Archives sur Internet: attention, sites mouvants!

Autor(en): **Giussani, Bruno**

Objekttyp: **Article**

Zeitschrift: **Horizons : le magazine suisse de la recherche scientifique**

Band (Jahr): **- (1998)**

Heft 36

PDF erstellt am: **13.05.2024**

Persistenter Link: https://doi.org/10.5169/seals-556027

Nutzungsbedingungen

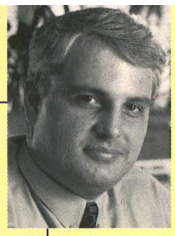
Die ETH-Bibliothek ist Anbieterin der digitalisierten Zeitschriften. Sie besitzt keine Urheberrechte an den Inhalten der Zeitschriften. Die Rechte liegen in der Regel bei den Herausgebern.
Die auf der Plattform e-periodica veröffentlichten Dokumente stehen für nicht-kommerzielle Zwecke in Lehre und Forschung sowie für die private Nutzung frei zur Verfügung. Einzelne Dateien oder Ausdrucke aus diesem Angebot können zusammen mit diesen Nutzungsbedingungen und den korrekten Herkunftsbezeichnungen weitergegeben werden.
Das Veröffentlichen von Bildern in Print- und Online-Publikationen ist nur mit vorheriger Genehmigung der Rechteinhaber erlaubt. Die systematische Speicherung von Teilen des elektronischen Angebots auf anderen Servern bedarf ebenfalls des schriftlichen Einverständnisses der Rechteinhaber.

Haftungsausschluss

Alle Angaben erfolgen ohne Gewähr für Vollständigkeit oder Richtigkeit. Es wird keine Haftung übernommen für Schäden durch die Verwendung von Informationen aus diesem Online-Angebot oder durch das Fehlen von Informationen. Dies gilt auch für Inhalte Dritter, die über dieses Angebot zugänglich sind.

Ein Dienst der *ETH-Bibliothek*
ETH Zürich, Rämistrasse 101, 8092 Zürich, Schweiz, www.library.ethz.ch

http://www.e-periodica.ch



Bruno Giussani (bg@giussani.com) est journaliste, spécialisé dans les nouveaux médias, et chroniqueur Internet du «New York Times» et de la «Neue Zürcher Zeitung».

Archives sur Internet: attention, sites mouvants!

Une partie des sources de la recherche scientifique sont menacées par une phrase que les utilisateurs du réseau internet connaissent fort bien: «404 document not found.» Il s'agit du message d'erreur qui s'affiche sur l'écran de l'ordinateur lorsqu'une information n'est plus disponible sur le web, la portion multimédia de l'internet. Cela peut signifier que le document en question a été déplacé ailleurs dans le réseau, dans un autre ordinateur. Le plus souvent toutefois, «404» indique que l'information a été effacée. Qu'elle est probablement perdue.

L'internet est un médium en constant renouvellement. On estime que la durée moyenne d'un document sur le réseau est de 75 jours. Des milliers de pages web sont mises à jour toutes les heures, ou même toutes les quelques minutes.

Cette nature fluide donne à l'internet sa vitalité et sa fraîcheur. Elle n'est toutefois pas sans conséquences pour la recherche, notamment en sciences sociales (mais pas uniquement).

Depuis un an et demi, Brewster Kahle, un Américain de 37 ans, dirige à San Francisco une organisation appelée Internet Archive **(www.archive.org)** dont l'activité consiste, comme le dit son nom, à figer le renouvellement continuel du réseau en archivant l'information publiquement accessible. «Ce qui se passe sur le réseau et dans le domaine des technologies numériques est tout simplement de la culture contemporaine», dit-il. Non seulement les textes, les images et les sons du web, mais également les documents disponibles dans les autres portions publiques de l'internet (pas de panique: votre boîte aux lettres électronique ne sera pas touchée), ainsi que les logiciels.

L'initiative a le mérite de soulever la question de l'urgence de la sauvegarde d'information et de documents qui, de plus en plus souvent, sont exclusifs à l'internet, c'est-à-dire qui n'ont pas d'«édition papier». Comme on a conservé les enregistrements vidéo des messages publicitaires politiques et des débats transmis durant la campagne présidentielle américaine de 1996, par exemple, l'Internet Archive a également pris soin de faire une copie complète des sites web de tous les candidats. Pour l'historien, ou le politologue, il est évident qu'il s'agit de deux sources également précieuses.

L'archivage de l'information disponible sur l'internet pose tout d'abord des problèmes techniques. Le réseau est immense, multiforme, et continue de croître. Son contenu n'est pas structuré. Kahle reconnaît qu'il lui sera impossible d'enregistrer tous les changements, mais il espère être en mesure de prendre au moins une première «photo» de l'état actuel de l'internet, pour ensuite la mettre à jour à intervalles réguliers.

Ensuite, si nous pouvons lire aujourd'hui sans difficulté un livre imprimé par Gutenberg, cela n'est souvent pas le cas pour une disquette informatique vieille de dix ans. La conversion régulière des fichiers dans de nouveaux formats devrait résoudre le problème en partie. La sauvegarde des logiciels qui permettent de lire ou interpréter ces informations, ainsi que la conservation ou la simulation des machines sur lesquelles ils fonctionnent, paraît également primordiale.

Avant de pouvoir ouvrir ses archives aux chercheurs et, éventuellement, au public, Brewster Kahle doit toutefois résoudre aussi toute une série de questions juridiques sur la protection de la sphère privée et le droit d'auteur, comme copier une page web ne signifie-t-il pas contrevenir au droit d'auteur? (Oui).

En attendant de pouvoir formuler des réponses, Kahle continue d'archiver – pour éviter que ces «traces précieuses de l'aventure humaine» ne soient perdues à jamais.

B.G.