



HAL
open science

Archives citoyennes du futur : l'archivage du numérique à la BDIC, chantier en cours

Julie Demange

► To cite this version:

Julie Demange. Archives citoyennes du futur : l'archivage du numérique à la BDIC, chantier en cours. Matériaux pour l'histoire de notre temps, 2017, Cent ans. De la BDIC à La contemporaine, 125-126, pp.52-55. 10.3917/mate.125.0052 . hal-03695271

HAL Id: hal-03695271

<https://hal.science/hal-03695271v1>

Submitted on 30 Oct 2022

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

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

Archives citoyennes du futur : l'archivage du numérique à la BDIC, chantier en cours

JULIE DEMANGE, BDIC

Depuis maintenant trois ans, la BDIC mène une réflexion sur l'archivage de ses collections numérisées ou nativement numériques. La question de l'archivage pérenne des collections numérisées à partir d'un support papier ou argentique se pose depuis longtemps dans les bibliothèques patrimoniales. Elle est aussi d'actualité dans les archives publiques ou dans les musées. Dans le cas de la BDIC, il s'agit de sécuriser la conservation des données de la bibliothèque numérique, l'Argonnaute. La même chose doit être faite pour les archives numériques natives au sens large qui sont progressivement collectées ou déposées dans l'établissement. La prise en considération des mutations liées à l'irruption et au développement des pratiques numériques n'est pas nouvelle pour l'institution. En 2005, un numéro de *Matériaux* interrogeait déjà l'impact du Web sur les pratiques militantes et les nouvelles sources qui en étaient issues¹. L'enjeu spécifique de l'archivage est en effet un défi autant pour les chercheurs que pour les professionnels des bibliothèques et des archives. Le numérique natif concerne plusieurs types de documents : la suite des dépôts papier des archives privées, comme par exemple les ressources désormais numériques de la Ligue des droits de l'homme ; les dépôts de producteurs audiovisuels (rushes de films documentaires) ; les productions audiovisuelles de la BDIC. Ces dernières sont souvent liées à des programmes de recherche, de plus en plus conséquents : d'une part, dans le cadre du Labex *Les Passés dans le présent*, l'archivage des matériaux sources accompagnant les produits de la recherche et surtout d'autre part, les co-productions d'archives orales, c'est-à-dire des entretiens filmés de témoins de l'histoire contemporaine, conduits avec des chercheurs. D'autres programmes constitués de ressources numériques natives sont également en cours de

réalisation et ont vocation à être archivés de manière pérenne. Cette dimension fait partie des axes que le nouveau réseau CollEx-Persée doit prendre en considération dans une perspective d'échange d'expériences et d'expertises. La BDIC, qui a été retenue comme bibliothèque délégataire dans la thématique « Mondes contemporains », entend mettre en avant cet enjeu, qui a pu être parfois négligé, des matériaux documentaires de la recherche ne se rattachant pas directement à des publications scientifiques. C'est un point d'étape sur la situation actuelle du traitement de ces archives natives numériques à la BDIC, qui est ici proposé.

Il est de plus en plus fréquent que les producteurs avec lesquels la BDIC est en contact pour la conservation d'archives « papier » traditionnelles, souhaitent compléter leurs versements par le don d'archives numériques. Le cas de figure n'est pas nouveau. Il n'est pas rare que l'archiviste qui pensait avoir à traiter uniquement un fonds papier découvre, dissimulés sous une liasse de documents ou dans une enveloppe, des disquettes, un cédérom, une clé USB, ou même, parfois, une unité centrale d'ordinateur... Jusqu'à peu, la BDIC n'avait guère de solutions pour récupérer les fichiers contenus sur ces différents supports, faute notamment de matériel approprié pour les lire. Des archives numériques ont ainsi été perdues, et lorsqu'il était possible d'accéder aux fichiers conservés sur ces différents supports, ceux-ci étaient simplement déposés sur un serveur de stockage, en attente d'un éventuel traitement.

Les questions liées aux formats informatiques des fichiers, qui évoluent rapidement et deviennent vite obsolètes, rendant les fichiers illisibles, se sont vite posées. Faute d'outils logiciels adéquats, les possibilités de traitement et de conversion de fichiers étaient limitées, et fastidieuses, car menées au cas par cas, c'est-à-dire fichier par fichier. Soucieuse de dépasser ces difficultés, la BDIC a, depuis 2015, décidé d'inscrire la réflexion sur l'archivage du numérique natif parmi ses priorités et lancé un projet appelé *Archives citoyennes du futur* (ACF). Ce terme fait référence aux caractéristiques d'une large



partie des archives constituées par la BDIC, qui émanent de producteurs issus souvent des milieux associatifs. Le travail a commencé par une étude juridique². Les conclusions du rapport ont permis de poser le cadre du projet : la logique juridique de traitement des archives privées est étendue aux archives numériques natives, le dépôt volontaire (par des individus ou des collectivités) de documents numériques natifs favorisé, les modalités de conservation, de mise en consultation et de réutilisation des données étant fixées au cas par cas, par convention. S'il est encore prématuré de parler de la mise en place d'un véritable programme d'archivage électronique, un travail de prospection et des premiers tests ont permis d'établir une typologie des documents numériques que la BDIC souhaite (et aura la capacité de) prendre en charge, et une chaîne de traitement propre à ces différents types de matériaux a été définie. En parallèle, une veille technique a été effectuée auprès d'établissements confrontés aux mêmes enjeux de chaîne de traitement et de mise au point de plan de gestion des données, notamment aux Archives nationales (programme VITAM) et au SIAF. Une analyse comparative des différentes solutions d'archivage « pérenne » a également été faite et un choix de stockage sécurisé historisé en partenariat avec le CNRS et la TGIR Huma-Num a été arrêté. La convention vient d'être signée : le déploiement de l'archivage entre en phase opérationnelle début 2018. Pour l'instant, quatre types de matériaux numériques entrent dans le périmètre du projet *Archives citoyennes du*

futur : les divers types documents numériques produits dans le cadre d'une activité privée « engagée », individuelle ou collective (fichiers de traitement de texte ou fichiers type Excel, courriers électroniques, newsletters, mailing-lists, etc.) ; les archives numériques (films montés et rushes) de producteurs audiovisuels ; les enregistrements numériques liés à des projets de constitution de fonds d'archives orales ; des sites Web ou les documents numériques publiés sur Internet, utilisés dans le cadre de programmes de recherche ou faisant partie des archives des donateurs privés. Précisons que pour les archives du Web (documents publiés en ligne), il ne s'agit pas de s'inscrire dans les logiques de collectes massives menées par la BnF et l'INA, qui se répartissent en France la responsabilité du dépôt légal de l'Internet, il est seulement envisagé de conserver les contenus publiés sur le Web par les donateurs avec lesquels elle aura signé une convention. Le choix a été fait de se concentrer, dans un premier temps, sur les trois premiers types de matériaux. Le traitement des archives audiovisuelles numériques natives est le chantier le plus avancé. La BDIC s'est lancée depuis longtemps dans la collecte de ce type de matériaux et dispose déjà d'un fonds conséquent d'archives orales. Depuis 2013, elle réalise aussi des entretiens filmés, produits avec un équipement vidéo numérique. Le service audiovisuel, rattaché depuis 2016 au département des archives, dispose de compétences solides pour le traitement de ces vidéos numériques (montage, fusion-



••• conversion, etc.), dont l'indexation est désormais effectuée en format EAD³ de manière à harmoniser les pratiques au sein du département. L'ensemble des fonds et collections d'archives (archives papier ou audiovisuelles) est donc décrit sous cette même forme normalisée, qui permet de produire des inventaires interoperables et compatibles avec le catalogue collectif national Calames⁴, outil de publication en ligne des inventaires de fonds d'archives. Des choix de formats (permettant d'assurer la qualité des documents sans pour autant être trop volumineux et « lourds ») ont été faits, aussi, pour assurer une bonne conservation des archives audiovisuelles.

Une architecture reposant sur plusieurs serveurs a été pensée afin que la BDIC puisse disposer d'espaces distincts pour le traitement et le stockage provisoire des fichiers collectés, la diffusion et la communication des fonds d'archives numériques et, enfin, leur conservation, sécurisée. Sur ce dernier point, crucial, le partenariat noué avec la TGIR Huma-Num permet à la BDIC, travaillant de concert avec les instances informatiques de l'Université Paris-Nanterre, d'intégrer le dispositif national de stockage distribué en réseau *Huma-Num Box*. Si les questions de formats et de volume se posent aussi pour les archives numériques produites par les associations et organisations déposant leurs archives à la BDIC, le premier point à résoudre pour entreprendre leur conservation reste cependant d'ordre organisationnel. Les premiers tests de collecte menés notamment en partenariat avec la Ligue des Droits de l'Homme⁵ ont montré que la mise en place d'une étape de pré-traitement des archives (consistant à préparer les lots de documents à archiver, avant même qu'ils n'intègrent la chaîne de traitement déployée par la BDIC) est indispensable, et doit absolument reposer sur une collaboration étroite avec les producteurs d'archives. Un protocole de « pré-versement » a donc été défini, distinguant plusieurs étapes. Il débute par une première rencontre avec les donateurs, lors de laquelle sont exposés et discutés les différents traitements auxquels seront soumis les archives. À l'issue de ces premiers échanges, les producteurs se voient proposer un ensemble de préconisations concernant les formats dans lesquels sauvegarder les fichiers, la manière de les nommer et de les organiser dans une arborescence de dossiers et de répertoires afin de préparer le lot d'archives (ensemble de fichiers) qu'ils souhaitent transmettre à la BDIC. Une fois ce lot d'archives constitué, la BDIC récupère les fichiers via un support de stockage externe, et engage une analyse de la structure du lot et des formats des fichiers, dont les résultats sont ensuite discutés avec les donateurs. La BDIC propose alors, dans la mesure du techniquement possible, des solutions de traitement pour chacun des différents types de fichiers identifiés et elle précise aussi, au besoin, les formats qu'elle n'a pas les moyens de convertir et de conserver. La liste des formats de conservation a été établie à partir

de la liste des formats de fichiers établis par le Centre Informatique National de l'Enseignement Supérieur (CINES) pour l'archivage des données de la recherche et en fonction des recommandations suivies par les principales institutions patrimoniales françaises, en conformité avec les choix de la direction des systèmes d'informations de l'Université Paris Nanterre. La liste des formats acceptés à l'entrée est pour sa part évolutive. Elle s'appuie sur les compétences techniques dont dispose la BDIC pour opérer de manière automatisée les conversions de fichiers par lots.

La préparation du « paquet de fichiers » à verser est alors finalisée et ceux-ci définitivement intégrés à la chaîne de traitement des archives numériques de la BDIC. Une copie de sauvegarde du « paquet » dans son état initial est effectuée sur un disque dur. On engage ensuite le traitement des fichiers, le but étant d'assurer ainsi aux documents les conditions d'une conservation sécurisée. À l'issue de ces opérations, des métadonnées de pérennisation sont associées aux paquets d'archives traités. Ils sont ensuite décrits de manière qualitative par l'archiviste, qui produit un inventaire servant aussi à l'indexation des données. Le paquet d'archives et les métadonnées qui lui sont associées (métadonnées de pérennisation, de gestion et de description) sont enfin déposés dans l'espace de conservation sécurisé, tandis qu'une copie est faite vers le dispositif de mise en ligne des données, qui prend en charge les opérations de conversion des fichiers, de formats de conservation vers des formats de diffusion adaptés aux spécificités du portail de communication des archives et documents, grâce auquel les lecteurs peuvent consulter les collections numériques. Pour assurer la sécurité et la protection des données privées, la consultation du portail est uniquement proposée sur les postes informatiques de la salle de lecture de la BDIC. Dans certains cas, en accord avec le producteur et ou le donateur, des documents peuvent aussi être mis en ligne sur la bibliothèque numérique de la BDIC, L'Argonnaute. Cette chaîne de traitement est en phase d'installation. La question de l'archivage des courriels et des listes de diffusion Internet est encore en cours d'instruction, tout comme celle de l'archivage des sites Web, une étude de prospection ayant été menée en 2016⁶. Elle concluait sur la nécessité d'un partenariat avec une institution ou un organisme disposant des moyens techniques suffisant pour réaliser la capture de sites et d'objets mis en ligne. La BDIC a eu l'occasion de tester le service « *Archive-It* » développé par le consortium international *Internet Archive*. Les résultats se sont avérés satisfaisants mais le déploiement de cette solution nécessite d'entreprendre au préalable un repérage des différents sites et documents numériques qui pourraient entrer dans le périmètre d'une collecte de la BDIC.

Pour conclure cet aperçu, il faut aussi évoquer les archives des chercheurs qui sont aujourd'hui des producteurs

d'archives numériques et des utilisateurs importants de données numériques, à des fins d'enquêtes ou de constitution de corpus d'étude. La BDIC s'est déjà engagée à conserver et rendre accessibles divers fonds d'archives numériques produits ou collectés dans le cadre de programmes de recherche collectifs, qui se caractérisent par un partenariat entre chercheurs et acteurs de la société civile. C'est par exemple le cas des archives du programme de recherche *Faire et arpenter l'histoire de Nanterre* (Labex *Les Passés dans le présent*). Il a été mis en œuvre par un collectif d'enseignants-chercheurs, de chercheurs, de personnels administratifs, d'étudiants et de partenaires associatifs et municipaux qui ont souhaité « faire à plusieurs voix » une histoire de l'Université de Nanterre (histoire du campus et de l'institution, histoire intellectuelle) et de sa relation au territoire. Il propose entre autres une forme originale de médiation à travers un parcours urbain et des dispositifs numériques mobiles *ad hoc*, pensés en amont de la recherche⁷. Un partenariat a été établi avec la BDIC pour la conservation des archives numériques produites au cours de cette recherche (fichiers bureautiques : notes, carnets de terrains, etc. ; images et documents audios et vidéos : entretiens, captations audiovisuelles ou sonores, etc.). Il s'inscrit dans la politique plus générale de l'Université Paris Nanterre de conservation et de valorisation de son propre patrimoine scientifique qui touche aussi bien à de la documentation qu'à des archives. Les opérations de pré-versements, menées en collaboration serrée avec les producteurs, sont ainsi effectuées sur un espace de travail partagé, un autre service proposée par Huma-Num⁸ : c'est tout l'intérêt de la collaboration avec la TGIR. Pour la suite de la prise en charge, les protocoles de traitement et de diffusion des archives collectées sont les mêmes que ceux mis en place pour les autres archives numériques natives.

A n'en pas douter, cet état des lieux méritera d'être révisé et actualisé d'ici quelques mois. *Work in progress* : chantier en cours ! Les premiers versements permettent de développer les outils et compétences nécessaires en toute transparence puisque les donateurs entreprennent une démarche volontaire et se montrent très collaboratifs et... patients. Il faut espérer que la BDIC dispose à l'avenir des moyens indispensables au bon développement de ce nouvel axe de sa politique documentaire, et puisse assurer de manière pérenne la collecte, le traitement et la conservation de ces documents d'archives, travaillant, ainsi, sur ce nouveau terrain, à la constitution du patrimoine de demain. L'ampleur de la tâche est telle que c'est dans une perspective de responsabilités partagées qu'il convient assurément de l'envisager. ■

Julie Demange est archiviste à la BDIC, elle est chargée des archives nativement numériques.

Notes

1. "Internet et mouvements sociaux : nouvelles pratiques militantes, nouvelles sources pour l'histoire" *Matériaux pour l'histoire de notre temps*, n°79, 2005. Voir aussi le programme mené dans le cadre du Labex *Les Passés dans le présent*, avec la BnF et Paris Tech, *Le devenir du patrimoine numérisé en ligne : l'exemple de la Grande Guerre*.
2. Anne-Laure Stérin, *Étude de faisabilité juridique du projet Archives citoyennes du futur à la BDIC*, Oct. 2015
3. L'EAD est une « grammaire » standard - c'est-à-dire qui repose sur des principes partagés et employés par un grand nombre de professionnels de la documentation - qui permet d'encoder en XML, un langage informatique, des instruments de recherches archivistiques (inventaires).
4. Calames - Catalogue en Ligne des Archives et des Manuscrits de l'Enseignement Supérieur - a été développé par les équipes de l'ABES. Il s'agit d'un catalogue collectif national dédié aux fonds de manuscrits et d'archives conservés dans les universités françaises et, plus largement, dans les établissements français d'enseignement supérieur et de recherche. Il est accessible à l'adresse : <http://www.calames.abes.fr> et le portail dédié aux collections de la BDIC à l'adresse : <http://www.calames.abes.fr/pub/bdic.aspx#culture=fr>
5. La BDIC conserve depuis le début des années 2000 les archives de la Ligue antérieures à 1940 (archives saisies par les Allemands en 1940, puis par l'Armée rouge, et finalement « revenues de Moscou »). Les archives postérieures à 1945, versées ces dernières années, sont en cours d'inventaire et seront ouvertes à la consultation en 2018. L'archivage des productions numériques de la LDH s'inscrit dans la continuité de ce chantier (voir article de T. Morel dans ce numéro).
6. Anne-Laure Stérin, 2015, op.cit.
7. <http://passes-present.eu/fr/faire-et-arpenter-lhistoire-de-luniversite-de-nanterre-2585>
8. <https://www.huma-num.fr/services-et-outils/stocker>