

Vers l'archivage du web social événementiel

Aida Chebbi, Mohamed Ben Romdhane

► **To cite this version:**

Aida Chebbi, Mohamed Ben Romdhane. Vers l'archivage du web social événementiel. Dans les mailles du filet : Révolution tunisienne et Web 2.0, Presses Universitaires de la Manouba, pp.119-140, 2013. <sic_00879380>

HAL Id: sic_00879380

https://archivesic.ccsd.cnrs.fr/sic_00879380

Submitted on 3 Nov 2013

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.

Vers l'archivage du web social événementiel

Aïda Chebbi & Mohamed Ben Romdhane
Institut Supérieur de Documentation
Université de la Manouba

Introduction

Le régime de Ben Ali exerçait un sévère contrôle sur l'utilisation d'Internet. Ce contrôle s'opérait tout d'abord par une censure très large de sites et pages Web d'opposants, mais aussi ceux de citoyens ordinaires (Mihoub, 2011). Les sites de partage de vidéos et de photos ainsi que les pages Facebook et les sites proposant des outils de contournement de la censure étaient également inaccessibles depuis la Tunisie.

C'est dans ce contexte contraignant qu'a émergé en ligne une nouvelle forme de dissidence : la cybercontestation (Lecompte, 2011; Mihoub, 2011). Des Tunisiens, pour la plupart jeunes et issus de la classe moyenne, résidant soit à l'étranger, soit en Tunisie, sont apparus de plus en plus nombreux à manifester sur la toile leur mécontentement de la situation politique, économique et sociale. Pour y arriver, ils recourent à des outils de contournement de la censure, tels que les serveurs mandataires (*proxy servers*)¹, ou ils s'approprient l'usage des réseaux sociaux, tels que blogues, Twitter et Facebook, à des fins contestataires (Bouzouita, 2011; Dakhli, 2011; Lecompte, 2011; Mihoub, 2011).

Dans ce texte, nous discuterons du rôle des réseaux sociaux dans le soulèvement populaire en Tunisie, de l'importance de préserver les traces de cet événement sur la toile mondiale et des défis aussi bien techniques que méthodologiques auxquels feront face les spécialistes de l'information quant à la sélection, au traitement et à la pérennisation des traces les plus représentatives de cet événement.

1. Le Web 2.0 et les réseaux sociaux numériques

Le concept de Web 2.0 est apparu lors d'une rencontre de « *brainstorming* » préparatoire à la conférence "*Web 2.0 Conference*" organisée par O'Reilly Media et Medialive International en 2004 (Boisseau, 2005). Selon Tim O'Reilly, le Web 2.0 est représenté par les sept principes suivants :

1. Le Web comme une plate-forme pour de nouvelles applications.
2. Les contributions des utilisateurs, sous forme de connaissances ou de commentaires et évaluations, forment la base d'une intelligence collective.
3. L'accessibilité et le partage des données permettent la création de nouvelles applications et connaissances en combinant plusieurs sources de données.

¹ « [...] serveur qui filtre toutes les connexions vers l'extérieur de façon qu'elles semblent provenir de la même machine afin de protéger le réseau interne sous-jacent contre les intrus. » (Office québécois de la langue française, 2006).

4. Les applications sont constamment améliorées et les changements sont disponibles en ligne presque immédiatement « versions bêta », sans avoir à attendre les changements de versions.
5. Les technologies Web 2.0 permettent la création d'applications plus flexibles.
6. Le Web sert de plateforme pour le développement des applications équivalentes à celles que l'on retrouvait autrefois sur les ordinateurs.
7. Des technologies, comme AJAX², permettent d'enrichir les interfaces-utilisateur disponibles sur le Web pour accomplir diverses tâches (Boisseau, 2005; O'Reilly, 2005; Paré et Chebbi, 2009).

Le Web 2.0 diffère donc du Web 1.0. Les internautes ne se contentent plus de naviguer pour repérer des informations mises à leur disposition. Ils éditent eux-mêmes les contenus, décident comment les visualiser et créent des communautés d'intérêts pour diffuser et partager ces contenus. Le Web 2.0 est né donc d'une évolution dans la manière dont les internautes font usage du Web. En effet, les technologies du Web 2.0 offrent cette possibilité de trouver, organiser, partager ou créer de l'information d'une manière à la fois personnelle, mais globalement accessible (Stenger et Coutant, 2010).

Alors que certains insistent sur ce changement au niveau des pratiques éditoriales et sur la dimension technologique du Web 2.0, d'autres mettent l'accent sur la dimension sociologique des applications du Web 2.0. Ainsi, le Web 2.0, dans sa dimension sociologique, représente une nouvelle forme de réseaux d'interaction sociale, soit les réseaux sociaux numériques (Stenger et Coutant, 2010; Quan, 2011).

Les réseaux sociaux numériques constituent des services Web 2.0 qui permettent à un usager de :

- construire un profil public ou semi-public au sein d'un système,
- gérer une liste d'utilisateurs avec lesquels il partage un lien,
- visualiser et naviguer sur leur liste de liens,
- assurer une présence en ligne en décrivant son statut et ses activités (Quan, 2011, p. 2).

Sous cette appellation, nous pourrions regrouper, entre autres, les cybercarnets ou blogues, les sites de partages de contenus multimédias (ex. : YouTube, Flickr), les sites de réseautage (ex. : Facebook, LinkedIn) et des sites de microblogage³ (ex. Twitter).

Soutenir ou être à l'origine d'un soulèvement populaire ne fut guère une des finalités de ces réseaux sociaux. Le Web 2.0 a toutefois joué un rôle non négligeable dans des mobilisations sociales récentes. Certes, ce rôle est souvent contesté voire même décrit comme un fantasme technophile (Ayari, 2011), d'aucuns ne pourraient nier l'importance des réseaux sociaux en 2001 lors des soulèvements contre le président Joseph Estrada aux Philippines ou

² « Ajax (acronyme d'Asynchronous Javascript and XML) est une manière de construire des applications Web et des sites web dynamiques basés sur diverses technologies Web, telles que JavaScript, CSS, XML, le DOM et le XMLHttpRequest, ajoutées aux navigateurs Web ce qui permet d'apporter de nouvelles fonctionnalités et d'augmenter la réactivité de sites et applications Web. » (Wikipédia, 2012).

³ « Le microblogage est en fait une forme de blogage permettant, notamment grâce à l'Internet mobile, de communiquer en temps réel par messages courts, limités à 140 ou 200 caractères, selon la plateforme utilisée. » (Office québécois de la langue française, 2009)

en 2003 quand des manifestations contre la guerre en Irak ont été organisées simultanément dans 600 villes de par le monde. En 2009, des mobilisations populaires en Iran ont été également orchestrées sur Twitter (Ayari, 2011; Granjon, 2010). Les révolutions arabes ont vu apparaître de nouvelles façons d'utiliser les réseaux sociaux à des fins contestataires ou militantes (Bouzouita, 2011; Richet 2011).

Nous proposons dans le point suivant une analyse plus approfondie du rôle des réseaux sociaux dans le soulèvement populaire en Tunisie. Il importe d'avertir le lecteur que cette analyse ne s'appuie pas sur des recherches bien fondées, mais sur quelques essais publiés sur le sujet et sur une lecture personnelle des derniers événements en Tunisie

2. Les réseaux sociaux et les mouvements sociaux : le cas de la révolution tunisienne

Médias et experts ont reconnu le rôle catalyseur des réseaux sociaux au mouvement de contestation tunisien né à Sidi Bouzid (Granjon, 2010; Mihoub, 2011). En effet, les réseaux sociaux ont facilité l'organisation de *flashmob*⁴, des émeutes, des grèves et plusieurs autres actes de désobéissance civile, et ce, malgré la surveillance d'*Ammar 404*, personnage incarnant la police de l'Internet (Ayari, 2011; Lecompte, 2011; Mihoub, 2011).

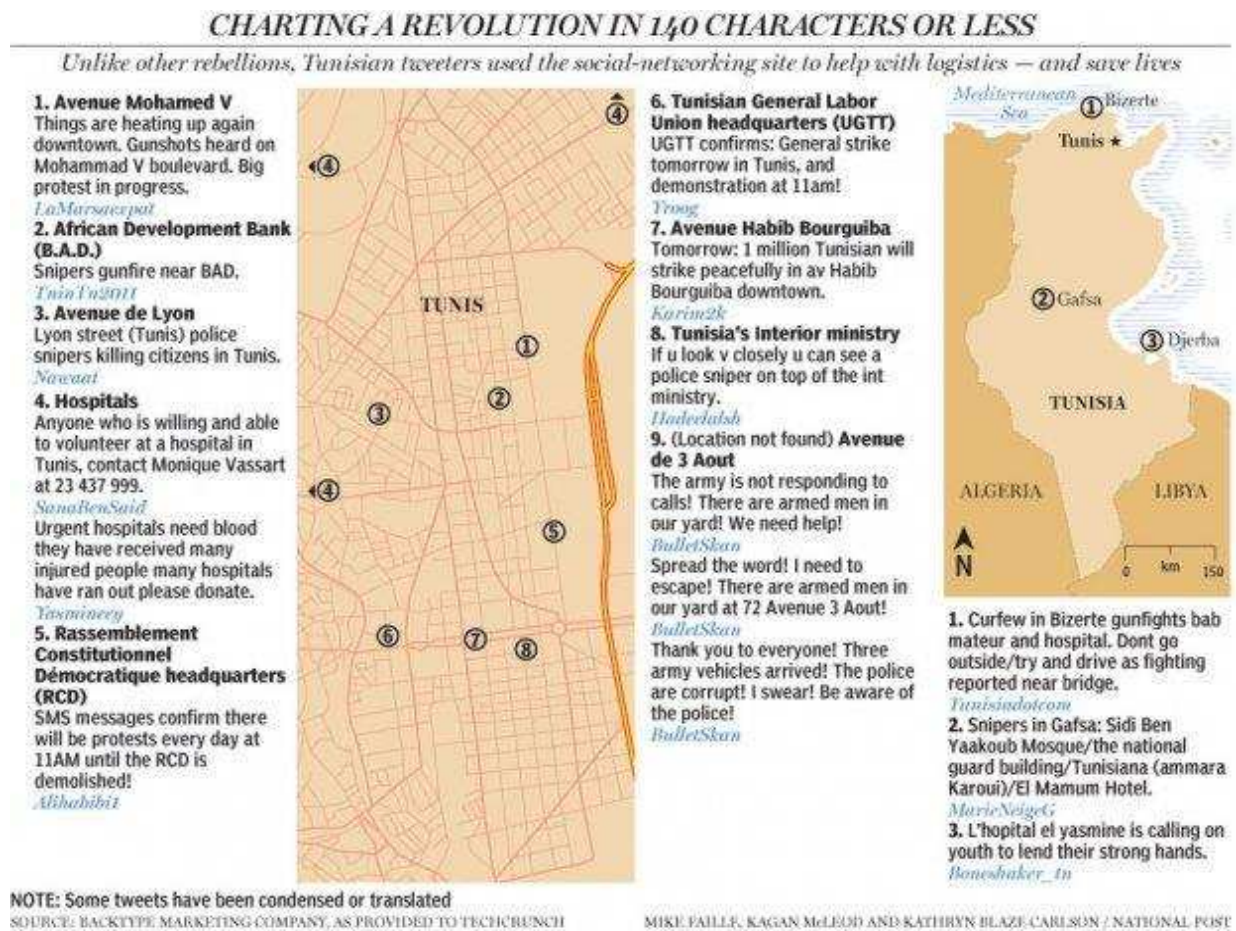
Dans un contexte où les médias traditionnels étaient sous la mainmise du pouvoir, les contestataires tunisiens, à l'aide de téléphones portables et d'appareils photo numériques, se sont improvisés journalistes citoyens en publiant sur Facebook, malgré le piratage et la censure, vidéos, photos et documents concernant, entre autres, les manifestations, la violence policière, les funérailles de victimes, la destruction d'archives, etc. La diffusion presque en temps réel de ces documents et témoignages a assuré une couverture des événements et leur transmission sur des chaînes satellitaires étrangères (Ayari, 2011; Farag, 2011; Granjon, 2010; Lecompte, 2012; Mihoub, 2011). Ces dernières ont même mis à la disposition des internautes des liens spéciaux qui permettent de poster directement leurs vidéos et leurs photos (Mihoub, 2011).

Un simple clic sur le bouton « J'aime », la publication d'une photo chargée en émotion sur son profil (ex. : drapeau tacheté de sang) devient une forme d'expression d'un mécontentement, de soutien à une revendication (Bouzouita, 2011). Facebook a contribué également « à la déterritorialisation des événements locaux, à la construction de l'indignation, à la convergence du sens sur les réseaux sociaux, mais aussi à la constitution d'un potentiel de mobilisation » dans un pays où corruption et censure sévissent (Granjon, 2010).

Au cœur de l'événement, Twitter a été utilisé pour rappeler un lieu de rendez-vous, rendre compte d'un événement en direct, orienter les trajectoires des manifestations en prévenant les emplacements de tireurs embusqués et des barrages policiers, demander de l'aide aux autres personnes de son quartier, organiser la logistique des manifestations ou demander des urgences médicales (Mihoub, 2011, Richet 2011). La figure 1 présente quelques *tweets* échangés lors de la « révolution tunisienne ».

⁴ Rassemblement, organisé par le moyen d'Internet, d'un groupe de personnes qui ne se connaissant pas pour la plupart, dans un lieu public, pour y effectuer des actions convenues d'avance, avant de se disperser rapidement.

Figure 1: Exemples d'usage de Twitter lors du soulèvement populaire en Tunisie, par le National Post (<http://nationalpostnews.files.wordpress.com/2011/01/twittermaplarge.jpg?w=620>), cité dans Richet, 2011



Des blogueurs cyberactivistes ont accompagné et amplifié ce mouvement de protestation populaire. Actifs pour la plupart bien avant la révolution⁵, parfois depuis l'étranger, les blogueurs tunisiens ont réalisé un travail important de contextualisation, de vérification et parfois même de traduction dans d'autres langues des informations qui circulaient en ligne (Lecompte, 2012; Mihoub, 2011).

La fuite d'informations sensibles grâce à Wikileaks et aux attaques du groupe Anonymous⁶ a contribué également à alimenter les réseaux sociaux en éléments de preuve sur la corruption du gouvernement incitant aussi les internautes à la mobilisation (Granjon, 2010; Mihoub, 2011).

Le Web 2.0 est devenu « [une] source de vérité » (Bouzouita, 2011), « un support significatif d'écriture et de stockage de la mémoire collective » (Mihoub, 2011). Les « archives » de la « révolution tunisienne » sont principalement consignées sur les réseaux sociaux (Dakhli,

⁵ Les cyberactivistes tunisiens sont bien présents sur la toile depuis les années 90. Ils s'expriment sur des blogues, des webzines et des sites individuels dont les plus connus : Nawaat, Takriz, Kalima, Réveil tunisien, TunEzine, Z : débat Tunisie (Bouzouita, 2011; Granjon, 2010; Lecompte, 2011; Mihoub, 2011).

⁶ Le 2 janvier 2011, Anonymous a lancé Opération Tunisie au cours de laquelle les sites de la présidence de la république et des principaux ministères ont été piratés, et ce, pour riposter à la censure du site Wikileaks (Mihoub, 2011).

2011). Les matériaux nécessaires à l'écriture et à la relecture de cet événement social de même que ceux relatifs à la naissance d'un usage contestataire de l'Internet devraient être préservés (Farag, 2011). Le Web 2.0 n'offre pas de garanties quant à la fixité et à la pérennisation des informations qu'il véhicule avant, durant ou bien après le soulèvement populaire en Tunisie. Comment sauvegarder alors ces traces pour pouvoir étudier un fragment de l'histoire de la Tunisie écrit, en partie, grâce à une appropriation des réseaux sociaux ?

3. L'archivage et la préservation du Web social événementiel

Même si les applications du Web 2.0 gèrent la temporalité des contenus (historique des changements et mises à jour des pages d'un Wiki, archives en ligne d'un blogue, téléchargement en local d'un profil Facebook), elles offrent rarement des fonctions de conservation à long terme (Bermès et Oury, 2009). De plus, le Web 2.0 apporte son lot de défis : accroissement du volume des informations, mise à jour et agrégation continues des informations, diversité des formats et des plateformes de diffusion, etc. Au-delà des ces difficultés techniques, se pose des questions sur le processus d'archivage, notamment celles relatives à l'évaluation et la sélection des matériaux les plus représentatifs de notre histoire contemporaine, la fréquence de capture, les méthodes de capture et de traitement des contenus issus du Web 2.0 ainsi que les méthodes d'accès et d'exploitation des contenus archivés.

Tout en présentant les propriétés de leurs composantes documentaires, nous décrirons dans ce qui suit les méthodes et techniques possibles pour l'archivage des blogues, des *tweets* et des pages Facebook. Nous discuterons ensuite des enjeux documentaires et éthiques de l'archivage du Web 2.0.

3.1 L'archivage des blogues

Selon les statistiques de la blogosphère en 2011, on estime à 164 millions le nombre de blogues dans le monde (Invesp, 2011). Les blogues deviennent très populaires dès 2003 (Daffini, 2010). Les résultats d'une récente étude de Technorati montrent que les blogues n'ont pas perdu de leur attrait malgré une forte concurrence avec d'autres applications de réseautage social (Technorati, 2011). L'étude révèle que ceux qui bloguent uniquement par passion sont largement majoritaires (60% des blogueurs amateurs, 21% de blogueurs professionnels). 55% des blogueurs mettent à jour leur blogue 4 à 12 fois par mois. 10 % utilisent WordPress comme plateforme de publication (Technorati, 2011)⁷.

Mais qu'est-ce qu'un blogue ?

Il existe différents types de blogues qui sont déterminés en fonction de leur thématique, objectif, audience ou format. Ces différents types de blogues se partagent par contre les mêmes caractéristiques techniques (Hank et al., 2007). Un blogue est un site Web personnel ou institutionnel tenu par un ou plusieurs auteurs (appelés blogueurs), qui s'expriment selon une certaine périodicité, sous la forme de billets (articles ou notes) sur un ou différents

⁷ Technorati s'est appuyé sur les témoignages de 4,114 blogueurs issus du monde entier.

sujets. Ces billets sont datés, à la manière d'un journal de bord, classés par ordre chronologique inverse, parfois enrichis d'hyperliens, d'images ou d'autres éléments multimédias, et pouvant faire l'objet de commentaires laissés par les lecteurs (Grand dictionnaire terminologique, 2009; Hank et al., 2007). Les billets peuvent être affichés par thème ou marqueur sémantique (*tag*), par date ou en effectuant une recherche par mots clés avec le moteur de recherche du blogue.

Chaque billet s'accompagne d'un lien propre et statique, appelé lien permanent. Les billets ou commentaires les plus récents sont rassemblés en fil de nouvelles (fil RSS ou fil Atom) grâce à un logiciel agrégateur afin de permettre aux abonnés d'un blogue de tenir à jour leur lecture des billets et commentaires de manière rapide et sélective.

De la même manière, des liens vers d'autres blogues (blogues « amis » ou *blogroll*) offrent une représentation communautaire des centres d'intérêt de l'auteur du blogue. Les billets et commentaires publiés sur ces blogues « amis » peuvent être intégrés, grâce à la technique de syndication des contenus (fil RSS), à son propre blogue (Daffini, 2010).

Les blogues peuvent être développés grâce à des plateformes dédiées à la publication de blogues. Deux types de plateformes coexistent, à savoir :

- les plateformes gérées automatiquement qui ne requièrent pas l'installation d'applications particulières. Les procédures de maintenance du blogue sont prises en charge par le fournisseur de la plateforme,
- Les plateformes à gérer soi-même. Le blogue doit être installé sur un serveur. Un minimum de compétences techniques est alors requis pour la maintenance du blogue.

Plusieurs systèmes de gestion de contenu (*Content Management System*) offrent également un module de blogue (Paré et Chebbi, 2009).

Les blogues sont de nature dynamique. La fréquence de leur mise à jour dépend des habitudes des blogueurs. Certaines mises à jour sont régulières, d'autres sont intermittentes (Hank et al., 2007).

Les projets d'archivage des blogues ne sont pas aussi nombreux et bien documentés tel le cas des projets d'archivage du Web public. La recension des écrits nous n'a pas permis de dresser un portrait exhaustif de l'état de la situation. Nous avons complété notre revue de littérature par l'exploration de quelques initiatives afin d'identifier les méthodes et techniques d'archivage des blogues et évaluer leurs avantages et inconvénients. Nous nous sommes basés sur la documentation et les archives Web des institutions suivantes :

- *Library of Congress*⁸: cette institution archive depuis 2007 une collection de 133 blogues professionnels concernant le domaine légal et juridique (*Legal Blawgs Web Archive*),
- Bibliothèque nationale de France (BnF)⁹ : en collaboration avec l'Association pour l'Autobiographie et le Patrimoine Autobiographique, la BnF collecte deux fois par an, depuis 2007, une liste de journaux personnels et littéraires. La collection compte quelque 400 sites et vise à être représentative de l'histoire du « diarisme

⁸ Adresse URL: <http://lcweb2.loc.gov/diglib/lcwa/html/lawlb/lawlb-overview.html>

⁹ Adresse URL: <http://blog.bnf.fr/lecteurs/index.php/2009/04/06/du-journal-intime-au-blog/>

- en ligne ». La BnF propose, pour explorer cette collection, un parcours guidé intitulé *(S')écrire en ligne* (disponible sur les postes de la salle de lecture).
- University of London Computing Center (ULCC), British Library et UK Joint Information systems Committee (JISC)¹⁰: Le JISC a mené, en collaboration avec l'ULCC et la British Library, un projet visant à expérimenter l'archivage des sites Web, *JISC PoWR project (Preservation of Web Resources)*. Le JISC a développé en 2008 un blogue institutionnel sur la plateforme WordPress pour documenter le projet, mais également pour tester une technique particulière d'archivage des blogues (archivage par fils RSS).
 - *National Library of Australia* et partenaires¹¹: PANDORA, les archives Web australiennes, comprennent une collection très sélective de 461 blogues académiques. La recherche dans PANDORA pourrait être limitée aux blogues.
 - Internet Archive : grâce à une collaboration avec la *Library of Congress* et la BnF, Internet Archive offre depuis le 19 janvier 2011 un accès en ligne à la collection "*Jasmine Revolution*"¹², une collection d'archives Web (235 ressources) constituée principalement de blogues (95 blogues) qui documentent le soulèvement populaire en Tunisie (voir Figure 2).

Figure 2 : Capture d'écran de la page d'accueil de la collection "*Jasmine Revolution*" d'Internet Archive



Dans une communication présentée en 2010 à l'*International Conference on Preservation of Digital Objects*, Kelly et Marieke discutent 3 stratégies possibles pour l'archivage des blogues, soient :

1. Production d'une version statique du contenu d'un blogue. Cette stratégie pourrait être pertinente dans le cas d'un blogue de projet. Par exemple, à la fin d'un projet, produire une version en HTML statique du blogue;

¹⁰ Adresse URL: <http://jiscpowr.jiscinvolve.org/wp/archived-blog/>

¹¹ Adresse URL: <http://pandora.nla.gov.au/subject/29>

¹² Adresse URL: <http://www.archive-it.org/collections/2323>

2. Production d'une copie de sauvegarde d'un blogue si l'hébergeur ou le fournisseur de la plateforme du blogue n'offre pas de garanties quant à la conservation à long terme du blogue;
3. Migration du blogue vers une autre plateforme pour maintenir ou enrichir ses fonctionnalités (Kelly et Marieke, 2010).

La première stratégie pourrait s'opérationnaliser en capturant un blogue à distance grâce à un robot de collecte. Cette technique a des limites telles que nous l'avons constaté en naviguant dans les archives des blogues de la *Library of Congress* ou celles de la *National Library of Australia*. Certains fragments des blogues n'ont pas pu être capturés. Dans d'autres cas, la visualisation de quelques éléments de la version archivée d'un blogue requiert l'installation de plugiciels ou des versions particulières de navigateurs Web. La figure 3 présente les conditions d'accès à une version archivée d'un blogue capturé par la *National Library of Australia*.

Figure 3 : Capture d'écran de la liste des captures d'un blogue archivé par la *National Library of Australia*

Bent lenses

Bent lenses was selected for preservation by the [National Library of Australia](#). This title is scheduled to be re-archived regularly. The [publisher's site](#) may provide more current information.

Bent Lenses blog ceased in Apr. 2008. Bent Lenses website will be archived in it's place.

Archived [03 Jun 2011 00:30](#)
 Archived [04 Jun 2009 00:01](#) ⓘ
 Archived [04 Jun 2008 15:50](#) ⓘ
 Archived [04 Jun 2008 09:13](#) ⓘ
 Archived [09 Oct 2006 00:00](#) ⓘ

This archive contains files that may need the following programs/plugins:

- Acrobat file(s) which can be read with Adobe Acrobat Viewer or Ghostview
- Portable Network Graphic file(s) which can be read with Internet Explorer 5+ or a PNG viewer or Netscape 6+
- Word file(s) which can be read with Microsoft Word or StarOffice

Create Persistent Identifier for URL:

Une autre technique consiste à produire une version PDF du blogue ou tout simplement une version papier. Cette technique ne permet pas de préserver l'aspect interactif et la structure du blogue et ne semble pas être adoptée comme stratégie de préservation numérique par aucune des institutions dont nous avons étudié les archives Web.

La production d'une copie de sauvegarde d'un blogue est une stratégie qui pourrait être opérationnalisée de deux manières. La première repose sur l'usage de la fonction d'exportation du blogue lui-même ou des applications API¹³ pour produire une copie du blogue. La deuxième consiste à réaliser des copies du blogue par extraction en format XML. Ces deux techniques présentent le risque de perdre l'aspect visuel et fonctionnel du blogue : tous les billets seront visualisés sur une même et unique page HTML.

La migration du blogue vers une autre plateforme pourrait avoir lieu par moissonnage (collecte) de fils RSS. Cette technique a été expérimentée par le JISC. Elle présente plusieurs inconvénients, dont la perte de billets ou de commentaires ou la perte de l'aspect visuel du

¹³ "An API is a set of commands, functions, and protocols which programmers can use when building software for a specific operating system." Dans: TechTerms, 2012. Adresse URL: <http://www.techterms.com/definition/api>

blogue et de sa structure. De même, les fils RSS ne permettent que la collecte des billets les plus récents (Kelly et Marieke, 2010).

Le recours aux services d'un tiers-archiviste tel qu'Internet Archive par le biais de son service Archive-It constitue une autre alternative. Certes, la collection "Jasmine Revolution" est une initiative qui devrait être saluée, toutefois, réalisée avec un regard externe sur le mouvement contestataire en Tunisie ne pourrait former une collection représentative des événements ayant lieu.

Les difficultés techniques ne sont pas les plus problématiques. L'identification des critères de sélection, des contenus à archiver et des responsabilités d'archivage est plus ardue à asseoir. Pour confirmer cette assertion, nous nous appuyons dans ce qui suit sur des résultats de quelques recherches qui cherchent à définir, selon différentes perspectives, des critères d'évaluation/sélection des contenus Web 2.0.

La recherche réalisée par Hank et ses collaborateurs analyse le point de vue des blogueurs sur leur propre production. Les blogueurs consultés sont d'avis que leurs blogs méritent d'être conservés pour les trois principaux critères suivants :

1. Originalité du blogue (contenu)
2. Signifiante culturelle du blogue
3. Popularité du blogue

D'autres critères, de moindre importance, ont été identifiés par les blogueurs. Ils sont indiqués dans la figure suivante (Figure 4).

Figure 4 : Critères pour la sélection des blogs, extrait de l'étude de Hank et al., 2007

Selection: Bloggers were asked to identify characteristics for appraisal.

Very Important or important	Characteristic	Very unimportant or unimportant
88.6%	Unique/innovative content	3.3%
84.4%	Cultural significance (blogger)	3.8%
75.3%	Popularity	8.7%
69.7%	Number of links to blog	8.6%
66.8%	Number of regular readers	9.8%
66.3%	Number of hits	14.2%
65.4%	Preservation preference	10.4%
63.4%	Unique/innovative functionality	12.6%
54.3%	Unique/innovative graphic attributes	16.3%
52.5%	Popularity of (other) blogs linking to blog	17.5%
33.9%	Number of occasional readers	20.7%
32.6%	Age of blog	26.5%
23.5%	Size of blog	30.8%

Hank et al. 2007.

Selon la même étude, les éléments textuels du blogue (billets et commentaires), les éléments multimédias et les mots clés (*tag*) constituent les éléments les plus importants à capturer sur un blogue (Davis, 2009; Hank et al. 2007). Selon les blogueurs, la responsabilité d'archivage revient aux créateurs de contenus, c'est-à-dire les blogueurs eux-mêmes (90% des répondants) et non aux fournisseurs de services (20%) ou encore aux institutions de mémoire (4%).

Capra et ses collaborateurs (2008) proposent une autre méthode pour identifier et sélectionner, par le biais d'applications API, les contenus les plus représentatifs d'un thème,

d'un événement, etc. Les auteurs suggèrent la méthode des requêtes, c'est-à-dire, lancer une recherche sur le Web avec une liste de mots clés en rapport avec un thème ou un sujet et collecter ensuite les résultats les plus pertinents. Cette méthode diffère de la logique du moissonnage (*crawling*) où l'on établit une liste d'URLs, c'est-à-dire une liste de sites Web à capturer et on paramètre le robot de collecte à aspirer ces sites et le tissu de liens qui les constituent. Lee et Thibo dressent les limites de la méthode de moissonnage :

« Web archiving based on recursive link following, however, faces two major challenges. First, link paths often do not map cleanly or directly to long-standing criteria for appraisal and collection development, e.g. topics, provenance, genres, dates. Second, a substantial portion of material on the Web is accessible through posing queries to databases, rather than following links. » (Lee et Thibo, 2008)

D'après les résultats préliminaires du *VidArch project*, projet mené par Lee et Thibo pour expérimenter l'archivage des vidéos YouTube relatifs aux élections présidentielles de 2008 aux États-Unis, l'archivage des blogues liés à cet événement permet non seulement de contextualiser les vidéos YouTube, mais également d'identifier d'autres ressources pertinentes à collecter (Lee et Thibo, 2008).

Il en ressort que différentes méthodes et techniques sont en cours d'évaluation. Chaque proposition présente des avantages et des inconvénients ce qui les rend complémentaires. Selon notre point de vue, la définition et la précision des objectifs d'archivage orienteront les responsables d'un projet d'archivage Web 2.0 vers un choix plus réfléchi des méthodes et techniques les plus appropriées au projet.

3.2 L'archivage de Twitter

Twitter occupe la deuxième place de l'activité des réseaux sociaux avec 500 millions d'abonnés générant 60 milliards de messages¹⁴.

Twitter est un site de microblogage où un abonné à ce service peut communiquer par messages courts limités à 140 caractères (des *tweets*) (Campbell et Dulabahn, 2010). Ces messages peuvent être diffusés en temps réel (*livetweet*). Les abonnés à Twitter ont également accès à un système de messagerie privée

Un abonné à Twitter, un *twitteur*, a la possibilité de partager son compte avec d'autres personnes (*followers*). La *timeline* constitue l'ensemble des messages qu'un abonné reçoit sur son compte. Des mots clés peuvent être associés à ces messages en ajoutant un *hashtag* (signe #). Le signe @ est utilisé dans certains *tweets* pour lier un de vos messages à un autre profil. Le signe RT est également utilisé pour indiquer que l'on recopie le *tweet* d'un autre abonné.

Le 14 avril 2010, Twitter a fait un don de ses tweets publics de 2006 à 2010 à la *Library of Congress*, environ 5 téraoctets de contenus (Raymond, 2010). Cette décision a été contestée. Twitter défend sa décision en se référant aux termes de sa politique de confidentialité:

¹⁴ SocialiteMedia (23 février 2012). Twitter hits the 500 million mark. Adresse URL: <http://socialitemedia.com.au/twitter-hits-the-500-million-mark/874/> (Consulté le 01 avril 2014).

« Lorsque vous utilisez un de nos Services, vous consentez à la récupération, le transfert, la manipulation, le stockage, la divulgation entre autres utilisations de vos informations comme décrit dans cette politique de confidentialité » (Twitter, 2011).

En même temps, Twitter rassure ses membres en expliquant que l'exploitation de ces archives est soumise à quelques conditions. En effet, la *Library of Congress* est tenue de limiter l'accès à ces archives. Les *tweets* ne sont accessibles que 6 mois après leur date de publication. Les données relatives à un compte Twitter, les *tweets* supprimés de même que les liens vers des photos et sites Web n'ont pas été transférés à la *Library of Congress*. De même, l'exploitation de ces archives serait limitée à des études de recherche à but non commercial (Campbell et Dulabahn, 2010). Ce corpus permettrait à la *Library of Congress* d'explorer de nouvelles approches de préservation des collections numériques. Les archives de Twitter sont d'une valeur culturelle, scientifique et historique inestimable. Les archives de Twitter sont une source primaire sur des événements de la vie quotidienne, mais aussi des événements marquants de notre histoire contemporaine.

“For historians, Twitter provides direct witness accounts of events in real time. It also serves as a virtual timeline of communications about events, people and places. This provides an enormous amount of raw unmediated primary source material for historical research.” (Campbell et Dulabahn, 2010).

Les chercheurs en sciences sociales puiseront certainement dans ces archives pour étudier les nouveaux modèles de socialisation, la réaction des individus à des événements historiques particuliers, etc.

En plus de la modalité de don ou dépôt volontaire par le fournisseur du service Web 2.0, plusieurs applications API permettent aux abonnés de Twitter de télécharger les traces de leurs activités sur Twitter, telles que Tweetnest¹⁵ et Thinkupapp¹⁶ et l'application API de Twitter. Internet Archive, par le biais de son service Archive-It, offre la possibilité d'archiver les *tweets* publics (Internet Archive, 2012). Google propose également une application *Google Replay*¹⁷ pour chercher des *tweets* sur un événement (Twitter blog, 2010).

Le 17 février 2012, Twitter a vendu les droits d'utilisation des données générées par ses utilisateurs à deux sociétés de forage de données (*data mining*), l'une américaine, Gnip Inc., l'autre anglaise, DataSift¹⁸. Cette fois-ci, les données seront utilisées à des fins commerciales et de marketing. Un exemple qui met en lumière la question nébuleuse des droits d'auteur et le droit au respect de la vie privée et des données confidentielles.

3.3 L'archivage de Facebook

Facebook est un réseau social qui a eu beaucoup de succès auprès des internautes. Le nombre total des usagers de Facebook est passé de 500 millions en 2011 à plus de 800

¹⁵ Adresse URL de l'application : <http://pongsocket.com/tweetnest/>

¹⁶ Adresse URL de l'application : <http://thinkupapp.com/>

¹⁷ Adresse URL de l'application: <http://googleblog.blogspot.com/2010/04/replay-it-google-search-across-twitter.html#!/2010/04/replay-it-google-search-across-twitter.html>

¹⁸ Fontaine, P. (02 mars 2012). Twitter a vendu vos tweets aux marketeurs. 01net. Adresse URL : <http://www.01net.com/editorial/559854/twitter-a-vendu-vos-tweets-aux-marketeurs/> (Consulté le 01 avril 2014).

millions en 2012. Sur Facebook, environ 3,5 milliards de contenus sont partagés chaque semaine¹⁹.

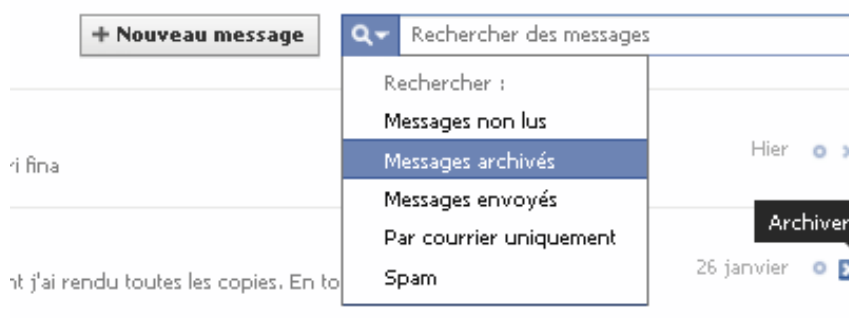
Facebook reste toutefois un réseau fermé (accès avec mot de passe). Une page Facebook d'un utilisateur nous permet d'avoir accès à ses informations personnelles (nom et prénom, date de naissance, adresse, numéro de téléphone, courrier électronique, occupation, passe-temps, etc.), à sa liste d'amis et groupes favoris, aux statuts de mur, informations, vidéos, photos ou documents de tout genre qu'il partage ou qu'il aime (ou n'aime pas), et, enfin, à ses commentaires (ce qu'il pense des publications des autres usagers voire même ses propres publications). Facebook offre à ses usagers, entre autres, un service de messagerie électronique, un moteur de recherche (contacts, messages, etc.), un service de discussion instantanée (tchat) et une profusion d'applications.

Est-il donc techniquement possible de capturer une myriade de contenus et d'applications, aussi variée que différente, qui circule sur Facebook? Est-il faisable de préserver l'aspect visuel et interactif d'une page Facebook et de garantir, en même temps, une collecte exhaustive des contenus et des documents qui y sont publiés ou partagés?

Les résultats des travaux de McCown et Nelson (2009) proposent quelques solutions que nous remettons en question. McCown et Nelson nous présentent tout d'abord les captures qu'elles soient manuelles permettant le téléchargement en local d'une photo, d'un document, etc., ou, automatiques en utilisant des extensions de navigateur qui rendent possible la réalisation de portraits « *snapshots* » de pages et sites Web telle que l'extension Scrapbook de FireFox.

Ils présentent ensuite la technique d'archivage des notifications Facebook par courrier électronique pour préserver les traces individuelles des changements qui affectent une page Facebook. Il est à noter qu'on peut désactiver l'envoi des notifications par courrier électronique et qu'il est possible de rechercher les messages archivés dans Facebook. La figure 5 présente l'interface Facebook qui permet la réalisation de ces opérations.

Figure 5 : Capture d'écran de l'archivage des notifications Facebook

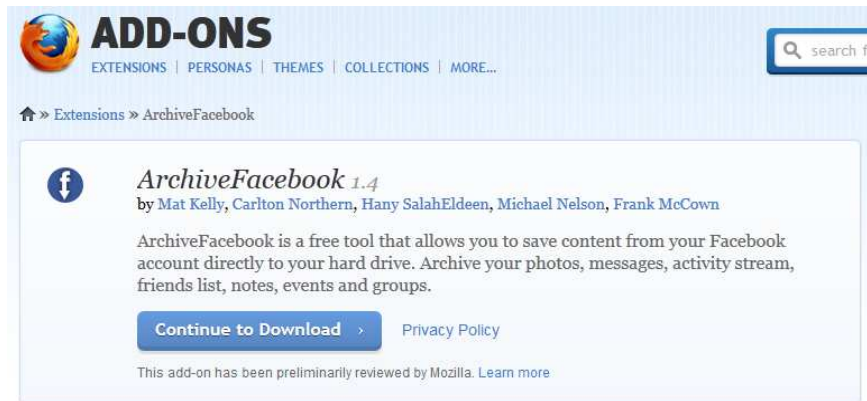


Les deux auteurs suggérèrent également les extensions de navigateurs conçues pour archiver Facebook, par exemple *ArchiveFacebook* de FireFox (voir Figure 6), ou encore, des applications API (Interface de programmation d'applications) développées par Facebook pour extraire des données à partir d'une page Facebook. Nous pouvons donner quelques

¹⁹ DigitalBuzz Blog (2012). Infographic: Social Media Statistics For 2012. Adresse URL: <http://www.digitalbuzzblog.com/social-media-statistics-stats-2012-infographic/> (Consulté le 28 janvier 2012).

applications API offertes par Facebook telles que Friends.get pour avoir la liste d'amis ou photos.get pour télécharger les photos partagées sur une page Facebook. Les auteurs, McCown et Nelson (2009), tentent eux-mêmes de développer Facebook Archiver, un prototype d'application API qui utilise une version modifiée de Scrapbook capable de performer des requêtes AJAX.

Figure 6 : ArchiveFacebook, une extension FireFox pour l'archivage des pages Facebook



Facebook offre à ses usages la possibilité de sauvegarder en local une copie de leur profil Facebook sous forme d'un dossier en format compressé envoyé par messagerie électronique. Pour ce faire, il suffit d'aller sur son profil, cliquer successivement sur « Modifier le profil », « Coordonnées » et « Ajouter/supprimer des adresses électroniques ». Une fois sur cette dernière page, il faut sélectionner « Télécharger une copie de vos données sur Facebook ». Ce lien se trouve complètement en bas de la page.

Cette solution permet de télécharger :

- les photos ou vidéos que vous avez publiées sur Facebook,
- vos publications, vos messages et vos conversations instantanées,
- la liste des contacts (le nom de vos amis) y compris leur adresse électronique,
- vos activités, vos notes, etc.

Sont exclues des archives, toutes informations qui concernent d'autres individus, c'est-à-dire informations personnelles, publications et mises à jour de vos amis de même que les commentaires sur le contenu d'autres individus).

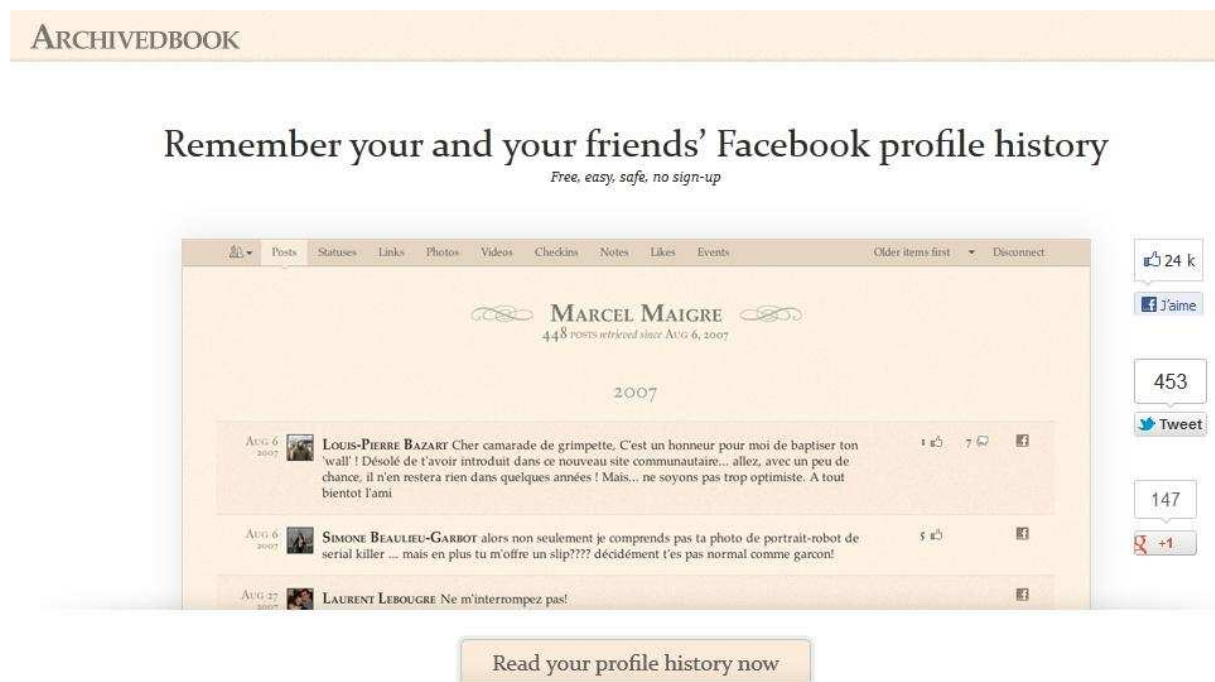
*Archivedbook*²⁰ est un autre exemple d'application API qui permet l'archivage d'un compte Facebook (voir Figure 7). Cette application n'a pas été, par contre, développée par Facebook. Elle permet non seulement de créer une archive en ligne de votre profil Facebook, mais également l'historique des profils de vos amis avant même la date d'entrée en contact sur Facebook. Des options d'affichage plus personnalisées sont également offertes. L'internaute pourrait visualiser ses publications par ordre chronologique croissant ou décroissant,

²⁰ Adresse URL de l'application : <http://www.archivedbook.com>

afficher l'historique du profil d'un ami, sélectionner une composante particulière d'une page Facebook (ex. : affichage des vidéos seulement), etc.²¹.

Malgré le potentiel de cette application, la lenteur d'accès aux données, la perte de données et l'absence d'un moteur de recherche forment les inconvénients majeurs d'*ArchivedBook*.

Figure 7 : Capture d'écran de la page d'accueil d'ArchivedBook



À la liste de solutions proposées par McCown et Nelson (2009) nous pourrions en citer d'autres telles que le service Archive-It d'Internet Archive, service offert avec abonnement, qui, grâce à une syntaxe particulière²², pourrait archiver des pages Facebook. Il est à noter que depuis 2010, Facebook bloque son site en entier aux robots de collecte.

En somme, ces solutions ne nous semblent pas pertinentes puisqu'elles produisent des archives incomplètes, dépourvues de leur aspect interactif, complètement décontextualisées ou qui ne reflètent pas l'expérience vécue par l'internaute. Ces solutions se heurtent également à la nature fermée de Facebook. L'application d'une de ces solutions n'est possible que si l'utilisateur détient les clés d'accès d'un compte Facebook. La capture de profil d'utilisateur à l'insu des détenteurs de comptes Facebook constitue une transgression aux règles de confidentialité et un non-respect des données privées des usagers de Facebook.

²¹ Informations supplémentaires sur ArchivedBook sont disponibles sur le blogue « Blog du modérateur » à l'adresse suivante : <http://www.blogdumoderateur.com/index.php/post/historique-compte-facebook-archivedbook> (Consulté le 28 janvier 2012).

²² La syntaxe d'archivage des pages Facebook d'Archive-it est disponible à l'adresse suivante : <https://webarchive.jira.com/wiki/display/ARIH/Archiving+Social+Networking+Sites+with+Archive-It> (Consulté le 14 avril 2012).

Les solutions présentées sont intéressantes pour un usage individuel sachant que, souvent, des applications et services Web 2.0 sont supplantés par d'autres, vendus à d'autres acteurs, fusionnés avec d'autres services ou disparaissent soudainement.

Discussion et conclusion

Le Web social a été approprié à des fins outre celles pour lesquelles il a été conçu, évolution naturelle de tout usage d'une nouvelle technologie. L'expérience d'usage des applications Web 2.0 et son évolution devraient être documentées, archivées à des fins scientifiques, historiques et culturelles. Les initiatives d'archivage du Web se démultiplient et le nombre de chercheurs dans les collections des archives Web ne cesse d'accroître (Aubry, 2010).

La capture et la pérennisation sont un défi permanent pour les institutions de mémoire. Le Web et les usages du Web ne cessent d'évoluer. Il a été démontré dans ce texte qu'aux difficultés techniques existent des solutions. Et puisque la démarche d'archivage ne peut être que sélective, la définition des critères de sélection et des territoires d'archivage constitue l'étape la plus difficile à franchir.

Le Web, notamment le Web 2.0, confronte le spécialiste d'information à de nouvelles problématiques et l'incite à se questionner sur des notions auparavant évidente dans un environnement documentaire papier : où se situe la limite entre le public et le privé? Quelles sont les frontières d'un document dans un environnement numérique réseauté? À qui revient la responsabilité de conserver le Web? Quand faut-il le faire? Selon quelle démarche? Sur la base de quels critères? Quel est le niveau acceptable de pertes d'intégrité de l'information à capturer? Comment assurer l'accès et l'exploitation de la masse archivée?

Pour pouvoir répondre à ces questions, Kelly et Merieke (2010) proposent une démarche systémique qui repose sur les actions suivantes :

- Définir une politique d'archivage qui explique la valeur de la collection Web, ses fins et utilités;
- Établir des critères d'évaluation et de sélection des ressources à archiver;
- Identifier et implémenter une stratégie d'archivage qui prend en considération les caractéristiques des ressources Web à collecter, les ressources disponibles et la mission de l'organisation;
- Préciser les droits d'auteur afin de minimiser les risques de litiges;
- Documenter les aspects technologiques de la collection Web (ex. configuration requise pour visualiser les archives Web);
- Déclarer les termes d'usage de la collection Web (Kelly et Merieke, 2010; Nash, 2005).

Cette démarche ne focalise pas uniquement sur les aspects techniques. La réussite d'un projet d'archivage du Web et son utilité ne dépendent pas de la performance des robots de collecte, mais de la qualité de la collection qu'il permet de constituer.

À travers l'exemple de la révolution tunisienne, nous avons tenté de décrire, d'abord, l'évolution de l'usage des technologies Web 2.0 à des fins contestataires. Ensuite, nous avons démontré l'importance de documenter et de garder la trace du Web social

événementiel. Nous avons exposé les difficultés techniques d'archivage du Web 2.0 à travers l'exemple des blogues, *tweets* et pages Facebook qui nous renvoient aux propriétés particulières des documents et contenus véhiculés sur Internet. Enfin, nous avons essayé de mettre en lumière les défis réels auxquels se confronte le spécialiste d'information lorsqu'il s'agit de mettre en place un projet d'archivage du Web, soient ceux liés à la sélection de matériaux pertinents pour l'écriture de notre histoire. Nous lui proposons à la fin de ce texte une démarche qui l'oriente dans ses actions.

Références bibliographiques

- Aubry Sara. « The experience of the National Library of France: Introducing web archives as a new library service », *Liber quarterly*, vol. 20, n°2, 2010, p.179–199.
- Ayari Michaël Béchir. « Non, les révolutions tunisienne et égyptienne ne sont pas des «révolutions 2.0» », *Mouvements*, vol. 2, n°66, 2011, p. 56-61.
- Bermès Emmanuelle et Oury Clément. « Web 2.0 et mémoire : De la conversation à la conservation », *Documentaliste-Sciences de l'Information*, vol. 46, n°1, 2009, p. 61-63.
- Boisseau Jean-Baptiste. *Qu'est-ce que le web 2.0 : Modèles de conception et d'affaires pour la prochaine génération de logiciels. Traduction française du texte What is Web 2.0 de Tim O'Reilly. 2005. En ligne] < http://www.eutech-sii.com/index.php?option=com_content&view=article&id=78 >.*
- Bouzouita Kerim. « Les coulisses de la révolution tunisienne : Au cœur de la cyber-guerre », *Géostratégiques-Turbulences maghrébines*, n°32, 2011, p.145 161.
- Campbell, Laura E. et Dulabahn Beth. « Digital preservation : The Twitter archives and NDIIPP », *iPRES 2010, Proceedings from the 7th International Conference on Preservation of Digital Objects*, Vienne, Autriche, 19-24 Septembre 2010, 4 p. En ligne] < <http://www.ifs.tuwien.ac.at/dp/ipres2010/papers/campbell-27.pdf> >.
- Capra Robert, Lee Christopher A., Gary Marchionini, Russell Terrell, Shah Chirag et Stutzman Fred. « Selection and context scoping for digital video collections : An investigation of YouTube and Blogs », *JCDL'08, Proceedings of the 8th ACM/IEEE-CS joint conference on Digital libraries*, Pittsburgh, États-Unis, juin 16-20 2008, p 211-220.
- Carou Alain. « Archiver la vidéo sur le Web » *BBF*, vol. 52, n°2, 2007, p. 56-60.
- Crook Edgar. « Web archiving in a Web 2.0 world », *Dreaming 08 – Australian Library and Information Association Biennial Conference*, Australie, 2-5 Septembre 2008, 9 p.
- Dakhli Leyla. « Une lecture de la révolution tunisienne », *Le Mouvement Social*, vol. 3, n°236, 2011, p. 89-103.
- Davis Richard M. « Practical blog preservation workshop », *Institutional Web Managers' Workshop (IWMW 2009)*, University of Essex, Royaume-Uni, 28-30 juillet 2009.
- Farag Iman. « En marge de la révolution égyptienne : Écrire l'histoire ou sauver la mémoire ? », *Mouvements*, vol. 2, n°66, 2011, p. 41-47.
- Granjon Fabien. « Le Web fait-il les révolutions? », *Nos vies numériques*, n°229, août-septembre 2011. [En ligne] < http://www.scienceshumaines.com/le-web-fait-il-les-revolutions_fr_27535.html >.
- Hank Carolyn, Choemprayong Songphan et Sheble Laura. « Blogger perceptions on digital preservation », *Proceedings of the 7th ACM/IEEE joint conference on digital libraries: building and sustaining the digital environment*. New York: Association for computing

- machiner, 2007, p. 77. [En ligne]
< http://ils.unc.edu/~hcarolyn/jcdl2007_bloggers_poster.pdf >.
- Internet Archive. *Archiving social networking sites with Archive-It*. 2012. [En ligne]
< <https://webarchive.jira.com/wiki/display/ARIH/Archiving+Social+Networking+Sites+with+Archive-It> >.
- Invesp. *How big is blogosphere?* 2011. [En ligne]
< <http://www.invesp.com/blog/business/how-big-is-blogsphere.html> >.
- Kelly Brian et Marieke Guy. « Approaches to archiving professional blogs hosted in the cloud », *iPRES 2010, 7th International Conference on Preservation of Digital Objects*, Vienne, Autriche, 19-24 Septembre 2010. [En ligne] < <http://blogs.ukoln.ac.uk/jisc-bgdp/2010/10/11/case-study-archiving-the-jisc-powr-blog/> >.
- Lecompte Romain. « Cybercontestation tunisienne », *Le magazine culturel en ligne de l'Université de Liège*, décembre 2011. [En ligne]
< <http://culture.ulg.ac.be/jcms/prod/769016/cybercontestation-tunisienne> >.
- Lee Christopher A. et Tibbo Helen. « Capturing the moment : Strategies for selection and collection of web-based resources to document important social phenomena », *Archiving 2008*, Berne, Suisse, 2008, p. 300-305.
- Marshall Catherine C., Bly Sara et Brun-Cottan. Françoise. « The long term fate of our digital belongings : Toward a service model for personal archives », *Proceedings of IS&T Archiving*, Ottawa, 23-26 mai 2006, p. 25-30.
- McCown Frank et Nelson Michael L. « What happens when Facebook is gone? », *JCDL'09, Proceedings of the 9th ACM/IEEE-CS joint conference on Digital libraries* Austin, Texas, USA, 14–19 juin 2009, p. 251-254.
- Mihoub Samia. « Le cyberactivisme à l'heure de la révolution tunisienne », *Archivo antropologico mediterraneo*, vol. 13, n°2, 2011, p. 17-31.
- Nash Michael. « L'archivage des sites web d'organisations ouvrières et de mouvements de gauche : État des lieux à la bibliothèque Tamiment de la New York University », *Matériaux pour l'histoire de notre temps*, n°79, juillet-septembre 2005, p. 87-91.
- Office québécois de la langue française (2002-2012). *Le grand dictionnaire terminologique*. 2002-2012. [En ligne] < <http://www.oqlf.gouv.qc.ca/ressources/gdt.html> >.
- O'Reilly Tim. *What is Web 2.0 : design patterns and business models for the next generation of software*. 2005. [En ligne]
< <http://www.oreillynet.com/pub/a/oreilly/tim/news/2005/09/30/what-is-web-20.html> >.
- Paré François-Xavier et Chebbi Aïda. *Intégration des outils du Web 2.0 dans les pratiques d'enseignement*. 2009. [En ligne]
< http://www.ebsi.umontreal.ca/clip/web2_accueil.html >.
- Quan Huangmao. *Online social networks & social network services : A Technical survey. Pervasive Communication Handbook*. CRC Press, 2011. [En ligne]
< <http://lucas.cis.temple.edu/~hquan/pubs/snsurvey.pdf> >.

- Rauber Andreas, Kaiser Max et Wachter Bernhard. « Ethical issues in Web archive creation and usage – Towards a research agenda », *IWAW08, 8th International Web Archiving Workshop*, Aarhus, Denmark, 18-19 septembre 2008, 7 p.
- Raymond Matt. *Twitter donates entire tweet archive to Library of Congress*. 2010. [En ligne] < <http://www.loc.gov/today/pr/2010/10-081.html> >.
- Richet Jean-Loup. « Internet et usages : Ce que la Libye a appris de la révolution en Égypte », *Chroniques : Moyen-Orient et Afrique du nord*, avril 2011. [En ligne] < http://www.dandurand.ugam.ca/uploads/files/publications/rflexions/Chronique_O MAN/Richet Internet MO260411.pdf >.
- Stenger Thomas et Coutant Alexandre. « Les réseaux sociaux numériques : Des discours de promotion à la définition d'un objet et d'une méthodologie de recherche », *Journal of language and communication studies*, n°44, 2010, p. 209-228.
- Technorati. *State of the Blogosphere 2011 : Introduction and Methodology*. 2011. [En ligne] < <http://technorati.com/social-media/article/state-of-the-blogosphere-2011-introduction/> >.
- Twitter. *Politique de confidentialité de Twitter*. 2011. [En ligne] < <https://twitter.com/privacy> >.
- Twitter Blog. *Tweet preservation*. 2010. [En ligne] < <http://blog.twitter.com/2010/04/tweet-preservation.html> >.