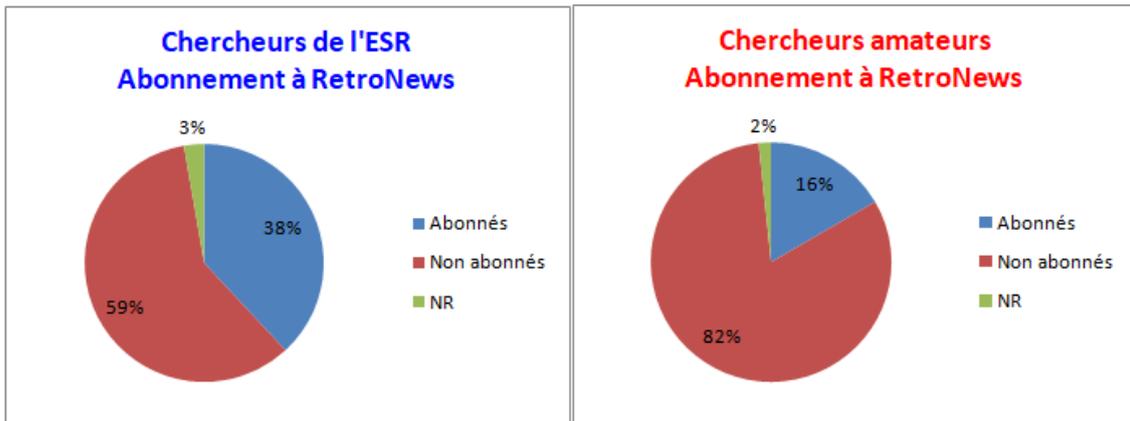


Figures 27 et 28 - Profils d'utilisateurs de RetroNews

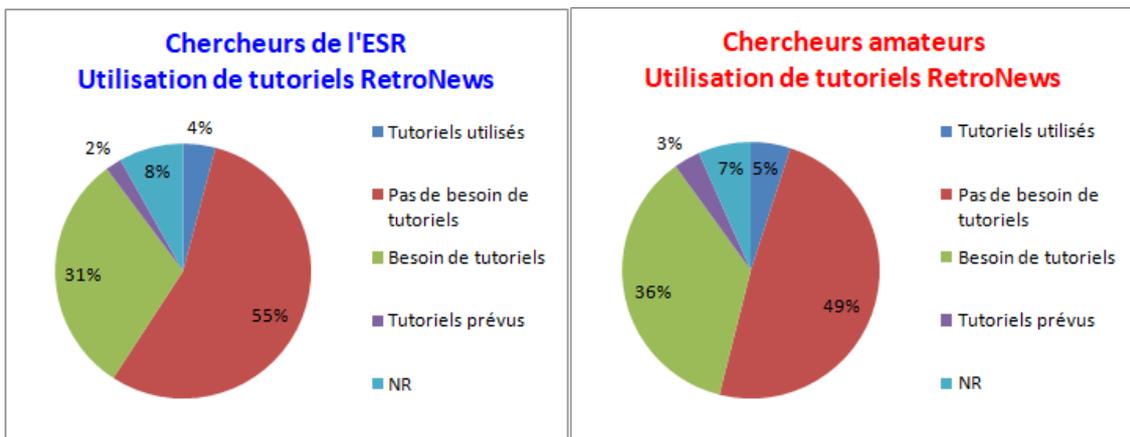
<sup>186</sup> *International Image Interoperability Framework* : standard permettant l'interopérabilité des ressources.

<sup>187</sup> *Application Programming Interface* : solution technique permettant à des applications de communiquer entre elles et d'échanger mutuellement des services ou des données.

**II. De la valorisation à l'exploitation des corpus : de nouveaux besoins liés à la numérisation de la presse ancienne**



**Figures 29 et 30 - Abonnement à RetroNews**

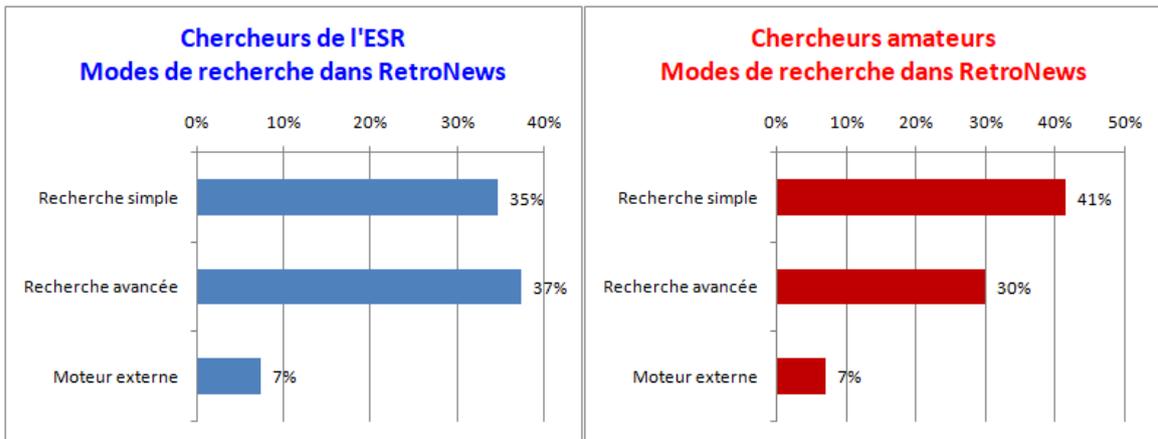


**Figures 31 et 32 - Utilisation de tutoriels RetroNews**

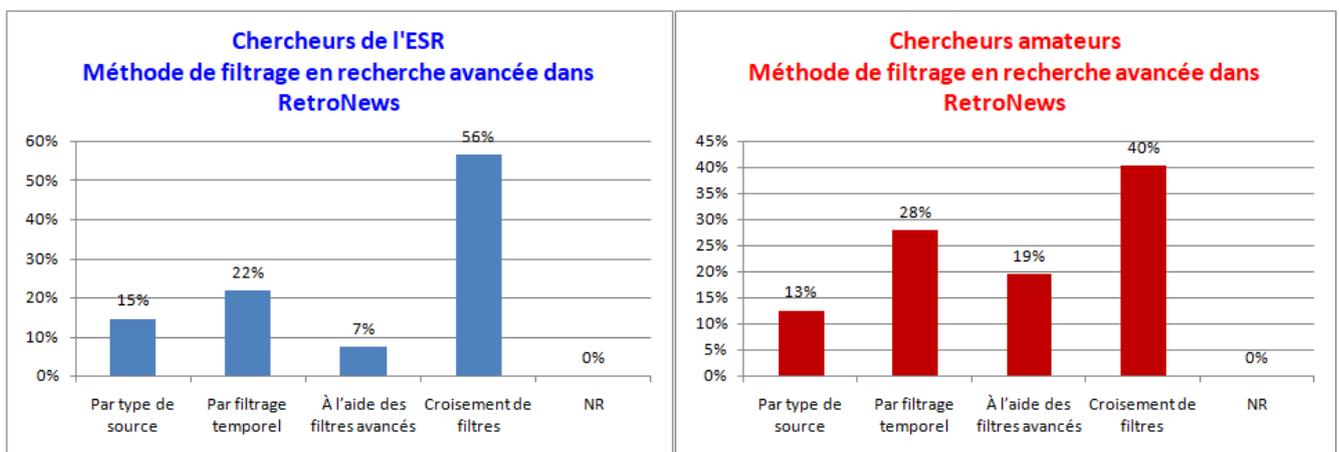
Dans RetroNews, c'est la recherche simple qui est le mode de recherche globalement dominant chez les participants (fig. 33 et 34). Le croisement de la recherche simple et de la recherche avancée est beaucoup moins fréquent que dans Gallica : seulement 14% des universitaires et 10% des généalogistes déclarent avoir recours aux deux méthodes<sup>188</sup>. Cela semble traduire un usage de Gallica plus orienté autour de la recherche experte que dans RetroNews, site média d'histoire à vocation plus grand public. En revanche, ceux qui font usage du moteur de recherche avancée sont nombreux à croiser différents filtres (fig. 35 et 36), ce qui témoigne d'une certaine familiarité de ces usagers avec la recherche dans les catalogues. Les universitaires sont proportionnellement plus nombreux à faire usage des sélections éditorialisées de RetroNews : dans le cadre de leurs recherches, mais aussi pour un usage pédagogique – usage évidemment beaucoup plus minoritaire chez les généalogistes et historiens amateurs (fig. 37 et 38). Parmi les autres types d'usages possibles sont mentionnées la rédaction d'articles pour la plateforme elle-même ainsi que la publication d'articles sur Wikisource. L'usage des sélections de RetroNews n'en demeure pas moins important chez les amateurs, principalement pour des recherches ou par simple curiosité.

<sup>188</sup> Cf. Annexe n°3 : « Détail des réponses ».

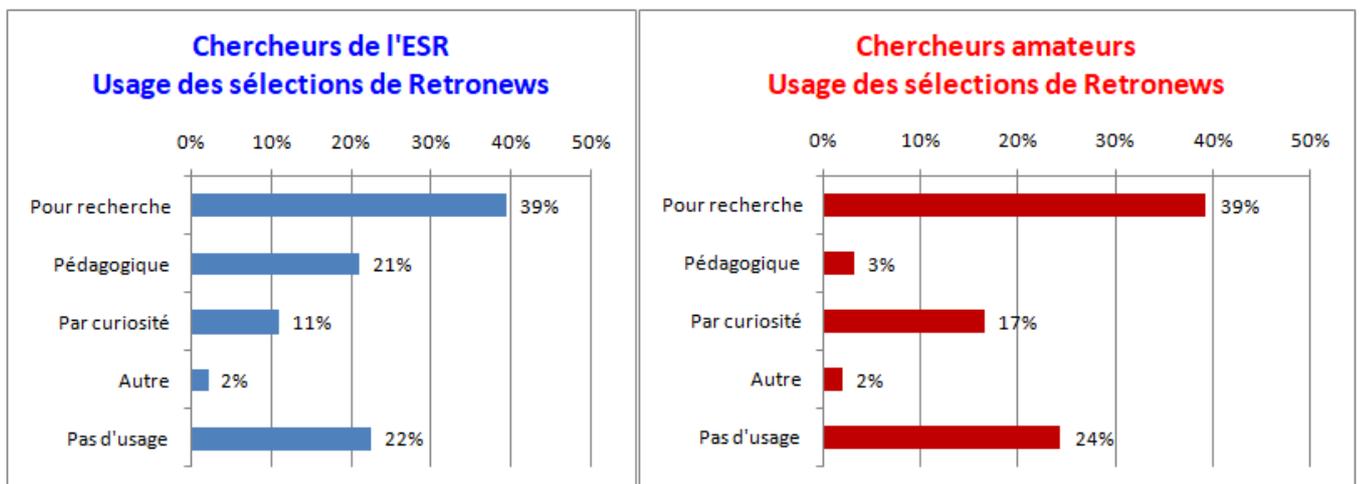
**II. De la valorisation à l'exploitation des corpus : de nouveaux besoins liés à la numérisation de la presse ancienne**



**Figures 33 et 34 - Modes de recherche dans RetroNews**



**Figures 35 et 36 - Méthode de filtrage en recherche avancée dans RetroNews**



**Figures 37 et 38 - Usage des sélections de RetroNews**

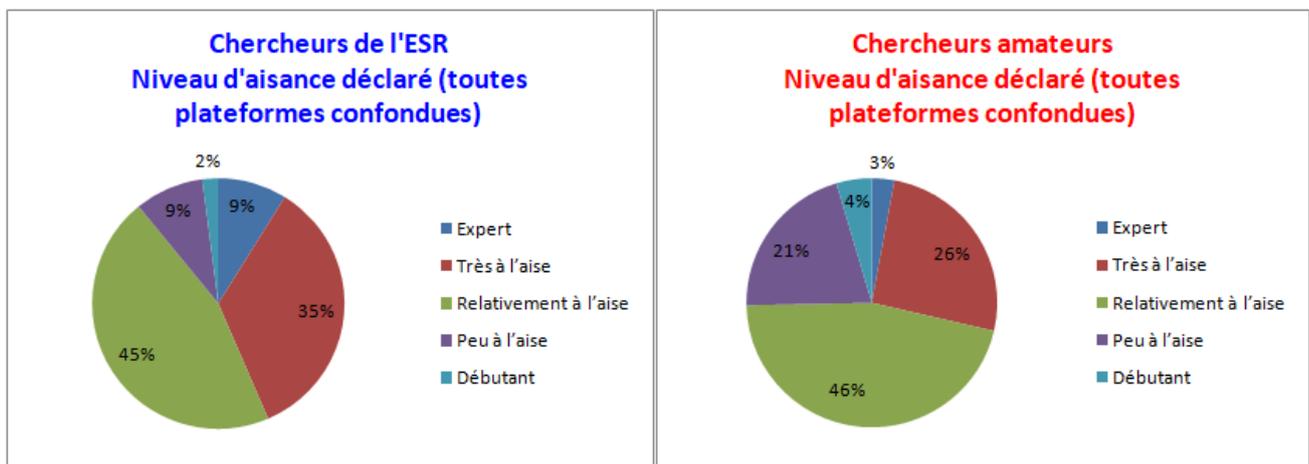
Concernant les fonctionnalités de recherche avancée, certains répondants pointent du doigt le caractère prohibitif de l'abonnement, vu comme un frein à la recherche. C'est notamment le cas dans le second panel, composé à majorité de particuliers ne disposant pas d'un abonnement institutionnel, ou pour les chercheurs d'un établissement non abonné. Plusieurs réclament la gratuité des fonctionnalités avancées – ou éventuellement des tarifs préférentiels, pour les généalogistes professionnels entre autres. Quelques réponses soulignent des problèmes d'OCR, causant parfois une recherche bruitée. Ceux-ci s'expliquent

## II. De la valorisation à l'exploitation des corpus : de nouveaux besoins liés à la numérisation de la presse ancienne

notamment par la qualité variable des microfilms qui ont servi de corpus initial à RetroNews. Plusieurs abonnés se disent néanmoins satisfaits par la granularité de recherche (page / paragraphe / article) et par diverses fonctionnalités offertes par la plateforme : fonctions d'annotation ou d'extraction des résultats, possibilité de croiser plusieurs filtres (filtrage temporel, par titre de journal, par mot-clé, etc.), analyses quantitatives avec visualisation graphique (fréquence d'un terme par titre ou type de journal, évolution d'un terme par année). Le fait que RetroNews autorise la recherche sur une période précise, voire dans un corpus précis de journaux, est apprécié dans le cadre d'une recherche universitaire en particulier. L'option de recherche par proximité de termes, telle que proposée dans Gallica, serait souhaitée de la même façon dans RetroNews par quelques répondants.

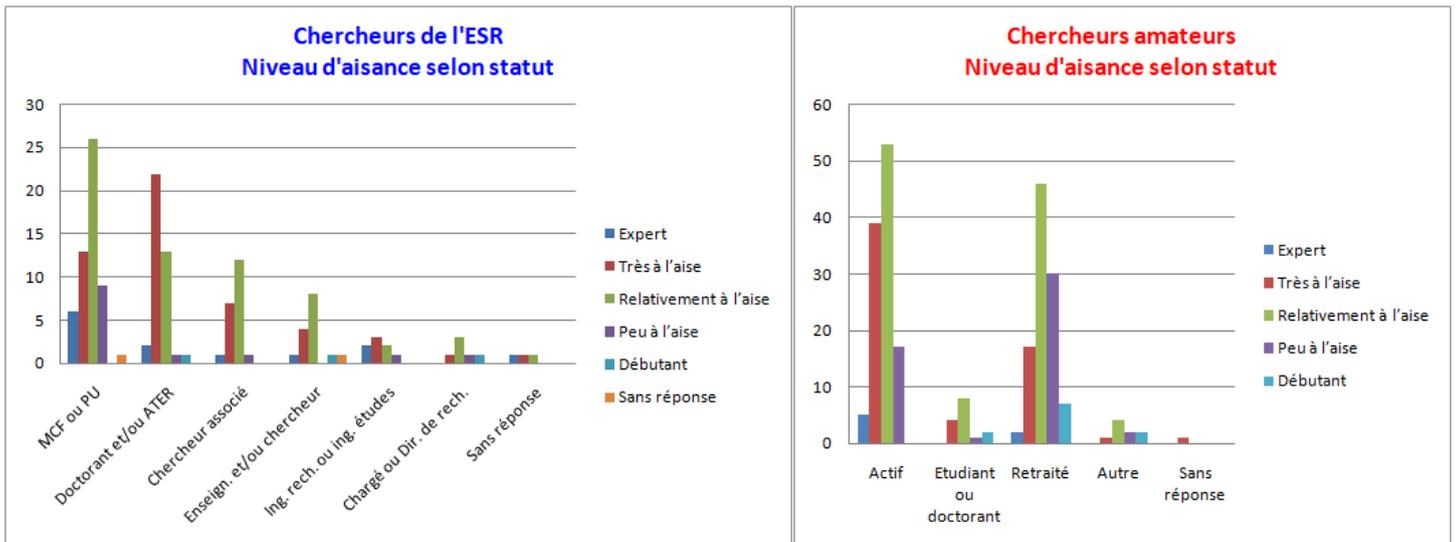
### *Besoins et manques identifiés*

Toutes plateformes confondues, les répondants des deux panels interrogés se disent en majorité (à 45% et 46%) relativement à l'aise dans la recherche. Si, logiquement, davantage de chercheurs de l'ESR s'estiment très à l'aise voire experts, la proportion de chercheurs amateurs aguerris voire très aguerris est significative (fig. 39 et 40). Cependant, la proportion de lecteurs se présentant comme peu à l'aise n'est pas non plus à négliger. Du côté de l'ESR, les 9% concernés ne sont pas que des jeunes chercheurs, mais également des chercheurs confirmés (fig. 41). Chez les amateurs, la proportion de personnes se décrivant comme peu à l'aise ou débutantes est plus élevée (25%). Les retraités, femmes sans emploi et étudiants sont les catégories les plus concernées par ces difficultés liées à l'usage du numérique (fig. 42). Le besoin de formation à Gallica et RetroNews se confirme donc, plus largement, à l'échelle de toutes les plateformes numériques.



Figures 39 et 40 - Niveau d'aisance déclaré (toutes plateformes confondues)

**II. De la valorisation à l'exploitation des corpus : de nouveaux besoins liés à la numérisation de la presse ancienne**



**Figures 41 et 42 - Niveau d'aisance selon le statut**

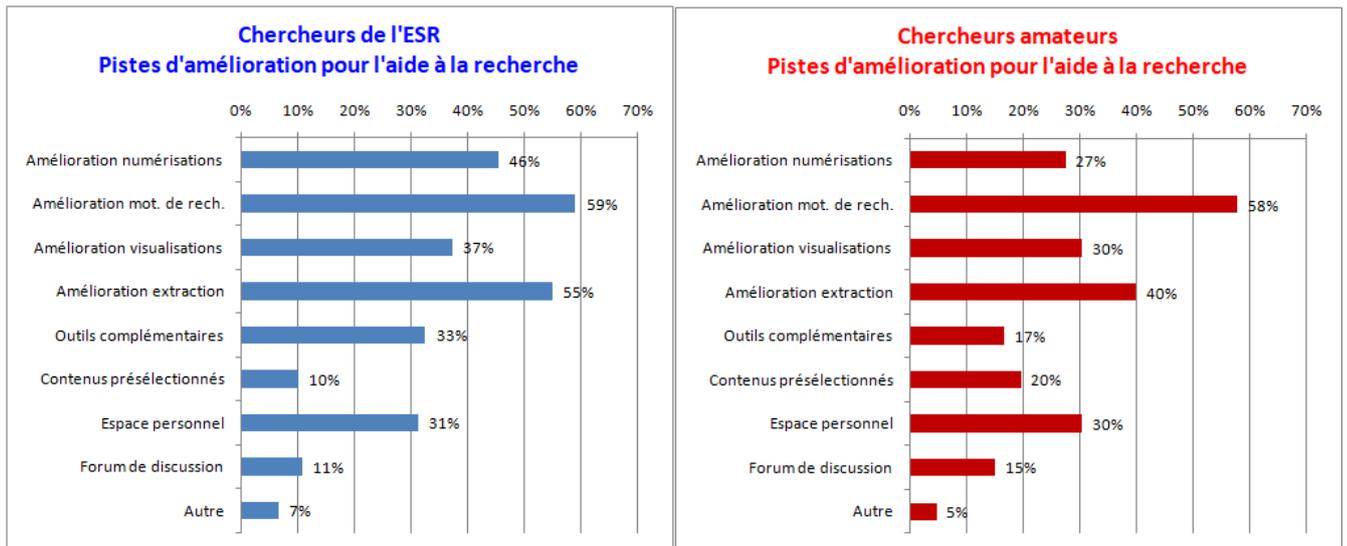
Paradoxalement, ce ne sont pas les publics les moins aguerris qui soulignent la complexité des moteurs de recherche – sans doute parce que le constat de ces difficultés implique déjà une certaine familiarité numérique. Deux généalogistes confirmés que nous avons interrogés résument bien ce sentiment :

Témoignage n°1 : « La plus grande difficulté, à mon sens, est que les moteurs de recherche sont conçus et élaborés par et pour des professionnels de la consultation informatique ou physique des archives. Heureusement, les relations écrites, téléphoniques, sur les lieux, avec les personnels des services des archives de France facilitent la quête de l'information par leur capacité d'aide et de soutien. Ils sont sans doute les premiers surpris des objets de recherche de leurs concitoyens, mais aussi heureux d'y apporter leur participation. »

Témoignage n°2 : « Quand on est un particulier, il faut déjà bien connaître les catalogues. Pour connaître toutes les subtilités de la recherche avancée, il faut vraiment être un chercheur. Il y a beaucoup de "petits trucs" qu'on apprend en cherchant : par exemple, écrire le numéro après le nom de la rue, car cela était souvent orthographié ainsi dans la presse d'époque. Idem pour les inversions nom/prénom : si l'on tape seulement le prénom suivi du nom, on risque de passer à côté de 30-40% des résultats. On n'a pas un manuel qui nous explique comment chercher dans la presse ancienne, ce sont des habitudes que l'on acquiert avec la pratique. »

L'amélioration des moteurs de recherche est désignée par les participants comme piste prioritaire pour favoriser la recherche : dans les deux panels, ils sont près de 60% à la réclamer. L'amélioration des fonctionnalités d'extraction est également demandée par un grand nombre de chercheurs, professionnels comme amateurs. L'amélioration des numérisations, des visualisations et la mise à disposition d'un espace personnel pour sauvegarder des documents sont fréquemment suggérées pour optimiser l'ergonomie des plateformes. Les universitaires sont plus nombreux à souhaiter la mise à disposition d'outils complémentaires pour l'exploitation des corpus, tandis que l'éditorialisation des contenus est davantage demandée par les amateurs (fig. 43 et 44).

## II. De la valorisation à l'exploitation des corpus : de nouveaux besoins liés à la numérisation de la presse ancienne



Figures 43 et 44 - Pistes d'amélioration pour l'aide à la recherche

### *Bilan de l'enquête : premières pistes*

Amenés à préciser les manques identifiés et leurs besoins, les enquêtés ont été nombreux à proposer des pistes d'amélioration concrètes pour favoriser la recherche dans les plateformes de presse. Si certaines demandes sont plus spécifiques aux universitaires et d'autres aux amateurs, nous constatons surtout la proximité des requêtes formulées par les deux panels. Nous les avons regroupées en cinq grands axes d'amélioration, dont nous présentons ici la synthèse (le détail des réponses à cette dernière question est donné en annexe)<sup>189</sup>. Chacune de ces pistes sera explorée et détaillée dans la suite de notre étude.

- **Axe n°1 - Amélioration des moteurs de recherche et des fonctionnalités de recherche** : Améliorer l'efficacité des moteurs de recherche est un axe jugé prioritaire par nombre de participants. Cela passe obligatoirement par une amélioration de la qualité de l'OCR (surtout sur les titres anciens et numérisations anciennes) et une généralisation de l'OCR à l'ensemble des collections mises en ligne, afin de permettre la recherche plein texte sur tous les titres numérisés. Cela passe aussi par un travail sur l'affichage des résultats (mieux triés, mieux contextualisés), la mise à disposition de filtres plus fins, ou encore d'options de visualisation et d'extraction mieux adaptées aux nécessités d'une recherche dans la presse.

- **Axe n°2 - Augmentation des numérisations et enrichissement des métadonnées** : L'intensification des programmes de numérisation est demandée afin de combler les lacunes des collections mises en ligne. Cela va de pair avec une concertation accrue entre les établissements de conservation, afin de permettre une meilleure centralisation des ressources, jugées actuellement trop dispersées. Les universitaires insistent sur l'importance de disposer de métadonnées de qualité afin de permettre un traitement plus fin de certaines collections (périodiques en volume notamment). Beaucoup réclament la gratuité d'accès à toutes les ressources de presse libres de droits.

<sup>189</sup> Cf. Annexe n°4 : « Typologie des besoins exprimés par les répondants ».

## II. De la valorisation à l'exploitation des corpus : de nouveaux besoins liés à la numérisation de la presse ancienne

- **Axe n°3 - Mise à disposition d'outils d'analyse et enrichissement sémantique des collections** : Ce type de requête est plus fréquent chez les universitaires, demandeurs de fonctionnalités supplémentaires pour l'exploitation et l'analyse des corpus : outils de textométrie, de repérage des concordances entre périodiques, de segmentation ou de recherche d'images, mise à disposition facilitée d'API, etc. Un travail sur l'enrichissement sémantique des collections (technique de traitement automatique du langage naturel<sup>190</sup>) est sollicité afin de permettre la recherche par nom de personne, de lieu, etc.

- **Axe n°4 - Partage des résultats de recherche** : Des outils de partage, espaces d'échange sur les pratiques et résultats de recherche, ou encore espaces de mise en commun (ex : correction d'OCR collaborative) sont demandés par plusieurs répondants. Il est intéressant de voir que cette dimension collaborative est plus prégnante chez les amateurs, sans doute parce qu'elle s'inscrit déjà dans les pratiques de recherche de très nombreux généalogistes et passionnés d'histoire<sup>191</sup>.

- **Axe n°5 - Formation** : Le besoin de formation se fait également plus prononcé du côté des amateurs, qu'il s'agisse de formations généralistes (bases de la recherche documentaire) ou de tutoriels répondant à des besoins précis (dans le cadre d'une recherche généalogique par exemple).

Les conclusions tirées de notre enquête – analyse des profils de répondants et de leurs pratiques de recherche, ainsi que les divers leviers d'amélioration soulevés par les participants eux-mêmes – constituent la base des préconisations que nous formulons dans le troisième temps de cette étude. Cette dernière partie est davantage prospective : il s'agit en effet d'offrir les services les plus adaptés aux besoins, en anticipant notamment sur ceux qui, accélérés par la numérisation massive, vont être amenés à se généraliser dans les années à venir. Nos recommandations intègrent plusieurs niveaux de familiarité numérique en couvrant un large spectre d'usages, des plus débutants aux plus experts.

---

<sup>190</sup> Le traitement automatique du langage naturel (TALN ou TAL) est domaine de recherche visant à développer des méthodes pour permettre aux ordinateurs de comprendre et de traiter le langage naturel.

<sup>191</sup> Voir des sites comme Geneanet ou Heredis, fonctionnant sur un principe de mise en commun des résultats pour enrichir les données généalogiques. Voir aussi l'exemple du *Forum Pages 14-18* cité plus haut (2.1.3., p. 46).

## **III. ADAPTER LES SERVICES AUX BESOINS POUR MIEUX VALORISER LA PRESSE ANCIENNE**

---

### **3.1. VERS PLUS DE MEDIATION NUMERIQUE ET DE FORMATION**

#### **3.1.1. La prise en compte du parcours utilisateur : une réflexion à poursuivre sur l'ergonomie des interfaces**

##### *Parler à tous les publics*

La réflexion sur l'ergonomie des interfaces n'est pas nouvelle. En 2018, une enquête menée à partir de l'analyse des logs de Gallica révélait que les stratégies de médiation proposées sur la plateforme (éditorialisation de la page d'accueil, pages Sélections, etc.) ne rencontraient pas le succès escompté, car l'utilisateur ne suivait pas nécessairement le parcours conçu pour lui :

Si la conception d'un site web induit toujours une présomption d'usage « normal » (par exemple : page d'accueil > moteur interne > consultation de document), [l'analyse met en évidence] la très grande diversité des logiques de parcours dans Gallica [...]. Il convient donc pour Gallica de ne pas concevoir la page d'accueil comme la porte d'entrée principale, mais de faire au contraire de toute page une porte d'entrée dans le site avec des propositions de parcours<sup>192</sup>.

Or, ce principe vaut pour n'importe quel portail patrimonial : la prise en compte du parcours utilisateur est déterminante dans toute pensée de la valorisation. Cela requiert, parfois, un effort de décentrement important au regard des habitudes professionnelles des personnels des bibliothèques. Il faut en effet proposer une interface qui puisse être prise en main de la manière la plus intuitive possible par les utilisateurs, en particulier les moins familiers avec le numérique et l'écosystème des ressources en ligne. On peut, à cet égard, s'appuyer sur l'exemple de L'Empreinte<sup>193</sup>, qui a fait le choix de limiter au maximum la terminologie spécialisée afin d'éviter tout ce qui était susceptible de gêner la compréhension de l'utilisateur. Cela a entièrement déterminé la manière de penser l'outil et de le mettre à disposition : langage utilisé, fonctionnalités mises en avant, etc. Évoquant la genèse du projet, Lionel Dujol explique :

Il était extrêmement important que notre proposition fasse sens tout de suite. Il a donc fallu abandonner notre terminologie archivistique : « périodiques », par exemple, ne parle pas du tout, alors que « journaux » ou « presse » parle immédiatement. Nous avons compris que tout ce qui allait structurer la pensée de la valorisation serait toujours, avant tout, le souci de la compréhension par le grand public. Mais cela ne devait pas se faire au détriment du projet scientifique : il ne fallait en aucun cas que cet effort de

---

<sup>192</sup> P. Chevallier, « Les données au service de la connaissance des usages en ligne », art. cit., p. 64.

<sup>193</sup> Portail patrimonial de la Drôme et de l'Ardèche dont nous avons parlé plus haut.

### III. Adapter les services aux besoins pour mieux valoriser la presse ancienne

vulgarisation ne dégrade l'attente scientifique du chercheur ou de la société savante<sup>194</sup>.

Or, ces deux logiques ne sont pas du tout incompatibles mais se superposent. Ainsi, les mentions inscrites dans la barre de recherche (« Explorez les histoires de Drôme et d'Ardèche », « Vous vous souvenez des lieux de votre jeunesse ? », « Essayez avec le nom de votre village », etc.) sont une manière de guider l'utilisateur dans son parcours. Pour autant, elles n'empêchent pas un usage plus scientifique de la plateforme. Dans cet exemple, la valeur ajoutée qu'offre la plateforme numérique est, selon les termes de Lionel Dujol, une « valeur conseil ». L'utilisateur choisit alors son parcours en fonction de ses attentes et de son niveau de familiarité numérique : ainsi le débutant se laissera-t-il guider par un parcours entièrement balisé (aide à la saisie dans la barre de recherche, contenus éditorialisés), tandis que lecteur plus autonome s'orientera davantage par lui-même, au gré des rebonds offerts par la plateforme.

#### *Prendre en compte le public généalogiste et les usagers non experts*

Ainsi, les portails de presse ont tout intérêt à flécher au mieux le parcours du visiteur, dès la page d'accueil mais aussi sur les autres pages du site, en anticipant tous les freins potentiels pour l'utilisateur le moins expérimenté. L'importance des généalogistes parmi le public de chercheurs amateurs est à prendre en compte : il est apparu dans notre enquête que ces lecteurs représentaient une catégorie très diverse d'utilisateurs, débutants ou relativement débutants pour certains, expérimentés voire experts pour d'autres. Leur ménager un point d'entrée spécifique, à la manière de la page « Ressources généalogiques » de Gallica<sup>195</sup> (comprenant notamment le *Journal officiel*), semble être une idée judicieuse. Cependant, comme cela est apparu dans l'enquête, cette page n'est pas forcément bien identifiée par les usagers qu'elle est susceptible d'intéresser<sup>196</sup>. Actuellement, elle est accessible seulement via la thématique « Histoire » dans Gallica ou par une recherche alphabétique. Il serait donc pertinent d'accroître sa visibilité, en favorisant par exemple une logique de rebond depuis d'autres pages afférentes du site. Les pages consacrées à la presse locale et régionale, les entrées géographiques (par département, région, etc.), ou encore la carte de Cassini<sup>197</sup>, pourraient être des points d'entrée privilégiés pour signaler l'existence de cette page (« Pour aller plus loin... »).

On pourrait aussi envisager qu'un lien soit fait avec l'outil Gallicarte, afin d'aider les généalogistes et historiens amateurs à géolocaliser les documents relatifs à un lieu donné. Pour l'heure, seules les images numérisées sont concernées, mais l'outil prévoit d'élargir cette géolocalisation aux documents

---

<sup>194</sup> Entretien avec Lionel Dujol, responsable prospective, innovation et accompagnement au changement (Valence Romans Agglo, Direction Lecture Publique), le 16/11/2022.

<sup>195</sup> Sélection de documents susceptibles d'intéresser les généalogistes : presse ancienne, revues savantes, photographies anciennes, cartes, gravures, etc. Voir *Gallica - Ressources généalogiques*, <https://gallica.bnf.fr/html/und/histoire/ressources-genealogiques> (consulté le 17 février 2023).

<sup>196</sup> Cf. supra (2.3.2., p. 61) : seuls deux participants l'ont mentionnée dans leurs réponses.

<sup>197</sup> La carte générale de la France de Cassini, et notamment l'outil Gallicarte, sont cités par Sophie Boudarel comme des ressources précieuses pour le généalogiste. Voir *La carte de Cassini*, <https://gallica.bnf.fr/html/und/cartes/france-en-cartes/la-carte-de-cassini> (consulté le 17 février 2023).

### III. Adapter les services aux besoins pour mieux valoriser la presse ancienne

écrits<sup>198</sup>. Une visualisation géographique serait alors rendue possible pour les titres numérisés disponibles dans Gallica. De manière générale, réfléchir à une approche géographique des fonds de presse, comme le propose notamment le site Presse locale ancienne sur sa page d'accueil, est une excellente manière de guider l'utilisateur dans sa navigation. Cette fonctionnalité est d'ailleurs souhaitée par plusieurs répondants. La facilitation graphique est le premier niveau de facilitation ; aussi plusieurs plateformes développent-elles aujourd'hui cette approche cartographique des fonds, ou envisagent de le faire<sup>199</sup>.

La récente refonte de l'interface de recherche proposée par le site des Archives départementales des Yvelines nous offre un bon exemple de ce qui peut être fait en matière de facilitation graphique pour un usager non expert. Marquant une étape importante dans la politique de diffusion à l'échelon local, cette nouvelle interface traduit plus largement une prise de conscience des services d'archives de la nécessité d'adapter les outils de recherche aux degrés d'expertise très variables de leurs publics. Depuis 2009, grâce à un important travail de numérisation mené en partenariat avec divers dépôts et établissements, les AD 78 ont entrepris de reconstituer une collection de référence de la presse ancienne des Yvelines et de l'ancienne Seine-et-Oise. Cette collection inclut aujourd'hui une centaine de titres, soit plus de 220 000 pages numérisées, et s'est encore agrandie de 22 nouveaux titres en 2022<sup>200</sup>. La nouvelle interface de recherche est pensée de manière à rendre l'exploration de cette collection la plus fluide et intuitive possible. L'entrée dans les fonds est facilitée par trois modes de visualisation : en mosaïque par ordre alphabétique, sous forme de frise chronologique, mais aussi grâce à une spatialisation des titres numérisés à l'aide de l'outil OpenStreetMap.

Ici, visualisation géographique et visualisation historique se complètent donc. Si l'on recherche par titre, différentes facettes permettent de filtrer la recherche par commune ou par lieu, tendance politique, thème, période de publication ou établissement détenteur des originaux. La recherche par numéros permet d'affiner encore l'interrogation de la collection, tout en affichant le nombre de numéros manquants pour chaque titre et en précisant si la recherche plein texte est applicable. L'interface s'adapte ainsi à plusieurs niveaux de familiarité numérique, en rendant accessibles, par les choix de présentation opérés, des critères de recherche pourtant avancés : opérateurs booléens, granularité fine de recherche, etc. Un tel exemple nous montre comment une même interface peut cumuler plusieurs logiques d'utilisation, de la plus débutante à la plus experte, sans pour autant que ces différents parcours utilisateur n'interfèrent. De manière générale, tout ce qui permet de faciliter l'expérience utilisateur et la lisibilité graphique (vue d'ensemble des numéros d'un même titre, carte, frise chronologique, possibilité de zoomer directement sur la zone de page ou le mot-clé recherché, etc.) est souhaitable. Améliorer l'ergonomie des interfaces est donc un premier stade – crucial – de médiation numérique.

---

<sup>198</sup> Sophie Boudarel, *Gallicarte, géolocalisez les collections de Gallica*, <https://la-gazette-des-ancetres.fr/gallicarte/>, 6 février 2020 (consulté le 17 février 2023).

<sup>199</sup> Entretien avec Matthieu Guenoux, bibliothécaire aux Archives départementales de Loire-Atlantique, le 26/10/2022.

<sup>200</sup> *Les Journaux de Seine-et-Oise dévoilent leurs caractères*, <https://archives.yvelines.fr/nous-connaître/nos-actualités/zoom-sur/les-journaux-de-seine-et-oise-devoilent-leurs-caracteres> (consulté le 17 février 2023).

- **Préconisation n°1** : Renforcer la prise en compte du parcours utilisateur, en tenant compte notamment des usagers les moins familiers du numérique et des besoins propres au public généalogiste.
- **Bonnes pratiques** : Guider la saisie dans la barre de recherche ; favoriser les rebonds ; prévoir une sélection de ressources pour le public généalogiste ; développer des outils de géolocalisation des titres numérisés (ex : OpenStreetMap, BatchGeo) ; améliorer la vue d'ensemble des numéros d'un même titre ; permettre une visualisation chronologique ; améliorer les fonctionnalités de zoom.

### 3.1.2. L'importance de la formation

Dans cette volonté de guider l'utilisateur dans son parcours, il ne faut pas négliger l'importance de la formation. De fait, nous avons pu observer dans notre enquête à quel point les freins susceptibles d'entraver une recherche interviennent souvent très tôt, dès les premières étapes : repérage dans le paysage complexe des portails de presse, identification des ressources pertinentes par rapport à l'objet de recherche, saisie des mots-clés dans la barre de recherche, tri des résultats. Les usagers qui font part de difficultés sont souvent ceux qui ne maîtrisent pas les « astuces de recherche » – ajout de guillemets pour une expression exacte, recherche par proximité de termes dans Gallica, inversion nom/prénom pour couvrir l'ensemble des occurrences dans la presse, etc.

Une solution possible serait de généraliser les tutoriels de formation à l'ensemble des plateformes. Les services d'archives l'ont d'ailleurs bien compris : nombreux sont ceux qui proposent des initiations à la généalogie ou mettent à disposition des tutoriels d'aide à la recherche, étant depuis longtemps habitués à recevoir un public amateur. Les bibliothèques numériques patrimoniales, qui partagent un public commun avec les Archives, pourraient s'inspirer davantage de ces dispositifs. Il est possible de les signaler dès la page d'accueil, comme le font notamment Gallica et RetroNews. Gallica propose, dans une vidéo explicative, un exemple de recherche complexe menée dans des journaux autour de la couverture du naufrage du Titanic dans la presse française<sup>201</sup>. Mais on pourrait envisager d'aller plus loin en proposant des modules gradués selon les niveaux, ou correspondant à différents types de requêtes : « Démarrer une recherche généalogique », « Rechercher une personne », « Retracer l'histoire d'un événement », « Que s'est-il passé le... ? », etc.

Ces tutoriels peuvent prendre plusieurs formes : fiches pratiques, foire aux questions, formulaire d'aide, capsules vidéo – voire pourquoi pas envisager, avec les évolutions technologiques actuelles, la mise en place de bots d'aide à la recherche. Un dispositif simple mais néanmoins efficace est proposé sur le site des Archives départementales de Loire-Atlantique, avec une page dédiée à la recherche dans les périodiques renvoyant vers un formulaire d'aide : le visiteur est guidé dans sa recherche en fonction des requêtes exprimées, ce qui affine au fur et à mesure les réponses apportées<sup>202</sup>.

<sup>201</sup> Chercher, trouver dans Gallica - Tutoriel n°2: « Percer les mystères du Titanic » (2021), <https://www.youtube.com/watch?v=Zi7csqvifHs> (consulté le 17 février 2023).

<sup>202</sup> Archives départementales de Loire-Atlantique, *La bibliothèque - Comment consulter une revue ou un journal ?*, [https://archives.loire-atlantique.fr/44/la-bibliotheque/c\\_20675](https://archives.loire-atlantique.fr/44/la-bibliotheque/c_20675) (consulté le 18 février 2023).

### III. Adapter les services aux besoins pour mieux valoriser la presse ancienne

Cependant, quel que soit le mode d'accompagnement choisi, un tutoriel ne remplit sa fonction que s'il est immédiatement visible par tous. Or, dans notre enquête, très peu de personnes affirment avoir eu recours à ce type de modules de formation en ligne. Est-ce par méconnaissance de ces outils, ou parce que ceux-ci ne sont pas nécessairement jugés utiles au regard du besoin ? Il nous semble en tout cas que les plateformes gagneraient à accroître encore davantage la visibilité de ces ressources de formation, par exemple en leur consacrant un onglet dédié. Mettre à disposition des tutoriels directement accessibles depuis les pages de presse pour guider l'utilisateur vers les ressources adéquates et lui dévoiler les astuces de la recherche avancée est un moyen de garantir un niveau de littératie numérique suffisant pour tous les utilisateurs<sup>203</sup>.

Cela vaut bien sûr pour les usagers débutants, mais peut tout aussi bien s'avérer utile pour les chercheurs, comme le met en évidence l'étude du consortium Couperin de 2020 : « une très forte demande de formation et d'accompagnement a été formulée, laissant entrevoir la possibilité d'un développement du rôle des professionnels de la documentation en tant qu'appui à la recherche »<sup>204</sup>. Les tutoriels en ligne sont donc une première étape pour répondre à ce besoin d'accompagnement, en complément d'éventuelles formations en présentiel ou en webinaire assurées par les bibliothèques – l'étude de Couperin soulignant notamment, de la part des chercheurs, le plébiscite des formations virtuelles<sup>205</sup>.

- **Préconisation n°2** : Renforcer la formation à la recherche dans les ressources de presse.
- **Bonnes pratiques** : Accroître l'offre de formation virtuelle (webinaires et tutoriels en ligne) ; augmenter la visibilité de ces outils d'aide sur les plateformes (onglet dédié, lien depuis les pages de presse) ; adapter les modules de formation aux différents besoins et au niveau des utilisateurs.

## 3.2. VERS UNE AMELIORATION DES OUTILS DE RECHERCHE ET D'EXPLOITATION DES CORPUS

### 3.2.1. La qualité de l'OCR, préalable indispensable à l'évolution des moteurs et fonctionnalités de recherche

Comme nous l'avons vu dans notre enquête, l'amélioration des moteurs de recherche est généralement la première requête formulée par les chercheurs, professionnels comme amateurs. La recherche plein texte sur l'ensemble des collections accessibles en ligne est la fonction la plus unanimement réclamée. La question de l'OCR est ici centrale : garantir un OCR de qualité sur les collections de presse ancienne constitue un préalable indispensable au développement d'outils

---

<sup>203</sup> La littératie numérique est définie par l'OCDE comme « l'aptitude à comprendre et à utiliser les technologies de l'information et de la communication dans la vie courante, à la maison, au travail et dans la collectivité en vue d'atteindre des buts personnels et d'étendre ses connaissances et ses capacités. » OCDE et Statistiques Canada, *La littératie à l'ère de l'information : Rapport final de l'Enquête internationale sur la littératie des adultes*, <https://www.oecd.org/fr/education/innovation-education/39438013.pdf>, 2000.

<sup>204</sup> M-P. Baligand et al., *Les pratiques de recherche documentaire des chercheurs français en 2020*, op. cit., p. 1.

<sup>205</sup> *Ibid.*, p. 35.

### III. Adapter les services aux besoins pour mieux valoriser la presse ancienne

de recherche performants. Or à l'heure actuelle, de très nombreux titres disponibles sur les plateformes en ligne ne sont pas océrisés. Parfois, leur état de conservation, typographie ou mise en page rendent très compliquée la reconnaissance des caractères ; c'est le cas notamment pour certains corpus anciens ou numérisations réalisées à partir de microfilms. Le taux de reconnaissance est encore trop souvent jugé insuffisant pour garantir la fiabilité de l'OCR. La généralisation de l'OCR à l'ensemble des collections mises en ligne, sa correction sur les titres affichant un taux de reconnaissance trop bas, ou l'océrisation rétrospective des collections numérisées, sont donc des axes d'amélioration décisifs pour favoriser l'exploitation des corpus.

Une piste possible pourrait être de développer la correction participative des collections océrisées, tel que cela a été suggéré par des répondants à notre enquête. Ces projets de *crowdsourcing* pour les journaux numérisés ont déjà cours dans plusieurs établissements à l'étranger, notamment en Australie avec le projet Trove (« *Become a Voluntrove* »<sup>206</sup>) et aux États-Unis (California Digital Newspaper Collection, Washington Digital Newspapers, ou encore Cambridge Public Library). Une expérimentation a récemment été menée en ce sens par les équipes de Gallica avec la plateforme ORCA : les lecteurs étaient invités à signaler les caractères erronés dans la version océrisée de 10 000 fascicules du quotidien *L'Excelsior*. Toutefois, appliquer ce genre d'outils de correction participative à l'ensemble de la collection de presse de Gallica semble difficilement envisageable, du fait de la quantité de pages océrisées (environ 150 millions actuellement). En revanche, une telle démarche collaborative pourrait s'envisager pour des corpus plus restreints, dans le cadre de projets de recherche spécifiques pour lesquels les gallicanautes pourraient être mis à contribution<sup>207</sup>. Des structures de taille plus modeste l'ont d'ailleurs envisagée pour leurs collections de presse ancienne ; c'est le cas de la bibliothèque numérique de Roubaix, qui a récemment mis en place ce service participatif gratuit :

Pour le moment, la Bn-R met à la disposition des internautes deux ensembles océrisés [...]. Mais c'est la presse ancienne, corpus de centaines de milliers de pages très nourries, qui a le plus besoin de votre relecture<sup>208</sup>.

Un OCR de qualité permet non seulement une recherche plein texte efficace, mais aussi le développement d'outils d'exploitation des collections (traitement automatique du langage, enrichissement sémantique, segmentation de la page, etc.). Or, avec les récentes avancées technologiques, ces outils d'humanités numériques arrivent aujourd'hui à maturité, ouvrant de riches perspectives pour la valorisation des collections de presse. Dans un article consacré aux services à la recherche à la BnF, Marie Carlin et Arnaud Laborderie expliquent qu'au-delà d'une simple transcription du texte, ces technologies sont destinées à améliorer substantiellement les fonctionnalités de consultation des collections, grâce à une meilleure identification et classification de leurs contenus. Elles permettront en effet « de naviguer à l'intérieur du document et de reconnaître les zones de contenu sur la page (segmentation), d'extraire la structure logique du document grâce à la technologie OLR (*Optical layout recognition*), d'identifier les entités nommées

---

<sup>206</sup> *Become a Voluntrove*, <https://trove.nla.gov.au/help/become-voluntrove> (consulté le 18 février 2023).

<sup>207</sup> Entretien avec Jean-Philippe Moreux, expert scientifique de Gallica à la BnF, le 14/11/2022.

<sup>208</sup> *Les services participatifs à Roubaix : comment ça marche ?*, <https://www.bn-r.fr/correction-ocr/index.php?p=tutoriel> (consulté le 18 février 2023).

(REN) »<sup>209</sup>. Ce tournant technologique est si important pour l'avenir de la valorisation de la presse qu'il mérite que l'on s'y arrête ; c'est pourquoi nous nous proposons de détailler ci-après ces technologies en cours de développement.

- **Préconisation n°3** : Garantir une bonne qualité d'OCR afin de permettre la recherche plein texte et le développement d'outils complémentaires d'exploitation des contenus.
- **Bonnes pratiques** : Généraliser l'océrisation à tous les nouveaux corpus numérisés ; intensifier les chantiers de conversion en mode texte des titres non océrisés ou le retraitement OCR de ceux qui n'atteignent pas un taux de reconnaissance suffisant ; envisager une correction participative de l'OCR sur des corpus adaptés (projets de *crowdsourcing*).

### 3.2.2. Un rôle accru à jouer pour les bibliothèques dans l'accompagnement à la recherche : les chantiers actuels

#### *Valoriser la double composante textuelle et illustrée des journaux*

On le voit, les bibliothèques ont donc un rôle de premier plan à jouer dans l'appui à la recherche, à l'heure où les humanités numériques renouvellent les approches des études médiatiques. Aujourd'hui, une demande régulièrement formulée par les chercheurs est de pouvoir passer de la recherche à la réutilisation, c'est-à-dire extraire des jeux de données des collections de presse<sup>210</sup>. On le constate d'ailleurs dans notre enquête : du côté des chercheurs de l'ESR en particulier, la demande pour ce genre d'outils est forte. La mise à disposition facilitée d'API permettant d'extraire les contenus à distance et de lancer des requêtes sur des corpus massifs est de plus en plus demandée.

En outre, des outils de recherche iconographique, d'analyse quantitative de données textuelles (textométrie) ou de rubriquage automatique (identification des zones de contenu de la page<sup>211</sup>) sont fréquemment sollicités. Une maîtresse de conférences ayant participé à notre enquête donne ainsi quelques exemples de fonctionnalités souhaitées :

- « Les outils qui me seraient utiles :
- des outils pour chercher dans des zones de texte spécifiques (par exemple dans le feuillet ou dans les annonces) ;
  - des outils pour établir des concordances entre plusieurs périodiques (identifier les reprises d'articles, par exemple) ;
  - un outil pour chercher les images dans les périodiques (par exemple les gravures de mode). »

---

<sup>209</sup> Marie Carlin et Arnaud Laborderie, « La BnF et les services à la recherche à l'heure des humanités numériques », *Arabesques*, 1 avril 2022, n° 105, p. 8-9.

<sup>210</sup> Entretien avec Jean-Philippe Moreux, expert scientifique de Gallica à la BnF, le 14/11/2022.

<sup>211</sup> Le rubriquage automatique, également appelé « catégorisation automatique » ou « classification automatique », est une technique de traitement automatique du langage naturel (TALN) qui vise à organiser automatiquement un grand nombre de documents en catégories ou en thèmes spécifiques, en se basant sur le contenu de ces documents. Le rubriquage automatique utilise des algorithmes de classification pour identifier les thèmes et les sujets abordés dans les documents, en se basant sur l'analyse de leur contenu textuel.

À l'heure actuelle, ce sont de plus en plus des approches hybrides, valorisant à la fois les illustrations et le contenu textuel des journaux, qui sont privilégiées dans les expérimentations menées. En témoigne l'outil GallicaPix développé par la BnF, prototype de recherche iconographique alliant reconnaissance optique de caractères et reconnaissance optique de la structuration des documents, pour permettre la recherche dans les éléments à la fois visuels et textuels des illustrations (le texte contenu dans un encart publicitaire par exemple)<sup>212</sup>.

#### *Enrichissement sémantique et rubriquage automatique des collections*

L'enrichissement des corpus de presse est un autre travail à mener pour accroître considérablement les possibilités d'exploitation de ces collections. L'axe à privilégier est celui de la reconnaissance des noms, qui concentrent l'essentiel des recherches professionnelles ou amateurs. Pour l'heure, une recherche de nom, aussi minutieuse soit-elle, ne peut être exhaustive. On trouve souvent dans les journaux des variantes orthographiques, dues notamment à des problématiques de translittération (« Rachmaninov » / « Rachmaninoff »), et qui ne sont donc pas détectables par les moteurs de recherche. Il existe aussi dans la langue de très nombreuses ambiguïtés sémantiques : « Washington » peut être à la fois un personnage historique, une ville ou un État. D'où les nombreuses limites de la recherche par mots-clés, qui génère tantôt trop de résultats, tantôt des résultats non pertinents, tantôt une absence de résultats. Aussi la reconnaissance d'entités nommées, qui permet de prendre en compte les différentes orthographes d'un mot et de désambiguïser les noms<sup>213</sup>, est-elle un moyen d'aller au-delà de l'OCR simple (reconnaissance d'une chaîne de caractères) pour permettre la recherche par nom de personne, lieu, organisation, etc.

Technique de traitement automatique du langage naturel, la reconnaissance d'entités nommées permet de structurer l'information contenue dans un texte et ouvre ainsi de nouvelles perspectives pour l'analyse textuelle. Ces données enrichies vont constituer dans les années à venir un apport décisif pour la recherche, en facilitant l'extraction d'informations sur un individu, un lieu ou un événement : informations contextuelles autour de cette entité, détection de l'opinion de l'auteur à son égard (analyse de sentiments), rapprochement de mots-clés sémantiquement proches (« Napoléon » et « Waterloo » par exemple), etc.

Cette caractérisation fine des contenus textuels est extrêmement précieuse pour un support aussi composite que la presse : dans une publication comprenant des articles aux sujets variés, la thématique de chaque article pourra être indexée de façon automatique, alors que jusqu'à présent, cela n'était guère envisageable manuellement. Désormais, les progrès réalisés en matière de création automatique de la structure thématique du journal permettent un accès au niveau de l'article, de la section ou de la rubrique. En permettant d'isoler les différentes composantes du fascicule (éditorial, faits divers, petites annonces, feuilleton, etc.), le rubriquage automatique offrira donc au chercheur une approche beaucoup plus fine de la presse ancienne : étude des typologies de quotidiens, de leur composante visuelle ou encore de leur structuration selon les époques, allant même jusqu'à un

---

<sup>212</sup> GallicaPix, un nouvel outil d'exploration iconographique - Le blog de Gallica, <https://gallica.bnf.fr/blog/21062021/gallicapix-un-nouvel-outil-dexploration-iconographique?mode=desktop> (consulté le 17 novembre 2022).

<sup>213</sup> En les liant à une base de connaissances : référentiels d'autorité, base de données géographiques, etc.

### III. Adapter les services aux besoins pour mieux valoriser la presse ancienne

niveau de granularité très précis (identification d'une signature, d'une légende, etc.)<sup>214</sup>.

#### *Une réalisation exemplaire : le projet NewsEye*

Projet coordonné par La Rochelle Université et mobilisant plusieurs équipes de recherche en Europe et bibliothèques nationales – dont la BnF –, NewsEye (2018-2022) a permis de développer des méthodes d'analyse pour répondre aux demandes réitérées des chercheurs sur la presse ancienne. L'objectif du projet était de créer une chaîne de traitement des journaux numérisés, de leur version scannée jusqu'à leur utilisation par les chercheurs en humanités numériques, en suivant plusieurs étapes :

1. Extraction de texte et extraction d'articles (à l'aide de technologies de reconnaissance de texte et de segmentation), afin de pouvoir exploiter les données textuelles, comprendre la structuration de la page du journal et accéder au grain de l'article en tenant compte des agencements d'informations potentiellement très différents d'un journal à l'autre ;

2. Enrichissement sémantique : reconnaissance des entités nommées à l'échelle de l'article ;

3. Analyse de texte dynamique : identification automatique de thèmes, d'opinions, de tendances ou exceptions dans un contexte donné ou dans une comparaison de contextes différenciés ;

4. Création d'un assistant personnel de recherche, collaborateur intelligent de l'utilisateur s'appuyant sur les données enrichies et outils d'analyse de texte dynamique pour l'aider à cibler sa requête – suggestion de mots-clés pertinents, identification de périodes temporelles, filtrage des articles par date ou type de point de vue, etc. –, à trier les articles et à les ajouter à son jeu de données<sup>215</sup>.

Au fur et à mesure, les fonctionnalités ont été testées par des équipes de chercheurs en sciences humaines pour affiner les méthodes et les algorithmes. La pertinence des outils a été évaluée par des cas d'études comme le port du pantalon pour les femmes<sup>216</sup> ou les interactions entre les journaux et leurs lecteurs<sup>217</sup>. La thématique du genre a d'ailleurs été valorisée à l'occasion d'une journée d'étude consacrée à la place des femmes dans la presse à la BnF<sup>218</sup>.

---

<sup>214</sup> Entretien avec Philippe Mezzasalma, chef du service Presse à la BnF, et Alexia Bauville, adjointe au chef du service Presse, le 07/11/2022.

<sup>215</sup> *Congrès Médias 19 - Numapresse : Presses anciennes et modernes à l'ère du numérique*, <https://www.bnf.fr/fr/agenda/congres-medias-19-numapresse-presses-anciennes-et-modernes-lere-du-numerique>, intervention d'Antoine Doucet et Axel Jean-Caurant (La Rochelle Université), 2-3 juin 2022 (consulté le 1 février 2023).

<sup>216</sup> *NewsEye, A Digital Investigator for Historical Newspapers - Women in Pants*, <https://www.newseye.eu/fr/case-studies/case-study-2-gender/women-in-pants/> (consulté le 19 février 2023).

<sup>217</sup> *NewsEye, A Digital Investigator for Historical Newspapers - Interaction between Newspapers and Readers*, <https://www.newseye.eu/fr/case-studies/case-study-4-media-and-journalism/interaction-between-newspapers-and-readers/> (consulté le 19 février 2023).

<sup>218</sup> *À la recherche des femmes dans la presse ancienne numérisée (1850-1950)*, <https://www.bnf.fr/fr/agenda/la-recherche-des-femmes-dans-la-presse-ancienne-numerisee-1850-1950>, 15 mars 2021 (consulté le 13 février 2023).

### *Élargir les usages de ces outils : un défi pour les bibliothèques numériques aujourd'hui*

Désormais, les outils permettent donc de visualiser, modéliser et analyser de très larges corpus de presse tout en contournant, grâce au traitement automatique, les difficultés d'indexation posées par la presse. Ces technologies sont destinées à changer en profondeur la manière d'étudier la presse ancienne. Mais la plus-value que représentent ces nouveaux outils ne concernera pas que les chercheurs professionnels : l'entrée de tout un chacun dans les fonds de presse – amateur d'histoire ou de généalogie, étudiant, ou toute personne faisant de la veille documentaire sans être experte de la recherche – s'en trouvera grandement facilitée. D'où l'intérêt de valoriser ces outils auprès de publics plus larges que les seuls spécialistes des humanités numériques : la communauté des chercheurs dans son ensemble, mais aussi tout utilisateur potentiel de presse ancienne. Aussi le projet NewsEye intègre-t-il un volet pédagogique auprès d'établissements scolaires, du grand public et des historiens amateurs.

Cependant, si des plateformes comme NewsEye, RetroNews ou Impresso en Suisse<sup>219</sup> ont pu être pensées dès leur conception pour ce type d'usages, les choses sont plus complexes avec Gallica : l'ancienneté de la plateforme, le caractère encyclopédique de ses collections et la diversité de ses publics rendent beaucoup plus difficile le passage à l'échelle. Le coût de la mise en œuvre de ces technologies est également un frein, en particulier pour la presse où l'hétérogénéité des collections complexifie l'application des modèles de reconnaissance automatique. Alors même que la demande est forte du côté des chercheurs, les moyens dégagés pour élargir le déploiement de ces outils ne sont pas suffisants. Le problème se pose de manière plus vive encore pour les portails régionaux : émerge en effet, aujourd'hui, la question de l'adaptation des moteurs de recherche lorsque d'ici quelques années, ces technologies à la fois complexes et coûteuses seront pleinement arrivées à maturité.

- **Préconisation n°4** : Renforcer les moyens accordés au développement d'outils d'enrichissement sémantique et de segmentation des collections.
- **Bonnes pratiques** : Généraliser les modèles de reconnaissance automatique pour les collections de presse (traitement OCR/OLR, reconnaissance d'images et de formes, traitement automatique du langage naturel) ; s'inspirer d'autres projets similaires en Europe (ex : Impresso) ; valoriser, par le biais de journées d'étude et autres opérations de médiation, les potentialités offertes par ces outils auprès de l'ensemble de la communauté des chercheurs mais aussi d'autres communautés utilisatrices (enseignants du secondaire, chercheurs amateurs, etc.).

### **3.2.3. Des approches innovantes pour améliorer la qualité de la médiation**

Le développement de ces outils d'humanités numériques n'aura pas seulement un impact sur la recherche : ces nouvelles technologies sont également vouées à changer la manière d'organiser la médiation autour des collections de presse ancienne. Réalisée automatiquement, l'indexation au grain de l'article

---

<sup>219</sup> *Impresso, Media Monitoring of the Past*, <https://impresso-project.ch/> (consulté le 5 décembre 2022).

### III. Adapter les services aux besoins pour mieux valoriser la presse ancienne

permettra de remédier aux difficultés d'éditorialisation auxquelles sont confrontées les bibliothèques numériques, en donnant la possibilité de générer des parcours thématiques dans les portails de presse. Des pistes prometteuses se dégagent pour la production de connaissances, et plus largement pour la valorisation des collections : « les méthodes et outils de visualisation de données sont en mesure d'aider les bibliothèques numériques à diversifier les modalités de parcours de leurs collections, en allant au-delà des classiques accès par recherche en texte intégral et feuilletage de pages », souligne Jean-Philippe Moreux<sup>220</sup>. L'article donne plusieurs exemples de visualisation interactive pouvant être utilisés pour faciliter l'entrée dans les collections ainsi que leur interprétation : représentation graphique du nombre moyen de mots par page ou du nombre d'illustrations par fascicule, mettant en évidence les événements singuliers de l'histoire d'un titre ; analyse statistique des formats de papier et de leur évolution en lien avec le contexte économique et les innovations technologiques ; frise chronologique pour enrichir les supports de médiation, etc. De même, la recherche automatique d'images pourra faciliter les actions de médiation numérique autour d'un type de périodique ou d'un titre donné. Les évolutions technologiques en cours invitent donc à réfléchir à d'autres champs d'application, dans le domaine muséal et les expositions en ligne notamment.

À l'avenir, on peut espérer que l'achèvement des projets de signalement et de numérisation de masse sur le territoire permettra aux établissements de se concentrer davantage sur des chantiers de valorisation des collections régionales. Des stratégies de fictionnalisation des archives de presse (production de podcasts, réécriture de feuilletons, mise en scène de faits divers anciens, accueil de résidences de journalistes ou d'écrivains autour d'enquêtes mobilisant des archives de journaux, conteurs, etc.<sup>221</sup>) pourront être étudiées pour en favoriser l'accessibilité et la lisibilité. Il sera possible de concevoir également des expositions thématiques ou des dispositifs utilisant des outils innovants, voire interactifs, d'exploitation des corpus. En d'autres termes, ces outils numériques ne sont pas destinés qu'à servir la communauté scientifique, mais gagnent à être employés beaucoup plus largement comme outils de découverte et d'exploration à destination de tous les publics.

Ces quelques réflexions autour de la diversification des parcours à travers les collections sont autant de pistes à explorer pour accroître le rayonnement de la future collection de référence<sup>222</sup>. Différentes opérations de valorisation sont en effet prévues autour de ces quelque 3000 titres emblématiques des collections françaises, avec une offre culturelle actuellement en cours de réflexion. Parmi les pistes évoquées : journées d'étude, cycles de manifestations sur l'histoire de la presse, accueils de classes autour de l'EMI, expositions hors les murs avec des partenaires sur l'ensemble du territoire, en particulier dans les Hauts-de-France<sup>223</sup>. Adossée à d'autres initiatives locales et nationales, comme celles menées par la future Maison du dessin de presse à Paris, cette programmation culturelle viendra

---

<sup>220</sup> J.-P. Moreux, « Approches innovantes pour la presse ancienne numérisée », art. cit.

<sup>221</sup> À ce sujet, voir les dispositifs proposés dans le cadre d'actions de médiation telles que *Les Chroniques retrouvées du Midi*, le procès du parricide de Leyr rejoué à Nancy, ou plus récemment, *Les Archichroniques* aux Archives départementales de l'Oise (dispositif de récit immersif autour de faits divers locaux).

<sup>222</sup> Cf. supra (1.1.3., p. 24).

<sup>223</sup> Archives départementales, bibliothèque municipale d'Amiens, Maison de Jules Verne, Musée de Picardie, Université de Picardie Jules Verne, etc.

renforcer encore la dynamique autour des collections de presse en France<sup>224</sup>. Dans ce contexte porteur, on peut envisager de croiser différentes modalités de médiation, à la fois physique et numérique : expositions hybrides en ligne et sur panneaux ou ateliers EMI utilisant des outils numériques pour interpréter les collections anciennes, par exemple.

- **Préconisation n°5** : Utiliser les outils d'humanités numériques au profit d'une meilleure valorisation et interprétation des collections de presse.
- **Bonnes pratiques** : S'appuyer sur des techniques de fouille, de visualisation de données et de traitement automatique du langage naturel pour améliorer la médiation numérique autour des collections (expositions virtuelles, parcours thématiques, modules éditoriaux) ; réfléchir à l'appropriation de ces nouveaux outils dans une perspective muséale (notamment autour de la future collection de référence) ; réfléchir à des approches croisées de médiation physique et numérique ; renforcer l'articulation des initiatives de valorisation nationales et locales.

### **3.3. VERS UNE COOPERATION RENFORCEE : DESENCLAVER POUR MIEUX VALORISER**

#### **3.3.1. Faciliter la recherche fédérée dans les corpus de presse : un horizon réaliste aujourd'hui**

##### *« Dérégionaliser » les corpus : une logique d'ouverture*

De plus en plus, les bibliothèques numériques sont incitées à sortir d'une logique de « silo » pour ce qui concerne la diffusion de la presse ancienne. Ce mouvement répond à un besoin exprimé par les chercheurs, qui déplorent la dissémination de l'information numérisée, les difficultés d'accès à certaines collections quand elles ne sont pas en ligne, ou des lacunes parfois dues à un manque de coordination entre établissements de conservation<sup>225</sup>. Lors des Journées du Patrimoine Écrit 2022, il a d'ailleurs été souligné que la multiplication des plateformes numériques engendrait pour l'utilisateur un certain manque de clarté. Ce sentiment est partagé par plusieurs des interlocuteurs que nous avons interrogés : à l'heure actuelle, force est de constater que cette inflation numérique ne sert pas toujours la valorisation des collections mises en ligne. L'utilisateur qui entreprend une recherche se doit de mener une veille documentaire active sur plusieurs portails, dont les moteurs de recherche ne donnent pas toujours entière satisfaction. Au sein de la sphère BnF, la distinction entre Gallica et RetroNews n'est pas forcément bien opérée par les utilisateurs, qui font parfois part de leur difficulté à appréhender ce paysage documentaire complexe. Certains insistent également sur l'importance d'une meilleure communication sur les calendriers de numérisation et de mise en ligne, ce qui implique un dialogue constant entre les établissements.

Cette logique de coopération et d'ouverture est également souhaitée par les chercheurs à l'échelle transnationale : améliorer la communication entre les institutions de conservation de différents pays favoriserait des recherches sur des

---

<sup>224</sup> Entretien avec Julie Ladant, directrice du département Droit, économie, politique à la BnF, le 13/02/2023.

<sup>225</sup> Cf. Annexe n°4 : « Typologie des besoins exprimés par les répondants ».

### III. Adapter les services aux besoins pour mieux valoriser la presse ancienne

thématiques beaucoup plus transversales. Pour Guillaume Pinson, le cloisonnement des corpus régionaux ou nationaux présente même un risque de biais pour la recherche, en « recréant l'illusion rétrospective de systèmes indépendants, strictement repliés à l'intérieur de frontières closes » :

La stratification régionale des bibliothèques numériques d'intérêt local (souvent associées à des sites d'archives départementales ou à des bibliothèques municipales) peut aussi conduire à des perceptions artificiellement compartimentées. [...] Paradoxalement, il y a donc un risque que l'accessibilité numérique engendre un phénomène de reconduction de la vieille perspective nationale de l'histoire littéraire<sup>226</sup>.

Or, cela va à l'encontre de la nature même du journal, objet de circulation par excellence. Étudier la presse, c'est aussi mettre en lumière ces logiques de réseaux, parfois sur de vastes échelles abolissant les frontières nationales. Plutôt que de reconduire ces frontières, le numérique devrait donc, au contraire, donner l'occasion au chercheur d'interroger ces circulations sur de larges corpus « dérégionalisés » voire « dénationalisés » :

Rien ne circulait davantage au XIXe siècle que les objets médiatiques. Utilisé correctement, le numérique offre une occasion inespérée de se replonger dans une culture du lien et d'interroger les imaginaires de la spatialité. Il ne s'agit donc pas de nier le caractère « national » ou « régional » des corpus, ce qui serait absurde, mais de comprendre [...] de quelle manière les journaux étaient fondés au croisement de multiples réseaux de circulations et d'échanges<sup>227</sup>.

#### *Vers la constitution d'une collection virtuelle unifiée ?*

Tout l'enjeu d'un tel désenclavement serait de sortir d'une logique de recherche titre à titre, afin de permettre une recherche fédérée sur des journaux de provenance diverse. On peut imaginer qu'une recherche de nom dans le cadre d'un travail généalogique est bien plus efficace si l'on est en mesure de consulter la presse de diverses régions, sans avoir à relancer plusieurs fois sa requête dans les portails locaux concernés. De même si l'on mène des recherches sur un journaliste en particulier : celui-ci peut très bien avoir publié dans plusieurs journaux différents. De manière générale, notre enquête révèle à quel point les objets d'étude actuels – surtout ceux des chercheurs de l'ESR, mais aussi dans une moindre mesure ceux des amateurs – sont transversaux<sup>228</sup>. Cela contraint donc les bibliothèques numériques à s'adapter à ces nouveaux besoins de recherche.

Or, les moyens technologiques permettent désormais d'envisager une telle consultation fédérée. La généralisation du standard IIIF, visant à améliorer l'accès et l'interopérabilité des images numériques, offre de riches perspectives de

---

<sup>226</sup> G. Pinson, « Lire la presse du XIXe siècle en contexte numérique », art. cit.

<sup>227</sup> *Ibid.* Voir aussi Micheline Cambron et Stéphanie Danaux, « La recherche sur la presse : nouveaux bilans nationaux et internationaux », *Médias 19*, <https://www.medias19.org/publications/la-recherche-sur-la-presse-nouveaux-bilans-nationaux-et-internationaux/presentation-du-dossier-problematiques-enjeux-outils-de-la-recherche-actuelle>, 2021.

<sup>228</sup> Voir par exemple l'étude de Jean-Luc Buard sur l'œuvre du feuilletoniste Marie Aycard et sa propagation internationale. L'auteur y explore non seulement les logiques de réseaux de la presse quotidienne du milieu du XIXe siècle et le statut précaire de l'auteur salarié, mais aussi les possibilités de la culture médiatique contemporaine, permettant la mise en évidence de ces logiques de séries sur une très large échelle de titres. Jean-Luc Buard, *Culture médiatique et presse numérisée : Médiasphère des feuilletons-nouvelles de Marie Aycard (1794-1859)*, Presses de l'Enssib, 2019, 234 p.

### III. Adapter les services aux besoins pour mieux valoriser la presse ancienne

reconstitution virtuelle de collections complètes<sup>229</sup>. Selon Emmanuelle Bermès, ce protocole d'interopérabilité est destiné à « bouleverser l'accessibilité des collections patrimoniales » et à « faire bouger les lignes dans la manière dont on construit des bibliothèques numériques partagées »<sup>230</sup>. Constituer virtuellement une collection de presse complète et homogène à l'échelle nationale devient donc une perspective atteignable.

Un tel objectif nécessite des efforts renforcés en faveur de la conservation partagée et de la numérisation concertée des titres en région. Cela implique également un important travail d'harmonisation des métadonnées, afin de faciliter l'agrégation des ressources au sein d'une interface unifiée – et *in fine*, de fluidifier la recherche au sein de ces collections. La collection de référence nationale entend être un premier pas vers cet idéal de collection unique et homogène, entièrement océrivée et alignée au niveau des métadonnées, afin d'offrir de meilleures possibilités d'exploitation et de valorisation. À terme, l'objectif de cette mutualisation documentaire est la création d'un réservoir de données dans lequel il sera possible, à l'aide de moteurs de recherche puissants, de proposer de la fouille de texte et d'images sur la totalité des corpus fédérés<sup>231</sup>. L'idée n'est pas nouvelle, puisqu'en 2009, le Schéma numérique des bibliothèques préconisait déjà de rassembler les titres numérisés au sein d'un portail spécifique pour « donner un accès global aux collections de presse disséminées dans diverses institutions »<sup>232</sup>.

Ce contexte explique le mouvement de fond que l'on observe actuellement en faveur de la plateforme nationale Gallica. En témoignage, notamment, le succès du dispositif des « marques blanches », qui englobe de plus en plus d'établissements – l'avantage étant pour eux de pouvoir disposer du moteur de recherche de Gallica, tout en gardant leur identité géographique propre. La subvention octroyée par la BnF à ses partenaires pour les chantiers de production numérique, rétroconversions et intégration dans Gallica favorise ce mouvement. De plus, les solutions techniques offertes aux collectivités optant pour l'intégration dans Gallica – en particulier l'archivage pérenne – donnent à ces dernières la possibilité de se consacrer à d'autres chantiers, comme le comblement des lacunes de leurs collections. Cette tendance à la mise en commun n'empêche pas que le portail régional conserve toute sa pertinence pour la valorisation patrimoniale des collections de presse locale – par le biais de l'éditorialisation et de la médiation notamment.

- **Préconisation n°6** : Favoriser le désenclavement, l'ouverture et la centralisation des ressources pour faciliter une recherche fédérée sur plusieurs titres.
- **Bonnes pratiques** : Utiliser les possibilités permises par les protocoles d'interopérabilité pour reconstituer virtuellement des collections complètes et homogènes ; favoriser l'agrégation des ressources au sein d'un portail national unique ; veiller à l'alignement et à l'enrichissement des métadonnées pour permettre une meilleure valorisation des collections.

---

<sup>229</sup> IIF a aussi l'avantage, en ces temps de sobriété numérique, d'éviter la diffusion d'un double corpus. *Journées du Patrimoine Écrit 2022 : Numériser les collections de presse*, intervention de Mathieu Stoll, chef du bureau de la diffusion et de la valorisation numérique au Service interministériel des Archives de France, 23-24 juin 2022.

<sup>230</sup> Conclusion de la journée d'étude « Les bases de données iconographiques à l'heure de l'intelligence artificielle : enjeux, état des lieux et perspectives » à l'École des chartes, le 1<sup>er</sup> décembre 2022.

<sup>231</sup> Entretien avec Arnaud Dhermy, chef de la mission de la coopération régionale à la BnF, le 29/09/2022.

<sup>232</sup> B. Racine, *Schéma numérique des bibliothèques*, op. cit., p. 65.

### 3.3.2. Intensifier les collaborations avec les partenaires académiques et les bibliothèques étrangères

#### *La nécessité d'un travail conjoint avec les chercheurs et structures de recherche*

Mais l'appui à la recherche dans les collections de presse ne pourra se faire sans un dialogue renforcé avec les partenaires académiques eux-mêmes. Cela implique, de la part des établissements de conservation comme des structures de recherche, de sortir de leurs silos professionnels respectifs. Il est donc crucial de créer ou de consolider des partenariats avec les laboratoires et équipes de recherche spécialisés sur la presse ancienne. Si la BnF est déjà très impliquée sur ces questions, ce type de collaborations mérite d'être engagé aussi en région. L'Occitanie explore actuellement cet axe de valorisation de ses journaux locaux grâce à la recherche et envisage une collaboration plus étroite avec le laboratoire RIRRA 21 de l'Université Paul-Valéry-Montpellier<sup>233</sup>. On pourrait tout à fait imaginer que d'autres régions prennent appui sur des structures de recherche soutenant le développement des humanités numériques, comme la MISHA à Strasbourg par exemple<sup>234</sup>. Organiser des rencontres entre professionnels de la documentation et de la recherche (à la manière de ce que propose l'agence de coopération Normandie Livre et Lecture autour de fonds littéraires<sup>235</sup>), favoriser les publications conjointes, sont des pistes à explorer. L'ouverture prochaine du Conservatoire national de la presse devrait favoriser ce type d'interactions, avec l'Université de Picardie Jules Verne en particulier<sup>236</sup>.

Dans le sillage du mouvement de la science ouverte, les bibliothèques gagneront à s'investir davantage sur le terrain de la valorisation des résultats de la recherche, notamment le travail des jeunes chercheurs. Les dispositifs d'accueil de chercheurs associés et chargés de recherche documentaire (CRD) à la BnF sont particulièrement propices à ce type de valorisation, comme en témoignent certains travaux de recherche ayant donné lieu à des contributions dans Gallica ou à des journées d'étude<sup>237</sup>. Capitaliser sur ce travail de recherche et de structuration de la collection de presse, donner à voir la production intellectuelle autour de ces archives, est un axe à consolider, à la BnF comme dans les territoires. L'octroi d'allocations doctorales financées en partie par les collectivités est un bon moyen de stimuler ces travaux de recherche – et, partant, de revaloriser les fonds régionaux. On peut en prendre pour exemple la région Occitanie, qui a récemment alloué une bourse doctorale à une jeune chercheuse afin de travailler, à l'aide des humanités numériques, sur la fabrique du roman-feuilleton occitan dans le corpus de presse régionale<sup>238</sup>.

<sup>233</sup> Entretien avec Mélanie Marchand, chargée de mission Bibliothèques & Patrimoine (Occitanie Livre & Lecture), le 27/10/2022.

<sup>234</sup> Plateforme Humanités Numériques | PHUN - MISHA - Maison Interuniversitaire des Sciences de l'Homme - Alsace - Université de Strasbourg, <https://www.misha.fr/plateformes/phun> (consulté le 12 février 2023).

<sup>235</sup> Voir à ce sujet la thématique « Recherches et valorisation des fonds littéraires » mise en avant à l'occasion des Journées du Patrimoine Écrit 2022.

<sup>236</sup> Entretien avec Julie Ladant, directrice du département Droit, économie, politique à la BnF, le 13/02/2023.

<sup>237</sup> Cf. supra (1.2.3., p. 30).

<sup>238</sup> Cette thèse en préparation de Jade Pétrault a pour titre « Le laboratoire du roman-feuilleton (1850-1940) en Occitanie ». Elle vise à « analyser les effets de reprise, de circulation et d'invention d'un siècle de romans-feuilletons en Occitanie, à partir d'un corpus de titres quotidiens de la région (*Le Messager du Midi*, *Le Courrier du Gard*, *Le Courrier de l'Aude*, *L'Éclair*, etc.) numérisés par la plateforme régionale RessourcesPatrimoines. [...] Il s'agit de mettre en

### III. Adapter les services aux besoins pour mieux valoriser la presse ancienne

À mesure qu'ils se complexifient, les projets d'exploitation de la presse ancienne requièrent un volet technique important et nécessitent des compétences de plus en plus exigeantes, que les établissements documentaires seuls ne sont pas en mesure d'assumer. D'où l'importance de s'appuyer davantage sur les expertises apportées par les équipes de recherche. C'est en ce sens que la BnF a développé un nouveau service d'appui à la recherche, le BnF DataLab, qui mutualise l'expertise documentaire de bibliothécaires et ingénieurs d'étude et celle des chercheurs soumettant leurs projets<sup>239</sup>.

#### *Favoriser l'ouverture des données : l'appui des infrastructures européennes*

Renforcer le dialogue avec les structures de recherche implique aussi d'œuvrer pour une mise à disposition facilitée et plus explicite de jeux de données, comme de plus en plus de chercheurs le demandent<sup>240</sup>. La BnF travaille actuellement à l'ouverture et à l'enrichissement des collections d'API de Gallica, notamment en utilisant le standard d'interopérabilité IIIF pour faciliter l'échange, l'extraction et la réutilisation de ces données<sup>241</sup>. Mais la rencontre entre ces services et leurs potentiels utilisateurs ne pourra se faire que si la bibliothèque va au-devant des chercheurs, en communiquant auprès des structures concernées et en formant efficacement les usagers. La réflexion doit se poursuivre hors les murs : référencement des jeux de données sur des plateformes de dépôt de données telles que Zenodo<sup>242</sup> (ou des entrepôts disciplinaires), mais aussi communication sur cette offre de services et formation des chercheurs à leur utilisation. Il est important que la formation vise en particulier les jeunes chercheurs, utilisateurs en puissance de ces nouveaux services à la recherche ; or pour l'heure, trop peu de doctorants y sont sensibilisés. Il serait donc judicieux de renforcer la communication auprès des écoles doctorales et collèges doctoraux.

De surcroît, pour un objet d'étude aussi transversal que la presse, il est plus que jamais nécessaire d'étendre la dynamique de coopération et l'ouverture des données à l'échelle européenne, comme cela a pu être testé dans le cadre du projet NewsEye. Afin de permettre à des équipes transnationales de travailler sur des corpus européens, il convient d'aller au-delà des simples opérateurs nationaux et de leurs silos technologiques. Il faudrait s'appuyer davantage sur des infrastructures européennes de support à la recherche telles que DARIAH ou CLARIN, qui permettent de partager des jeux de données, des corpus textuels massifs, des savoir-faire et des méthodologies. Spécialisée dans les ressources textuelles, l'infrastructure CLARIN semble particulièrement adaptée pour le partage de données et de méthodes autour de la presse ancienne – d'autant plus

---

évidence les logiques des circuits de reprise nationaux et internationaux de ces romans-feuilletons ; d'étudier la poétique, les conditions d'autonomisation et la manière dont cette production occitane investit et met en scène le territoire régional. » *Theses.fr - Le laboratoire du roman-feuilleton (1850-1940) en Occitanie*, Thèse en préparation, Montpellier 3, 2021.

<sup>239</sup> BnF DataLab, <https://www.bnf.fr/fr/bnf-datalab>, 15 octobre 2021 (consulté le 20 février 2023).

<sup>240</sup> Cf. Annexe n°4 : « Typologie des besoins exprimés par les répondants ».

<sup>241</sup> API et jeux de données, <https://api.bnf.fr/> (consulté le 6 mai 2022).

<sup>242</sup> Zenodo est un répertoire de travaux de recherche, de logiciels et de données créé par le CERN en 2013, grâce au programme-cadre européen pour la recherche et le développement Horizon 2020. Par exemple, les jeux de données produits dans le cadre du projet NewsEye sont disponibles sur Zenodo et archivés dans le référentiel OpenAIRE de la Commission européenne.

### III. Adapter les services aux besoins pour mieux valoriser la presse ancienne

depuis que l'exception en faveur de la fouille de textes et de données à des fins de recherche a été introduite dans la loi française, ouvrant de nouvelles perspectives d'exploitation des corpus, y compris pour des contenus encore sous droits<sup>243</sup>. Les bibliothèques doivent désormais porter cette dynamique d'ouverture à l'échelon international.

- **Préconisation n°7** : Intensifier les collaborations avec les équipes de recherche spécialisées sur la presse ancienne, valoriser les résultats des travaux et favoriser la réutilisation des jeux de données produits dans le cadre de ces recherches.
- **Bonnes pratiques** : Stimuler les travaux de recherche, notamment par l'accueil de chercheurs et l'octroi de bourses doctorales ; valoriser les résultats de la recherche par l'organisation de journées d'étude ou publications coécrites par des chercheurs et professionnels de la documentation ; décrire et mettre à disposition des jeux de données sur des plateformes académiques dédiées ; mieux communiquer autour des nouveaux services à la recherche, notamment auprès des écoles doctorales ; renforcer la formation des chercheurs à ces nouveaux outils et méthodes.

#### *Coopérer avec les bibliothèques européennes*

Le renforcement de la coopération internationale peut aussi se faire à la faveur de projets Europeana, tels qu'*Europeana Newspapers* ou *Rise of Literacy*. Nous retiendrons pour exemple le programme *Europeana Collections 1914-1918 : Remembering the First World War*, projet de numérisation concertée entre une dizaine de bibliothèques européennes<sup>244</sup> lancé dans le cadre des commémorations du centenaire de la Grande Guerre, et auquel la Bibliothèque nationale et universitaire de Strasbourg a pris part. Pour des corpus intrinsèquement liés à l'histoire de l'Europe comme les collections de guerre liées à l'histoire allemande de la BNU (*Kriegssammlungen*), on voit bien la pertinence d'une coopération à plusieurs niveaux et la nécessité de sortir des frontières nationales. Les documents numérisés dans le cadre d'*Europeana Collections 1914-1918*, sous licence ouverte, sont désormais hébergés par chacune des institutions partenaires mais aussi accessibles via un service commun afin de permettre leur libre réutilisation. Les ressources peuvent être consultées selon leur nature, leur thématique (les prisonniers de guerre, les femmes, la propagande, etc.) ou selon les fronts (Front de l'Est, front italien, etc.).

De tels exemples de coopération européenne sont à encourager, car ils permettent d'enrichir la valorisation des collections grâce à l'expertise des différents établissements partenaires. On pourrait, par exemple, envisager de renforcer les collaborations entre la BNU et les bibliothèques allemandes et autrichiennes, afin de s'appuyer sur leur expertise en matière de numérisation et d'océrisation des caractères gothiques (*Fraktur*). Une autre piste serait d'envisager une double diffusion des corpus de presse de régions transfrontalières, à la fois sur une plateforme française et étrangère, pour en accroître la visibilité auprès des chercheurs comme du grand public. La BNU réfléchit à un tel partenariat avec la

---

<sup>243</sup> *Légifrance*, Ordonnance n° 2021-1518 du 24 novembre 2021 complétant la transposition de la directive 2019/790 du Parlement européen et du Conseil du 17 avril 2019 sur le droit d'auteur et les droits voisins dans le marché unique numérique, <https://www.legifrance.gouv.fr/jorf/id/JORFTEXT000044362034> (consulté le 20 février 2023).

<sup>244</sup> Notamment la *Stadtbibliothek* de Berlin, la *British Library*, la Bibliothèque royale de Belgique, la Bibliothèque nationale serbe, ou encore la Bibliothèque nationale autrichienne.

### III. Adapter les services aux besoins pour mieux valoriser la presse ancienne

*Deutsche Digitale Bibliothek*, la presse alsacienne et mosellane étant susceptible d'intéresser un public outre-Rhin autant qu'un public français<sup>245</sup>. Certains portails étrangers diffusent d'ailleurs des journaux en français : c'est le cas notamment de *Cordela*, la bibliothèque numérique de la Vallée d'Aoste, qui contient tous les périodiques publiés dans la région depuis le XIXe siècle, dont de nombreux titres en français<sup>246</sup>. Des rapprochements transfrontaliers pourraient ainsi être étudiés.

- **Préconisation n°8** : Intensifier la coopération à l'échelle européenne.
- **Bonnes pratiques** : Collaborer dans le cadre de projets de recherche, de numérisation ou de valorisation européens tels que ceux portés par Europeana ; s'appuyer sur les expertises des bibliothèques étrangères ; envisager une double diffusion pour les corpus transfrontaliers.

### 3.3.3. S'appuyer sur les communautés de pratique

Enfin, il nous semble important d'intensifier les coopérations avec les communautés de pratique, qu'il s'agisse de communautés de chercheurs professionnels ou amateurs. Définie par Étienne Wenger et Jean Lave comme un groupe de personnes partageant « un intérêt commun envers la même pratique, ainsi qu'une volonté partagée d'améliorer leur expertise et leur expérience de celle-ci », la notion de communauté de pratique repose sur le concept d'apprentissage social, qui relève d'un « processus participatif et collectif au cours duquel les interactions sociales, le sentiment d'appartenance et l'engagement réciproque se développent »<sup>247</sup>. Lors de notre enquête, nous avons constaté à quel point certaines de ces communautés avaient joué un rôle important en s'impliquant activement dans l'étude et sa diffusion. Plusieurs répondants nous ont ainsi indiqué avoir eu connaissance de notre enquête via des associations d'entraide généalogique locales – le relais s'étant fait le plus souvent via les services d'archives, ou via des réseaux sociaux de généalogistes. Cela confirme l'importance de tels relais, en particulier auprès du public amateur.

Les bibliothèques ont donc tout intérêt à s'appuyer sur ces communautés très structurées à l'échelon local : sociétés savantes, associations de généalogistes et d'historiens amateurs, associations d'amis d'écrivains, etc. Du fait de la dimension souvent très spécialisée de leurs sujets d'étude, les membres de ces associations disposent d'une réelle expertise sur certains fonds de presse, et leur concours pourrait être davantage sollicité pour enrichir les métadonnées des collections à la faveur de campagnes participatives (indexation, correction d'OCR, etc.). Certains établissements n'hésitent pas non plus à solliciter la contribution de particuliers pour compléter les lacunes de leurs collections afin de numériser les titres manquants. C'est le cas des Archives des Yvelines avec leur campagne « À la

---

<sup>245</sup> Entretien avec Jérôme Schweitzer, directeur du pôle Partage, diffusion et réseaux à la Bibliothèque nationale et universitaire de Strasbourg, le 07/12/2022.

<sup>246</sup> La plupart de ces journaux, jusqu'au fascisme, était en français car la Vallée d'Aoste faisait partie des États de Savoie. Nous remercions Donato Arcaro de nous avoir signalé l'existence de cette plateforme. *Cordela*, <http://cordela.regione.vda.it/> (consulté le 20 février 2023).

<sup>247</sup> Oihab Allal-Chérif et Lise Arena, « Étienne Wenger - Les communautés de pratique au service de la connaissance située et de l'apprentissage social » dans *Les Grands Auteurs en Systèmes d'information*, Caen, EMS Editions, 2018, p. 462-483.

### III. Adapter les services aux besoins pour mieux valoriser la presse ancienne

recherche de la presse perdue », appelant les citoyens à participer au projet de constitution d'une collection numérique de référence<sup>248</sup>.

Mobiliser ces communautés autour de la valorisation de leur patrimoine local passe aussi par une logique d'ouverture et de dissémination du fonds patrimonial. Les communautés d'amateurs passionnés sont, de fait, souvent très actives sur des plateformes participatives extérieures aux portails régionaux : Flickr pour le partage de photographies, Heredis pour la construction d'arbres généalogiques, etc. L'exemple du programme *Flickr Commons*, qui permet aux organismes institutionnels de diffuser leurs contenus iconographiques auprès des communautés utilisatrices du site, peut tout à fait s'envisager pour la presse ancienne, en particulier pour valoriser sa composante illustrée. C'est le choix qu'a fait La contemporaine, qui a versé dans des albums Flickr un certain nombre de pages de journaux et de Unes illustrées<sup>249</sup>. Les membres peuvent dès lors contribuer à l'indexation de ces documents à l'aide de « tags » (étiquettes de mots-clés) et de commentaires. Les bibliothèques gagneraient à s'inscrire davantage sur ce terrain de la mise en commun, de la participation et de l'élaboration collective de la connaissance, afin de permettre aux individus de s'approprier les collections et de concevoir eux-mêmes des outils de valorisation : identification de certains contenus, diffusion de photographies de presse ou coupures de presse via des blogs d'amateurs, etc.<sup>250</sup>.

Dans *Les bibliothèques et la médiation des connaissances*, Bertrand Calenge invite à inscrire les dispositifs de médiation des connaissances « dans les pratiques [des personnes concernées] et dans la communauté qu'elles constituent »<sup>251</sup>. On peut prendre comme exemple de ce type de dispositif le Challenge AZ lancé par la généalogiste Sophie Boudarel et désormais géré par l'association Geneatech : ce défi mêlant généalogie et écriture invite les généalogistes amateurs à présenter leur généalogie sous la forme de billets de blog, en suivant les lettres de l'alphabet. Les portails régionaux pourraient valoriser ces productions éditoriales en encourageant leurs lecteurs à participer à de tels défis d'écriture, voire en publiant les publications de qualité sur leur site. Selon Lionel Dujol, « la bibliothèque participative est participante », et l'engouement autour du patrimoine et des collections de presse ancienne est un bon moyen de retendre ce lien entre la bibliothèque et ses communautés d'utilisateurs<sup>252</sup>. La bibliothèque peut demander aux citoyens de participer à ce qu'elle détient dans ses collections, tout autant qu'elle peut participer en retour à ce qui anime les communautés.

Tout l'enjeu est alors la mise à disposition, pour ces publics, de lieux d'échange et de partage entre pairs : forums de discussion ou espaces collaboratifs sont des idées à explorer, comme cela a été demandé par certains participants à notre enquête. Dans *L'impératif participatif*, Marta Severo encourage ainsi le développement, aux côtés des plateformes institutionnelles et amateurs, de

---

<sup>248</sup> À la recherche de la presse perdue, <https://archives.yvelines.fr/participer/a-la-recherche-de-la-presse-perdue> (consulté le 12 février 2023).

<sup>249</sup> Flickr - Albums de La contemporaine, <https://www.flickr.com/photos/98069937@N05/albums/> (consulté le 20 février 2023).

<sup>250</sup> Lionel Dujol, *Communs du savoir et bibliothèques*, Paris, France, Éditions du Cercle de la librairie, 2017.

<sup>251</sup> B. Calenge, *Les bibliothèques et la médiation des connaissances*, op. cit.

<sup>252</sup> Entretien avec Lionel Dujol, responsable prospective, innovation et accompagnement au changement (Valence Romans Agglo, Direction Lecture Publique), le 16/11/2022.

### III. Adapter les services aux besoins pour mieux valoriser la presse ancienne

« plateformes contributives culturelles »<sup>253</sup>. L'étude de publics menée par le SIAF et la BnF dresse aussi ce constat :

En dépassant la communauté des professionnels, la question est celle de la mise en œuvre d'espaces de travail réunissant les différentes expertises des amateurs, des chercheurs académiques et des professionnels, garantissant une montée en compétences de tout un chacun pour l'usage des « communs » que sont les patrimoines numériques et numérisés<sup>254</sup>.

Ménager des espaces d'entraide, qu'il s'agisse d'ateliers de formation par les pairs, de groupes sur les réseaux sociaux ou de forums d'échange via les portails patrimoniaux, est essentiel pour servir de relais aux actions de médiation numérique et renforcer l'aide à la prise en main des outils. L'accompagnement doit se prolonger hors les murs, en misant sur la puissance des communautés pour encourager le partage de méthodes et d'expériences entre pairs et, par là même, favoriser les interactions autour de ce patrimoine commun que constitue la presse ancienne.

- **Préconisation n°9** : Renforcer la prise en compte des communautés de pratique dans la valorisation des collections de presse, en particulier à l'échelon local.
- **Bonnes pratiques** : Identifier les communautés d'« amateurs éclairés » pouvant servir de relais aux actions de valorisation menées par les établissements en région ; généraliser les campagnes participatives d'enrichissement des métadonnées (notamment grâce à la mise en place de plateformes contributives dédiées) ou de comblement des lacunes des collections locales ; disséminer les collections libres de droits sur des plateformes de partage prisées par les amateurs ; mettre en place des forums de discussion et d'entraide (cf. forum Pages 14-18).

---

<sup>253</sup> Marta Severo, *L'impératif participatif: institutions culturelles, amateurs et plateformes*, Bry-sur-Marne, France, INA, 2021, p. 8.

<sup>254</sup> N. Couillard et I. Bastard, *Qui sont les usagers en ligne communs entre les Archives et la BnF ?*, op. cit., p.14.

## CONCLUSION

---

La numérisation massive des corpus de presse offre aujourd’hui des possibilités inédites, pour les bibliothèques, de travailler main dans la main avec les communautés professionnelles et amateurs. Si le numérique marque un tournant dans les usages des lecteurs, il bouleverse également la manière d’organiser la médiation et l’offre de services autour de la recherche dans ces collections. Les bibliothèques numériques doivent s’adapter à de nouveaux modes de consultation et d’analyse de la presse : la « lecture distante », telle que définie par Franco Moretti dans les années 2000, constitue désormais une tendance significative. Pour autant, les modes de lecture plus traditionnels – recherche d’un nom dans quelques journaux, travail sur des textes singuliers, etc. – restent encore très prégnants. Les usages se sont diversifiés et avec eux, les besoins. Il est donc crucial d’accompagner ces évolutions en dégagant les moyens nécessaires pour accroître encore la numérisation et la diffusion de ces corpus.

Au-delà, il convient aussi de réfléchir à l’efficacité des modes de diffusion. La réflexion autour de l’ergonomie des interfaces est à poursuivre : favoriser la rencontre entre l’usager et la presse ancienne passe aussi par des dispositifs de médiation adaptés à ces collections complexes par nature. L’accès cartographique dans les fonds pour mettre en avant l’approche territoriale de ces ressources, l’adaptation des filtres de recherche aux spécificités de la presse et la mise en place de tutoriels sont des pistes à explorer. En matière de diffusion, nous avons vu que l’échelon à privilégier dépendait largement de l’objectif de valorisation. Dans une optique de mise en valeur d’un patrimoine local, le portail régional est une porte d’entrée pertinente pour organiser une médiation numérique autour des collections et produire des contenus éditoriaux propres à intéresser les habitants. En revanche, l’échelon local n’est pas le plus adapté pour la recherche. Plus que jamais aujourd’hui, il est important d’aider le chercheur à se retrouver dans ce paysage complexe, caractérisé par l’abondance informationnelle et la dispersion des ressources. Selon l’objectif visé, deux logiques peuvent alors être adoptées : d’un côté la dissémination des fonds de presse, qu’il peut être pertinent de « désenclaver » pour aller au-devant de communautés susceptibles de s’y intéresser voire de contribuer à leur enrichissement ; de l’autre, la mutualisation et la centralisation des ressources par moissonnage et intégration dans le portail national.

Outre un enjeu de clarification du paysage de recherche, disposer d’un tel réservoir de ressources permettrait d’obtenir une collection de presse complète et homogène, et de rendre plus opérantes les techniques de fouille. Ce mouvement va de pair avec l’amélioration constante des outils favorisant l’extraction de données à une échelle de plus en plus fine, mais aussi leur analyse et leur interprétation. Si les prototypes se heurtent encore actuellement à la difficulté du passage à l’échelle, les projets en cours et leurs récentes avancées permettent d’espérer une démocratisation de ces technologies dans les années à venir. L’horizon ultime serait de pouvoir envisager l’interrogation des moteurs de recherche non plus seulement en mode texte mais aussi à la manière d’une banque d’images, ce qui valoriserait ainsi le double contenu textuel et illustré du journal.

Ces nouveaux outils ouvrent de très larges perspectives pour l’indexation et l’enrichissement des fonds de presse. Tout l’enjeu, désormais, est de les faire

connaître à un plus vaste public en mettant en évidence leur potentiel en matière de médiation numérique autour de ces collections. Comme le soulignait dès 1997 Roland Schaer, premier directeur de la diffusion culturelle à la BnF, avec le numérique, « le souci éditorial traverse plus profondément qu'autrefois l'activité centrale de la bibliothèque »<sup>255</sup>. La bibliothèque ne doit pas être qu'un lieu de recherche érudite : plus que jamais, son rôle d'éditorialisation doit être revalorisé. La future collection de référence sera l'occasion de repositionner la singularité de la bibliothèque comme lieu : lieu de médiation physique et numérique, lieu de rencontre autour de ces collections qui, depuis le passé, éclairent l'époque contemporaine.

---

<sup>255</sup> Roland Schaer, « Sur la politique culturelle de la Bibliothèque nationale de France », *Bulletin des bibliothèques de France* (BBF), 1997, n° 6, p. 18-22.

## SOURCES

---

### Liste des entretiens menés

- Morgane AVELLANEDA, Attachée temporaire d'enseignement et de recherche à l'Université de Strasbourg / IHRIM, ancienne Chargée de recherche documentaire au service Presse, Bibliothèque nationale de France (entretien du 3 juin 2022) ;
- Irène BASTARD, Cheffe de projet Publics et usages à la BnF, Délégation à la stratégie et à la recherche, Bibliothèque nationale de France (entretien du 5 octobre 2022) ;
- Alexia BAUVILLE, Adjointe au chef du service Presse, Bibliothèque nationale de France (entretien du 7 novembre 2022) ;
- Nicolas BEAUPRÉ, Professeur en histoire contemporaine et chercheur rattaché au centre Gabriel Naudé, Enssib (entretien du 7 mars 2022) ;
- Frédéric BLIN, Responsable du pôle Services et Collections, Bibliothèque nationale et universitaire de Strasbourg (entretien du 29 juillet 2022) ;
- Sophie BOUDAREL, Généalogiste professionnelle et créatrice du blog *La Gazette des Ancêtres* (entretien du 11 octobre 2022) ;
- Gaétan CARLIER, Conservateur stagiaire en charge de la préparation à la numérisation du *Progrès*, Bibliothèque municipale de Lyon (entretien du 14 novembre 2022) ;
- Philippe CHEVALLIER, Adjoint au responsable de la coordination de la recherche, Délégation à la stratégie et à la recherche, Bibliothèque nationale de France (entretien du 31 janvier 2022) ;
- Émilie CHEVALME, Chargée de mission Patrimoine Écrit, Mobilis - Pôle régional de coopération des acteurs du livre et de la lecture en Pays de la Loire (entretien du 12 octobre 2022) ;
- Noémie COUILLARD, Co-fondatrice et directrice scientifique de l'agence de recherche Voix/Publics (entretien du 18 novembre 2022) ;
- Jacques DEVILLE, Conseiller Livre et Lecture, Patrimoine écrit et archives, DRAC Nouvelle-Aquitaine (entretien du 16 mai 2022) ;
- Arnaud DHERMY, Chef de la mission de la coopération régionale, Bibliothèque nationale de France (entretien du 29 septembre 2022) ;
- Lionel DUJOL, Responsable prospective, innovation et accompagnement au changement, Valence Romans Agglo, Direction Lecture Publique (entretien du 16 novembre 2022) ;
- Mathilde DUTERTRE, Adjointe au chef du service Numérisation, Bibliothèque nationale de France (entretien du 23 novembre 2022) ;
- Nathalie FABRY, Chargée du traitement rétrospectif des collections de périodiques et responsable du site Presse locale ancienne, Bibliothèque nationale de France (entretien du 4 octobre 2022) ;
- Marie-Noëlle GOURDEL, Cheffe de projet numérisation, Bibliothèque nationale de France (entretien du 23 novembre 2022) ;

- Matthieu GUENOUX, Bibliothécaire, Archives départementales de Loire-Atlantique (entretien du 26 octobre 2022) ;
- Pierre GUINARD, Responsable du fonds ancien, Bibliothèque municipale de Lyon (entretien du 28 novembre 2022) ;
- Cindy HOPFNER, Chargée de valorisation et médiation numérique des collections patrimoniales, Bibliothèques de Nancy (entretien du 22 novembre 2022) ;
- Arnaud LABORDERIE, Chef de projet Gallica chargé de l'exploitation des données pour la recherche, Bibliothèque nationale de France (entretien du 9 juin 2022) ;
- Julie LADANT, Directrice du département Droit, économie, politique, Bibliothèque nationale de France (entretien du 13 février 2023) ;
- Étienne MANCHETTE, Responsable Partenariats et contenus, BnF-Partenariats, (entretien du 30 septembre 2022) ;
- Mélanie MARCHAND, Chargée de mission Bibliothèques et Patrimoine, Occitanie Livre & Lecture (entretien du 27 octobre 2022) ;
- Philippe MEZZASALMA, Chef du service Presse, Bibliothèque nationale de France (entretien du 7 novembre 2022) ;
- Sarah MOMBERT, Maîtresse de conférences en littérature du XIXe siècle, École Normale Supérieure de Lyon / IHRIM (entretien du 3 juin 2022) ;
- Jean-Philippe MOREUX, Expert scientifique de Gallica, Bibliothèque nationale de France (entretien du 14 novembre 2022) ;
- Sophie ROBERT, Chargée de collections au sein du service Presse, Bibliothèque nationale de France (entretien du 3 juin 2022) ;
- Jérôme SCHWEITZER, Directeur du pôle Partage, diffusion et réseaux, Bibliothèque nationale et universitaire de Strasbourg (entretien du 7 décembre 2022) ;
- André-Pierre SYREN, Directeur de la valorisation, Enssib (entretien du 31 mai 2022).

### **Liste des portails consultés**

***Tous les liens ont été vérifiés le 24 février 2023.***

*Archives départementales des Deux-Sèvres et de la Vienne*, <https://archives-deux-sevres-vienne.fr/>

*Archives départementales de Loire-Atlantique*, [https://archives.loire-atlantique.fr/44/accueil-archives/j\\_6](https://archives.loire-atlantique.fr/44/accueil-archives/j_6)

*Archives départementales des Yvelines*, <https://archives.yvelines.fr/>

*Aurelia*, <https://mediatheques.orleans-metropole.fr/patrimoine/collections-patrimoniales-numerisees>

*Bibliothèque Méjanès*, <https://bibliotheque-numerique.citedulivre-aix.com/presse-ancienne-aixoise>

*Bilketa*, <https://www.bilketa.eus/>

*Commun Patrimoine*, <https://www.communpatrimoine.fr/communpatrimoine/>

*Cordela*, <http://cordela.region.vda.it/>

*ENPC*, <https://heritage.ecoledespoints.fr/enpc/fr/content/les-collections-numerisees-types-de-documents-presse>

*Europeana Newspapers*, <http://www.europeana-newspapers.eu/>

*FranceArchives, Presse, journaux et imprimés numérisés : Liste des portails numériques des services d'archives en France*, <https://francearchives.fr/fr/article/27699901>

*Geneanet*, <https://www.geneanet.org/>

*Impresso - Media Monitoring of the Past*, <https://impresso-project.ch/>

*Lectura Plus*, <https://www.lectura.plus/Presse/>

*Le Gazetier révolutionnaire*, <https://gazetier-revolutionnaire.gazettes18e.fr/>

*Le Gazetier universel*, <https://gazetier-universel.gazettes18e.fr/>

*L'Empreinte*, <https://www.lempreinte.valenceromansagglo.fr/>

*Limédia Kiosque*, <https://kiosque.limedia.fr/>

*Marius*, <http://marius.marseille.fr/marius/jsp/site/Portal.jsp>

*Médias 19 - Littérature et culture médiatique*, <https://www.medias19.org/>

*Memonum*, <https://memonum-mediatheques.montpellier3m.fr/memonum/fr/content/accueil-fr>

*NewsEye, Enquêteur numérique pour la presse ancienne*, <https://www.newseye.eu/>

*Normannia*, <https://www.normannia.info/>

*Numapresse - Du papier à l'écran. Mutations culturelles, transferts génériques, poétiques médiatiques de la presse française*, <http://www.numapresse.org/>

*Numelyo*, <https://numelyo.bm-lyon.fr/>

*Numistral*, <https://www.numistral.fr/fr>

*Nutrisco*, <https://nutrisco-patrimoine.lehavre.fr/nutrisco/?mode=desktop>

*Overnia*, <https://overnia.bibliotheques-clermontmetropole.eu/>

*Pagella*, <https://pagella.bm-grenoble.fr/pagella/fr/content/accueil-fr>

*Pireneas*, <https://www.pireneas.fr/pireneas/fr/>

*Portail Presse et médias : L'actualité de la presse à la BnF... et ailleurs*, [https://bnf.libguides.com/presse\\_medias/actualite\\_collections](https://bnf.libguides.com/presse_medias/actualite_collections)

*Presse locale ancienne*, <https://presselocaleancienne.bnf.fr>

*Ressources Patrimoines en Occitanie*, <https://ressourcespatrimoines.laregion.fr/>

*Retronews*, <https://www.retronews.fr/>

*Rosalis*, <https://rosalis.bibliotheque.toulouse.fr/rosalis/fr/>

*Rotomagus*, <https://www.rotomagus.fr/rotomagus/?mode=desktop>

*Séléné*, <https://selene.bordeaux.fr/>

*Yroise*, <https://yroise.biblio.brest.fr/yroise/fr/>



## BIBLIOGRAPHIE

---

*Tous les liens ont été vérifiés le 24 février 2023.*

### **Culture numérique et médiation du patrimoine**

CALENGE Bertrand, *Les bibliothèques et la médiation des connaissances*, Paris, France, Éditions du Cercle de la librairie, 2015.

CHAUMIER Serge et MAIRESSE François, « Chapitre 2 - Les fonctions du médiateur » dans *La médiation culturelle*, Paris, Armand Colin (coll. « Collection U »), 2013, p. 61-114.

DIALLO Malik, « Limédia, prix de l'innovation numérique 2019 », *Bulletin des bibliothèques de France* (BBF), 16 janvier 2020.

DOUEIHI Milad, *La grande conversion numérique*, trad. Paul Chemla, Paris, France, Éditions du Seuil, 2008.

DOUEIHI Milad, *Digital cultures*, Cambridge, Mass., Harvard University Press, 2011.

DUFRENE Bernadette, IHADJADENE Madjid, BRUCKMANN Denis, BARBIER Benjamin, *Numérisation du patrimoine : quelles médiations ? quels accès ? quelles cultures ?*, Paris, France, Hermann, 2013.

DUJOL Lionel, *Communs du savoir et bibliothèques*, Paris, France, Éditions du Cercle de la librairie, 2017.

DUJOL Lionel et MERCIER Silvère, *Médiation numérique des savoirs : des enjeux aux dispositifs*, Montréal, Canada, Les éditions Asted, 2017.

HURARD François et PHOYU-YEDID Nicole, *L'écosystème de l'audio à la demande (« podcasts ») : enjeux de souveraineté, de régulation et de soutien à la création audionumérique*, rapport de l'Inspection générale des affaires culturelles, n°2020-11, <https://www.culture.gouv.fr/Espace-documentation/Rapports/L-ecosysteme-de-l-audio-a-la-demande-podcasts-enjeux-de-souverainete-de-regulation-et-de-soutien-a-la-creation-audionumerique>, 19 novembre 2020.

JEANNENEY Jean-Noël, *Quand Google défie l'Europe : plaidoyer pour un sursaut*, Paris, France, Mille et une nuits, 2005.

OCDE et Statistiques Canada, *La littératie à l'ère de l'information : Rapport final de l'Enquête internationale sur la littératie des adultes*, 2000.

SCHAER Roland, « Sur la politique culturelle de la Bibliothèque nationale de France », *Bulletin des bibliothèques de France* (BBF), 1997, n° 6, p. 18-22.

### **Histoire de la presse**

#### *Ressources pour l'étude de la presse*

DELPORTE Christian, BLANDIN Claire et ROBINET François, *Histoire de la presse en France : XXe-XXIe siècles*, Malakoff, France, Armand Colin, 2016.

DEVREUX Lise et MEZZASALMA Philippe, *Des sources pour l'histoire de la presse : guide*, Paris, France, Bibliothèque nationale de France, 2011.

ENCYCLOPÆDIA UNIVERSALIS, « La Gazette (Théophraste Renaudot) », <https://www.universalis.fr/encyclopedie/la-gazette/>.

HATIN Eugène, *Bibliographie historique et critique de la presse périodique française ou Catalogue systématique et raisonné de tous les écrits périodiques de quelque valeur publiés ou ayant circulé en France depuis l'origine du journal jusqu'à nos jours* (1866), Torino, Italie, Bottega d'Erasmus, 1960.

KALIFA Dominique, REGNIER Philippe, THERENTY Marie-Ève et VAILLANT Alain, *La civilisation du journal : histoire culturelle et littéraire de la presse française au XIXe siècle*, Paris, France, Nouveau monde éditions, 2011.

LEVRIER Alexis et WRONA Adeline, *Matière et esprit du journal : du Mercure galant à Twitter*, Paris, France, PUPS, 2013.

THERENTY Marie-Ève, VAILLANT Alain, *Presse et plumes : journalisme et littérature au XIXe siècle*, Paris, France, Nouveau Monde éditions, 2004.

### ***Étudier la presse en contexte numérique***

BUARD Jean-Luc, *Culture médiatique et presse numérisée : Médiasphère des feuillets-nouvelles de Marie Aycard (1794-1859)*, Presses de l'Essai, 2019.

CAMBRON Micheline et DANAUX Stéphanie, « La recherche sur la presse : nouveaux bilans nationaux et internationaux », *Médias 19*, <https://www.medias19.org/publications/la-recherche-sur-la-presse-nouveaux-bilans-nationaux-et-internationaux/presentation-du-dossier-problematiques-enjeux-outils-de-la-recherche-actuelle>, 2021.

CLAVERT Frédéric, « Une histoire par les données ? Le futur très proche de l'histoire des relations internationales », *Bulletin de l'Institut Pierre Renouvin*, 2016, vol. 44, n° 2, p. 119-130.

CLAVERT Frédéric et MULLER Caroline (dir.), *Le goût de l'archive à l'ère numérique* (blog), <https://gout-numerique.net/>.

DE MAEYER Juliette, « Traverser les corpus de presse numériques », intervention à la Maison Européenne des Sciences de l'Homme et de la Société (MESHS), Lille, France, 2019.

GAILLARD Claire-Lise, « Feuilletter la presse ancienne par Giga Octets », dans Frédéric Clavert et Caroline Muller (dir.), *Le goût de l'archive à l'ère numérique*, <https://gout-numerique.net/table-of-contents/archives-numerisees-la-salle-de-lecture-virtuelle-et-physique/feuilletter-la-presse-ancienne-par-giga-octets>, 4 juin 2018.

KAPFERER Benedikt, « Interaction between Newspapers and Readers », *NewsEye, A Digital Investigator for Historical Newspapers*, <https://www.newseye.eu/fr/case-studies/case-study-4-media-and-journalism/interaction-between-newspapers-and-readers/>.

LANGLAIS Pierre-Carl, *La formation de la chronique boursière dans la presse quotidienne française (1801-1870) : métamorphoses textuelles d'un journalisme de données*, Thèse de doctorat, Université Paris-Sorbonne, 1970-2017, France, 2015.

LANGLAIS Pierre-Carl, « Identifier les rubriques de presse ancienne avec du *topic modeling* », <https://numapresse.hypotheses.org/11>, 31 mars 2017.

LANGLAIS Pierre-Carl, « Lire Gallica à distance : pour un usage panoramique de l'archive numérisée », *I2D - Information, données & documents*, 2021, vol. 1, n° 1, p. 104-108.

LEU Philipp, « RetroNews pour les historiens », <https://histoirebnf.hypotheses.org/1933>, 7 mars 2018.

MORETTI Franco, *Graphes, cartes et arbres : modèles abstraits pour une autre histoire de la littérature*, Paris, France, Les Prairies ordinaires, 2008.

MOREUX Jean-Philippe, « Approches innovantes pour la presse ancienne numérisée : fouille et visualisation de données », <https://bnf.hypotheses.org/208>, 30 décembre 2017, mis à jour 6 septembre 2017.

OMARI Nejma, trad. Adrienne Ummels, « Women in Pants », *NewsEye, A Digital Investigator for Historical Newspapers*, <https://www.newseye.eu/fr/case-studies/case-study-2-gender/women-in-pants/>.

PETRAULT Jade, *Le laboratoire du roman-feuilleton (1850-1940) en Occitanie*, Thèse en préparation (Theses.fr), Montpellier 3, 2021.

PINSON Guillaume, « Lire la presse du XIXe siècle en contexte numérique : vers une nouvelle historiographie des objets médiatiques ? », *Dix-Neuf*, 2 octobre 2017, vol. 21, n° 4, p. 378-388.

RYGIEL Philippe, *Historien à l'âge numérique*, Villeurbanne, Presses de l'Enssib (coll. « Papiers »), 2019.

### ***Colloques, journées d'étude et journées professionnelles***

À la recherche des femmes dans la presse ancienne numérisée (1850-1950), <https://www.bnf.fr/fr/agenda/la-recherche-des-femmes-dans-la-presse-ancienne-numerisee-1850-1950>, 15 mars 2021.

*Ces journaux des diasporas qui ont fait la presse parisienne*, colloque international Transfopress à la BULAC, <https://www.bulac.fr/ces-journaux-des-diasporas-qui-ont-fait-la-presse-parisienne>, 9-11 mars 2022.

*Couper / coller. Le journalisme aux ciseaux du XIXe siècle à Internet*, colloque Numapresse / ENS Lyon, <https://www.fabula.org/actualites/109625/colloque--coupercoller-le-journalisme-aux-ciseaux-du-xixe-siecle.html>, 9 février 2022.

*Crimes fictionnels / crimes factuels. Fictions criminelles européennes et discours médiatiques du crime*, <https://www.medias19.org/actualites/crimes-fictionnels-crimes-factuels-fictions-criminelles-europeennes-et-discours-mediatiques-du-crime-7-8-octobre-2021>, 7 octobre 2021.

*Droit(s) et presse*, <https://www.bnf.fr/fr/agenda/droits-et-presse>, 18 novembre 2022.

*Journées du Patrimoine Écrit 2022 : À la une ! Numériser, valoriser et conserver la presse ancienne*, Rouen, 23-24 juin 2022.

*Les feuilletons dans la presse*, <https://www.bnf.fr/fr/agenda/les-feuilletons-dans-la-presse>, 10 avril 2021.

*Presses anciennes et modernes à l'ère du numérique*, colloque international Médias 19 / Numapresse au Centre culturel canadien et à la BnF, <https://www.bnf.fr/fr/agenda/congres-medias-19-numapresse-presses-anciennes-et-modernes-lere-du-numerique>, 30 mai - 3 juin 2022.

*Programmes Europeana de commémoration du centenaire de la Première Guerre mondiale (Grande Collecte etc.)*, présentation de Jérôme Schweitzer à l'Enssib, <https://www.enssib.fr/bibliotheque-numerique/document-64009>, 5 novembre 2013.

## **Conservation, numérisation et valorisation de la presse ancienne**

### ***Rapports et articles***

GIRARD Aline, *La numérisation concertée de corpus d'imprimés. État des lieux des programmes et des partenaires*, rapport de la Bibliothèque nationale de France, [https://www.bnf.fr/sites/default/files/2018-11/num\\_concertee\\_impr\\_progr\\_partenaires.pdf](https://www.bnf.fr/sites/default/files/2018-11/num_concertee_impr_progr_partenaires.pdf), décembre 2014, version du 28 avril 2015.

GIRARD Aline et AURERIN Catherine, *La presse physique et numérique à la BnF : état des lieux et perspectives*, rapport de la Bibliothèque nationale de France, mars 2018.

LE MAITRE Stéphanie, « La valorisation numérique du patrimoine photographique de presse », *Médias 19*, <https://www.medias19.org/publications/presse-et-patrimoine/la-valorisation-numerique-du-patrimoine-photographique-de-presse>, 20 juin 2022.

LOUISON-LOUDOT Valérie, « Presse locale ancienne : un site de la BnF exclusivement dédié à la presse locale ancienne et d'intérêt général », <https://histoirebnf.hypotheses.org/817>, 13 juillet 2017.

MANCHETTE Étienne et THOUNY Nathalie, « RetroNews : éditorialiser la presse ancienne », *Bulletin des bibliothèques de France (BBF)*, 2018, n° 15, p. 32-35.

MISTRAL Julie et DESRICHARD Yves, *Les plans de conservation partagée des périodiques : un état de l'art quantitatif et qualitatif*, <https://abes.fr/wp-content/uploads/2020/01/Etude-PCPP-2019.pdf>, décembre 2019.

NGUYEN Thi-Phuong et VALLAS Philippe, « La conservation des documents papier : point sur l'évolution des techniques et des stratégies », *Bulletin des bibliothèques de France (BBF)*, 2006, n° 4, p. 11-21.

POUCHARD David, « La valorisation des fonds photographiques, ou comment concilier le droit d'auteur et l'accès au patrimoine culturel », *In Situ. Revue des patrimoines*, 15 octobre 2018, n° 36.

PREMEL Benjamin, « La numérisation de la presse rétrospective à la BnF », *Le Temps des médias*, 2007, vol. 8, n° 1, p. 267-278.

RACINE Bruno, *Schéma numérique des bibliothèques*, rapport élaboré dans le cadre du Conseil du Livre, décembre 2009.

SCHWEITZER Jérôme, « Autre chose que des livres... : les entrées du dépôt légal dans les collections de la BNU », *La Revue de la BNU*, 1 novembre 2018, n° 18, p. 26-33.

SCHWEITZER Jérôme, « La presse : axe majeur de la numérisation en 2022 », <https://bnu.hypotheses.org/9893>, 10 février 2022.

TESSIER Marc et BAFFERT Maxime, *La presse au défi du numérique*, rapport au ministère de la Culture et de la communication, février 2007.

### ***L'action territoriale de la BnF (site institutionnel)***

*Cartographie de la coopération territoriale*, <https://www.bnf.fr/fr/cartographie-de-la-cooperation-territoriale>.

*Coopération régionale et action territoriale de la BnF*, <https://www.bnf.fr/fr/cooperation-regionale-et-action-territoriale-de-la-bnf>.

*Gallica marque blanche*, <https://www.bnf.fr/fr/gallica-marque-blanche>.

*La constitution d'une collection de référence des titres de presse*, <https://www.bnf.fr/fr/la-constitution-dune-collection-de-referance-des-titres-de-presse>.

*Les partenaires de Gallica en 2017*, [https://www.bnf.fr/sites/default/files/2018-11/partenaires\\_gallica.pdf](https://www.bnf.fr/sites/default/files/2018-11/partenaires_gallica.pdf), février 2018.

*Rapport d'activité BnF 2021*, <https://www.bnf.fr/fr/bnf-rapport-dactivite-2021>.

### ***Quelques actualités autour de la presse ancienne en France***

*ActuaLitté*, « À la BnF, la correction collaborative au service de la recherche », <https://actualitte.com/article/103262/numerisation/a-la-bnf-la-correction-collaborative-au-service-de-la-recherche>, 8 novembre 2021.

*ActuaLitté*, « Après la presse ancienne, la BnF numérise les titres modernes », <https://actualitte.com/article/108013/numerisation/apres-la-presse-ancienne-la-bnf-numerise-les-titres-modernes>, 30 septembre 2022.

*ActuaLitté*, « La BnF lance un concours d'architecture pour son pôle d'Amiens », <https://actualitte.com/article/110197/bibliotheque/la-bnf-lance-un-concours-d-architecture-pour-son-pole-d-amiens>, 13 février 2023.

*AUPLA : Application sur les Unica et la Presse Locale Ancienne*, <http://bibliotheque-blogs.unice.fr/sudoc-ps/application-unica-presse-locale-ancienne/>, 6 novembre 2019.

*Bilketa, le portail numérique des fonds documentaires basques fait peau neuve*, <https://www.baskulture.com/article/bilketa-le-portail-numrique-des-fonds-documentaires-basques-fait-peau-neuve-5054>, 29 juillet 2022.

*Le Bien Public (1868-1944) en ligne sur Gallica*, Agence Livre & Lecture Bourgogne-Franche-Comté, <https://www.youtube.com/watch?v=gLTa1miz0us>, 9 février 2022.

*Les Journaux de Seine-et-Oise dévoilent leurs caractères*, <https://archives.yvelines.fr/nous-connaître/nos-actualités/zoom-sur/les-journaux-de-seine-et-oise-devoilent-leurs-caracteres>, décembre 2022.

*Occitanie Livre & Lecture - Compte rendu du groupe de travail « Numérisation »*, <https://www.occitanielivre.fr/groupe-projet-numerisation>.

*Occitanie Livre & Lecture - La numérisation pour mieux partager*, <https://www.occitanielivre.fr/la-numerisation-pour-mieux-partager>.

### ***Quelques actions de valorisation***

*Agir ensemble pour les patrimoines - Les Chroniques retrouvées du Midi : la presse ancienne préservée et révélée*, <https://www.dailymotion.com/video/x2x83uh>, 8 juillet 2015.

À la recherche de la presse perdue, <https://archives.yvelines.fr/participer/a-la-recherche-de-la-presse-perdue>.

Albums de La contemporaine sur Flickr, <https://www.flickr.com/photos/98069937@N05/albums/>.

Archichroniques : Revivez les grands évènements et faits divers de l'Oise!, <https://actu.oise.fr/actualites/les-archichroniques-revivez-les-grands-evenements-et-faits-divers-de-loise-45403>, 17 janvier 2023.

De la Gazette à Internet, exposition BnF / CLEMI en prêt en 2022-2023, <https://essentiels.bnf.fr/fr/societe/medias/b301c6f3-73f9-4882-81c8-5b81d897a876-gazette-internet>.

Focus sur l'activité européenne de la BnF et publication du corpus « Presse européenne » sur Gallica, <https://www.bnf.fr/fr/actualites/focus-activite-europeenne-bnf-et-publication-corpus-presse-europeenne-gallica>, 9 mai 2022.

Grande Collecte : élargir les sources de notre histoire, <https://www.culture.gouv.fr/Actualites/Grande-Collecte-elargir-les-sources-de-notre-histoire>.

Histoires de fausses nouvelles, exposition BnF / CLEMI en prêt en 2022-2023, <https://essentiels.bnf.fr/fr/enseignants/b64967cc-0065-4fe1-9490-9b1c7d51250b-fausses-nouvelles>.

Joséphine Baker, portraits dessinés dans la presse - Le blog de Gallica, <https://gallica.bnf.fr/blog/30112021/josephine-baker-portraits-dessines-dans-la-presse?mode=desktop>, 30 novembre 2021.

L'âge d'or de la presse algérienne : 1880-1914 - Le blog de Gallica, <https://gallica.bnf.fr/blog/23012023/lage-dor-de-la-presse-algerienne-1880-1914?mode=desktop>, 23 janvier 2023.

Les premiers titres de presse en Algérie : 1830-1880 - Le blog de Gallica, <https://gallica.bnf.fr/blog/16012023/les-premiers-titres-de-presse-en-algerie-1830-1880?mode=desktop>, 16 janvier 2023.

Les services participatifs à Roubaix : comment ça marche ?, <https://www.bn-r.fr/correction-ocr/index.php?p=tutoriel>.

L'Exprimante, un distributeur de presse ancienne, <https://www.lectura.plus/5486-lexprimantenbsp-un-distributeur-de-presse-ancienne.html>.

Limédia - Le procès du parricide de Leyr, ou comment parler de la presse ancienne, <https://epitome.hypotheses.org/4288>, 3 décembre 2019.

Limédia - Rire, réagir, réfléchir, <https://kiosque.limedia.fr/histoire-presse/rire-reagir-reflechir/>.

Quand Gérard Noirielle redonne vie à Chocolat - Le blog de Gallica, <https://gallica.bnf.fr/blog/08112017/quand-gerard-noirielle-redonne-vie-chocolat?mode=desktop>, 8 novembre 2017.

Quand la France découvre la photographie (Léa Bello feat. Nota Bene), RetroNews-BnF, <https://www.youtube.com/watch?v=bs4nT7Fj9NM>, 2020.

Quand une gallicanaute étudie le marché de la rencontre - Le blog de Gallica, <https://gallica.bnf.fr/blog/06112015/quand-une-gallicanaute-etudie-le-marche-de-la-rencontre?mode=desktop>, 6 novembre 2015.

« *Séries noires à la Une* » : la presse et le crime, un podcast de RetroNews, <https://www.retronews.fr/histoire-de-la-presse/evenement/2021/01/25/series-noires-la-une-la-presse-et-le-crime-un-podcast-de>, 25 janvier 2021.

## **Les lecteurs de presse ancienne : connaissances des publics**

### *Enquêtes de publics*

BALIGAND Marie Pascale, COLCANAP Grégory, HARNAIS Vincent, ROUSSEAU-HANS Françoise, WEIL-MIKO Christine, *Les pratiques de recherche documentaire des chercheurs français en 2020 : étude du consortium Couperin*, rapport Couperin n°2, Couperin.org. (hal-03148285), 2021.

BEAUDOUIN Valérie et DENIS Jérôme, *Observer et évaluer les usages de Gallica. Réflexion épistémologique et stratégique*, rapport de recherche, BnF ; Telecom ParisTech, 2014.

CHEVALLIER Philippe, « Les données au service de la connaissance des usages en ligne : l'exemple de l'analyse des logs de Gallica », *Les Enjeux de l'information et de la communication*, 2018, vol. 19/2, n° 2, p. 57-67.

CHEVALLIER Philippe, « Les évolutions de l'usage savant du patrimoine documentaire en ligne », *Biens Symboliques / Symbolic Goods. Revue de sciences sociales sur les arts, la culture et les idées*, 1 novembre 2020, n° 7.

COUILLARD Noémie, « Qui et comment fait-on des recherches historiques dans les archives en ligne ? », <https://polipubli.hypotheses.org/678>, 31 août 2022.

COUILLARD Noémie et BASTARD Irène, *Qui sont les usagers en ligne communs entre les Archives et la BnF ?*, <https://francearchives.gouv.fr/fr/article/677845362>, septembre 2022.

GUIGUENO Brigitte et PENICAUT Emmanuel, *Qui sont les publics des archives ? Enquêtes sur les lecteurs, les internautes et le public des activités culturelles dans les services publics d'archives (2013-2014)*, <https://francearchives.gouv.fr/fr/article/38196>, 2015.

*Actualité des études - Synthèse de l'Observatoire des publics 2020 (Une « photo de famille »)*, <https://www.bnf.fr/fr/lactualite-des-etudes>, 2020.

*Gallicanautes : loin des clichés du numérique. Enquête auprès des usagers de la bibliothèque numérique Gallica*, <https://www.bnf.fr/fr/gallicanautes-loin-des-cliches-du-numerique>, avril 2017.

*Guide méthodologique sur les études de publics*, <https://www.culture.gouv.fr/Thematiques/Musees/Pour-les-professionnels/Rendre-les-collections-accessibles-aux-publics/Assurer-la-mediation-aupres-des-visiteurs/Guide-methodologique-sur-les-etudes-de-publics>, 5 octobre 2020.

### *Études sur la participation des publics amateurs*

ALLAL-CHERIF Oihab et ARENA Lise, « Étienne Wenger - Les communautés de pratique au service de la connaissance située et de l'apprentissage social » dans *Les Grands Auteurs en Systèmes d'information*, Caen, EMS Editions, 2018, p. 462-483.

ANDRO Mathieu et SALEH Imad, « La correction participative de l'OCR : le *crowdsourcing* au profit des bibliothèques numériques », *Bulletin des bibliothèques de France* (BBF), 17 juin 2015.

BEAUDOUIN Valérie, « Le forum Pages 14-18 et sa communauté de pratique », dans Philippe Chevallier et Lionel Maurel (dir.), *Le web français de la Grande Guerre : Réseaux amateurs et institutionnels*, Nanterre, Presses universitaires de Paris Nanterre (coll. « Les passés dans le présent »), 2022, p. 207-256.

BEAUDOUIN Valérie, « Les usages sociaux du patrimoine en ligne : l'exemple de la Grande Guerre », dans Étienne Anheim, Anne-Julie Etter, Ghislaine Glasson-Deschaumes et Pascal Liévaux (dir.), *Les patrimoines en recherche(s) d'avenir*, Nanterre, Presses universitaires de Paris Nanterre (coll. « Les passés dans le présent »), 2022, p. 265-278.

FLICHY Patrice, *Le sacre de l'amateur : sociologie des passions ordinaires à l'ère numérique*, Paris, France, La République des idées : Seuil, 2010.

SALOMON Jean-Jacques et SALOMON Igor, *Panorama des associations d'amis d'écrivains*, Paris, France, Éditions du Palio, 2021.

SEVERO Marta, *L'impératif participatif : institutions culturelles, amateurs et plateformes*, Bry-sur-Marne, France, INA, 2021.

### ***Ressources pour la généalogie***

BOUDAREL Sophie, « Gallicarte, géolocalisez les collections de Gallica », *La Gazette des ancêtres* (blog), <https://la-gazette-des-ancestres.fr/gallicarte/>, 6 février 2020.

BOUDAREL Sophie, « Gallica 1 - Présentation des ressources », *La Revue française de Généalogie*, <https://www.youtube.com/watch?v=GSh5jLe8vnI>, 18 septembre 2020.

NEULAT Tony, *Gallica et RetroNews: deux eldorados généalogiques*, Paris, France, Archives & Culture, 2021.

*Gallica - Ressources généalogiques*, <https://gallica.bnf.fr/html/und/histoire/ressources-genealogiques>.

*Gallica - La carte de Cassini*, <https://gallica.bnf.fr/html/und/cartes/france-en-cartes/la-carte-de-cassini>.

## **Science ouverte et Humanités numériques**

### ***Ressources générales***

CARLIN Marie et LABORDERIE Arnaud, « La BnF et les services à la recherche à l'heure des humanités numériques », *Arabesques*, 1 avril 2022, n° 105, p. 8-9.

CHAGUE Alix et ROMARY Laurent, « L'intelligence artificielle, une ouverture du champ des possibles », *Arabesques*, 1 octobre 2022, n° 107, p. 4-5.

DARDENNE Nadine, « DARIAH : une infrastructure numérique au service des sciences humaines et sociales », <https://humanum.hypotheses.org/155>, 6 février 2014.

PASTORE Graziella, *Les coopérations entre chercheurs et bibliothécaires dans le cadre des projets de numérisation de corpus documentaires*, mémoire d'étude Enssib, mars 2018.

*Légifrance*, « Ordonnance n° 2021-1518 du 24 novembre 2021 complétant la transposition de la directive 2019/790 du Parlement européen et du Conseil du 17 avril 2019 sur le droit d'auteur et les droits voisins dans le marché unique numérique », <https://www.legifrance.gouv.fr/jorf/id/JORFTEXT000044362034>.

*Ouvrir la science ! - Ministère de l'Enseignement supérieur et de la recherche*, « La fouille de textes et de données à des fins de recherche : une pratique confirmée et désormais opérationnelle en droit français », <https://www.ouvrirlascience.fr/la-fouille-de-textes-et-de-donnees-a-des-fins-de-recherche-une-pratique-confirmee-et-desormais-operationnelle-en-droit-francais>.

### ***Divers projets et outils***

SCHUH Julien, *Pyllica : un outil de récupération automatisée de données sur gallica.bnf.fr*, <https://numapresse.hypotheses.org/14>, 29 mars 2017.

*API et jeux de données*, <https://api.bnf.fr/>.

*API et jeux de données - Projet Europeana Newspapers*, <https://api.bnf.fr/fr/node/190>.

*API et jeux de données, Pyllica, extracteur Python de corpus de périodiques de Gallica*, <https://api.bnf.fr/fr/extracteur-python-de-corpus-de-periodiques>.

*BnF DataLab*, <https://www.bnf.fr/fr/bnf-datalab>, 15 octobre 2021.

*Gallicagram*, outil de lexicométrie développé par Benjamin Azoulay (Université Paris-Saclay) et Benoît de Courson (ENS-PSL), <https://shiny.ens-paris-saclay.fr/app/gallicagram>.

*GallicaPix*, outil de recherche iconographique développé par Jean-Philippe Moreux (BnF, Gallica) et Guillaume Chiron (Université de La Rochelle), <https://gallicapix.bnf.fr/rest?run=findIllustrations-form.xq>.



## ANNEXES

---

### *Table des annexes*

<b>ANNEXE N°1 - DETAIL DE LA METHODOLOGIE D'ENQUETE .....</b>	<b>106</b>
<b>ANNEXE N°2 - QUESTIONNAIRE D'ENQUETE.....</b>	<b>111</b>
<b>ANNEXE N°3 - DETAIL DES REPONSES .....</b>	<b>117</b>
<b>ANNEXE N°4 - TYPOLOGIE DES BESOINS EXPRIMES PAR LES REONDANTS.....</b>	<b>126</b>

# ANNEXE N°1 - DETAIL DE LA METHODOLOGIE D'ENQUETE

## CONNAISSANCE DES PUBLICS

Cerner les caractéristiques des publics auprès desquels nous allions enquêter était une première étape indispensable avant l'élaboration de notre questionnaire d'enquête. Une fois nos deux panels choisis – chercheurs de l'ESR et chercheurs amateurs –, nous avons donc affiné notre connaissance de ces publics à l'aide de lectures et d'entretiens. Ces entretiens à visée méthodologique nous ont permis non seulement de mieux structurer notre méthodologie d'enquête, mais aussi d'identifier les ressources les plus pertinentes sur lesquelles nous appuyer, en particulier de précédentes études de publics menées par le SIAF et la BnF : *Qui sont les publics des archives ? Enquêtes sur les lecteurs, les internautes et le public des activités culturelles dans les services publics d'archives (2013-2014)*, de Brigitte Guigueno et Emmanuel Pénicaut ; *Les publics de la BnF : Synthèse de l'Observatoire 2020* ; *Deux milliards de clics : Enquête sur les usagers en ligne des archives (2021)*, de Noémie Couillard et Maylis Nouvellon, et le prolongement de cette enquête, *Qui sont les usagers en ligne communs entre les Archives et la BnF ? (2022)*, de Noémie Couillard et Irène Bastard. Les ressources de formation transmises par la généalogiste professionnelle Sophie Boudarel nous ont également permis de mieux cerner le type de recherches menées par les généalogistes amateurs dans la presse ancienne.

## ÉLABORATION DU QUESTIONNAIRE D'ENQUETE

Pour l'élaboration du questionnaire d'enquête, nous avons adopté une démarche comparatiste. Afin de rendre pleinement exploitables les résultats et de faciliter les comparaisons, nous avons fait le choix d'aligner les questions entre les deux panels sondés. Seule la première section (« Profil général ») diffère ; le reste du sondage est identique. Nous avons par ailleurs décidé de nous limiter à une vingtaine de questions afin que la participation à l'enquête n'excède pas les cinq minutes. Nous avons indiqué aux participants que l'enquête était menée dans le respect du RGPD et que les données personnelles n'étaient pas destinées à être conservées. Nous avons d'ailleurs veillé, dans notre synthèse, à anonymiser les réponses.

### Première partie : Cerner globalement les usages

La section « Votre usage de la presse ancienne numérisée » visait à estimer la place occupée par la presse ancienne dans les recherches de l'utilisateur : s'agit-il pour lui d'un corpus essentiel ou d'un type de ressources parmi d'autres ? À quelle fréquence est-il amené à consulter ces collections en ligne, à l'aide de quels outils et pour quels usages ? Cette première étape du questionnaire devait ainsi nous permettre de cerner plus précisément le profil du chercheur et nous donner un indice de son degré d'expertise, en distinguant le lecteur utilisant la presse ancienne comme une ressource peu voire non prioritaire du chercheur expert appliquant à ces corpus des opérations complexes d'extraction de données.

Cette série de questions offrait différents niveaux de granularité permettant de couvrir des usages allant du plus ou moins expert : ainsi, les répondants qui indiquaient pratiquer l'extraction de données se voyaient proposer une question subsidiaire, pour préciser quels étaient les outils de traitement et d'analyse des données utilisés. À l'inverse, nous avons envisagé la possibilité qu'un lecteur de presse ancienne puisse ne jamais avoir recours aux corpus numérisés, privilégiant la consultation de collections physiques en salle de lecture ; nous avons donc laissé cette possibilité ouverte dans le questionnaire – bien que celui-ci, n'étant diffusé que sous forme électronique, ait été *a priori* destiné à des usagers familiers du numérique.

Dans cette section, nous souhaitions également savoir à quelles plateformes l'utilisateur avait recours, soit pour la consultation (Gallica, RetroNews, Geneanet, etc.), soit pour trouver des documents en utilisant des portails agrégateurs de ressources de presse, tels que Presse locale ancienne ou FranceArchives. Il nous intéressait de savoir si l'utilisateur consultait de la presse uniquement sur les grands sites nationaux tels que Gallica et RetroNews, ou également sur des portails de bibliothèques numériques en région. Nous avons ici laissé la question ouverte, afin de permettre aux sondés d'indiquer le cas échéant le nom des portails régionaux consultés.

## Deuxième partie : Connaître les pratiques de recherche des usagers

Dans la suite du questionnaire, nous nous sommes concentrés en priorité sur les deux plateformes nationales Gallica et RetroNews, qui constituent les deux ressources majeures pour la presse ancienne en France. Étant nécessairement contraints par le format de l'enquête, et ne pouvant couvrir l'ensemble des portails existants, nous avons fait le choix de nous limiter à ces deux sites vers lesquels nous supposons que le chercheur professionnel se tournait en priorité – ce qui nous a été confirmé par des entretiens informels que nous avons pu avoir avec différents chercheurs. Quant au public des généalogistes et amateurs d'histoire, nous avons supposé qu'il était susceptible d'explorer aussi bien les collections de presse de Gallica que celles de RetroNews, où certains des contenus sont éditorialisés de manière à attirer des publics *a priori* moins experts. Les deux plateformes sont d'ailleurs toutes deux présentées par les généalogistes professionnels comme des ressources de premier plan pour la recherche généalogique<sup>256</sup>.

Afin, toutefois, que le focus sur ces deux plateformes ne constitue pas un biais de notre enquête, nous avons choisi de ne proposer aucune question obligatoire, afin de permettre au répondant de passer à la question suivante s'il souhaitait ne pas répondre. De fait, il nous fallait envisager la possibilité qu'un usager n'utilise que Gallica et non RetroNews, ou vice-versa – voire aucune des deux plateformes si ses recherches étaient limitées à un portail régional par exemple. Aussi avons-nous décidé d'intégrer les non-réponses comme une modalité de réponse à part entière.

Dans cette partie de l'enquête, nous souhaitions mettre en lumière les fonctionnalités de recherche utilisées en priorité par les usagers et les éventuelles différences dans la manière d'appréhender l'un et l'autre site. L'utilisateur passe-t-il

---

<sup>256</sup> Voir Tony Neulat, *Gallica et RetroNews: deux eldorados généalogiques*, Paris, France, Archives & Culture, 2021, 111 p.

aisément, par exemple, du moteur de recherche avancée de Gallica à celui de RetroNews et vice-versa ? L'usage des modules éditoriaux de RetroNews est-il comparable à celui des « Sélections » de Gallica ? Quelle est la part d'utilisateurs qui n'utilisent ni la recherche simple ni la recherche avancée, mais entrent dans les collections via un moteur de recherche externe ? Ces questions étaient importantes pour évaluer non seulement la familiarité numérique des usagers de ces plateformes, mais aussi les fonctionnalités les plus plébiscitées. À partir du double exemple de Gallica et de RetroNews, nous souhaitions ainsi réfléchir à l'adéquation des fonctionnalités de recherche et modes de valorisation des collections de presse avec les besoins des utilisateurs, pour pouvoir proposer des pistes de réflexion et d'amélioration transposables à d'autres portails.

### Troisième partie : Connaître les besoins des usagers

La dernière section du questionnaire interrogeait plus précisément les participants sur leurs besoins en matière d'aide à la recherche et à l'exploitation des corpus. Une question sur leur propre perception de leur niveau d'expertise nous permettait d'abord de vérifier leur degré de familiarité avec les plateformes numériques, déjà déduit des réponses précédentes, mais ici envisagé sous l'angle du ressenti personnel des usagers. Nous espérions ainsi pouvoir mieux apprécier leurs éventuels besoins de formation.

Nous leur avons également demandé de pointer du doigt les manques constatés sur les différentes plateformes de consultations utilisées, mais aussi de réfléchir à des pistes d'amélioration potentielles. Différentes suggestions étaient esquissées afin d'aider les répondants à formuler des axes d'amélioration : amélioration des fonctionnalités de recherche et de visualisation, des outils d'exploitation et d'extraction des corpus, de l'éditorialisation des contenus, etc. Un champ de commentaire libre était laissé à l'utilisateur pour faire part de remarques complémentaires. Cette ultime section du questionnaire devait ainsi nous permettre de réfléchir à des préconisations en réponse aux éventuelles lacunes ou difficultés soulevées par les lecteurs.

### DIFFUSION DE L'ENQUETE

L'enquête a été menée sur un mois et demi, de mi-octobre à fin novembre. Pour la diffusion des deux questionnaires, nous avons eu recours à de multiples canaux en ciblant les réseaux qui nous paraissaient les plus pertinents au regard des publics visés. Pour celui à destination des chercheurs de l'ESR, nous avons pu d'abord bénéficier des réseaux professionnels de notre directeur, Frédéric Clavert : réseaux sociaux d'une part – Facebook et Twitter – et listes de diffusion d'autre part : la liste H2C de l'association « Historiennes et historiens du contemporain » et la liste DH (*Digital Humanities*), liste de discussion francophone concernant les Humanités numériques pour toutes les disciplines des sciences humaines et sociales. Le questionnaire a également été transmis à l'équipe d'*Impresso-Media monitoring of the past*, projet porté par un consortium composé de spécialistes des humanités numériques, historiens, linguistes, bibliothécaires et archivistes, et spécialisé sur la fouille de texte dans les archives de presse.

L'objectif était de viser une diffusion assez large auprès des chercheurs en SHS : non seulement auprès de spécialistes des humanités numériques, à même d'enrichir notre réflexion par leur expertise technique, mais aussi auprès

d'historiens non spécialistes des humanités numériques, dont nous estimions que les usages reflétaient les pratiques de recherche d'une majorité de chercheurs. Dans un second temps, nous avons donc ciblé des laboratoires et équipes de recherche spécialisés dans les études médiatiques, certains porteurs de projets en humanités numériques et d'autres non. Un repérage de ces équipes nous a permis de diffuser l'enquête auprès des réseaux suivants<sup>257</sup> :

- **CELSA** : Groupe de recherches interdisciplinaires sur les processus d'information et de communication (GRIPIC) ;
- **IHRIM - Institut d'Histoire des Représentations et des Idées dans les Modernités** : UMR 5317 (CNRS, ENS de Lyon, Université Lumière Lyon 2, Université Lyon 3 Jean Moulin, Université Jean Monnet Saint-Étienne, Université Clermont-Auvergne), dont le groupe « Presse18 » à Lyon 2 ;
- **La Rochelle Université** : équipe du projet NewsEye ;
- **Médias 19** : plateforme scientifique coordonnée par Guillaume Pinson (Université Laval, Québec) et Marie-Ève Thériault (RIRRA 21, Université Paul-Valéry - Montpellier 3) ;
- **Numapresse** : projet ANR ;
- **Sciences Po Paris** : Département d'histoire et Département archives ;
- **Université Paris 1 Panthéon-Sorbonne** : Centre d'histoire du XIXe siècle et Centre d'histoire sociale du XXe siècle ;
- **Université Paris 8** : Centre d'études sur les médias, les technologies et l'internationalisation (CÉMTI) ;
- **Université Paris Nanterre** : équipe « Poétique historique des textes modernes » (PHisTeM : poétique historique et histoire de l'imprimé au XIXe siècle) ;
- **Université Paris-Panthéon-Assas** : Institut français de presse et Centre d'analyse et de recherche interdisciplinaires sur les médias (CARISM) ;
- **Université Paul-Valéry - Montpellier 3** : laboratoire « Représenter, inventer la réalité, du Romantisme au XXIe siècle » (RIRRA 21) ;
- **Université Sorbonne Nouvelle** : Institut de Recherche Médias, Cultures, Communication et Numérique (IRMÉCCEN) ;
- **Université Sorbonne Paris Nord** : Laboratoire des Sciences de l'Information et de la Communication (LabSIC) ;
- **Université de Versailles Saint-Quentin-en-Yvelines** : Centre d'histoire culturelle des sociétés contemporaines (CHCSC) et Transfopress (réseau transnational pour l'étude de la presse en langues étrangères).

En complément de ces différents réseaux, nous avons également contacté la Société pour l'histoire des médias (SPHM) ainsi qu'un certain nombre de chercheurs individuels dont nous avons repéré les travaux. Plusieurs d'entre eux ont aimablement accepté de se faire le relais de notre enquête.

---

<sup>257</sup> Nous tenons à remercier Sophie Robert, du service Presse de la BnF, de nous avoir permis de nous appuyer sur son propre travail de repérage de ces équipes, qui nous a été particulièrement précieux.

Pour le second questionnaire, nous avons d'abord pris appui sur les réseaux de la généalogiste professionnelle Sophie Boudarel, animatrice du blog *La Gazette des ancêtres* et formatrice à la recherche dans Gallica et RetroNews. Cela nous a permis, dans un premier temps, de diffuser le questionnaire auprès d'une importante communauté de généalogistes amateurs.

Toutefois, nous souhaitons ne pas nous cantonner à cette seule catégorie d'utilisateurs mais viser plus largement un public de passionnés d'histoire menant des recherches à titre non professionnel, avec des usages plus ou moins experts. À ce titre, nous avons jugé pertinent de nous rapprocher des réseaux des services d'archives et des sociétés savantes. Nous avons alors diffusé notre questionnaire auprès des listes de discussion de l'association BiblioPat et de l'AAF (Association des archivistes français), en demandant aux professionnels concernés de bien vouloir relayer l'enquête auprès de leurs publics. Si contacter les sociétés savantes s'avérait plus complexe en raison de l'absence d'une liste de diffusion commune, le relais s'est fait naturellement avec divers services d'archives départementales, associations d'entraide généalogique et autres associations ayant assuré la diffusion du questionnaire auprès de leurs membres. Via le CTHS (Comité des travaux historiques et scientifiques), nous avons en outre pu cibler quelques sociétés historiques et archéologiques précises.

**L'enquête a finalement recueilli un total de 388 participations, réparties comme suit :**

- 241 réponses pour le questionnaire à destination des chercheurs de l'ESR ;**
- 147 réponses pour le questionnaire à destination des généalogistes et amateurs d'histoire.**

## ANNEXE N°2 - QUESTIONNAIRE D'ENQUETE

### USAGES DE LA PRESSE ANCIENNE NUMERISEE : ENQUETE AUPRES DES CHERCHEURS DE L'ENSEIGNEMENT SUPERIEUR

#### Profil général

**Vous êtes :**

- Un homme  
 Une femme  
 Je préfère ne pas répondre

**Votre code postal :**

**Votre établissement de rattachement :**

**Votre statut :**

- Doctorant et/ou ATER  
 Enseignant et/ou chercheur contractuel (y compris contrats post-doctoraux)  
 Chercheur associé  
 Maître de conférences ou Professeur des universités  
 Chargé de recherche ou Directeur de recherche  
 Ingénieur de recherche ou Ingénieur d'études

**Sur quel type de projet de recherche travaillez-vous actuellement ?**

*(plusieurs choix possibles)*

- Recherche individuelle  
 Projet au sein d'une équipe mono-disciplinaire  
 Projet au sein d'une équipe pluridisciplinaire  
 Projet pluridisciplinaire avec volet technique impliquant un ou plusieurs ingénieur(s) de recherche

⇒ Précisez si besoin :

### USAGES DE LA PRESSE ANCIENNE NUMERISEE : ENQUETE AUPRES DES GENEALOGISTES ET AMATEURS D'HISTOIRE

#### Profil général

**Vous êtes :**

- Un homme  
 Une femme  
 Je préfère ne pas répondre

**Votre code postal :**

**Vous êtes :**

- En activité professionnelle

*Êtes-vous un professionnel de la recherche ?*

Oui Non

*Votre activité est-elle en lien avec l'histoire ?*

 Oui

⇒ Précisez :

 Non Retraité

*Êtes-vous un ancien professionnel de la recherche ?*

 Oui Non

*Votre ancienne activité est-elle en lien avec l'histoire ?*

 Oui

⇒ Précisez :

 Non Étudiant ou doctorant

*Discipline d'études :*

*Niveau d'études :*

 Licence Master DUT, IUT, BTS Préparation de concours Doctorat Autre Autre

⇒ Précisez :

**À titre personnel, vous menez principalement des recherches  
(plusieurs choix possibles) :**

 Généalogiques Historiques sur un sujet précis (personne, lieu, événement...) Sur la littérature Sur divers sujets, selon votre curiosité du moment Autre

⇒ Précisez si besoin :

## SUITE DU QUESTIONNAIRE (COMMUNE AUX DEUX PANELS)

### **Votre usage de la presse ancienne numérisée**

**Dans le cadre de vos recherches, vous êtes amené à consulter de la presse ancienne numérisée :**

 Intensivement Fréquemment Ponctuellement Rarement Jamais

**Pour vous, la presse ancienne numérisée est :**

- Votre corpus d'étude principal
- Une ressource importante
- Une ressource parmi d'autres, mais non prioritaire
- Vous privilégiez plutôt la consultation de collections physiques

**Quels usages avez-vous des collections de presse numérisées ?  
(plusieurs choix possibles)**

- Essentiellement de la consultation / recherche d'informations
- Consultation et extraction de données  
*Quelles méthodes de traitement et d'analyse de données employez-vous ?*
  - Tableur (ex : Excel, Calc...)
  - Base de données (ex : FileMaker, MySQL, MS Access...)
  - Logiciel de préparation de données / nettoyage et mise en forme (ex : OpenRefine)
  - Logiciel de fouille de texte ou de données / text- ou datamining (ex : Iramuteq, Alceste, Mallet...)
  - Autres méthodes  
⇒ Précisez :
- Autres usages  
⇒ Précisez :

**À quelle(s) plateforme(s) en ligne avez-vous recours pour trouver de la presse ancienne ?***(plusieurs choix possibles)*

- Gallica
- RetroNews
- Presse locale ancienne
- FranceArchives
- Geneanet
- Un ou des portail(s) en région  
⇒ Précisez le(s)quel(s):
- Autre  
⇒ Précisez :

**Naviguer / rechercher dans Gallica****Êtes-vous un utilisateur régulier de Gallica ?**

- Oui
- Non

**Avez-vous déjà eu recours à des tutoriels ou bénéficié d'une formation à la recherche dans Gallica ?**

- Oui

*Si oui, votre formation a été assurée :*

- Par la BnF
- Par une autre bibliothèque
- Par des pairs
- Autre

⇒ Précisez :

- Non, je n'en ai pas besoin
- Non mais j'en aurais besoin
- C'est prévu

**Pour rechercher dans Gallica, vous utilisez :**  
(plusieurs choix possibles)

La recherche simple

La recherche avancée (avec filtrage des résultats)

*Dans le moteur de recherche avancée, comment filtrez-vous prioritairement vos résultats ?*

- Par notice et texte intégral (auteur/contributeur, titre, cote, terme du texte, langue...)
- Par proximité de termes
- Par date de fascicule
- Par collection ou thème
- Par zone géographique
- Vous croisez plusieurs de ces filtres

Les onglets « Sélections » de la page d'accueil

*Parmi les « Sélections » de Gallica, où cherchez-vous principalement de la presse ancienne ?*

- En fonction du type de presse ou de revues (Principaux quotidiens, Hebdomadaires et presse magazine, Journaux de tranchées, Journaux clandestins, Revues savantes...)
- En fonction de la zone géographique (Presse locale et régionale, Presse par pays, Accès par régions, départements, grandes villes, outre-mer...)
- En fonction d'une thématique particulière
- Autre

⇒ Précisez :

La recherche via un moteur externe

**Quelles fonctionnalités de Gallica utilisez-vous le plus ? Sur ce thème, pensez-vous avoir quelque chose à rajouter ?**

**Naviguer / rechercher dans RetroNews**

**Êtes-vous un utilisateur régulier de RetroNews ?**

- Oui
- Non

**Disposez-vous d'un abonnement (personnel ou via votre institution) à RetroNews ?**

- Oui  
*Si oui, vous êtes abonné :*  
 A titre personnel  
 Via votre institution  
 Non

**Avez-vous déjà eu recours aux tutoriels mis en ligne sur le site RetroNews ?**

- Oui  
 Non, je n'en ai pas besoin  
 Non mais j'en aurais besoin  
 C'est prévu

**Pour rechercher dans RetroNews, vous utilisez :  
*(plusieurs choix possibles)***

- La recherche simple  
 La recherche avancée (avec filtrage des résultats)  
*Dans le moteur de recherche avancée, comment filtrez-vous prioritairement vos résultats ?*  
 Par filtrage temporel (intervalle de dates, période historique ou date de publication)  
 Par type de source (titre de journal, type de presse ou périodicité)  
 À l'aide des filtres avancés (événement, sujet, personne, organisation, lieu)  
 Vous croisez plusieurs de ces filtres  
 La recherche via un moteur externe

**Quel usage faites-vous des sélections de RetroNews (cycles thématiques, séquences pédagogiques, vidéos, diaporamas...) ?  
*(plusieurs choix possibles)***

- Usage pour des recherches  
 Usage pédagogique  
 Consultation par simple curiosité  
 Autre  
     ⇒ Précisez :   
 Je n'en ai pas l'usage

**Quelles fonctionnalités de RetroNews utilisez-vous le plus ? Sur ce thème, pensez-vous avoir quelque chose à rajouter ?**

**Vos besoins**

**Toutes plateformes confondues, comment évalueriez-vous votre niveau d'aisance pour mener une recherche dans les collections de presse numérisées ?**

- Expert  
 Très à l'aise

- Relativement à l'aise
- Peu à l'aise
- Débutant

**En matière d'aide à la recherche dans ces collections, auriez-vous des besoins spécifiques ? Pistes d'amélioration possibles :**

*(plusieurs choix possibles)*

- Amélioration des numérisations
- Amélioration des moteurs de recherche
- Amélioration des fonctionnalités de visualisation
- Amélioration des fonctionnalités d'extraction
- Mise à disposition d'outils complémentaires pour l'exploitation des corpus
- Mise à disposition de contenus préselectionnés (en fonction d'une thématique précise, d'un événement, d'une personne...)
- Mise à disposition d'un espace personnel pour sauvegarder des documents
- Mise à disposition d'un forum de discussion pour échanger avec des pairs
- Autre

**Précisez vos besoins / les manques identifiés :**

**(Facultatif) - Indiquez votre adresse électronique si vous souhaitez pouvoir éventuellement être recontacté à l'issue de cette enquête :**

## ANNEXE N°3 - DETAIL DES REPONSES

## SECTION 1 : PROFIL GENERAL

## Genre

## Chercheurs de l'ESR

Hommes	73	50%
Femmes	71	48%
Préfèrent ne pas rép.	3	2%
	<b>147</b>	<b>100%</b>

## Chercheurs amateurs

Hommes	92	38%
Femmes	144	60%
Préfèrent ne pas rép.	5	2%
	<b>241</b>	<b>100%</b>

## Localisation

## Chercheurs de l'ESR

Ile de France	55	38%
Autres départements	81	55%
Etranger	6	4%
Sans réponse	5	3%
	<b>147</b>	<b>100%</b>

## Chercheurs amateurs

Ile de France	58	24%
Autres départements	133	55%
Etranger	3	1%
Sans réponse	47	20%
	<b>241</b>	<b>100%</b>

## Chercheurs de l'ESR

## Etablissements de rattachement

<b>Chercheurs de l'ESR</b>	<b>147</b>	
Université, labo ou UMR	101	69%
<i>Paris/Rég.paris.</i>	30	20%
<i>Hors Paris</i>	71	48%
Grande école	15	10%
<i>ENS (Paris/Lyon)</i>	3	2%
<i>Sciences Po</i>	5	3%
<i>EHESS</i>	7	5%
Institut de recherche	9	6%
<i>CNRS</i>	6	4%
<i>INSERM</i>	1	1%
<i>INHA</i>	1	1%
<i>INRAP</i>	1	1%
Autre (INSPE, ingé, véto...)	4	3%
Etablissement à l'étranger	8	5%
Sans réponse	10	7%

## Statut - Activité principale

### Chercheurs de l'ESR

MCF ou PU	55	37%
Doctorant et/ou ATER	39	27%
Chercheur associé	21	14%
Enseign. et/ou contract.	15	10%
Ing.Rech. ou Ing.Etud.	8	6%
Charg. ou Dir.Rech.	6	4%
NR	3	2%
	<b>147</b>	100%

### Chercheurs amateurs

Actifs	114	47%
Retraités	102	42%
Etudiants/Doctorants	15	6%
Autre	9	4%
NR	1	1%
	<b>241</b>	100%

\*Actifs ou retraités 216  
 dont Profess.archiv. ou biblioth. 30  
 Généalogistes professionnels 9

## Chercheurs amateurs

### Activité (actuelle ou ancienne) en lien avec la recherche et/ou l'histoire

<b>Chercheurs amateurs</b>	<b>241</b>	
Activité en lien avec rech.	46	19%
Pas d'act. en lien avec rech.	170	71%
Activité en lien avec hist.	72	30%
Pas d'act. en lien avec hist.	144	60%

## Disciplines et niveaux d'études

### Amateurs : Etudiants/Doct.

LSH	7	47%
Sciences po-Droit	3	20%
Hist. du livre-Archiv.	2	13%
NR	3	20%
	<b>15</b>	100%

### Amateurs : Etudiants/Doct.

Licence-Master	12	79%
Doctorat	1	7%
Prépa concours	1	7%
NR	1	7%
	<b>15</b>	100%

## Types de recherche

<b>Chercheurs de l'ESR</b>	<b>147</b>	
Recherche individuelle	121	82%
Projet équipe monodis.	23	16%
Projet équipe pluridis.	47	32%
Projet pluri. + volet techn.	17	12%

<b>Chercheurs amateurs</b>	<b>241</b>	
Généalogiques	192	80%
Historiques	148	61%
Sur la littérature	9	4%
Selon curiosité	71	29%
Autre	6	2%
+ croisent généal./hist.	112	46%

## SECTION 2 : VOTRE USAGE DE LA PRESSE ANCIENNE NUMERISEE

## Fréquence d'utilisation de la presse ancienne numérisée

## Chercheurs de l'ESR

Intensivement	41	28%
Fréquemment	67	46%
Ponctuellement	33	22%
Rarement	4	3%
Jamais	2	1%
	<b>147</b>	<b>100%</b>

## Chercheurs amateurs

Intensivement	13	5%
Fréquemment	107	44%
Ponctuellement	105	44%
Rarement	14	6%
Jamais	2	1%
	<b>241</b>	<b>100%</b>

## Place de la presse ancienne numérisée dans la recherche

## Chercheurs de l'ESR

Corpus d'étude principal	21	14%
Ressource importante	78	53%
Ressource parmi autres	41	28%
Consultation collections	4	3%
NR	3	2%
	<b>147</b>	<b>100%</b>

## Chercheurs amateurs

Corpus d'étude principal	9	4%
Ressource importante	143	59%
Ressource parmi autres	88	37%
Consultation collections	1	0%
NR	0	0%
	<b>241</b>	<b>100%</b>

## Usages des collections de presse numérisée

## Chercheurs de l'ESR

	<b>147</b>	
Consultation/recherche	102	69%
Consultation extraction	72	49%
Autres	9	6%

## Chercheurs amateurs

	<b>241</b>	
Consultation/recherche	183	76%
Consultation extraction	95	39%
Autres	4	2%

## Méthodes de traitement et d'analyse

## Chercheurs de l'ESR

	<b>147</b>	
Tableur	50	34%
Bases de données	17	12%
Logiciel prépa/mise en f.	4	3%
Logiciel data mining	17	12%
Autres	30	20%

## Chercheurs amateurs

	<b>241</b>	
Tableur	36	15%
Bases de données	11	5%
Logiciel prépa/mise en f.	5	2%
Logiciel data mining	8	3%
Autres	36	15%

## Méthode de traitement et d'analyse selon Statut

### Chercheurs de l'ESR

Étiquettes de lignes	Valeurs				
	Tableurs	Bases de données	Logiciels prépa/mise en forme	Logiciels datamining	Autre
MCF ou PU	19	8	2	11	13
Doctorant et/ou ATER	18	3		4	6
Chercheur associé	6	2		1	1
Enseign. et/ou chercheur	4	1	1	1	4
Ing. rech. ou ing. études	2	2	1		3
Chargé ou Dir. de rech.	1				1
Sans réponse		1			2
<b>Total général</b>	<b>50</b>	<b>17</b>	<b>4</b>	<b>17</b>	<b>30</b>

### Chercheurs amateurs

Étiquettes de lignes	Valeurs				
	Tableurs	Bases de données	Logiciels prépa/mise en forme	Logiciels datamining	Autre
Actif	12	4		5	8
Etudiant ou doctorant	5	1	1		
Retraité	18	6	4	3	24
Autre	1				4
Sans réponse					
<b>Total général</b>	<b>36</b>	<b>11</b>	<b>5</b>	<b>8</b>	<b>36</b>

## Plateformes utilisées

### Chercheurs de l'ESR

	<b>147</b>	
Gallica	137	93%
RetroNews	71	48%
Presse locale ancienne	32	22%
France Archives	24	16%
Geneanet	21	14%
Portail(s) en région	15	10%
Autre	46	31%

### Chercheurs amateurs

	<b>241</b>	
Gallica	224	93%
RetroNews	133	55%
Presse locale ancienne	105	44%
France Archives	40	17%
Geneanet	141	59%
Portail(s) en région	68	28%
Autre	34	14%

## SECTION 3 : NAVIGUER / RECHERCHER DANS GALLICA

## Profils d'utilisateurs de Gallica

## Chercheurs de l'ESR

Utilisateurs réguliers	120	81%
Utilisateurs occasionnels	26	18%
NR	1	1%
	<b>147</b>	<b>100%</b>

## Chercheurs amateurs

Utilisateurs réguliers	182	76%
Utilisateurs occasionnels	59	24%
NR	0	0%
	<b>241</b>	<b>100%</b>

## Formation à Gallica

## Chercheurs de l'ESR

Formation reçue	10	7%
Pas besoin de formation	98	67%
Besoin de formation	37	25%
Formation prévue	2	1%
	<b>147</b>	<b>100%</b>

## Chercheurs amateurs

Formation reçue	26	11%
Pas besoin de formation	101	42%
Besoin de formation	110	45%
Formation prévue	4	2%
	<b>241</b>	<b>100%</b>

## Besoin en formation Gallica selon Profil utilisateur

## Chercheurs de l'ESR

Formation Gallica ?	Étiquettes	Formation reçue	Pas besoin de formation	Besoin de formation	Formation prévue	Sans réponse	Total général
Étiquettes de lignes							
Utilisateur régulier		8	82	29		1	120
Utilisateur occasionnel		2	15	7	1	1	26
Sans réponse			1				1
<b>Total général</b>		<b>10</b>	<b>98</b>	<b>36</b>	<b>1</b>	<b>2</b>	<b>147</b>

## Chercheurs amateurs

Formation Gallica ?	Étiquettes de	Formation reçue	Pas besoin de formation	Besoin de formation	Formation prévue	Total général
Étiquettes de lignes						
Utilisateur régulier		20	82	79	1	182
Utilisateur occasionnel		6	19	31	3	59
<b>Total général</b>		<b>26</b>	<b>101</b>	<b>110</b>	<b>4</b>	<b>241</b>

## Types de recherche dans Gallica

### Utilisation des différents modes de recherche

Chercheurs de l'ESR			Chercheurs amateurs		
Recherche simple	85	58%	Recherche simple	143	59%
Recherche avancée	116	79%	Recherche avancée	164	68%
Pages Sélections	19	13%	Pages Sélections	22	9%
Moteur externe	31	21%	Moteur externe	39	16%

### Utilisation conjointe de la recherche simple et avancée

Chercheurs de l'ESR			Chercheurs amateurs		
Recherche simple & avancée	59	40%	Recherche simple & avancée	83	34%
Pages Sélections	19	13%	Pages Sélections	22	9%
Moteur externe	31	21%	Moteur externe	39	16%

## Méthode de filtrage des résultats en recherche avancée dans Gallica

Chercheurs de l'ESR			Chercheurs amateurs		
Par collection ou thème	1	1%	Par collection ou thème	10	6%
Par date de fascicule	19	16%	Par date de fascicule	10	6%
Par notice et texte intégral	26	22%	Par notice et texte intégral	29	18%
Par proximité de termes	15	13%	Par proximité de termes	30	18%
Par zone géographique	1	1%	Par zone géographique	16	10%
Croisement de filtres	54	47%	Croisement de filtres	68	41%
NR	0	0%	NR	1	1%
	<b>116</b>	100%		<b>164</b>	100%

## SECTION 4 : NAVIGUER / RECHERCHER DANS RETRONEWS

### Profils d'utilisateurs de RetroNews

Chercheurs de l'ESR			Chercheurs amateurs		
Utilisateurs réguliers	53	36%	Utilisateurs réguliers	82	34%
Utilisateurs occasionnels	93	63%	Utilisateurs occasionnels	159	66%
NR	1	1%	NR	0	0%
	<b>147</b>	100%		<b>241</b>	100%

## Abonnement à RetroNews

### Chercheurs de l'ESR

Abonnés	56	38%
Non abonnés	87	59%
NR	4	3%
	<b>147</b>	<b>100%</b>

### Chercheurs amateurs

Abonnés	40	16%
Non abonnés	197	82%
NR	4	2%
	<b>241</b>	<b>100%</b>

## Utilisation de tutoriels sur RetroNews

### Chercheurs de l'ESR

Tutoriels utilisés	6	4%
Pas de besoin de tutoriels	81	55%
Besoin de tutoriels	45	31%
Tutoriels prévus	3	2%
NR	12	8%
	<b>147</b>	<b>100%</b>

### Chercheurs amateurs

Tutoriels utilisés	12	5%
Pas de besoin de tutoriels	118	49%
Besoin de tutoriels	87	36%
Tutoriels prévus	8	3%
NR	16	7%
	<b>241</b>	<b>100%</b>

## Types de recherche dans RetroNews

### Utilisation des différents modes de recherche

#### Chercheurs de l'ESR **147**

Recherche simple	51	35%
Recherche avancée	55	37%
Moteur externe	11	7%

#### Chercheurs amateurs **241**

Recherche simple	100	41%
Recherche avancée	72	30%
Moteur externe	17	7%

### Utilisation conjointe de la recherche simple et avancée

#### Chercheurs de l'ESR **147**

Recherche simple & avancée	21	14%
Moteur externe	11	7%

#### Chercheurs amateurs **241**

Recherche simple & avancée	24	10%
Moteur externe	17	7%

## Méthode de filtrage des résultats en recherche avancée dans RetroNews

### Chercheurs de l'ESR

Par type de source	8	15%
Par filtrage temporel	12	22%
À l'aide des filtres avancés	4	7%
Croisement de filtres	31	56%
NR	0	0%
	<b>55</b>	<b>100%</b>

### Chercheurs amateurs

Par type de source	9	13%
Par filtrage temporel	20	28%
À l'aide des filtres avancés	14	19%
Croisement de filtres	29	40%
NR	0	0%
	<b>72</b>	<b>100%</b>

## Usage des sélections de Retronews

Chercheurs de l'ESR			Chercheurs amateurs		
		147			241
Pour recherche	58	39%	Pour recherche	95	39%
Pédagogique	31	21%	Pédagogique	8	3%
Par curiosité	16	11%	Par curiosité	40	17%
Autre	3	2%	Autre	5	2%
Pas d'usage	33	22%	Pas d'usage	59	24%

## SECTION 5 : VOS BESOINS

### Niveau d'aisance déclaré (toutes plateformes confondues)

Chercheurs de l'ESR			Chercheurs amateurs		
Expert	13	9%	Expert	7	3%
Très à l'aise	52	35%	Très à l'aise	62	26%
Relativement à l'aise	66	45%	Relativement à l'aise	111	46%
Peu à l'aise	13	9%	Peu à l'aise	50	21%
Débutant	3	2%	Débutant	11	4%
	<b>147</b>	<b>100%</b>		<b>241</b>	<b>100%</b>

### Niveau d'aisance selon Statut

#### Chercheurs de l'ESR

Niveau d'aisance	Étiquette						Total
Étiquettes de lignes	Expert	Très à l'aise	Relativmnt à l'aise	Peu à l'aise	Débutant	Sans réponse	général
MCF ou PU	6	13	26	9		1	55
Doctorant et/ou ATER	2	22	13	1	1		39
Chercheur associé	1	7	12	1			21
Enseign. et/ou chercheur	1	4	8		1	1	15
Ing. rech. ou ing. études	2	3	2	1			8
Chargé ou Dir. de rech.		1	3	1	1		6
Sans réponse	1	1	1				3
<b>Total général</b>	<b>13</b>	<b>51</b>	<b>65</b>	<b>13</b>	<b>3</b>	<b>2</b>	<b>147</b>

## Chercheurs amateurs

Niveau d'aisance	Étiquettes					Total général
Étiquettes de lignes	Expert	Très à l'aise	Relativmnt à l'aise	Peu à l'aise	Débutant	
Actif	5	39	53	17		114
Etudiant ou doctorant		4	8	1	2	15
Retraité	2	17	46	30	7	102
Autre		1	4	2	2	9
Sans réponse		1				1
<b>Total général</b>	<b>7</b>	<b>62</b>	<b>111</b>	<b>50</b>	<b>11</b>	<b>241</b>

## Pistes d'amélioration pour l'aide à la recherche

Chercheurs de l'ESR			Chercheurs amateurs		
	147			241	
Amélioration numérisations	67	46%	Amélioration numérisations	66	27%
Amélioration mot. de rech.	87	59%	Amélioration mot. de rech.	139	58%
Amélioration visualisations	55	37%	Amélioration visualisations	73	30%
Amélioration extraction	81	55%	Amélioration extraction	96	40%
Outils complémentaires	48	33%	Outils complémentaires	40	17%
Contenus présélectionnés	15	10%	Contenus présélectionnés	47	20%
Espace personnel	46	31%	Espace personnel	73	30%
Forum de discussion	16	11%	Forum de discussion	36	15%
Autre	10	7%	Autre	11	5%

## ANNEXE N°4 - TYPOLOGIE DES BESOINS EXPRIMES PAR LES REpondANTS

Axes d'amélioration	Problèmes ou manques identifiés	Solutions préconisées	Verbatims questionnaire n°1 (chercheurs de l'ESR)	Verbatims questionnaire n°2 (chercheurs amateurs)
<b>Axe n°1 : Amélioration des moteurs de recherche et des fonctionnalités de recherche</b>				
<b>Moteurs de recherche</b>	<p>Manque de fluidité dans la navigation : résultats non triés, recherche simple ou via moteur externe jugée parfois plus efficace</p> <p>Problème de mémorisation des résultats de recherche</p> <p>Résultats de recherche partiels ou bruités</p> <p>Ressources inexploitable quand recherche plein texte impossible</p> <p>Inadaptation de certains moteurs de recherche trop peu performants</p>	<p>→ Simplifier le tri des résultats de recherche, éditorialiser/prioriser les résultats</p> <p>→ Faciliter la mémorisation des résultats et la possibilité de relancer une recherche en signalant les ajouts depuis la dernière recherche</p> <p>→ Améliorer la qualité de l'OCR</p> <p>→ Rendre l'ensemble des collections numérisées disponibles en mode texte</p> <p>→ Améliorer la finesse des moteurs de recherche et des outils d'analyse</p>	<p><i>- Gallica : amélioration de la navigation dans les résultats de recherche : une fois un corpus identifié par une recherche de terme, par exemple, il faudrait pouvoir travailler sans risque de le perdre ou de se perdre.</i></p> <p><i>- Certaines plateformes d'archives numérisées ont un moteur de recherche textuel peu performant qui loupe des occurrences quand on recherche un mot clé et parfois une OCR de qualité limitée en amont crée des problèmes de lecture (ponctuation qui saute, espaces qui disparaissent...).</i></p> <p><i>- Les moteurs de recherche sont encore à développer. L'équivalent Belge de Gallica, Belgica Press est beaucoup plus précis, permet des recherches avancées, en plein texte etc.</i></p> <p><i>- Les principaux manques concernent des collections qui sont mises en ligne sans possibilité de recherche dans le texte.</i></p> <p><i>- Problème du "bruit" dans la recherche par mot de Gallica.</i></p> <p><i>- Occurrences mieux contextualisées dans la liste de résultats.</i></p> <p><i>- Sur les sites d'archives numérisées notamment par les archives départementales le moteur de recherche est simple</i></p>	<p><i>- Possibilité de relancer une recherche déjà utilisée en voyant les ajouts depuis la dernière recherche (nouveaux documents numérisés ou mieux indexés).</i></p> <p><i>- Selon les sites, les moteurs de recherche sont plus ou moins adaptés. Idem pour les fonctions de visualisation et/ou extraction.</i></p> <p><i>- Souvent sur les sites des archives départementales il n'y a pas de possibilité de recherche par mot-clé.</i></p> <p><i>- En ce qui me concerne, l'amélioration prioritaire serait celle des moteurs de recherche, afin de simplifier le repérage préalable dans l'arborescence riche mais touffue des nombreuses ressources documentaires de la BnF.</i></p> <p><i>- Un accès indexé simplifié pour certaines périodes notamment sur Retronews et Gallica.</i></p> <p><i>- Dans Retronews la recherche par proximité est indispensable, de même que la recherche géographique.</i></p>

	<p>Insuffisance ou absence, selon les plateformes, de certains filtres de recherche (par mots-clés, titre exact, proximité de termes, zone géographique, période historique, etc.)</p>	<p>Généraliser à l'ensemble des plateformes la recherche par mots-clés, par titre de journal et par proximité de termes. Permettre un filtrage historique et/ou géographique</p>	<p><i>et assez inefficace. Ces sources sont donc quasiment inexploitable pour une analyse scientifique pointue.</i></p> <p><i>- Sur Gallica, la recherche simple est souvent plus opératoire que la recherche avancée en utilisant les guillemets pour obtenir une expression. Certains documents ne sont pas disponibles en mode texte (même s'il y a des progrès évidents), et font encore l'objet d'un seul pdf (par exemple la revue Mon Camarade pour laquelle Gallica a les numéros 1935-1939 en un seul fichier compact).</i></p> <p><i>- Dans Gallica, si vous tapez le titre exact, il est très fréquent que ce titre ne sorte pas. En revanche, si vous sélectionnez le nom de l'auteur, dans le menu déroulant, l'ouvrage s'y trouve. Paradoxalement, en faisant une recherche Google, on tombe sur la référence qu'on n'a pas réussi à acquérir via le moteur de recherche Gallica. C'est pour moi le sujet le plus handicapant.</i></p> <p><i>- Pour la recherche en plein texte, il manque parfois la possibilité de rechercher des mots à la proximité d'autres mais sans être adjacents, un booléen du type NEAR qui serait suffisamment puissant.</i></p>	
OCR	<p>OCR non systématique, ce qui ne permet pas toujours la recherche plein texte</p> <p>Qualité de l'OCR insuffisante, surtout pour les titres anciens et/ou les numérisations anciennes</p> <p>Correction manuelle de l'OCR =&gt; perte de</p>	<p>Généraliser l'OCR à l'ensemble des collections mises en ligne</p> <p>Améliorer la qualité de l'OCR, surtout sur les titres anciens et numérisations anciennes</p> <p>Envisager une correction d'OCR automatisée</p>	<p><i>- Amélioration de l'OCR sur les corpus XVIIe.</i></p> <p><i>- Amélioration de la qualité des océrisations.</i></p> <p><i>- Amélioration de la qualité de l'OCR. Beaucoup de titres ont été numérisés avant l'apparition de cette technologie et ne sont pas "cherchables" en plein texte.</i></p> <p><i>- Une OCRisation systématique des documents (c'est très lacunaire sur Gallica).</i></p> <p><i>- Surtout une amélioration de la reconnaissance des mots dans les textes anciens.</i></p>	<p><i>- Qualité de l'OCR à améliorer.</i></p> <p><i>- Le point le plus important pour moi est l'amélioration des moteurs de recherche, passant par une amélioration de la reconnaissance de caractères. Mais pour avoir travaillé sur des plateformes étrangères (Belgique, Italie, Malte), les plateformes françaises me semblent assez bien placées sur ce point.</i></p> <p><i>- Pages pas toujours pourvues d'un OCR et donc pas toujours possible d'effectuer une recherche plein texte, d'où une perte de temps.</i></p>

	temps		- Plus de textes accessibles en OCR pour améliorer les recherches.	- La reconnaissance OCR pêche un peu parfois sur les impressions anciennes, c'est dommage car cela fausse le résultat de recherche.  - La LAD* est parfois très hasardeuse (non reconnaissance de plusieurs mots simples) car les numérisations sont dans ces cas lisibles pour un œil humain mais de mauvaise qualité pour une IA. Beaucoup de travail de correction à effectuer. (* LAD : Lecture Automatique des Documents)
<b>Téléchargement / extraction des résultats</b>	Manque de souplesse des fonctionnalités de téléchargement d'un document / extrait de document  Difficulté d'extraction des données à partir des résultats de recherche	Adapter la granularité des fonctionnalités de téléchargement aux besoins du chercheur (document intégral ocrisé, page, entrefilet...)  Faciliter l'extraction des données (listes de dépouillement) à partir des résultats de recherche	- Certaines plateformes d'archives numérisées ne permettent pas le téléchargement intégral d'un document mais seulement le téléchargement page par page, cela provoque une perte de temps importante et un travail fastidieux de recollement des documents. De manière générale, tout ce qui permet de travailler plus vite sur la documentation est une bonne idée.  - J'ai remarqué qu'on ne pouvait plus sélectionner et télécharger des zones de texte (correspondant à un article ou à un entrefilet) sur Gallica. Je trouvais cette fonctionnalité extrêmement utile.  - Pouvoir télécharger l'intégralité de la version OCRisée du document. Avoir un moteur de recherche efficace dans Gallica.	- Possibilité de transformer des résultats de recherche en listes de dépouillement.  - La seconde amélioration à prendre en compte est celle touchant aux fonctionnalités d'extraction, pour rendre plus aisée l'exploitation des documents récupérés.
<b>Options de visualisation</b>	Difficulté pour visualiser et accéder à l'ensemble des numéros d'un même titre  Zoom peu commode/lisible voire impossible	Améliorer la vue d'ensemble des numéros d'un même titre  Systématiser le zoom, améliorer le cas échéant sa qualité et son ergonomie (ex : zoom ciblé sur la zone comprenant l'article ou le mot-clé recherché)	- Sur Gallica il est difficile d'avoir accès à tous les numéros, toutes les années à la suite, il faut à chaque fois arriver sur une sorte de calendrier pour voir les numéros qui existent par mois. Cela demande beaucoup de manipulation/temps pour chaque fois revenir au départ. En plus on a peu de visibilité pour voir d'un seul coup d'œil tous les numéros qui existent. Sur Retronews, une fois qu'on a trouvé le mot-clé surligné en jaune (ça c'est pratique) il est souvent peu pratique de zoomer, on se perd dans la page de journal scannée.	- Que la qualité de la numérisation permette d'accéder à un petit entrefilet au sein, par exemple, de la grande page d'un quotidien. Même chose pour les images.  - Gallica surligne le mot de la recherche, mais le zoom de la page auquel on accède en premier lieu est à 100% il faut en suite sélectionner le zoom et lorsque le mot est court il est difficile de le retrouver dans la page. Un souci qui se poserait moins si l'on arrivait dans la page avec un zoom concentré sur le quart de la page correspondant à la sélection.

	<p>Manque d'options de visualisation cartographique pour les journaux locaux</p> <p>Différents problèmes de visualisation selon les plateformes</p>	<p>Permettre la géolocalisation des titres numérisés (en particulier en région)</p> <p>Améliorer la qualité de visualisation des résultats / optimiser l'ergonomie des plateformes (défilement des images, surlignage...)</p>	<p>- Zoom pour lecture possible de tous les documents.</p> <p>- Pouvoir géolocaliser les titres de journaux locaux.</p>	<p>- Il manque un ciblage direct sur l'article repéré par la recherche, car la visualisation montre la page entière.</p> <p>- Parfois difficile de retrouver une donnée sur une page.</p> <p>- Avoir la zone géographique avec degré d'importance que couvrent les journaux locaux.</p> <p>- Stabilité du défilement des images parfois difficile.</p> <p>- Surlignage jaune des mots clés de recherche impossible à effacer.</p>
<b>Espaces personnels</b>	<p>Manque de possibilités d'annotation des documents sauvegardés</p>	<p>Ajouter, au sein des espaces personnels, un champ d'annotation des documents sauvegardés</p>		<p>- J'utilise déjà les espaces personnels mais j'aimerais pouvoir ajouter des notes en regard des documents sélectionnés.</p>

## Axe n°2 : Augmentation des numérisations et enrichissement des métadonnées

<b>Programmes de numérisation</b>	<p>Lacunes dans les collections numérisées (en particulier pour la presse ancienne locale)</p> <p>Difficulté d'accès à la presse contemporaine sous droits</p>	<p>Accroître les programmes de numérisation (notamment dans les centres d'archives en région) pour combler les lacunes existantes</p> <p>Favoriser autant que possible l'accès à la presse sous droits</p>	<p>- Problème des collections incomplètement numérisées.</p> <p>- Certaines revues non numérisées mais gros effort depuis 5 ans.</p> <p>- Nouveaux programmes de numérisation.</p> <p>- Numérisation plus large.</p> <p>- Mise à disposition de plus de titres de presses locaux.</p> <p>- Encourager autant que possible les numérisations des collections de la presse contemporaine (19e-20e siècle) de la part de la BNF.</p> <p>- Favoriser la numérisation de la part des centres d'archives (privés, publics ; parisiens ou régionaux) disposant des fonds de presse.</p>	<p>- Dommage que certains départements ne proposent aucune presse ancienne !</p> <p>- Insuffisance de journaux locaux numérisés sur les sites des archives départementales et ou municipales. Insuffisance d'informations numérisées alors que les musées, les archives départementales ou municipales regorgent d'informations précieuses qui dorment dans des cartons.</p> <p>- Numérisations de nombreux titres non disponibles en ligne (presse régionale) exemples : L'Ariégeois, le Journal de Saint Giron, les journaux associatifs et religieux (paroisses). Ils ne sont disponibles qu'aux AD de chaque département. Enfin les journaux clandestins de la Seconde Guerre mondiale.</p> <p>- Il faudrait plus de presses anciennes et locales numérisées.</p>
-----------------------------------	--	--	--	---

			- Même s'il y a beaucoup de choses, on ne trouve pas tout sur Gallica (en particulier pour la presse régionale).	
<b>Gratuité d'accès aux ressources</b>	<p>Problème des ressources (notamment étrangères) disponibles uniquement en accès payant</p> <p>Système Freemium (ressources en accès libre / fonctionnalités de recherche avancées) =&gt; frein à la recherche</p>	<p>Mettre en place une réglementation (consortium européen ?) contre le monopole de plateformes étrangères sur certaines ressources numérisées</p> <p>Limiter les ressources / fonctionnalités accessibles uniquement sur abonnement / négocier des tarifs préférentiels pour certaines professions</p>	<p>- Tout n'est pas numérisé, hélas ; et dans les domaines étrangers, beaucoup de matériel numérisé est difficilement accessible, car payant, et fort cher. Les bibliothèques françaises font des efforts louables pour les garder accessibles, mais manquent de moyens. Il faudrait un consortium européen, et aussi une politique de réglementation des entreprises étrangères de numérisation, qui se comportent en monopoles abusifs ; les menacer d'investir dans la mise en ligne gratuite aiderait probablement. Dans mon domaine (histoire de l'Amérique du Nord), le principal problème est économique, pas technique !</p> <p>- La recherche dans Retronews est super, mais l'accès sous paywall à notre patrimoine et à des outils avancés est un principe affligeant.</p> <p>- Barrière à l'entrée de Retronews, et absence d'abonnement institutionnel par l'université pour y accéder.</p>	<p>- J'ai besoin que Retronews ne soit pas hors de ma portée étant donné son coût prohibitif.</p> <p>- Des plateformes en accès libre surtout lorsqu'il s'agit de collections publiques qui ont été numérisées.</p> <p>- Libre réutilisation des contenus (pour la plupart dans le domaine public) et gratuité d'accès (Retronews).</p> <p>- Je ne veux pas payer, surtout pour chercher "à l'aveugle"... sans savoir si je vais trouver quelque information utile à la compréhension d'un événement local touchant l'un des ancêtres familiaux...</p> <p>- Besoin de baisser le prix de l'abonnement à RetroNews, ce qui permettrait à un plus large public de l'utiliser.</p> <p>- Gratuité de RetroNews.</p> <p>- Retronews devrait avoir un tarif préférentiel pour les généalogistes professionnels.</p>
<b>Indexation / métadonnées</b>	<p>Problème des périodiques numérisés en volumes</p> <p>Problème des journaux comprenant plusieurs éditions</p>	<p>Affiner les métadonnées pour les publications en volume / permettre la séparation des fascicules</p> <p>Préciser l'édition numérisée pour les titres avec plusieurs éditions</p>	<p>- Sur Gallica, problème des périodiques numérisés en volumes, comme des monographies.</p> <p>- Amélioration des métadonnées (pour les journaux qui disposent de plusieurs éditions, on ne sait pas quelle édition est numérisée).</p> <p>- Meilleur traitement des journaux anciens reliés sous forme de volume, afin de pouvoir séparer les fascicules et ne pas avoir affaire à un ensemble de 1000 pages sans autre précision (ex. La Mode dans Gallica).</p>	
<b>Coordination / centralisation</b>	Dissémination des ressources entre divers	Favoriser les programmes de	- Collections trop disséminées.	- Dispersion de l'information numérisée. Au sein d'un même département tout devrait être centralisé au sein d'une même

des ressources	établissements de conservation / lacunes  Manque de visibilité sur les calendriers de numérisation / mise en ligne	numérisation concertée et la centralisation des ressources (à l'échelle d'un département / à l'échelle nationale)  Mieux communiquer sur les calendriers de numérisation / mise en ligne	- Favoriser la mise en relation des centres d'archives (privés, publics) disposant de fonds de presse auprès de la BNF pour faciliter le processus de numérisation. En effet, il existe parfois des manques et des lacunes dans les catalogues et dans les collections de Gallica / Retronews (il manque des années ou des numéros dans les collections qui sont parfois consultables physiquement car les fonds conservés mais difficiles d'accès car disséminés dans des centres d'archives parisiens ou régionaux).  - Communication sur les calendriers de numérisation et de mise en ligne.	entité.  - Faire en sorte de compléter les lacunes dans les corpus de journaux en nouant des partenariats avec d'autres bibliothèques possédant les collections.
<b>Axe n°3 : Mise à disposition d'outils d'analyse et enrichissement sémantique des collections</b>				
Analyse de données textuelles (textométrie)	Outils d'analyse/fouille de texte et de visualisation de données inopérants quand taux d'OCR insuffisant  Manque d'outils de textométrie	Améliorer la qualité de l'OCR (en particulier sur les corpus anciens) pour favoriser l'analyse des corpus textuels  Mettre à disposition davantage d'outils de textométrie	- Mise à disposition d'outils de textométrie.	- Au niveau des fonctionnalités d'extraction, la principale difficulté réside dans les taux, souvent faibles, d'ocrisation en ce qui concerne la presse provinciale antérieure aux années 1870-1880. Ce qui complique, voire rend impossible, l'utilisation de logiciel type Voyant Tools. Le recours à la base de données pour analyser le contenu du journal reste donc à l'heure actuelle la seule solution pertinente.
Outils complémentaires	Difficulté pour chercher dans des zones spécifiques de la page  Absence d'outils permettant d'établir des concordances entre plusieurs périodiques  Difficulté pour chercher des images dans les périodiques	Développer l'OLR pour permettre la segmentation des titres numérisés  Identifier et signaler les reprises d'articles et autres concordances entre périodiques  Mettre à disposition des outils de recherche d'image	- Les outils qui me seraient utiles : ▪ des outils pour chercher dans des zones de texte spécifiques (par exemple dans le feuilleton ou dans les annonces) ; ▪ des outils pour établir des concordances entre plusieurs périodiques (identifier les reprises d'articles, par exemple) ; ▪ un outil pour chercher les images dans les périodiques (par exemple les gravures de mode).  - Je cherche des informations ponctuelles donc je me satisfais à peu près des outils sommaires mis à disposition. Mais des outils plus travaillés m'inciteraient probablement à recourir encore plus à ce type de sources.	

	<p>Difficulté pour obtenir toute la presse sur un jour en particulier</p> <p>Difficulté d'accès à certaines API</p>	<p>Donner accès à toute la presse disponible pour un jour donné (cf. RetroNews : « Que s'est-il passé le... ? »)</p> <p>Faciliter l'accès aux API pour favoriser l'échange de données et de fonctionnalités</p>	<p>- Par rapport aux plateformes comparables à l'international, la seule chose qui manque est la possibilité d'obtenir toute la presse pour tel ou tel jour. Mais est-ce que bonne idée ? Cela favorise le côté événementiel pur et dur par rapport à des recherches de long terme.</p> <p>- Existence plus explicite et facilitée d'API.</p>	
<b>Enrichissement sémantique</b>	<p>Absence de distinction entre sujet et objet pour les titres recherchés</p> <p>Impossibilité actuelle de rechercher par personne ou par lieu</p>	<p>Améliorer l'enrichissement sémantique des collections</p> <p>Travailler sur les entités nommées / entités géographiques pour permettre l'analyse sémantique des corpus</p>	<p>- Il est fréquent que le titre recherché soit présent en citation (de dicto) et non comme objet cible (de re). Il manque à Gallica un bouton permettant de choisir.</p> <p>- Analyse sémantique, recherche par personne, lieu, plutôt que par chaîne de caractères.</p>	
<b>Axe n°4 : Partage des résultats de recherche</b>				
<b>Espaces collaboratif</b>	<p>Manque d'espaces d'échange pour partager ses résultats entre pairs</p> <p>Manque de coordination sur la correction des OCR</p>	<p>Mise en place de forums de discussion</p> <p>Prévoir un espace collaboratif permettant une correction participative des textes ocrésés</p>		<p>- La création d'un forum de discussion me semble être une excellente idée.</p> <p>- Pourquoi pas aller jusqu'à prévoir un espace collaboratif où l'on pourrait déposer les textes corrigés afin d'en faire profiter la collectivité.</p>
<b>Outils de partage</b>	<p>Difficulté pour identifier, citer et partager des coupures de presse (lien mais aussi contenu textuel / iconographique)</p>	<p>Favoriser l'identification, la citation et le partage d'articles ou d'extraits (texte, image...)</p>		<p>- Possibilité d'identifier aisément une coupure de presse pour la retrouver, la citer et la partager.</p> <p>- Pouvoir partager non seulement la référence ou le lien, mais aussi extraire le contenu référencé, sous un format texte ou un format image.</p>

## Axe n°5 : Formation

<b>Repérage et usage des ressources</b>	Difficulté pour appréhender les ressources disponibles →	Améliorer (à tous les niveaux) la formation à la recherche documentaire et à l'usage des ressources en ligne	<p>- <i>Besoin de formation en ligne.</i></p> <p>- <i>D'une manière générale, j'ai un peu de mal à trouver ce que je cherche.</i></p>	<p>- <i>Je ne sais pas utiliser les ressources mises à disposition.</i></p>
<b>Recherche avancée et tri des résultats</b>	Problèmes pour affiner les recherches : saisie des mots-clés et tri des résultats (notamment quand homonymie) →	Améliorer l'aide à la recherche (formations, tutoriels en ligne) pour répondre aux difficultés récurrentes des usagers		<p>- <i>La recherche avancée est compliquée.</i></p> <p>- <i>J'ai du mal, concernant les recherches sur les noms propres homonymes de noms communs, à indiquer des mots complémentaires qui permettraient de mieux cibler les termes recherchés.</i></p> <p>- <i>Besoin d'aide à la saisie de "mots-clés" pertinents.</i></p> <p>- <i>Gallica : j'ai des difficultés à trouver ce que je cherche bien souvent. Surtout quand le mot-clé est en même temps un nom commun ou a plusieurs utilisations et quand dans ma recherche j'utilise un groupe de mots. Exemples : les noms de famille « Barousse » et la vallée. Ou l'Ariège (rivière). Ou « les moulins sur le Gers » ou « la vallée de la Save » : si j'écris « les grottes préhistoriques de la vallée de la Save », je ne sais pas trouver les réponses alors que je sais qu'il y a de très nombreux articles depuis plus de 170 ans.</i></p> <p>- <i>Quand je lance une recherche, j'obtiens des résultats qui prennent en compte chaque mot mais pas l'ensemble ce qui donne beaucoup trop de résultats.</i></p> <p>- <i>Les résultats de recherches sur Retronews sont trop larges ou alors je n'arrive pas à bien faire mes recherches. Quand je tape un prénom plus nom, j'aimerais bien qu'ils ne me proposent pas les résultats de tous les noms de famille identiques d'un côté et tous les prénoms identiques de l'autre mais juste prénom et nom ensemble.</i></p>



## TABLE DES ILLUSTRATIONS

---

Figure 1 - Statut des chercheurs de l'ESR .....	51
Figure 2 - Activité des chercheurs amateurs.....	51
Figures 3 et 4 - Types de recherche .....	53
Figures 5 et 6 - Fréquence d'utilisation de la presse ancienne numérisée ....	55
Figures 7 et 8 - Place de la presse ancienne numérisée dans la recherche....	55
Figures 9 et 10 - Détail des usages.....	56
Figures 11 et 12 - Méthodes de traitement et d'analyse .....	57
Figures 13 et 14 - Méthodes de traitement et d'analyse privilégiées selon le statut.....	57
Figures 15 et 16 - Plateformes utilisées .....	58
Figures 17-18 (gauche) et 19-20 (droite) - Profils d'utilisateurs de Gallica et Formation à Gallica .....	60
Figures 21 et 22 - Besoin en formation à Gallica selon le profil d'utilisateur.....	60
Figures 23 et 24 - Modes de recherche dans Gallica.....	61
Figures 25 et 26 - Méthode de filtrage en recherche avancée dans Gallica ..	61
Figures 27 et 28 - Profils d'utilisateurs de RetroNews.....	62
Figures 29 et 30 - Abonnement à RetroNews .....	63
Figures 31 et 32 - Utilisation de tutoriels RetroNews .....	63
Figures 33 et 34 - Modes de recherche dans RetroNews .....	64
Figures 35 et 36 - Méthode de filtrage en recherche avancée dans RetroNews .....	64
.....	64
Figures 37 et 38 - Usage des sélections de RetroNews .....	64
Figures 39 et 40 - Niveau d'aisance déclaré (toutes plateformes confondues) .....	65
.....	65
Figures 41 et 42 - Niveau d'aisance selon le statut .....	66
Figures 43 et 44 - Pistes d'amélioration pour l'aide à la recherche .....	67



# TABLE DES MATIERES

---

<b>SIGLES ET ABBREVIATIONS .....</b>	<b>7</b>
<b>INTRODUCTION.....</b>	<b>9</b>
<i>Cadre général et problématique d'étude.....</i>	<i>9</i>
<i>Périmètre de l'étude et de l'enquête de publics.....</i>	<i>11</i>
<b>I. PRESERVER, VALORISER : LE PAYSAGE DE LA PRESSE ANCIENNE EN FRANCE.....</b>	<b>15</b>
<b>1.1. Numériser pour sauvegarder un patrimoine en danger : état des lieux, constats et premières réponses .....</b>	<b>15</b>
<i>1.1.1. La presse ancienne, un objet complexe .....</i>	<i>15</i>
De quelle presse parle-t-on ? .....	15
Des collections immenses et disséminées, difficiles à signaler .....	17
Des collections hétérogènes, posant des difficultés d'indexation .....	19
<i>1.1.2. Un patrimoine en péril .....</i>	<i>19</i>
Un état de conservation des collections alarmant.....	19
Numériser pour sauvegarder : une volonté politique réaffirmée, mais des moyens à renforcer .....	20
<i>1.1.3. Les solutions mises en œuvre pour améliorer le signalement et la conservation des collections .....</i>	<i>21</i>
Le rôle de Presse locale ancienne, héritier de la BIPFPIG .....	21
Le futur Conservatoire national de la presse, une réponse apportée aux problèmes de conservation et de stockage.....	22
<b>1.2. Numériser pour mieux valoriser : les axes de valorisation de la presse ancienne .....</b>	<b>23</b>
<i>1.2.1. Programmes de numérisation à l'échelle nationale et régionale : les grands jalons .....</i>	<i>23</i>
<i>1.2.2. Une coopération renforcée pour mieux valoriser la presse ancienne régionale .....</i>	<i>25</i>
L'action territoriale de la Bibliothèque nationale de France.....	25
Constituer une collection de référence complète et homogène : l'objectif du marché « Presse 2021 » .....	26
<i>1.2.3. Gallica et RetroNews, deux plateformes nationales pour diffuser les collections numérisées .....</i>	<i>27</i>
La mise en valeur de la presse dans Gallica : différents modes d'accès pour différents publics .....	27
RetroNews, le « site média d'histoire » de la BnF : l'accent mis sur l'éditorialisation .....	29
<b>1.3. Différentes stratégies de valorisation selon les territoires.....</b>	<b>31</b>

1.3.1. <i>Des modes de diffusion pluriels : de l'indépendance à l'intégration dans Gallica, différents positionnements à l'égard de l'opérateur national</i> .....	31
1.3.2. <i>Exemples de valorisation numérique en région</i> .....	34
Limédia Kiosque : une expérimentation innovante de mutualisation des ressources d'un territoire .....	34
Impliquer les habitants dans la redécouverte de leur patrimoine .....	35
Opérations en lien avec l'EMI et l'EAC.....	37
1.3.3 <i>Un écosystème en pleine mutation : vers une mise en commun des ressources ?</i> .....	38
<b>II. DE LA VALORISATION A L'EXPLOITATION DES CORPUS : DE NOUVEAUX BESOINS LIES A LA NUMERISATION DE LA PRESSE ANCIENNE</b> .....	<b>41</b>
<b>2.1. Le lecteur de presse ancienne, usager aux mille visages</b> .....	<b>41</b>
2.1.1. <i>L'exploitation des collections de presse, enjeu-clé de leur valorisation</i> .....	41
2.1.2. <i>Diversité et porosité des usages</i> .....	42
2.1.3. <i>Le « sacre de l'amateur » : vers un public amateur de plus en plus éclairé ?</i> .....	43
<b>2.2. Ce que le numérique fait au journal : un rapport nouveau à la presse ancienne en tant que source</b> .....	<b>45</b>
2.2.1. <i>Presse numérisée et « remédiation » : de nouvelles approches de recherche</i> .....	45
2.2.2. <i>« Feuilletter la presse par Giga Octets » : les enjeux de la fouille et de la visualisation de données</i> .....	47
L'émergence de nouveaux besoins de recherche : text-mining, datamining, datavisualization .....	47
La démocratisation de ces usages experts à l'heure de la massification des données .....	48
<b>2.3. Étude de publics : mieux connaître les usages et les besoins de deux catégories de lecteurs</b> .....	<b>50</b>
2.3.1. <i>Cadrage de l'étude</i> .....	50
2.3.2. <i>Synthèse des résultats</i> .....	51
Profils des répondants .....	51
Types de recherche.....	52
Usages de la presse ancienne numérisée.....	54
Plateformes utilisées.....	57
Utilisation de Gallica et RetroNews .....	59
Besoins et manques identifiés .....	65
Bilan de l'enquête : premières pistes.....	67
<b>III. ADAPTER LES SERVICES AUX BESOINS POUR MIEUX VALORISER LA PRESSE ANCIENNE</b> .....	<b>69</b>

<b>3.1. Vers plus de médiation numérique et de formation .....</b>	<b>69</b>
3.1.1. <i>La prise en compte du parcours utilisateur : une réflexion à poursuivre sur l'ergonomie des interfaces .....</i>	69
Parler à tous les publics .....	69
Prendre en compte le public généalogiste et les usagers non experts	70
3.1.2. <i>L'importance de la formation .....</i>	72
<b>3.2. Vers une amélioration des outils de recherche et d'exploitation des corpus.....</b>	<b>73</b>
3.2.1. <i>La qualité de l'OCR, préalable indispensable à l'évolution des moteurs et fonctionnalités de recherche .....</i>	73
3.2.2. <i>Un rôle accru à jouer pour les bibliothèques dans l'accompagnement à la recherche : les chantiers actuels .....</i>	75
Valoriser la double composante textuelle et illustrée des journaux...	75
Enrichissement sémantique et rubriquage automatique des collections .....	76
Une réalisation exemplaire : le projet NewsEye .....	77
Élargir les usages de ces outils : un défi pour les bibliothèques numériques aujourd'hui .....	78
3.2.3. <i>Des approches innovantes pour améliorer la qualité de la médiation .....</i>	78
<b>3.3. Vers une coopération renforcée : désenclaver pour mieux valoriser .....</b>	<b>80</b>
3.3.1. <i>Faciliter la recherche fédérée dans les corpus de presse : un horizon réaliste aujourd'hui .....</i>	80
« Dérégionaliser » les corpus : une logique d'ouverture .....	80
Vers la constitution d'une collection virtuelle unifiée ? .....	81
3.3.2. <i>Intensifier les collaborations avec les partenaires académiques et les bibliothèques étrangères .....</i>	83
La nécessité d'un travail conjoint avec les chercheurs et structures de recherche .....	83
Favoriser l'ouverture des données : l'appui des infrastructures européennes.....	84
Coopérer avec les bibliothèques européennes .....	85
3.3.3. <i>S'appuyer sur les communautés de pratique.....</i>	86
<b>CONCLUSION .....</b>	<b>89</b>
<b>SOURCES.....</b>	<b>91</b>
<b>BIBLIOGRAPHIE.....</b>	<b>95</b>
<b>ANNEXES.....</b>	<b>105</b>
<b>TABLE DES ILLUSTRATIONS .....</b>	<b>135</b>
<b>TABLE DES MATIERES .....</b>	<b>137</b>