



19 / 20 / 21 novembre 08

Cinéma et audiovisuel :  
quelles mémoires numériques pour l'Europe ?

## DES PROJETS EN EUROPE, DES PROJETS EUROPÉENS ?

### Peut-on confier la mémoire aux machines ?

**Claude MUSSOU,**

*Chef de projet Dépôt légal du web, Ina.*

Contrairement à ce que peut laisser penser l'intitulé de la session, cette intervention ne s'inscrit pas dans une dynamique de projet européen, mais davantage dans la problématique d'une posture européenne et publique en matière de conservation des patrimoines et notamment du Web.

Quelles mémoires numériques pour l'Europe ? À cette question qui donne son titre à l'Édition 2008 de la conférence, Jean-Michel Rodes, directeur du dépôt légal à l'Ina, que je remplace aujourd'hui, répond par une autre question : peut-on confier la mémoire aux machines ? Ou plutôt, à quelles machines peut-on confier notre mémoire ?

Il a formulé cette question avant que *Le Monde 2* en date du 15 novembre 2008 ne titre « Peut-on tout confier à Google ? »

C'est le résultat d'une analyse et d'une réflexion menées à partir de l'étude des évolutions de l'informatique et de ses composantes au cours des 40 dernières années. Aujourd'hui, dans le secteur du patrimoine audiovisuel, il est difficile de parler de mémoire sans évoquer les dimensions d'outil et techniques de l'informatique.

Nous retracerons brièvement la petite histoire de cette évolution pour mieux souligner les enjeux de mémoire patrimoniale qui se jouent aujourd'hui, alors que l'information, les contenus numériques circulent sur les réseaux et sont accessibles depuis une grande diversité de plates-formes. Voilà pour cette petite histoire de ce que Jean-Michel appelle la galaxie informatique : ce sont donc les différentes composantes de l'informatique qui ont structuré et jalonné l'économie et les usages de l'informatique ; à savoir, les processeurs pour la puissance des calculs, la mémoire pour le stockage des données et les réseaux pour leur circulation. Selon l'état d'avancement de tels ou tels de ces composants, la géométrie de cette galaxie s'est soit contractée, soit dilatée, soit centralisée, soit distribuée.

Dans les années 1960-1970, les ordinateurs coûtent cher pour ce qui concerne les composantes, à savoir les processeurs, les disques durs, mais aussi pour ce qui relève de tout ce qui est programmation et administration des systèmes. En revanche, les données sont peu volumineuses et elles peuvent circuler à moindres frais sur des réseaux téléphoniques à bas débit. On compte à l'époque en bauds/seconde et non pas en bits/seconde. Les entreprises externalisent leur informatique dans de grands centres serveurs et offrent ce que l'on appelle alors les Services Bureau. Au sein d'une même entreprise, les ordinateurs sont de simples terminaux relayés à un serveur central.

À la fin des années 1970, cette architecture se trouve inversée par l'arrivée sur le marché des PC. Les coûts des processeurs et des disques durs ont baissé. Les périphériques, les supports de stockage et d'échange des données (CD, disquettes) sont facilement manipulables, et le marché grand public se développe. Les entreprises ré-internalisent leurs machines. Les logiciels sont distribués sur des supports et l'architecture informatique est donc décentralisée, très distribuée, et très peu communicante.

Au tournant des années 2000, ce cycle d'expansion/contraction de l'architecture est une fois encore inversé en raison du déploiement d'Internet. La normalisation des protocoles d'échanges, l'accroissement rapide des débits favorisent le développement du réseau mondial des réseaux. Du fameux modem 1 500 bauds/seconde, on passe à des connexions ADSL de plusieurs mégabits pour les particuliers. Dans les entreprises, les débits des réseaux sont plus que décuplés. Les entrées et sorties des données via l'Internet se généralisent.

Le développement d'Internet a entraîné l'interconnexion et la circulation d'informations entre des machines qui sont dispersées. L'architecture décentralisée s'est recentralisée, mais cette fois, au point d'entraîner une concentration quasi-monopolistique, à l'échelle mondiale, des entreprises de services des internautes qui ont installé un gigantesque maillage d'entrepôts de serveurs de stockage aux puissances de calcul phénoménales, et que l'on appelle des *Data Center* ou même des fermes de serveurs.

Cette architecture, qui permet l'accès à des millions d'octets de données stockées sur des serveurs disséminés sur toute la surface du globe, est désignée sous le vocable poétique de *Cloud Computing* ou informatique en nuages. Voilà pour ce qui est des infrastructures.

Qu'en est-il des données et des contenus ? Le mouvement et la circulation des contenus épousent les architectures que nous avons précédemment décrites. Elles sont stockées en local et non partagées quand les infrastructures sont décentralisées.

Avec la généralisation des usages d'Internet, le développement des données produites et diffusées est exponentiel. On parle aujourd'hui de deux milliards de photos mises en ligne sur *Flickr*, d'autant de boîtes de courriels ouvertes pour la seule année 2008, de 100 millions de clips vidéo disponibles sur YouTube. La majeure partie de ces données sont des données personnelles. La polémique a fait florès autour de la disponibilité de ce que le journaliste américain John Battle appelle « une base de données d'intentions humaines ».

Internet se nourrissant d'Internet, les données et les contenus publiés, qu'ils soient ou non privés, sont collectés, indexés, archivés et centralisés par les moteurs de recherche, Google au premier chef. Ce sont de nouveaux gisements de données, produits par les internautes, et de plus en plus par les entreprises, qui sont donc rendus disponibles sur des applications centralisées. Concernant les contenus qui nous intéressent ici aujourd'hui, la consultation en ligne, à la demande ou non, permet de nouveaux usages de communication et de consommation de la télévision, de la vidéo, du cinéma, de la musique, qui participent de la création et de l'alimentation de ces gigantesques serveurs. Les majors du Web rivalisent pour numériser les fonds d'archives et de bibliothèques afin de les rendre accessibles en ligne.

Je cite Jean-Michel : « *Ces évolutions technologiques et économiques successives ont ouvert la voie à une appropriation mondiale privée et oligopolistique des capacités de publication et de conservation des patrimoines* ». Je ne crois pas qu'il s'agisse ici, pour Jean-Michel Rodes, de jouer les Cassandre, et d'agiter le drapeau rouge de l'antiaméricanisme. Au même titre que grand nombre d'entre nous, il est tout prêt à se féliciter de la satisfaction que l'on peut avoir à accéder à tous types de documents, d'informations, très vite et partout, grâce à l'efficacité et aux performances en termes de vitesse et de continuité de service de ces géants de l'industrie de l'information.

Néanmoins, la prudence est de mise, puisque cette nouvelle architecture mondiale de l'informatique est récente. Ces entreprises ont tout juste 10 ans. Un retour de situation est donc tout à fait envisageable. Par exemple, prenons Google à nouveau : il gère 900 000 serveurs, tous connectés en réseau et soumis à des dizaines de millions de requêtes simultanées. Cette taille peut s'avérer être

un important facteur de fragilité technique qui va s'accroître au fur et à mesure que l'architecture informatique de l'entreprise va continuer de se développer. À cette fragilité technique s'ajoutent les incertitudes sur la pérennité d'une bonne santé économique de l'entreprise. Le modèle actuel de Google, financé exclusivement par la publicité, peut se trouver mis à mal dans le contexte de crise que nous traversons aujourd'hui.

Dans *Le Monde 2* cité précédemment, Bernard Benhamou, délégué interministériel aux usages de l'Internet affirme : « *Les prochains mois diront si son modèle économique protège l'entreprise d'une crise d'ampleur* ». Il précise : « *Qu'il serait par trop hasardeux de se reposer sur Google, en particulier pour ce qui touche le patrimoine culturel* ».

Ces entreprises ne sont pas philanthropes, même si la légende affirme en général que tout a commencé dans un garage et pour le plus grand bien de l'humanité. C'est sans doute la motivation initiale, mais le marché réclame du profit. Dans ce contexte, rien ne pourrait s'opposer au pillage ou à la destruction des données, si les critères de rentabilité ne sont pas atteints. On se souvient que la cause la plus répandue de la disparition des films muets fut leur destruction intentionnelle au moment de l'arrivée du cinéma parlant. Robert HARRIS, un conservateur qui a témoigné devant une commission du congrès, le dit : « *La plupart des films n'a pas survécu à cause de la vente en masse par les studios. Il n'y avait aucune raison de sauver ces films, ils avaient simplement besoin d'une chambre forte, mais le matériel était cher à l'époque* ».

De fait, ces nouveaux dispositifs de stockage et de diffusion n'offrent d'assurance, ni en matière de garantie pour la pérennité du patrimoine, ni en termes de protection de l'intégrité des données privées ou collectives. Les moteurs de recherche du Web ne constituent pas, loin de là, des systèmes stables d'archivage de l'information. Des sites Web naissent et meurent chaque jour. Les blogs s'effacent des serveurs selon la logique du *first-in/first-out*. Les coûts et la rentabilité économique des stockages, de l'indexation, et de la communication des données, prévalent pour leur accessibilité.

Dans ce contexte de profusion des informations publiées et de mainmise industrielle sur le stockage de données, la maîtrise de la mémoire pour une collectivité devrait, selon Jean-Michel, faire partie des libertés fondamentales. La Révolution française s'est accompagnée d'ailleurs d'une importante institutionnalisation des politiques de mémoire avec la création de centres d'archives, de bibliothèques, de musées. Je précise à nouveau qu'il ne s'agit pas ici de prêcher pour que l'effort public se substitue à une initiative privée dont les qualités et les atouts ont contribué aux spectaculaires avancées dans la diffusion des connaissances, la mise en place de nouveaux territoires de sociabilité et de savoirs. Cette mémoire est volatile et flottante. Il faut garder à l'esprit le fait qu'il a fallu des siècles de savoirs accumulés dans les bibliothèques, dans les archives publiques, pour que le marché trouve, à ces pratiques, une certaine solvabilité. Ces politiques patrimoniales au long cours ont permis la numérisation des fonds.

Pour conclure, je vais citer Jean-Michel Rodes, *in extenso*. Cette conclusion est, en quelque sorte, une profession de foi : « *Il est donc nécessaire que l'État, les États maintiennent cette fonction de dernier recours, de garants de mémoire collective, de garants de liberté individuelle, et protecteurs de la vie privée. Au sens noble, une action de police et de civilisation de la mémoire et de ses usages. Dans le domaine du livre, du disque, du cinéma, de la radio, de la télévision, cette activité de dernier recours, de garde-fou, s'appelle le dépôt légal. Il faut le développer également, et à son juste niveau pour le Web comme une garantie démocratique et citoyenne essentielle. C'est la condition d'exercice des autres droits et libertés. C'est d'autant plus urgent que l'on commence à mieux connaître les turbulences et les mouvements erratiques de ces nuages planétaires* ».

Le dépôt légal du Web, c'est une mission que la BnF et l'Ina ont désormais en partage et qu'un certain nombre d'institutions en Europe et de par le monde ont également entrepris de mettre en place. Je vous remercie.

\*\*\*\*\*

Suivi éditorial : Loraine Pereira – chargée de mission pour le patrimoine cinématographique / INP.