

## Numériser la presse ancienne

### 1. Qu'est-ce que la presse ancienne ?

L'histoire de la presse écrite débute au XVII<sup>e</sup> siècle avec l'apparition des premières gazettes dont celle de Théophraste Renaudot en 1631. Toutefois, la presse ne connaît un véritable essor qu'au XIX<sup>e</sup> siècle du fait de l'industrialisation et la démocratisation de la société, ainsi que de la fin de la censure ([loi du 29 juillet 1881](#)). Se développent alors la presse quotidienne, hebdomadaire, mensuelle, la presse généraliste et spécialisée, ainsi que la presse locale. On peut situer **la fin de la presse ancienne en 1944**, date d'un renouvellement complet de la presse française. C'est la date qui semble avoir été retenue, le plus souvent, dans les plans de numérisation de la presse ancienne.

De nombreuses institutions ont initié des projets de numérisation de leurs collections de périodiques anciens. Il s'agit, en particulier, des archives départementales et municipales, des agences de coopération du livre et de la lecture et des bibliothèques territoriales.

### 2. Quelques questions préalables

#### Les objectifs

La définition des finalités de la numérisation et du public visé demeure un préalable incontournable. Le choix du corpus à numériser détermine ici grandement les objectifs du projet.

- **Les impératifs de conservation**

La numérisation des documents de presse est une réponse aux problématiques posées par la conservation des collections physiques :

- **la dégradation du papier** : les papiers sur lesquels avaient été imprimés les journaux jusqu'à une période récente sont de mauvaise qualité et se dégradent rapidement. La lignine, un constituant du bois, contient des groupements chimiques qui s'oxydent. Cela conduit à un jaunissement du papier.
- **les grands formats** : les documents de presse ont souvent des grands formats. La numérisation permet d'externaliser le stockage des documents originaux, source d'encombrement, et de conserver uniquement, *in situ*, les versions numérisées.

- **Les impératifs pédagogiques et culturels**

Les campagnes de numérisation de ces documents sont une manière d'élargir la diffusion de ces documents précieux et leur visibilité. La presse est, en effet, un matériau d'exploitation pédagogique irremplaçable. Dans l'enseignement primaire et secondaire, la [Semaine de la presse et des médias dans l'École®](#) est largement plébiscitée. La numérisation de la presse régionale et locale a aussi pour finalité de valoriser un patrimoine spécifique, souvent peu connu.

- **Les impératifs scientifiques**

La numérisation de la presse ancienne permet aux chercheurs de vérifier plus aisément et plus rapidement leurs données (faits, citations, etc.) et de renouveler leurs objets d'étude, en particulier concernant l'histoire des médias.

#### Les aspects juridiques

- **La protection du droit d'auteur**

Les journaux sont des créations intellectuelles originales et sont donc protégées par le droit d'auteur. Ce sont aussi des œuvres collectives. La durée des droits patrimoniaux est de 70 ans à compter de la date de publication des journaux. De ce fait, la grande majorité de la presse ancienne appartient au domaine public. Concernant les périodiques de moins de 70 ans, il est possible de recourir à une exception au droit d'auteur dans le cadre de la numérisation de ces documents : l'exception à des fins de conservation ([article L 122.5 du Code de la propriété intellectuelle](#)). Cette exception permet de reproduire et de communiquer sans autorisation expresse, mais uniquement au sein de l'établissement, une œuvre du fonds encore protégée par le droit d'auteur, lorsque celle-ci est abîmée ou figure sur un support obsolète. [Mais de récents développements pourraient changer la donne.](#)

- **La protection des données personnelles**

Le délai de protection des données personnelles est de 100 ans. Bien que les journaux contiennent de nombreuses données personnelles, l'anonymisation du fonds n'est pas indispensable dans la majorité des cas. Toutefois, tout projet de numérisation doit faire l'objet d'une déclaration à la Commission nationale de l'informatique et des libertés (CNIL) en vertu de la [Loi n°78-17 relative à l'informatique, aux fichiers et aux libertés du 6 janvier 1978](#) (articles 36, 40, et 68 à 70).

### Les aides financières nationales

La bibliothèque peut se tourner vers le plan national de numérisation du ministère de la Culture et de la Communication, lancé en 1996. Il propose des [appels à projets annuels de numérisation](#) et soutient les [services numériques innovants](#). L'appel à projet 2013 permettait le financement de projets de numérisation de collections entrant dans le champ de l'« histoire locale » et donc de titres de la presse régionale et locale. Des opérations planifiées peuvent également être mises en place dans le cadre des [partenariats de la Bibliothèque nationale de France \(BnF\)](#) avec ses « pôles associés régionaux » (PAR). Ceux-ci pilotent et coordonnent des actions de signalement et de valorisation numérique de documents issus de fonds patrimoniaux, en concertation avec la BnF, et avec son soutien. À titre d'exemple, voir [le pôle associé Région Languedoc-Roussillon](#). [Voir un autre article](#).

## 3. Les problématiques techniques

### Les contraintes matérielles

La presse ancienne présente des spécificités contraignantes : faible contraste, grand format, taches, impression irrégulière, report d'encre entre les pages, papier acide et parfois déchiré, juxtaposition de textes et d'illustrations. La numérisation étant de meilleure qualité si elle est effectuée à partir des originaux et non de microfilms, il est ainsi souvent indispensable de dérelier et consolider les pages pour les numériser à plat par balayage. Une résolution de 300 dpi (valeur qui exprime le nombre de pixels composant une ligne d'un pouce) paraît amplement suffisante, quel que soit le format du document. Le choix du mode niveaux de gris s'avère le plus pertinent car restituant mieux les nuances colorimétriques. Le mode image présentant cependant peu d'intérêt en termes de navigation dans les documents, une reconnaissance optique de caractères du texte ([OCR](#)) est incontournable.

### Les fonctionnalités

- **Une consultation rapide et intuitive**

Afin de permettre des opérations de dépouillement, la navigation titre à titre est essentielle, comme le recours à un menu déroulant du sommaire. Pour chaque numéro, plusieurs types d'affichage doivent être possibles (page à page, mosaïque, zoom). Pour affiner la recherche, des filtres (date, titre du journal, etc.) doivent être mis en place, ainsi que différents ordres de classement des résultats (pertinence, titre de presse, date d'édition, etc.).

- **La recherche plein texte**

La recherche plein texte grâce à l'OCRisation demeure une plus-value essentielle pour permettre un gain de temps dans l'étude des fonds. Afin d'exploiter les résultats de l'OCR, un format basé sur le standard [ALTO](#) (*Analyzed Layout and Text Object*) peut être utilisé. Celui-ci permet la construction de différents types de fichiers de diffusion dont le PDF multicouche et la surbrillance des mots recherchés.

- **Une description fine des objets numérisés**

Plusieurs niveaux de granularité (volume, fascicule, article, etc.) et donc de description (métadonnées sur le journal en général, sur une page, un article, etc.) peuvent être envisagés. Certains formats de métadonnées sont particulièrement adaptés à cet enjeu, comme le format [METS](#) (*Metadata Encoding and Transmission Standard*), plébiscité par la BnF et utilisé par les bibliothèques numériques [Numelyo](#) ou [Aurelia](#).

## 4. Les étapes de la numérisation d'un fonds de presse ancienne

### 1) Choix des fonds à numériser :

Public visé ? Redondance avec d'autres fonds ? Partenariats possibles ? Projet thématique ?

### 2) Choix du mode de numérisation :

Sur site ou chez le prestataire ? Numérisation directe ou indirecte (à partir des microfilms) ?

### 3) Passation du marché public de numérisation :

Allotissement ? Parmi les critères du marché : matériel utilisé, traitement des documents, traitement des métadonnées, capacité à gérer l'OCR, METS et ALTO, protocole et échantillonnage de contrôle ?

### 4) Mise en ligne du fonds :

Achat d'un logiciel ou solution hébergée ?

### 5) Valorisation des fonds numérisés :

Référencement dans d'autres bases de données, [expositions virtuelles](#), etc. [Envisager le signalement, la diffusion et la communication de ces fonds](#).



Liens	Ressources
	<ul style="list-style-type: none"><li>• <a href="#">Numériser en bibliothèque</a> : fiches pratiques de l'enssib.</li><li>• <a href="#">Enquête sur la numérisation de la presse ancienne locale et régionale : rapport de synthèse</a>, Aline Girard, BnF, avril 2014.</li><li>• <a href="#">Manuel de constitution de bibliothèques numériques</a>, Thierry Claerr, Isabelle Westeel, Cercle de la Librairie, 2013. (Critique de l'ADBS)</li><li>• <a href="#">Numériser les publications en série</a>, Journée d'étude, Centre régional des lettres Midi-Pyrénées, 6 décembre 2012.</li><li>• <a href="#">Le droit, un frein ? Mise en ligne et valorisation de la presse</a>, Michèle Battisti, Paralipomènes, 7 mai 2010.</li><li>• <a href="#">Mise en ligne et la valorisation de la presse XIX-XXI<sup>e</sup> : 1/4, 2/4, 3/4 et 4/4</a>, conférences filmées de la Bibliothèque municipale de Lyon, 6 et 7 mai 2010.</li><li>• <a href="#">Écrire un cahier des charges de numérisation et de conversion en mode texte de collections de presse</a>, Marie-Élise Fréon et Catherine Mocellin, BnF, BMC d'Orléans, MCC, mars 2010.</li><li>• <a href="#">Hier à la une : les collections de presse dans les bibliothèques françaises</a>, Journées du patrimoine écrit, SLL, 13-14 septembre 2007.</li><li>• <a href="#">La Presse ancienne</a>, dossier de Patrimoine numérique.</li></ul> <p><b>Exemples</b></p> <ul style="list-style-type: none"><li>• <a href="#">Association pour la recherche et l'archivage de la mémoire arménienne ( ARAM )</a> : bibliothèque numérique contenant entre autres des séries de journaux anciens en arménien.</li><li>• <a href="#">Aurelia</a> : Bibliothèque numérique d'Orléans</li><li>• <a href="#">Centre d'histoire du travail de Nantes</a> : numérisation et mise en ligne de sept périodiques anciens locaux relevant du secteur social</li><li>• <a href="#">Gallica</a> : bibliothèque numérique de la BnF</li></ul> <p><a href="#">Presse numérisée</a>, BnF, 28 octobre 2013. <a href="#">Politique de numérisation à la BNF</a>, Guillaume Doizy, EIRIS, février 2010. <a href="#">La Bibliographie de la presse française politique et d'information générale</a>, BnF, 7 mars 2014.</p> <ul style="list-style-type: none"><li>• <a href="#">Léon Vivien</a> : le quotidien d'un poilu sur Facebook</li></ul> <p><a href="#">Au cœur du quotidien d'un poilu</a>, Musée de la Grande Guerre du Pays de Meaux, 2014. <a href="#">Un poilu sur la Toile : Facebook 1914 : une expérience digitale du musée de la Grande Guerre du Pays de Meaux</a>, Michel Rouger, BBF, 2013, n° 5. <a href="#">Léon Vivien, un poilu sur Facebook !</a>, Bruno Texier, Archimag, 23 mai 2014.</p> <ul style="list-style-type: none"><li>• <a href="#">Michael</a> : inventaire multilingue du patrimoine numérique</li><li>• <a href="#">Mémoire et actualité en Rhône-Alpes</a> : Portail collaboratif de ressources régionales</li><li>• <a href="#">Numelyo</a> : bibliothèque numérique de Lyon</li><li>• <a href="#">Numelyo, la bibliothèque numérique de Lyon : une multitude de fonds offerts au public</a>, Nicolas Gros et Pierre Guinard, BBF, 2013, n° 5.</li><li>• <a href="#">Overnia</a> : bibliothèque du patrimoine de Clermont-Ferrand</li><li>• <a href="#">Patrimoine numérique</a> : catalogue des collections numérisées</li><li>• <a href="#">Rosalis</a> : bibliothèque numérique de Toulouse</li><li>• <a href="#">Rosalis, à la croisée des chemins : une décennie d'explorations</a>, Patrick Hernebring, BBF, 2013, n° 5.</li><li>• <a href="#">Les Tablettes rennaises</a> : Patrimoine numérisé de la bibliothèque de Rennes Métropole</li><li>• <a href="#">TROVE</a> : portail des revues numérisées de la bibliothèque nationale d'Australie.</li></ul>

Fiche créée par Antoine Barthélémy, Mathilde Cailliet, Solen Costaouëc, Rémi Gaillard, le 13 janvier 2014  
Modifiée par Amandine Jacquet, le 10 septembre 2014