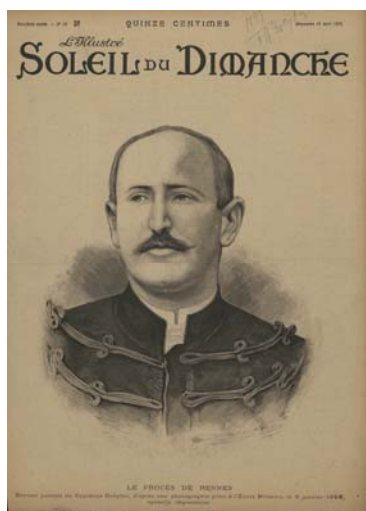


**Dates & lieu de la mission :**

26-27 juin 2014, Poitiers (MSH)

**Objectifs de la mission :**

La raison première de la mission était de numériser la collection de journaux de Henri Mitterrand (numérisation prévue dans le projet ArchiZ).



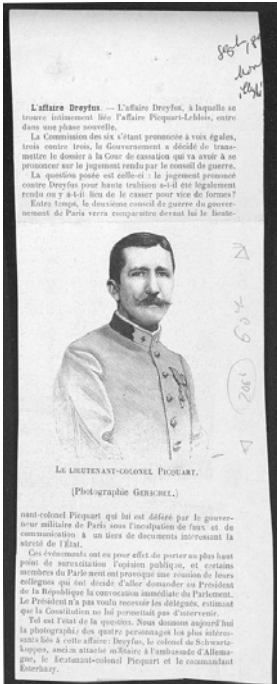
Ces journaux français & anglais couvrent l'affaire Dreyfus (1884-1906). Ils étaient classés par titre de journaux, avec des numéros complets, des pages isolées, des coupures de page datées ou non (149 exemplaires de journaux, complets ou en extraits). Le format des documents à numériser allait d'une coupure de presse de quelques centimètres à des exemplaires en A1.

Le second objectif de la mission était de tester grandeur nature la station de numérisation gérée par la MSH de Poitiers et à laquelle nous avons accès. Cette expérimentation devait permettre d'établir des liens étroits avec la plate-forme technologique de cette MSH et d'envisager des collaborations à plus grand échelle.

**Déroulement**

Jeudi matin, nous sommes accueillis par le responsable de la plate-forme, David Chesnet. Il nous explique le fonctionnement des deux scanners que nous allons utiliser. La prise en main a duré une heure, avec test sur des exemples de notre corpus. Nous avons passé aussi un moment à déterminer le plan de nommage et les étapes de la chaîne de numérisation :

- Choix des pages à numériser : faut-il numériser tout un journal même si tout n'y est pas en rapport avec l'affaire Dreyfus, ainsi que les versos des pages sélectionnées par Henri Mitterrand et qui, elles aussi, n'ont pas de rapport avec l'Affaire ? Certains numéros sont déjà numérisés sur Gallica ; ici notre plan de numérisation a pour vocation de rendre compte de l'état d'une collection, donc des choix opérés par un universitaire, autour de l'affaire Dreyfus : nous ne numériserons que les pages sélectionnées par cet universitaire. Nous avons choisi aussi de numériser dans l'ordre les documents trouvés dans les chemises.
- Partage des journaux entre les deux scanners (l'un plus rapide mais limité au double A3, l'autre plus lent mais A1, voire A0). Ces journaux ont des formats plus grands que ce que nous connaissons aujourd'hui dans la presse quotidienne et nous avons numérisé en une seule page les doubles pages, ce qui induit des grands formats.
- Définition préalable des paramètres des machines pour optimiser le temps de traitement et avoir des images homogènes en terme de résolution, contraste, qualité du lettrage, format de sauvegarde :



- Format : images haute qualité en JPEG sans compression, pour différents usages (dont la publication web).
- Couleurs : après plusieurs essais, nous avons aussi choisi de numériser en couleurs, malgré un temps de numérisation plus élevée : les journaux sont jaunis (nous les numérisons dans l'état), certains ont des illustrations en couleurs (pour une édition ultérieure, il y a intérêt à avoir les mêmes paramètres pour l'ensemble des pages numérisées). Il pourra toujours être possible d'avoir une version noir et blanc par la suite.
- Cadrage : les journaux et les coupures de presse n'ont pas des bords parfaitement rectilignes. Nous avons choisi de numériser avec des marges en fond noir, pour garder l'aspect du document.
- Les images sont traitées brutes, sans retouches.

- Établissement des métadonnées (titre, date, numéro...) pour chaque image numérisée, pour éviter de les saisir à chaque numérisation.

Jeudi après-midi et vendredi matin, nous avons utilisés les 2 scanners à plein et nous avons réussi à numériser l'ensemble des documents amenés.



Vendredi après-midi nous avons fait le bilan et la liste des choses à faire ou à éviter pour une prochaine expérience.

## Résultats

Ces documents ont une particularité par rapport à une collection et un projet de recherche : ils forment un ensemble hétérogène dont le traitement est forcément très long. Nous avons tout numérisé : 502 fichiers, 8,6 gigas.

Comme première expérience, elle fut réussie en terme quantitatif. Nous avons tiré de nombreux enseignements pour accélérer par la suite une nouvelle numérisation de masse.

L'étape technique de numérisation est très courte et très simple avec les machines utilisées (interface intuitive et agréable). Par contre, nous avons constaté qu'il faut effectuer une analyse fine des collections en amont, afin de définir les objets à numériser et leur ordre. Nous avons perdu du temps devant les machines à trier les doublons, classer les documents, ce qui a nuit à la fluidité du travail sur la machine et a demandé de nombreuses vérifications supplémentaires. Notre chaîne de traitement était souvent arrêtée par des décisions à prendre sur des documents, il faut alors impérativement vérifier le nombre de pages du journal par rapport au nombre de fichiers produits.

Il est important de prévoir les traitements qu'on fera de documents particuliers : que fait-on quand une image est agrafée sur la page à numériser ? Faut-il numériser ensemble toutes les petites coupures qui sont regroupées avec un trombone ? Ces décisions ne peuvent être prises que par des experts du projet, ce qui écarte la solution d'une sous-traitance à des étudiants ou stagiaires locaux non accompagnés. On ne peut envisager l'emploi de personnes sur place que si le corpus a été correctement préparé.

Un autre temps de traitement est à prévoir sur les fichiers résultats : vérifier le nommage et le classement des fichiers (remettre les fichiers des double-pages numérisées sur la grande machine avec les fichiers numérisés sur la petite), repositionner ou recadrer des images (mise en place de scripts avec Photoshop), vérifier et compléter les métadonnées des images produites.



Le temps de numérisation du document sur la petite machine était de 2 secondes, sur la grande d'une dizaine de secondes. De nombreuses opérations sont automatisées (par exemple, la vitre du scanner qui se lève après la numérisation). La hauteur de la vitre qui protège le document peut être réglée en fonction de l'état du document (pliure, bout corné, etc.) mais c'est un réglage manuel qui demande du temps et de la précision.

Les documents sont numérisés par le haut, ce qui a été parfois source d'erreurs et de mauvaises manipulations de notre part, car nous avons l'habitude avec les scanners bureautiques de numériser par le bas. Numériser par le haut permet de moins manipuler les documents parfois très fragiles.

Notre plan de nommage était de prendre les initiales des journaux, le numéro puis l'ordre de numérisation. Nous avons dû souvent indiquer « sn » pour les documents sans numéro, voire « sd » pour les documents sans date. A chaque changement de numéro de journal, nous avons dû changer les paramètres de nommage, ce qui prend du temps et peut être source d'erreurs. Les images sont automatiquement mises sur un répertoire d'un serveur de la MSH. Pour optimiser les temps de traitement, nous avons fait une structure simple de ce répertoire, en créant *a posteriori* un sous-répertoire par journal et y plaçant à la fin les

documents numérisés. Un jeu des fichiers est resté sur place, un autre a été copié sur un disque dur externe et un autre sur l'ordinateur de JSM : vu la volumétrie, ces temps de copie sont à prévoir, surtout si un train est à prendre après.

Faute d'exemple, nous n'avons pas testé la numérisation d'ouvrage avec la tranche qui peut être fixée par le scanner, pour éviter d'aplatir le livre ou d'avoir un effet bombé.

L'accueil fut excellent et le personnel compétent et disponible, les machines en très bon état de marche et performantes. Nous avons rencontré en début et fin de mission la directrice adjointe de la MSH, Lydie Bodiou, avec qui nous avons dressé un bilan de ces deux jours de numérisation, en vue d'une possible coopération ultérieure.

Jean-Sébastien Macke & Richard Walter

1<sup>er</sup> juillet 2014

