

L'archivage du Web ou le Web comme mémoire des sociétés contemporaines

Par **Alexandre CHAUTEMPS**

Chef du service Dépôt légal numérique,
Bibliothèque nationale de France

Bref historique de l'archivage du Web

Le Web a été inventé en 1989 par Tim Berners-Lee puis a été rendu librement utilisable en 1993 par son employeur de l'époque, le CERN. A peine plus de dix ans plus tard, le Web est devenu un outil de communication très largement employé. En 2004, 50 % des Français disposent d'un accès à Internet ⁽¹⁾. Les individus, la société et les institutions ont déjà produit un contenu nativement numérique, et sans équivalent imprimé, d'un volume non négligeable. Pourtant, l'idée de collecter le Web pour créer des archives pérennes paraît encore étrange.

Seuls quelques visionnaires s'y consacrent. Dès 1997, Brewster Kahle, créateur de la fondation Internet Archive, avait attiré l'attention dans un article paru dans une grande revue de vulgarisation scientifique (Kahle, 1997) qui posait déjà les grands principes de ce qui allait devenir l'archivage du Web. Les premières initiatives d'archivage du Web avaient vu le jour en 1996, portées par Internet Archive et par quelques institutions : Bibliothèque royale de Suède (projet KulturarW3) et Bibliothèque nationale d'Australie (projet PANDORA). En France, la BnF avait, dès 1999, mené des collectes expérimentales du Web (cf. Masanès, 2002), la première collecte d'envergure étant réalisée à l'occasion de la campagne électorale de 2002. A la suite du travail d'information réalisé par la Bibliothèque nationale de France et le ministère de la Culture, le parlement a décidé d'inscrire dans la loi, en août 2006, une mission de dépôt légal (cf. Stirling *et al.*, 2011) portant sur « les signes, signaux, écrits, images, sons ou messages de toute nature faisant l'objet d'une communication au public par voie électronique » (Code du Patrimoine, art. L131-2). Cette mission est confiée conjointement à trois institutions : l'Institut national de l'audiovisuel (INA) pour les programmes de télévision et de radio diffusés *via* l'internet, le Centre national de la cinématographie (CNC) pour les films diffusés en salle et la BnF pour tout le reste du Web français.

Le positionnement des activités d'archivage du Web dans le cadre du dépôt légal est un élément particulièrement important en ceci qu'il donne une assise juridique à ces activités, mais aussi par le processus de patrimonialisation qu'il implique (cf. Bermès, 2020, p. 68 *et sq.*). Les collections nativement numériques ainsi constituées s'inscrivent en effet dans la continuité des collections de documents imprimés entrés par dépôt légal, dont les plus anciens remontent au XVI^e siècle.

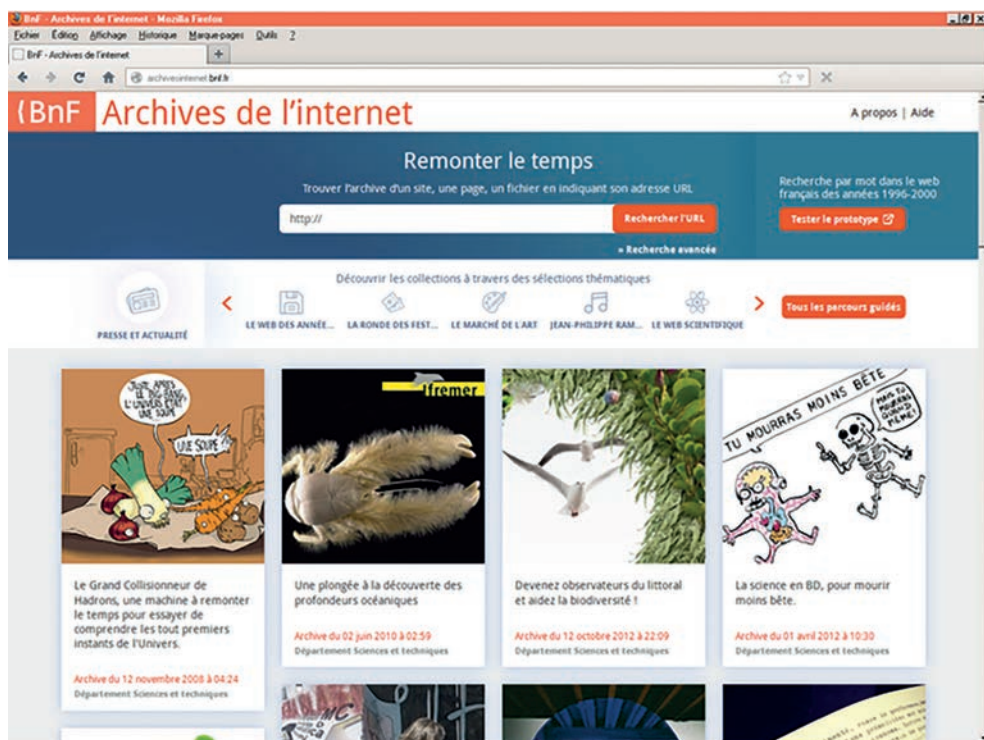
Chaque année, la BnF effectue une collecte large qui moissonne l'ensemble des domaines identifiés comme faisant partie du Web français. Il s'agit de tous les domaines en .fr ainsi que des autres domaines dont le producteur est localisé en France, quelle que soit leur extension (.com, .org, .net mais aussi les domaines de haut niveau liés aux territoires d'outre-mer et les « nouvelles »

(1) Cf. BIGOT R. (2004), *La Diffusion des technologies de l'information dans la société française*, Paris, Credoc, p. 83 et sq., <https://bit.ly/2HAKt5X> (consulté le 11 mars 2020).

extensions telles que .paris, .bzh ou .immo, .tools, etc.⁽²⁾). Cette collecte large est enrichie par des collectes ciblées, plus fréquentes et/ou plus profondes, donc plus complètes, qui portent sur des sites sélectionnés par les bibliothécaires de la BnF et d'établissements partenaires en régions, en fonction de critères documentaires. Certaines collectes ciblées sont réalisées à l'occasion d'événements : campagnes électorales, événements sportifs, mais aussi attentats, épidémies, célébrations, etc., de manière à garder une trace numérique desdits événements.

Depuis 2010, la BnF réalise par ses propres moyens l'ensemble de ses collectes (pour la description de l'architecture technique voir Le Follic *et al.*, 2012) et garantit la préservation des collections patrimoniales numériques ainsi constituées (cf. Derrot *et al.*, 2012).

A la fin de l'année 2019, la BnF avait constitué une collection d'archives Web d'un volume d'environ 1 200 téraoctets, regroupant près de 35 milliards de fichiers, dont les éléments les plus anciens remontent à 1996 et dont l'accroissement se poursuit quotidiennement.



Page d'accueil de l'application de consultation des archives
© Bibliothèque nationale de France

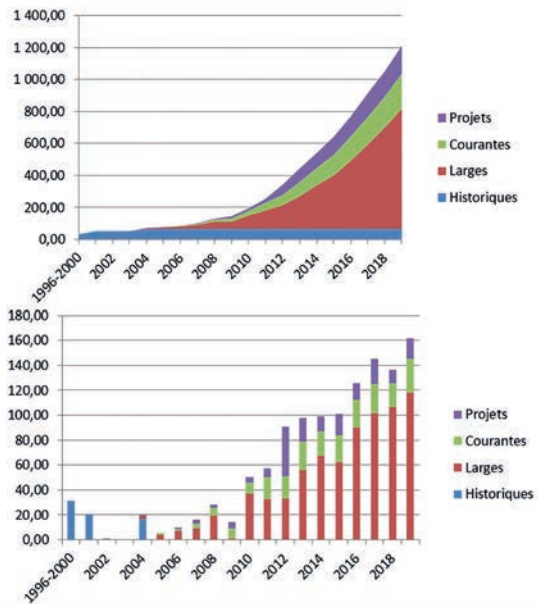
L'accès aux Archives de l'Internet de la BnF est possible dans les locaux de l'établissement, en bibliothèque de recherche⁽³⁾ ainsi que dans vingt bibliothèques partenaires en régions⁽⁴⁾.

(2) Pour clarifier les notions de nom de domaine et d'extension, liée à un domaine de haut niveau, voir Bortzmeyer, 2018, chapitre « Noms de domaine et DNS », pp. 89 *et sq.*

(3) Voir description des Archives de l'Internet : <https://www.bnf.fr/fr/archives-de-linternet> et les conditions d'accès à la bibliothèque de recherche : <https://www.bnf.fr/fr/bibliotheque-de-recherche> (liens consultés le 11 mars 2020).

(4) Voir la localisation des différents points d'accès, en Métropole et en outre-mer : <https://bit.ly/2PjOpfF> (consulté le 11 mars 2020).

- 1,2 pet-octet de données
- De 1996 à nos jours
- Collecte large annuelle
- Collectes ciblées (sélection par bibliothécaires BnF et partenaires)



Archives de l'Internet BnF : volumétrie des collections début 2020
© Bibliothèque nationale de France

La collecte du Web par robot permet aussi de moissonner de nombreux livres et périodiques numériques, principalement aux formats EPUB et PDF. Néanmoins, les documents numériques distribués *via* des boutiques en ligne ne sont pas collectés par ce moyen. Des conventions sont conclues avec les éditeurs et distributeurs afin de procéder par la voie du dépôt, techniquement plus appropriée. Des dépôts de documents numériques en ligne sont ainsi réalisés à titre expérimental⁽⁵⁾.

De nombreuses autres institutions de par le monde remplissent une mission d'archivage du Web analogue à celle de la BnF. Il s'agit de bibliothèques nationales mais aussi de centres d'archives, voire de fondations indépendantes. La plupart de ces institutions sont regroupées au sein de l'Internet International Preservation Consortium (IIPC)⁽⁶⁾, créé en 2003, dont la BnF est membre fondateur (cf. Illien, 2011). IIPC regroupe à l'heure actuelle 57 institutions appartenant à 35 pays, parmi lesquels l'Europe, l'Amérique du Nord et dans une moindre mesure l'Asie sont les régions du monde les plus représentées. L'archivage du Web reste un chantier à ouvrir en Afrique, en Amérique latine et même en Inde et il s'agira d'un enjeu majeur des décennies à venir.

Si les institutions patrimoniales se sont donné pour mission de préserver le Web pour l'avenir, la collecte du Web est également utilisée pour servir les besoins immédiats de la recherche. Elle est alors pratiquée par un éventail plus large d'institutions, dont la plupart ne font pas partie du consortium IIPC : bibliothèques universitaires et laboratoires de recherche principalement. Dans la partie suivante de cet article, nous allons nous intéresser plus particulièrement aux utilisations possibles des archives Web dans le cadre de projets de recherche.

(5) Dans ce cadre, des conventions de dépôt ont été conclues par la Bibliothèque nationale de France avec l'éditeur de livres numériques *publie.net* et le distributeur de musique en ligne *Idol*.

(6) Voir <http://netpreserve.org> (consulté le 11 mars 2020).

Usage(s) des archives du Web

Depuis le milieu de la dernière décennie, les sources nativement numériques, et singulièrement les archives Web, deviennent des sources légitimes pour la recherche dans de nombreuses disciplines des sciences humaines et sociales, et sont même déjà devenues des sources nécessaires et incontournables pour les travaux se focalisant sur la période récente (après l'an 2000).

Le savoir-faire du chercheur se modifie dans le contexte de l'essor des humanités numériques⁽⁷⁾. L'historien, le sociologue, l'anthropologue ou le linguiste, sans renoncer aux compétences propres à leur discipline, deviennent aussi des praticiens, et parfois des spécialistes des données numérisées et nativement numériques, de leur identification, de leur sélection, de leur manipulation, tout en étant conduits à s'intéresser aux circonstances de leur production et de leur collecte (voir Milligan, 2019).

Dans ce cadre, les archives du Web constituent un terrain d'exploration des groupes sociaux et des comportements.

Un premier exemple est constitué par le projet mené par Sophie Gebeil (Université Aix-Marseille) sur la mémoire de l'immigration maghrébine en France de 1999 à 2014 (cf. Gebeil, 2015). L'objet du travail de Sophie Gebeil n'était pas centré sur les événements eux-mêmes mais sur le processus mémoriel dont ils faisaient l'objet. Son approche s'inscrivait dans le cadre de l'histoire mémorielle, dans la continuité des travaux de Pierre Nora (voir Nora, 2008). Il s'agit de retracer la manière dont les Français issus de l'immigration nord-africaine produisent différents récits mémoriels en lien avec leur propre histoire individuelle ou familiale. Les archives Web de la BnF ont constitué le matériau principal de ces travaux. Le Web permet de mettre en rapport, sur un même plan, différents discours : celui des institutions (notamment le Musée de l'histoire de l'immigration), celui des médias traditionnels, celui des acteurs associatifs et/ou militants, celui des personnes s'exprimant à titre individuel (témoignages). L'essor du Web a en effet facilité l'expression des deux dernières catégories, et surtout, la diffusion de leur point de vue. Par ailleurs, le travail de Sophie Gebeil met en évidence une démarche de patrimonialisation de la mémoire des populations immigrées : Chibanis, mémoire des quartiers, mémoire des bidonvilles constituent autant d'objets mémoriels souvent laissés de côté ou traités de manière stéréotypée par les médias traditionnels. Sophie Gebeil a choisi de travailler sur le Web archivé et non sur le Web vivant car son besoin méthodologique était de disposer d'un corpus stable : les comptages d'occurrences, l'analyse des liens entre sites, la constitution de réseaux internes à la communauté étudiée, rendaient nécessaire le fait de disposer d'un matériau textuel figé dans le temps et protégé de la volatilité intrinsèque au Web vivant⁽⁸⁾.

Autre exemple : le projet *La Grande Guerre sur le Web*, dirigé par Valérie Beaudouin (Télécom Paris-Tech) et soutenu par le Labex *Les passés dans le présent* (cf. Beaudouin *et al.*, 2018 et Bermès, 2020, p. 80⁽⁹⁾). Ce projet s'est déroulé de 2013 à 2018. L'objectif initial consistait à étudier la manière dont les ressources patrimoniales numérisées mises en ligne par les bibliothèques étaient diffusées et réutilisées sur le Web, notamment dans le cadre de la commémoration. La première étape a été de sélectionner les sites en rapport avec le sujet et de les décrire par un jeu de métadonnées propres au projet. Sur cette base, les sites sélectionnés ont été collectés par la BnF à un rythme

(7) Lire à ce sujet le *Manifeste des Digital humanities*, <https://journals.openedition.org/jda/3652> (consulté le 11 mars 2020).

(8) On trouvera davantage d'informations méthodologiques dans GEBEIL (2017). Voir également le parcours guidé *Les mémoires de l'immigration maghrébine (1999-2014)* sur le site de la BnF : <https://www.bnf.fr/fr/centre-d-aide/parcours-guides-dans-les-archives-de-linternet#step03>

(9) Voir le parcours guidé « Mémoire et histoire de la Grande Guerre sur le web » : <https://www.bnf.fr/fr/centre-d-aide/parcours-guides-dans-les-archives-de-linternet#step01> ainsi que le billet de Philippe Chevallier, « La Grande Guerre sur le web : un éclairage inédit », publié en 2017 sur le carnet de recherche Hypothèses de la BnF : <https://bnf.hypotheses.org/1588>

semestriel, constituant ainsi le corpus de travail (volumétrie totale : 6,7 téraoctets). L'équipe projet s'est attachée à cartographier, à l'aide du logiciel Gephi, les relations que les sites Web consacrés à l'histoire de la Grande Guerre tissaient entre eux. Les sites dédiés à la commémoration s'inscrivaient dans un ensemble plus vaste, où les sites institutionnels côtoient les sites d'amateurs, d'associations, qu'ils soient centrés sur la généalogie, l'histoire des régiments, la mémoire familiale. L'approche cartographique et statistique mise en œuvre par l'équipe projet permet de mettre en évidence une polarisation marquée en deux ensembles, sites institutionnels et sites d'amateurs, correspondant à deux réseaux humains bien distincts, mais montre aussi l'existence d'une zone frontière où les deux réseaux communiquent ponctuellement entre eux.

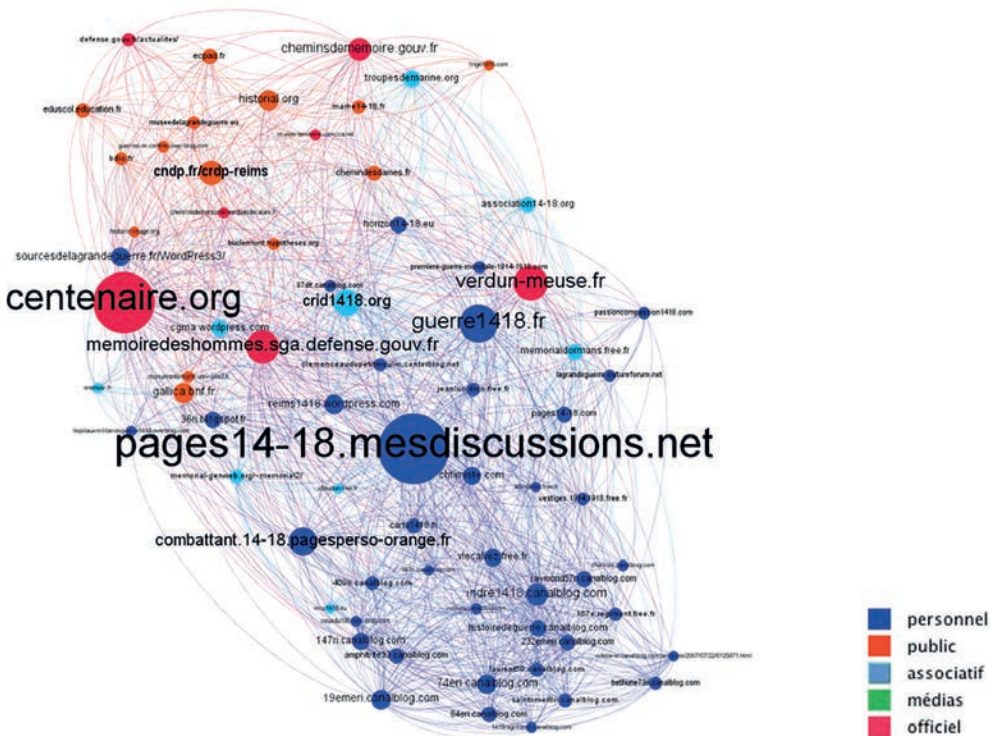


Figure produite par Valérie Beaudouin, Télécom Paris, à partir des données de la « Collecte Grande Guerre » de la BnF (archives du Web).

© Télécom Paris Tech

Le forum 14-18⁽¹⁰⁾, qui était l'un des lieux privilégiés de ces échanges, a d'ailleurs fait l'objet d'une vaste étude par fouille de données dans le cadre du deuxième volet du projet.

Le Web est également un terrain privilégié pour étudier l'évolution de la langue. Le projet Néonaute (cf. Aubry, 2018), mené de 2016 à 2018 sous la direction d'Emmanuel Cartier (Université Paris 13), avait pour objet l'étude du cycle de vie de trois ensembles de mots : 30 000 néologismes, un ensemble de mots français dont l'emploi est recommandé de préférence à des anglicismes, et un ensemble de termes féminisés. Le corpus choisi était celui des collectes Actualités réalisées par la BnF entre 2010 et 2017 (volumétrie totale : 13 téraoctets). Il est constitué des sites de presse français, dans leur partie accessible gratuitement, collectés à une fréquence quotidienne. Le

(10) Voir <https://forum.pages14-18.com/> (ressource consultée le 29 mars 2020).

corpus a été indexé en texte intégral avec le moteur d'indexation Apache Solr puis a fait l'objet de différents traitements destinés à « nettoyer » les pages Web en en retirant tout ce qui n'est pas le texte des articles (boutons et liens de navigation, en-têtes, pieds de page, messages publicitaires, etc.). Ensuite, l'équipe projet a procédé à différents traitements d'analyse morpho-syntaxique, de lemmatisation et d'extraction de fragments de texte comportant chaque occurrence des termes analysés avec un environnement textuel suffisant pour permettre l'analyse de leur rôle dans la phrase. L'évolution de l'emploi des termes est étudiée en tenant compte de leur environnement d'apparition, en distinguant notamment presse d'information générale, presse magazine, presse technique, presse professionnelle, etc.



Projet Néonaute : visualisation multidimensionnelle interactive des occurrences.

© Université Paris 13

Le Web est aussi un gisement très riche d'articles scientifiques, dont une part importante est diffusée en accès ouvert⁽¹¹⁾. De nombreux documents PDF sont collectés dans les différentes archives Web, mais il n'est pas possible de savoir de manière simple lesquels correspondent à des publications scientifiques. Un projet récent, mené par Internet Archive, a consisté à utiliser des technologies d'intelligence artificielle pour identifier des articles scientifiques au sein de la masse des documents PDF archivés. La première étape a été de définir, avec l'aide de bibliothécaires humains, un corpus d'entraînement constitué de sources importantes de publications scientifiques en accès ouvert sur le Web vivant. L'algorithme de *machine learning* a ensuite été entraîné à reconnaître un article scientifique. Après cela, il a été lancé sur les archives afin d'identifier les articles scientifiques parmi les PDF archivés, d'évaluer la qualité de la version archivée, de procéder à un archivage dans le cas où le document était inexistant ou de mauvaise qualité dans les archives, et enfin d'effectuer le signalement des versions archivées dans un catalogue en ligne de publications scientifiques : Fatcat⁽¹²⁾. Ce projet montre de manière convaincante les possibilités offertes par l'intelligence

(11) Pour une définition des différentes formes d'accès ouvert, voir le site web de l'École des Ponts Paris-tech : <https://espacechercheurs.enpc.fr/fr/open-access> (lien consulté le 14 mars 2020).

(12) Voir <https://fatcat.wiki> (lien consulté le 14 mars 2020).

artificielle pour explorer les vastes gisements d'informations constitués par les archives Web, tâche qui, du fait de l'importance de la volumétrie, demeure irréalisable par des humains⁽¹³⁾.

Enfin, les archives du Web constituent une source incontournable pour une discipline émergente : l'histoire du Web. Celle-ci, en effet, est d'abord racontée par le Web lui-même. C'est par les archives du Web que l'on peut reconstituer la succession des technologies (langages de programmation, protocoles, formats...) qu'il a utilisées. C'est également par les archives du Web que l'on peut appréhender les événements fondateurs de son évolution. Les différentes périodes et les moments-clés de l'histoire du Web y ont laissé de nombreuses traces : passage du Web 1.0 au Web 2.0, essor des blogs, des réseaux sociaux, des plateformes de mise en relation, mais aussi mouvements récurrents comme le Web activisme ou le *net art* (voir Brügger, 2017 et Schafer, 2018a et b⁽¹⁴⁾).

En conclusion : quelques perspectives

Nous l'avons vu, les archives du Web offrent un nouveau terrain d'investigation aux chercheurs et mobilisent tout un arsenal de techniques spécifiques, dont l'enjeu est de faire face à une masse de données d'une ampleur sans précédent.

Comment connaître ces techniques et apprendre à les utiliser ? De nombreuses ressources existent, bien entendu, sur le Web⁽¹⁵⁾. Ces connaissances et leur mise en pratique font également l'objet d'une transmission par des humains, dans le cadre d'ateliers dédiés à la prise en main et à l'exploitation des outils de manipulation de données, de fouille de texte et de *data visualization* : les datathons. On suivra avec intérêt les activités du groupe d'historiens canadiens Archives Unleashed, particulièrement investi dans cette transmission, tout à la fois par la création d'outils logiciels⁽¹⁶⁾ et l'organisation régulière de datathons.

Les bibliothèques développent également de nombreux services dédiés aux chercheurs travaillant sur des données numérisées et nativement numériques, parmi lesquelles les archives du Web tiennent une place importante. La Bibliothèque nationale de France travaille actuellement à la mise en œuvre d'une telle offre de service, BnF Data Lab, qui devrait voir le jour à l'automne 2020 (cf. Eloi *et al.*, 2019).

Si la mise en œuvre d'outils techniques est une étape indispensable, c'est cependant avant tout sur l'accompagnement humain que reposera la large diffusion des humanités numériques et l'utilisation de ces vastes collections dans de nombreux travaux de recherche.

Bibliographie

AUBRY S., CARTIER E. & STIRLING P. (2018), "Néonaute: mining Web archives for linguistic analysis", communication à l'IIPC Web archiving conference, Wellington (Nouvelle-Zélande), http://netpreserve.org/ga2018/wp-content/uploads/2018/11/IIPC_WAC2018-Sara_Aubry_Emanuel_Cartier_Peter_Stirling-Neonaute-mining_Web_archives_for_linguistic_analysis.pdf

BEAUDOUIN V. & PEHLIVAN Z. (2017), « Cartographie de la Grande Guerre sur le Web » : Rapport final de la phase 2 du projet *Le devenir en ligne du patrimoine numérisé : l'exemple de*

(13) Pour davantage d'informations sur ce projet, voir Praetzelis (2019).

(14) Voir également le site Web90 – Patrimoine, Mémoires et Histoire du Web dans les années 1990 : <https://web90.hypotheses.org> (consulté le 11 mars 2020).

(15) Un exemple parmi beaucoup d'autres : le site The Programming historian, <https://programminghistorian.org> (consulté le 11 mars 2020).

(16) Voir la page du site d'Archives Unleashed consacrée au toolkit : <https://archivesunleashed.org/aut/> (consultée le 11 mars 2020), ainsi que Milligan (2016).

la Grande Guerre, Paris, Bibliothèque nationale de France; Bibliothèque de documentation internationale contemporaine ; Télécom ParisTech, 2017, <https://hal.archives-ouvertes.fr/hal-01425600>

BEAUDOUIN V., CHEVALLIER Ph. & MAUREL L. (dir.) (2018), *Le Web français de la Grande Guerre : réseaux amateurs et institutionnels*, Nanterre, Presses universitaires de Paris Nanterre, <https://catalogue.bnf.fr/ark:/12148/cb45699767f>

BERMES E. (2020), « Le numérique en bibliothèque : naissance d'un patrimoine : l'exemple de la Bibliothèque nationale de France (1997-2019) », thèse de doctorat, Paris, Ecole nationale des Chartes, 107 p., <https://tel.archives-ouvertes.fr/tel-02475991>

BORTZMEYER S. (2018), *Cyberstructure : l'internet, un espace politique*, Caen, C&F éditions, « Société numérique », <https://catalogue.bnf.fr/ark:/12148/cb45637569v>

BRÜGGER N. (dir.) (2017), *Web 25: histories from the first 25 years of the World Wide Web*, New York, Peter Lang publishing, "Digital formations", <https://catalogue.bnf.fr/ark:/12148/cb453621987>

BRÜGGER N. & LAURSEN D. (dir.) (2019), *The historical Web and digital humanities: the case of national Web domains*, London, Routledge, "Digital research in the arts and humanities", <https://catalogue.bnf.fr/ark:/12148/cb457626551>

DERROT S., FAUDUET L., OURY C. & PEYRARD S. (2012), "Preservation Is Knowledge: A community-driven preservation approach", communication à la 9^e International Conference on Preservation of Digital Objects (iPRES), Canada, <https://hal.archives-ouvertes.fr/hal-00925315>

ELOI C., MOIRAGHI E. & ROSE V. (2019), « Un espace pour les humanités numériques à la BnF », *Bulletin des bibliothèques de France (BBF)*, n°17, p. 90-95. <http://bbf.enssib.fr/consulter/bbf-2019-17-0090-009>

GEBEIL S. (2015), « La fabrique numérique des mémoires de l'immigration maghrébine sur le Web français (1999-2014) », thèse de doctorat, Université Aix-Marseille, 2 vol. (1 109 p.), <http://www.theses.fr/fr/2015AIXM3119#>

GEBEIL S. (2017), « Quand l'historien rencontre les Archives du Web », billet publié sur le blog Web Corpora, 10 novembre 2017, <https://Webcorpora.hypotheses.org/380>

ILLIEN G. (2011), « Une histoire politique de l'archivage du Web », *Bulletin des bibliothèques de France*, n°2, pp. 60-68, <http://bbf.enssib.fr/consulter/bbf-2011-02-0060-012>

KAHLE B. (1997), "Preserving the Internet", *Scientific American*, Vol. 276, N°3, March, pp. 82-83, <https://www.jstor.org/stable/24993660?seq=1> (accès payant)

LE FOLLIC A., STIRLING P. & WENDLAND B. (2012), "Putting it all together: creating a unified Web harvesting workflow at the Bibliothèque nationale de France", communication au workshop IIPC How to fit in? Integrating a Web archiving program in your organisation, Paris, Bibliothèque nationale de France, novembre, <https://hal-bnf.archives-ouvertes.fr/hal-00873759>

MASANES J. *et al.* (2002), "A first experience in archiving the French Web", Actes de la 6^e European conference on Research and advanced technology for digital libraries (ECDL), Rome (Italie), <https://bit.ly/39JLpRv>

MILLIGAN I. (2016), "Lost in the Infinite Archive: The Promise and Pitfalls of Web Archives", UWSpace, <http://hdl.handle.net/10012/10322>

MILLIGAN I. (2019), “Historians’ archival research looks quite different in the digital age”, *The Conversation*, <https://theconversation.com/historians-archival-research-looks-quite-different-in-the-digital-age-121096>

NORA P. (2008), « Entre mémoire et histoire » : [article rédigé en 1984], *in* *Les Lieux de mémoire*, vol. I, Paris, Gallimard, « Quarto », pp. 23-43, <https://catalogue.bnf.fr/ark:/12148/cb36695433c>

PRAETZELLIS M. (2019), “From Open Access to Perpetual Access: Archiving Web-Published Scholarship”, communication à l’IIPC Web archiving conference, Zagreb, http://netpreserve.org/ga2019/wp-content/uploads/2019/07/IIPCWAC2019-JEFFERSON_BAILEY__MARIA_PRAETZELLIS-From_open_access_to_perpetual_access-archiving_Web-published_scholarship.pdf

SCHAFER V. (2018a), *En construction : la fabrique française d’internet et du Web dans les années 1990*, Bry-sur-Marne, INA, « Études et controverses », <https://catalogue.bnf.fr/ark:/12148/cb45632257t>

SCHAFER V. (dir.) (2018b), *Temps et temporalités du Web*, Nanterre, Presses universitaires de Nanterre, « Intelligences numériques », <https://catalogue.bnf.fr/ark:/12148/cb45466310p>

STIRLING P., ILLIEN G., SANZ P. & SEPETJAN S (2011), « La situation du dépôt légal de l’Internet en France : retour sur cette nouvelle législation, sur sa mise en pratique depuis cinq ans, et perspectives pour le futur », communication au 77^e congrès de l’International federation of library associations (IFLA), Porto Rico, <https://www.ifla.org/past-wlic/2011/193-stirling-fr.pdf>