

Ctrl Print Delete

Flânerie spéculative en niveaux de gris

Introduction

Cet essai approfondit certains sujets traités lors de la présentation donnée le 9 décembre 2011 à l'EESAB Lorient, autour de la plate-forme de publication Greyscale Press.

Le travail d'édition de Greyscale Press s'inscrit dans un contexte de changements radicaux opérant depuis quelques années: (1) Le passage au support numérique, qui reconfigure en profondeur tous nos médias : l'audiovisuel, les systèmes de communication, le livre. Concernant le livre en particulier, des opérations d'acquisition par scannage à grande échelle, entamées durant la dernière décennie, commencent à impacter les voies d'accès que nous avons aux archives de nos civilisations de l'écrit. Quelques protagonistes: Google, Internet Archive, les hackerspaces. (2) Les technologies de production, en particulier l'édition à petite échelle, révolutionnées par l'impression à la demande et le livre électronique, offrent de nouvelles perspectives aux éditeurs, créateurs et... spammeurs. (3) Politique des contenus: une offre accrue d'œuvres dans le domaine public (via Project Gutenberg, Google Books ou Internet Archive) ou sous licence libre (Wikipédia), autant de matériaux ouverts à tous usages, mais pouvant aussi tendre à la dévaluation de la production intellectuelle. (4) Politique de l'accès: l'organisation WikiLeaks expose les enjeux du libre-accès à l'information, diffusant les secrets militaro-étatiques à la vitesse du web. À y regarder de plus près, il s'agit encore et toujours d'archives, d'amoncellements de chiffres et de données défiant nos capacités d'analyse, et qui appellent à de nouveaux modes de lecture transversaux et algorithmiques.

Mutations (1) :

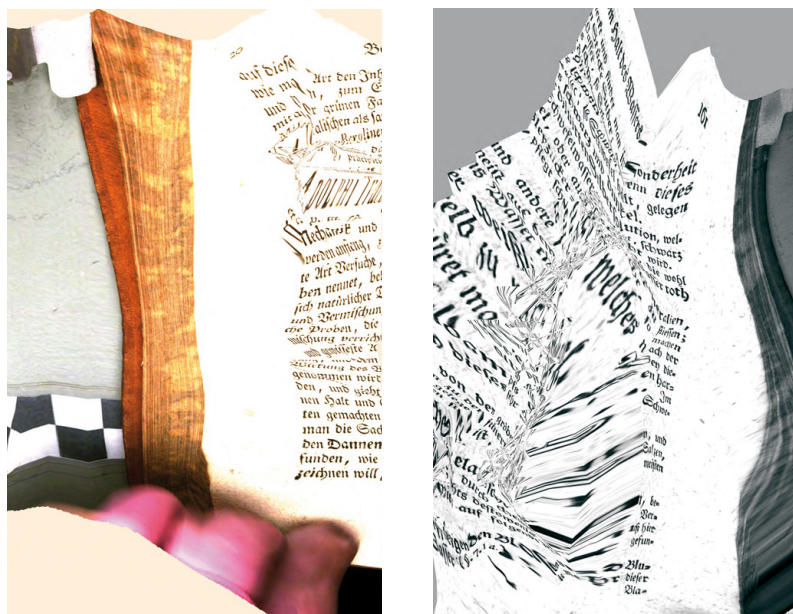
Le Grand Archivage

Au courant des dernières années, plusieurs événements parallèles ont concouru à modifier en profondeur notre rapport aux systèmes d'archivage, de lecture et de publication.

Le projet *Google Books*, préparé dans le plus grand secret pendant deux ans, fut révélé au public en 2004 (lors de la foire du livre de Francfort). L'idée originelle remonterait à 1996, année où les futurs fondateurs Sergey Brin et Larry Page travaillent sur le fonctionnement de la Stanford Digital Library – un projet qui les guidera dans la conception de leur moteur de recherche (GOOGLE,

2011).

En mars 2012, le nombre de livres scannés par Google se monte à 20 millions, sur les 130 millions existant dans le monde – selon une estimation calculée, elle-aussi, par Google (TAYCHER, 2010). En 2010, Jon Orwant (engineering manager de Google Books) annonce que l'intégralité des livres produits par l'humanité sera numérisée au courant de la décennie à venir (JACKSON, 2010). Autant les proportions de ce projet sont gargantuesques, autant son exécution demeure invisible – à l'exception de quelques doigts captés ici ou là par l'œil du scanner.



Légende : Erreur de calibration commise par Google Books en 2008, découverte par des internautes en 2011.

Plaquées brutalement contre la surface de l'écran, ces formes organiques inattendues mettent (littéralement) le doigt sur cette invisibilité, sur le degré d'abstraction automatisée dont Google se plait à entourer ses services. Ce n'est pas un hasard si ces erreurs – dévoilant la main d'œuvre humaine intervenant dans le processus – ont suscité un large écho sur la toile. À titre d'exemple, la découverte en août 2011 d'un ouvrage particulièrement maltraité par le logiciel de reconnaissance optique, «*Wohlgemeynte Gedanken...*» (le titre lui-même est erroné, l'original s'intitulant «*Hydrologie*»), a provoqué nombre de réactions. Le hacker Jamie « jzw » Zawinski n'a pas hésité à le hisser au rang de "classique dans le champ des *glitch studies*" (ZAWINSKI, 2011), tandis que l'artiste Greg Allen l'a détourné en *ready-made*, par la publication d'un fac-similé en print-on-demand (ALLEN, 2011).

C'est également un artiste, Andrew Norman Wilson, travaillant comme opérateur vidéo aux

Corporate Headquarters de Google, qui leva le voile sur la réalité sociale du processus d'archivage, dans sa vidéo "Workers Leaving the Googleplex" (WILSON, 2011). On y découvre que les opérateurs du scannage – surnommés "ScanOps" – constituent la caste la plus méprisée dans la hiérarchie du géant de Mountain View, et que toute référence à leur existence est sévèrement sanctionnée (par licenciement immédiat dans le cas de Wilson).

Mutations (2) :

DIY Bookscanning

Une autre opération de scannage poursuivant le même objectif, mais dotée de moyens plus limités, et pratiquant une politique d'ouverture diamétralement opposée, est menée depuis 2005 par l'Internet Archive (IA). Organisation à but non lucratif consacrée à l'archivage du Web, l'IA opère des centres de scannage dans 25 lieux du globe, en collaboration avec des institutions publiques (notamment la *Bibliotheca Alexandrina* en Egypte). En parallèle à ce travail de numérisation – quelques 3 millions de livres scannés en 2011 – l'IA s'est également attelé à la création d'une archive physique, visant à conserver un exemplaire de toute œuvre publiée sur la planète (KAHLE 2011). Rangés dans des conteneurs industriels, les livres sont stockés selon un principe visant le long terme, en prévision d'un futur où l'accès à l'original imprimé pourrait s'avérer crucial.



Légende : Stations de scannage («Scribe Stations»), Internet Archive, San Francisco, 2011. Photo : Jason "Textfiles" Scott, CC-BY-2.0.

Proche de cet idéal de préservation, une communauté d'amateurs s'est formée autour de la pratique du "DIY Bookscanning". Ce mouvement, bien que décentralisé, a mis au point des modèles de scanners réalisés en "open hardware" avec du matériel peu onéreux, ainsi que des

logiciels de reconnaissance optique open-source. Le slogan de ce mouvement, “one book scanner in every hackerspace”, situe ce projet dans la dynamique des machines de réplique 3D, en pleine expansion.

Mais pour quelle raison voudrait-on conserver une copie physique ? D’une part, le processus de numérisation, lié à la technique du moment, est par définition imparfait. Les machines évoluant, les résolutions qui nous paraissent aujourd’hui « bonnes » seront jugées insatisfaisantes d’ici quelques années, et l’opération devra être répétée. D’autre part, étant donné la facilité de manipulation du média électronique, une copie physique peut servir d’ultime garant matériel pour assurer la véracité d’un texte.

Fissures (1)

Ce dernier aspect, la difficulté à garantir l’authenticité d’une archive numérique, a été relevé par Julian Assange, fondateur de l’organisation WikiLeaks, dans une interview menée par Hans Ulrich Obrist au printemps 2011. Assange y décrit un projet ambitieux, visant à perfectionner le fonctionnement des hyperliens qui forment la trame d’Internet. Ce système assignerait une adresse unique à toute forme de contenu (texte, enregistrement audiovisuel), liant l’adresse aux données numériques du document. Reposant sur des modèles mathématiques, ce système empêcherait toute falsification, puisque la moindre modification du fichier entraînerait une modification de son adresse. Selon Assange, ce système, conçu pour garantir l’intégrité des données publiées par WikiLeaks, pourra s’appliquer à toute création existant sous forme numérisée – “un système d’indexation pour la Tour de Babel” (OBRIST, 2011).

Fissures (2)

Dans l’hypothèse qu’un tel système d’authentification puisse se mettre en place, il est important de relever la volatilité immanente de toute archive numérique. S’il fallut plusieurs guerres pour anéantir la bibliothèque d’Alexandrie, il suffit d’une situation économique morose pour qu’un géant technologique envoie aux oubliettes un pan singulier de l’histoire du web. Un cas exemplaire s’est déroulé en 2009 : GeoCities, un service d’hébergement fondé en 1994, passé en 1999 aux mains de la société Yahoo, annonçait en avril 2009 sa fermeture “dans le courant de l’été”. Jason Scott, archiviste chevronné de l’âge informatique, déclarera par la suite que “Yahoo a montré comment détruire un maximum de matériau historique en un temps minimum” (SCOTT, 2011). Farouchement décidés à transmettre aux futurs historiens ce monument du «web vernaculaire» (pour reprendre l’expression de l’artiste Olia Lialina), un groupe d’activistes nommé

«*The Archive Team*» parvint à sauvegarder une fraction considérable des quelque 38 millions de sites que comportait GeoCities. Suite à la fermeture définitive, le groupe mit à disposition sur des réseaux de téléchargement peer-to-peer une archive compressée de 641.32 gigabytes, pouvant tenir sur un petit disque dur portable.

Mutations (3) :

Les Machines à Livres

Si les opérations d'archivage à grande envergure peuvent paraître vertigineuses, l'évolution des modes de production dans le domaine de l'imprimé ne l'est pas moins. Un bel exemple est l'*Espresso Book Machine*, une super-photocopieuse combinant impression, imposition et reliure, pouvant générer un livre d'aspect professionnel en un temps record. Installée dans quelques bibliothèques et librairies dans le monde (dont la *Bibliotheca Alexandrina* citée plus haut), elle donne accès aux catalogues de quelques éditeurs partenaires, ainsi qu'aux vastes archives du domaine public. Là encore, cet appareil ne constitue que la partie visible de l'iceberg, la majeure partie de la production PoD (*print-on-demand*) se déroulant à l'abri des regards. Les six dernières années ont vu une avancée exponentielle de ce secteur: partant de 20'000 titres en 2006, passant à 120'000 l'année suivante, le PoD a atteint en 2010 une production de 2.7 millions (BOWKER, 2011). C'est une masse 8 fois supérieure à l'édition classique, en terme de titres disponibles. Une analyse plus détaillée montre qu'une large partie de ces titres est constituée de rééditions d'ouvrages en domaine public, rendus disponibles grâce aux opérations de numérisation que nous venons de décrire, ou encore de contenus copyleft, tels que des articles Wikipédia agrégés (ROMERO, 2011). Mitchell Davis, le fondateur de BiblioBazaar (qui a généré 1.4 millions de titres en 2010) n'hésite pas à déclarer : «Nous sommes avant tout une compagnie informatique, même si à l'issue du processus nous produisons des livres» (ALBANESE, 2010). On peut dès lors s'interroger si le terme de «livre» est approprié pour désigner ces objets éditoriaux: ne s'agit-il pas en premier lieu d'ectoplasmes numériques, de simples numéros ISBN attribués à un fichier PDF généré sans intervention humaine, attendant une hypothétique matérialisation dans un centre d'impression robotisé?



Légende : Deux ouvrages de l'éditeur Alphascript/Betascript Publishing, contenant des articles Wikipédia assemblés aléatoirement, devant la tombe de Jorge Louis Borgès (Cimetière des Rois, Genève). Photo : Manuel Schmalstieg.

Chronologie:

2002: Google débute secrètement le projet Google Book Search.

2005: Amazon acquiert la compagnie de print-on-demand (PoD) BookSurge.

2007: Amazon introduit la tablette Kindle. – L'Espresso Book Machine est inaugurée à New York. – Lulu.com a imprimé plus d'un million de livres.

2008: Le nombre de titres publiés en PoD dépasse les titres traditionnels. – L'Internet Archive a scanné 300'000 livres. – Google a scanné 7 millions de livres.

2009: La production PoD dépasse 1 million de titres, elle est trois fois supérieure au marché traditionnel. – L'éditeur-spammeur BiblioBazaar publie 270'000 titres à lui seul. – Fermeture de GeoCities.

2010: La production PoD atteint 2.7 millions. – Apple introduit la tablette iPad. – Google annonce vouloir scanner tous les livres de la planète (près de 130 millions). – Archive Team diffuse les archives de GeoCities via PirateBay.

2011: Amazon annonce que ses ventes e-books dépassent les ventes papier. – Internet Archive débute l'archivage de livres physiques (statut: 500'000), et a scanné 3 millions de livres. – L'éditeur-spammeur VDM Publishing publie 500'000 titres sur Amazon.

2012: Google a scanné 12 millions de livres...

A propos de Greyscale Press

Conçu comme un simulacre de maison d'édition, Greyscale Press consiste en un processus d'élaboration d'ouvrages papier, utilisant des technologies d'impression à la demande, et focalisé sur l'archive numérique - qu'il s'agisse de blogs abandonnés, de wikis et autres machines d'écritures collaboratives, de bases de données légales (les procès-verbaux du Tribunal pénal international pour l'ex-Yougoslavie) ou techniques (les RFC publiés par l'Internet Engineering Task Force, officialisant les standards de l'Internet).