



Francesca Musiani, Camille Paloque-Bergès, Valérie Schafer et Benjamin G. Thierry

Qu'est-ce qu'une archive du web ?

OpenEdition Press

Introduction

DOI : 10.4000/books.oep.8737
Éditeur : OpenEdition Press
Lieu d'édition : OpenEdition Press
Année d'édition : 2019
Collection : Encyclopédie numérique
ISBN électronique : 9791036504709



<http://books.openedition.org>

Référence électronique

MUSIANI, Francesca ; et al. *Introduction* In : *Qu'est-ce qu'une archive du web ?* [en ligne]. Marseille : OpenEdition Press, 2019 (généré le 24 avril 2019). Disponible sur Internet : <<http://books.openedition.org/oep/8737>>. ISBN : 9791036504709. DOI : 10.4000/books.oep.8737.

INTRODUCTION

Pour les vingt ans de la fondation Internet Archive, créée en 1996 et pionnière dans l'archivage du Web aux États-Unis, ainsi que pour les dix ans du dépôt légal du Web en France, la Bibliothèque nationale de France (BnF) et l'Institut national de l'audiovisuel (Ina) coorganisaient en 2016 un colloque intitulé « Il était une fois dans le Web : 20 ans d'archives de l'internet en France ». S'il fallait conter l'épopée de l'archivage du Web, sans doute y croiserait-on quelques preux protagonistes partis en quête d'un archivage du Web mondial ou exhaustif, des défis à affronter – humains, législatifs ou techniques – où cohabiteraient notamment captcha, robots.txt et droits d'auteurs, des issues heureuses aussi, tel le mariage prometteur de l'archive du Web et de la recherche. Roman chevaleresque, conte de fées, livre dont vous êtes le héros ? Si sans doute l'archive du Web pourrait s'y prêter, ce livre se contentera d'une mise en intrigue plus classique et d'initier le lecteur aux enjeux des archives et de l'archivage de la Toile. Il raconte l'archive de manière pragmatique, en essayant de rendre palpable à la fois sa fabrique (technique, institutionnelle, juridique), les évolutions qu'a connues cet archivage en une vingtaine d'années et la relation que les acteurs de l'archivage du Web entretiennent avec les publics. Ce tour d'horizon ne serait pas complet sans s'intéresser à la manière dont ces archives, outre leurs qualités patrimoniales, peuvent aujourd'hui être exploitées par le monde de la recherche. Il sera donc question d'archivage, d'outils et de métadonnées, mais aussi d'héritage culturel, de géopolitique ou encore d'éthique, tant les enjeux qui touchent au patrimoine et à l'archive ne peuvent être dissociés d'enjeux de mémoire et d'histoire aux multiples parties prenantes.

Organisé en quatre parties, notre propos suit l'archive du Web depuis sa naissance et sa conservation, dans les deux premières parties, jusqu'à son exploitation.

Lorsque l'on évoque l'archive du Web, il faut se figurer un objet singulier, interactif, fluide et non figé. Mais aussi une archive qui, bien qu'elle ressemble au Web du passé, n'en est pas la copie conforme et peut selon les fonds prendre des formes distinctes, enchâssées dans des interfaces, supportées par des techniques qui livrent des résultats visuellement différents. L'exemple le plus évident, sur lequel nous reviendrons, est celui de l'archivage du réseau social numérique Twitter : ici les différences d'archivage entre les deux institutions françaises en charge du dépôt légal du Web en France, la BnF et l'Ina, sont visibles à l'œil nu. À la BnF, les archives de Twitter s'apparentent à des captures d'écran, tandis que l'Ina a fait le choix d'une collecte fondée sur des données d'avant-garde brutes, sans capturer les images de fond. Mais au-delà, toute archive du Web véhicule des choix, des arbitrages. Ces choix ne sont pas seulement techniques, mais aussi profondément humains et sociétaux, voire politiques, ce qui est sans aucun doute le lot traditionnel d'autres types d'archives.

Avec le patrimoine pléthorique du Web en cours de constitution, outre la question de la masse (plus de 345 milliards de pages web archivées depuis 1996 par la fondation Internet Archive¹) se pose la question de la collecte, largement automatisée. Car les archives du Web introduisent bien des ruptures, que ce soit dans la notion même d'archive, ou dans les pratiques des archivistes et des chercheurs, même si on peut également y voir des continuités, comme nous le montrerons.

Automatisée, la collecte des archives du Web l'est à partir de périmètres négociés, et donc de choix humains. De cette curation, au moins initiale, dépend la représentativité de l'archivage, par elle se lit aussi l'inégale valeur accordée aux matériaux nativement numériques, archivés ou exclus de la collecte. Outillée au besoin, l'analyse des archives du Web l'est au service de questions posées par le chercheur, et là encore de choix humains. Ce sont aussi ces agentivités et interactions humaines et techniques que cet ouvrage propose de découvrir.

Ce projet est né d'une double volonté collective : celle de prolonger une initiative pensée avec Louise Merzeau quelques mois

1. Voir <https://archive.org/web/>.

avant sa disparition, elle qui, dans les ateliers du DL Web Ina² avec Claude Mussou ou encore dans notre projet ANR Web90³ et au fil de ses écrits, a tant fait pour penser, mais aussi pour faire connaître l'archive du Web. Le désir également de partager notre « goût » de l'archive du Web, alors que nous la prenons depuis plusieurs années comme source et objet d'étude. Nous espérons que cet ouvrage convaincra d'ailleurs le lecteur de considérer de manière indissociable la création de l'archive et son analyse et qu'il lui fournira des clés pour cela. Et peut-être sera-t-il même tenté de reconsidérer ensuite avec un regard décalé d'autres archives que celles nativement numériques ?

2. Voir le blog des ateliers : <http://atelier-dlweb.fr/blog/>.

3. <https://web90.hypotheses.org>.