

Bibliothèque universelle et culture de l'éphémère

Sergei SOLOVIEV¹

Je sentais de plus en plus le besoin de rassembler et de conserver les volumes anciens, de charger les scribes consciencieux d'en tirer les copies nouvelles. Cette belle tâche ne me semblait pas moins urgente que l'aide aux vétérans ou les subsides aux familles prolifiques et pauvres ; je me disais qu'il suffit de quelques guerres, de la misère qui suit celles-ci, de la période de grossièreté ou de sauvagerie sous quelques mauvais princes, pour que périsse et jamais les pensées venues jusqu'à nous à l'aide de ces frêles objets de fibres et d'encre.

M. YOURCENAR,
Mémoires d'Hadrien

Si la publication électronique ne cesse d'être valorisée par les producteurs de logiciels, les journalistes, les hommes politiques, et si les tenants de l'imprimé (éditeurs, bibliothécaires, et de nombreux universitaires) s'opposent à ce type de publication sous prétexte qu'elle n'a pas de légitimité, force est de constater le silence presque absolu autour d'un autre débat : celui de la pérennité de l'information.

Autant les partisans du tout électronique n'évoquent jamais la brève durée de vie des formats de fichiers utilisés pour de telles publications, *off line* ou *on line*, autant leurs détracteurs essayent rarement d'insister sur ce caractère éphémère de l'archive ; pourtant, leurs arguments pourraient alors avoir beaucoup plus de valeur.

Une fois de plus, on vit une situation paradoxale, considérablement amplifiée par le développement de l'Internet.

1. IRTT, université Paul-Sabatier, Toulouse, équipe « Réseaux, Savoires & Territoires ».
soloviev@irtt.fr

Tout d'abord, on assiste à une multiplication de l'information numérisée, qui nécessite des médiations autrement plus considérables que celles imposées par l'imprimé : on ne peut pas accéder à l'information numérisée sans passer par des supports matériels et logiciels ; or les uns comme les autres ont une durée de vie limitée, qui, sauf exception, ne dépasse pas dix ans : une disquette 5"1/4 est aujourd'hui aussi peu lisible qu'une disquette standard de 400 Ko, et plus personne ne se rappelle l'existence de disquettes 8". Les textes écrits aux formats Word 3 ou MacWrite pour Macintosh, indépendamment du support, sont eux aussi considérés comme illisibles, puisque ces logiciels ne fonctionnent plus avec les systèmes d'exploitation actuels, et puisque les successeurs de ces logiciels ne sont pas capables — ou même autorisés (souvent par la volonté des éditeurs exprimée dans la loi du copyright) d'interpréter ces anciennes versions.

Si, comme l'explique D. Norman², cette instabilité des supports comme des logiciels témoigne de l'« adolescence » de l'industrie informatique, force est de constater que les monopoles en voie de constitution, mais toujours fragiles, font tout pour rendre périssables les formats de fichiers ou les protocoles d'archivages qui ont plus de deux ans d'âge. Cette surenchère technologique va jusqu'à rendre inutilisable un ordinateur de plus de quatre ans, et contribue encore plus à cette absence de pérennité de l'information.

Enfin, l'Internet contribue de façon considérable à la mise en forme de l'information sous forme électronique, et à son échange, tout comme ses protocoles relancent le débat sur les formats de fichiers : les internautes s'échangent des documents « attachés », qui sont des transcriptions ou des inédits, mais qui utilisent des formats normatifs : ou propriétaires (donc instables), ou libres (un peu plus stables) ; mais la démocratie, au sens du règlement d'une question après débat entre toutes les personnes concernées, perd ses droits au profit de la statistique primaire : si, aujourd'hui, 90% de mes interlocuteurs utilisent Word XY pour rédiger des documents, et si moins de 10% d'entre eux savent sauvegarder un fichier au format HTML pour le lire avec un navigateur, je vais acheter le logiciel Word XY, et l'an prochain, le logiciel Word YZ, afin de limiter mon « coût de communication » avec mes collègues, même si je fais partie de l'infime minorité qui utilise le seul système éditorial relativement stable depuis plus de 15 ans, à savoir LaTeX. Ainsi, les « normalisations », promues par les pères fondateurs de l'Internet, attachées à la notion de format universel, comme à celle de « compatibilité ascendante », semblent être condamnées dès leur publication, puisqu'incompatibles avec le besoin des entreprises de valoriser des formats propriétaires, et sans cesse évolutifs, dans le seul but de s'accorder une rente à vie.

2. *The Invisible Computer*, Cambridge, MA, MIT Press, 1998.

OÙ EST DONC LE PARADOXE ?

Ce dernier est double : alors que les experts en sciences humaines, souvent attachés à l'imprimé, n'insistent que sur l'absence de validation par les pairs sur l'Internet, alors que la plupart des érudits, toutes disciplines confondues, s'accordent pour estimer à au moins 30% le taux d'articles publiés dans des revues imprimées de qualité, faux ou mensongers (volontairement ou involontairement), on ne peut qu'être impressionné par le manque de réaction des uns et des autres devant ce caractère éphémère de l'information.

Autrement dit, les experts qui passent des années à distinguer un ouvrage de 1545 publié à Paris de sa version « piratée » publiée à Genève deux ans après, ou qui travaillent sur les manuscrits des auteurs incontournables des XIX^e et XX^e siècles, semblent incapables de s'inquiéter, non seulement de la genèse des textes des futures célébrités de la littérature française de la fin du XX^e siècle, mais aussi de la relecture électronique de leurs propres publications. D'autre part, ces champions de la culture, et de sa transmission, font fi de son statut actuellement éphémère, donc promu à une mort certaine, alors qu'on aurait pu imaginer qu'ils choisiraient la préservation de toutes les formes, passées comme actuelles, de cette culture. On pourrait alors se demander si nos érudits sont des défenseurs de la « culture », comme ils le proclament si souvent dans les médias, ou des rentiers de la culture passée, se désintéressant complètement de la culture présente, qui ne sera valorisable que pour les universitaires des générations futures.

D'où la seconde face du paradoxe : les premiers à défendre cette notion de culture accessible dans les décennies à venir sont les informaticiens, et les seuls à témoigner du débat sont les journalistes. On doit sans doute mentionner ici l'article d'H. Fischer, président de la Fédération internationale des associations de multimédia³. On doit aussi mentionner quelques pages consacrées au débat par le même journal à l'été 2000 (22-22 juillet), des contributions comme celle de J.-P. Smets à l'occasion du colloque « Comprendre les usages de l'Internet » (« Les brevets pour nuire à la société d'information »), et l'article « Le droit et la mémoire à l'ère numérique », que ce dernier a rédigé avec A.-L. Sibony dans *Le Monde* du 14 septembre 2000. Toutes ces publications, ainsi que la version préliminaire de cet article, s'accordent à rappeler les dangers de l'éphémère dans la publication électronique.

Le problème est que le débat (comme le constat des dangers) ne suffit pas, l'enjeu est grand, et l'action est nécessaire.

On ne s'inquiéterait pas de la tournure des événements si elle n'était que localement située dans le temps, et indépendante des choix socio-politiques. Or, cette évolution des formats de fichiers et des supports oblige les archivistes

3. « Le paradoxe du numérique et de l'oubli », *Libération* du 24 novembre 1999.

des temps modernes, qu'ils soient, comme on l'a vu bibliothécaires, ou plus fréquemment, informaticiens sensibilisés au problème, à réaliser systématiquement des conversions de documents, d'anciens formats, oubliés, en de nouveaux, actuels mais périssables. Il est clair que dans une telle situation, souvent urgente autant que sous-estimée (personne n'est payé pour cela), la majorité des documents électroniques n'est pas convertie dans les formats d'actualité. Aussi, c'est la littérature la plus consultée-consommée, et non pas la plus importante à long terme, qui est mise à jour. Dans cette logique de consommation, qui vaut autant pour les textes littéraires, actuels ou retrancris, que pour les articles scientifiques les plus pointus, il est clair que les pays et les langues s'engagent dans un combat inéquitable : un texte perse de l'Antiquité, un théorème publié dans une revue russe, n'ont aucune chance face aux délires de Bill Gates sur l'avenir électronique de l'humanité, garanti par cette notion bizarrement démocratique de « portefeuille électronique pour enfants » géré grâce à la double attention des parents et de Windows 2000 !

De la part de la communauté scientifique, l'action scientifique (c'est-à-dire, l'effort de compréhension et de la recherche) n'est pas moins nécessaire.

On peut dresser la longue liste des axes qui restent inexplorés. On cherche des solutions — mais presque à l'aveuglette. On propose des solutions qui sont tout de suite perverties en pratique. On peut évoquer de nouvelles normes appliquées à l'hypertexte, comme le standard SML et son satellite XSL. Leur but était de terrasser les risques de volatilité de l'information numérisée évoqués plus haut. Mais déjà, avec les logiciels propriétaires implémentant XSL, les risques réapparaissent (par exemple, les applications de LotusXSL et de Sablatron donnent des résultats incompatibles).

Face à de tels projets, dont la réussite est statistiquement peu garantie, ne vaudrait-il pas mieux penser dès maintenant la perte et la dégradation d'informations, sous plusieurs formes, *a priori* éloignées (culturellement) mais dans les faits très proches ?

D'une part, en créant les conditions d'une sauvegarde muséologique de ce qui s'écrit sur le Web. Diverses universités américaines, souvent pour des raisons purement « théoriques » (topologie du Web, etc.) s'y lancent déjà. Peut-être même faut-il intégrer les conditions d'une sauvegarde des logiciels, des systèmes d'exploitation, et des ordinateurs du passé. Il y a fort à parier que si l'Europe ne s'y met pas, dans trente ans, nos historiens seront condamnés à payer tribut à ces universités pour avoir accès à leurs archives.

D'autre part, en développant des modèles mathématiques qui permettent non seulement d'étudier les problèmes de gestion d'information dans les archives électroniques, mais aussi les points faibles de la réalité éditoriale actuelle (analyse logique de documents, vérification des liens, etc.), et modéliser son développement futur (par exemple, à base des réseaux d'automates ou de théorie des tenseurs).

Enfin, en donnant plus de poids économique à la pérennisation de l'information, par la mise en évidence du problème de l'assurance d'information.

Restera bien sûr à confronter de tels modèles avec la réalité sociale, fortement mise à contribution actuellement avec les nouvelles règles économiques et juridiques qui se mettent en place, et qui correspondent plus à une logique de modernisation coercitive que de rationalisme scientifique⁴.

4. Je tiens à remercier Louis Féraud, Éric Guichard et Clarisse Herrenschnidt pour nos discussions constructives et leur aide à la mise en forme de la version française de ce texte.