



19 / 20 / 21 novembre 08

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

CHANGEMENTS DANS LES PÉRIMÈTRES DE LA CULTURE

Nouvelles questions, nouveaux objets du dépôt légal : le cas des œuvres audiovisuelles sur le Web.

Alain CAROU,
BnF.

Au cours des dernières années, on a mis le doigt sur la manière dont le numérique renouvelle les problématiques d'accès aux archives audiovisuelles constituées, aux fonds d'images déjà là, et dont les usages sont transformés par le passage de l'analogique au numérique. Les journées Archimages, depuis plusieurs années, en sont le reflet et la traduction.

Nous avons encore à prendre la mesure, c'est plus neuf, d'un autre bouleversement, celui qui a commencé à toucher la collecte et donc la constitution même des collections. Les principes qui y président se voient soumis à un travail de réexamen.

La production et la diffusion d'images animées en croissance exponentielle devenues proliférantes effrayent souvent la communauté des archivistes. Le problème doit être affronté et je crois que cela a été fait sous des angles différents au cours de précédentes interventions lors de ces journées 2008.

Je prendrai comme premier exemple la croissance très rapide du nombre de vidéos proposées chaque année aux grands festivals. Dans tous les festivals, le nombre de films candidats est chaque année en augmentation exponentielle. Que sauvons-nous de tout cela ? Autre exemple (l'exemple par excellence ?) : des dizaines de milliers de nouvelles vidéos sont mises quotidiennement sur une plateforme comme *YouTube*. Que fait-on de tout cela ?

Il y a peut-être lieu de reposer à la base la question des objets de la collecte : **que voulons-nous collecter à présent ?**

Permettez-moi d'envisager la question selon une approche nécessairement particulière qui n'est ni celle qu'aurait, par exemple, une archive axée sur un domaine particulier du social, ou qu'aurait une archive axée sur l'identité d'une région. Notre optique à la Bibliothèque nationale de France est celle du dépôt légal.

Permettez-moi un bref rappel de ce que nous sommes, de ce qu'est le dépôt légal comme donnée constitutive de l'identité de l'établissement.

Très schématiquement, les origines du dépôt légal sont historiquement liées à l'invention de l'imprimerie et du coup à l'émergence de la notion de publication. Le principe du dépôt légal appliqué

à la notion de publication s'est étendu au fil du temps aux estampes, aux cartes et plans, à la photographie, etc., et de manière systématique, dès 1925, aux médias non-livres. Dans le principe, c'était quelque chose d'assez révolutionnaire. Vous savez qu'en même temps, ce principe va mettre un certain temps, voire énormément de temps, à entrer dans les faits.

La logique d'extension du dépôt légal aux médias émergents conduit, en 1977, à l'institution du dépôt légal des vidéogrammes. La Bibliothèque nationale de France assure la collecte, le catalogage et la conservation des vidéogrammes mis à disposition du public en France depuis quasiment cette époque.

Les vidéogrammes en question sont toutes les formes de publications audiovisuelles qui passent par d'autres canaux de diffusion que la projection de pellicule en salle ou que la télévision : l'édition vidéo commerciale en VHS et DVD, mais aussi l'édition non commerciale, pédagogique, la vidéo associative et militante, la production de communication des entreprises, des ministères, des collectivités locales, etc. Un domaine nettement plus étendu que l'édition vidéographique à laquelle on nous réduit parfois par simplification. Tout vidéogramme porté à la connaissance d'un public en France est assujéti au dépôt légal. Nous recevons quotidiennement des vidéos diffusées par le biais de représentations publiques de tout type et, pas seulement par la mise en circulation de supports produits en nombre, que ce soit des vidéos destinées à être diffusées dans des lieux de patrimoine, dans des assemblées générales d'associations d'organisations non gouvernementales, dans des commerces, etc... Aujourd'hui, nous conservons une collection de 170 000 vidéogrammes et nous recevons à peu près 10 000 vidéogrammes nouveaux par an.

Aujourd'hui, la notion de publication, fondatrice du dépôt légal, déborde largement les frontières des formes de diffusion traditionnelles, soumises au filtre d'intermédiaires (producteurs, diffuseurs). Ces formes (l'exploitation cinématographique, la diffusion télévisuelle, l'édition vidéo, la communication institutionnelle) sont encore dominantes, mais elles ne sont plus hégémoniques. Nous devons nous organiser pour faire face au risque de perdre un grand nombre d'œuvres audiovisuelles qui n'ont de fenêtres de visibilité que dans des festivals, dans des rassemblements militants ou dans des contextes locaux encore plus informels comme le bistrot du coin ou sur le Web.

Je ne dirais pas que ces réalités sont totalement nouvelles pour nous. La vidéo a signifié dès ces débuts, dans les années 1970, la possibilité d'ouvrir des espaces informels d'expression audiovisuelle. À côté de l'édition vidéo, nous avons déjà travaillé depuis de nombreuses années à collecter la vidéo militante ou les « télé-brouettes », ces journaux vidéo locaux, sur cassette vidéo, qui diffusaient de l'information à un niveau local quand l'Internet ou des solutions télévision numérique n'existaient pas. La nouveauté est la technologie employée, ainsi que l'échelle et les volumes concernés qui ont totalement explosé.

Les vidéogrammes sur le Web sont un des grands cas à traiter.

Depuis deux ans, nous avons beaucoup avancé sur le dépôt légal des vidéos sur Internet. C'est une extension de nos missions, au départ un peu intimidante, d'une part à cause de la masse de vidéos en ligne et du rythme de croissance de cette masse absolument colossale, et d'autre part à cause du caractère problématique de ce travail au regard des notions qui sont traditionnellement associées au dépôt légal et qui en délimitent les frontières :

- premièrement, la fameuse notion de publication définissant le dépôt légal : tout un chacun peut publier tout ce qu'il veut sur le Web, plus de filtrage.
- deuxièmement, la notion de territoire. Que veut dire diffuser en France puisque tout ce qui est publié sur le Web est consultable de la même manière dans le monde entier ?

Cette problématique, néanmoins, remarquons-le, ne nous est pas spécifique. Elle est celle de tout le dépôt légal Web, que l'expression y passe par des mots, des sons ou des images. Cette problématique a donc été un peu celle de tous les services de la BnF depuis que l'établissement, en

2000 pour les premières instructions, s'est saisie de la question de l'archivage du Web au titre du dépôt légal.

Disons quelques mots de l'activité d'archivage de l'Internet à la BnF dans sa globalité. Après six ans d'expérimentation, la loi du 1^{er} août 2006 a institué le principe d'un dépôt légal de l'Internet partagé entre les compétences de l'Ina et celles de la BnF. Le maître mot est l'échantillonnage, car nous ne pouvons pas prétendre enregistrer toutes les modifications instantanées de millions de sites en même temps. La notion de couverture totalement exhaustive est explicitement évacuée des termes de la loi.

Schématiquement, la technique de collecte repose sur un robot qui explore les pages du Web et qui en découvre sans cesse de nouvelles en cliquant sur les liens qui sont dans les pages. De page en page, le robot avance. À chaque nouveau fichier accédé par le robot, une copie en est enregistrée avec des métadonnées indiquant la date et l'heure où l'on a trouvé ce fichier et où on l'a collecté. C'est une sorte de tampon d'enregistrement du dépôt légal comme on en met physiquement sur les livres. Ces fichiers représentent aujourd'hui plus de 100 téraoctets (100 000 gigaoctets) de données collectées. Ces fichiers sont stockés sur une structure de disques durs qui sont eux-mêmes régulièrement sauvegardés.

Le Web est mondial et, à terme, l'entreprise d'archivage du Web ne prendra tout son sens que si elle a aussi une couverture mondiale, qu'il puisse y avoir des interconnexions entre les archives. Cela ne peut être que le fruit d'une coopération internationale. La BnF a fondé, avec d'autres grandes archives du monde, un consortium international pour la préservation d'Internet (IIPC) qui tente de répondre à cet enjeu. IIPC développe des outils et des pratiques communes assurant l'interopérabilité des archives membres. La BnF a fait le choix d'articuler deux grands types de collecte, d'une part une collecte de type entièrement automatique sur les sites français qui permet de couvrir très largement le Web français, et d'autre part des collectes ciblées, approfondies, de grands sites identifiés par des bibliothécaires spécialisés dans la discipline concernée - cette collecte ciblée pouvant être aussi une collecte plus fréquente que la collecte automatique qui a lieu, pour l'instant, une fois par an. Il y a enfin des collectes organisées autour de projets spécifiques et cohérents, comme cela s'est passé avec les campagnes électorales (qui ont été en 2007, entre autres, un grand moment de décollage de l'expression audiovisuelle sur le Web).

La pertinence de l'archive repose à la fois sur des paramétrages techniques, un socle technique fort, mais aussi sur des priorités documentaires fixées par les bibliothécaires et non pas sur un mode tout automatique qui serait en partie inefficace. La BnF a déjà travaillé, pour ces choix documentaires, en coopération avec les bibliothèques qui reçoivent le dépôt légal en région, de manière à assurer un maillage territorial plus fin dans le cadre de certains projets comme les élections.

De la même manière que tous les départements de collections de la Bibliothèque concourent à la collecte dans leur domaine de spécialité, le Département de l'Audiovisuel s'inscrit dans ce dispositif d'ensemble à ce titre de fixation des priorités documentaires et des projets dans son domaine.

Deux grands ordres de contenus nous intéressent : d'une part, les contenus de nature audiovisuelle et d'autre part, les contenus qui documentent le monde du cinéma et de l'audiovisuel (des sites de producteurs, des sites de distributeurs de films, des sites d'éditeurs vidéo, des blogs de cinéphiles...), ce à l'exclusion de tout ce qui relève des médias radio et télévision qui sont du ressort de l'Ina.

Pour les contenus de nature audiovisuelle proprement dits, je reviens à ma question initiale : que choisir dès lors qu'il est matériellement exclu de tout collecter, de tout archiver, tout en restant fidèle à l'esprit du dépôt légal ? La réflexion a eu lieu en trois temps.

Premier temps : faire en sorte de collecter les sites des éditeurs, des commanditaires ou des producteurs de contenus audiovisuels que nous collections déjà, de façon régulière et massive sur supports, depuis les débuts du dépôt légal, et qui sont passés en ligne. Je prends comme exemple, ce que l'on appelle de la communication institutionnelle, la « Web TV » de la présidence française de l'Union européenne. Elle est vouée à disparaître, sans doute, au début de l'année prochaine. Archiver

les « Web TV » des grandes entreprises, les sites d'enseignement par l'audiovisuel (conférences, etc), les sites de vidéo à la demande (VOD) qui se substitue en partie à l'offre DVD, voilà ce qu'on appelle assurer la continuité des collections, faire en sorte que les contenus collectés sur supports continuent à l'être en ligne.

Deuxième temps : faire en sorte de collecter les sites édités individuellement ou dans un cadre collectif par des vidéastes qui s'inscrivent dans une démarche artistique ou dans une démarche d'intervention dans le champ social. Ainsi, archiver les sites de festivals, qui publient les vidéos qu'ils sélectionnent, comme les festivals de vidéos tournées avec un téléphone. Vous voyez ici une archive du 14 novembre 2007 d'un vidéaste qui dispose de son propre site.

Troisième temps et pas le moins ardu : collecter sur les grandes plates-formes de publication de vidéos. Nous nous sommes occupés, à ce jour, de Dailymotion, une plate-forme de publication de vidéos éditées par une société française et où les usagers français sont majoritairement représentés. Au sein d'IIPC, plusieurs membres travaillent sur la collecte des vidéos qui sont sur *YouTube*. Il nous a paru assez naturel, en tant qu'établissement français, de nous orienter prioritairement sur Dailymotion. Vous avez ici la page d'accueil de Dailymotion du 16 novembre 2006, qui a pas mal évolué depuis.

Échantillonner une plate-forme comme celle-là, sur quels critères ? Pour mieux le comprendre, nous allons aller en ligne. Ce sera plus parlant pour ceux qui ne connaissent pas Dailymotion.

De toute évidence, la priorité est de nous intéresser aux contenus créés pour le Web, non aux contenus qui ont été créés pour d'autres médias et qui vivent une seconde vie sur le Web, légalement ou illégalement : les films, les séries TV, les bandes-annonces créées pour d'autres médias que le Web, les publicités repiquées à la télévision, les journaux télévisés, les vidéo-clips de l'industrie musicale, etc.

Les portes d'entrée sur Dailymotion pour trouver des critères sont multiples. Les tags, les mots-clés sont apposés par le membre qui a posté la vidéo. Ce critère permet d'envisager des collectes thématiques. Cela a été fait pour les présidentielles 2007 ou après les grèves concernant les retraites. Cependant, l'usage abusif des mots-clés, des tags populaires par les usagers, par exemple, l'association du nom d'une chanteuse à la mode à un film sur votre belle-mère peut rendre le résultat peu pertinent. L'idée de privilégier un thème sur un autre n'a rien d'évident dans l'optique qui est la nôtre, celle du dépôt légal. À l'inverse, elle serait parfaitement légitime pour un fonds documentaire spécialisé. Les collections issues du dépôt légal sont souvent le dernier recours des chercheurs qui s'intéressent précisément à des publications qui étaient dépourvues de toute légitimité documentaire à l'époque de leur parution.

Les groupes de membres réunis par un même centre d'intérêt, c'est une autre porte d'entrée possible et pouvant déterminer les critères (des personnes réunies autour de la musique rap). Cette piste permet d'introduire un élément fort à prendre en compte. Sur les plates-formes d'échange, nous n'avons pas simplement affaire à des vidéos, à des stocks d'images flottantes, un peu atomisées, mais à des réseaux sociaux proliférants, des groupes, des amis d'un membre (tout membre a des amis qui eux-mêmes ont des amis), des commentaires croisés entre membres. Tout un réseau d'interrelations s'est mis en place au fil du temps par les membres actifs sur la plate-forme et est tracé et affiché dans les pages du site lors de la consultation.

Mais le parasitage des groupes par des membres hyperactifs et envahissants qui s'inscrivent à tous les groupes qui bougent nous a finalement incités à ne pas faire des groupes un critère d'échantillonnage majeur.

Je pourrais continuer un moment à passer en revue les critères possibles, leur intérêt et leurs limites. Mais je veux en venir à ce que nous avons finalement décidé de faire.

Nous avons décidé de privilégier le critère membre, autrement dit l'individu ou le collectif qui poste les vidéos. Je pourrais appeler cela une approche de type généalogique (par opposition à une approche thématique). Une fois que l'on a sélectionné un membre, on collecte tout de lui, toutes ses

vidéos et à chaque collecte, on le recollecte sans procéder à un tri arbitraire, avec une idée de continuité dans le temps et de continuité du *corpus* des vidéos d'un membre. Cette approche est également contextuelle. On sait de quel groupe le membre fait partie sur la plate-forme, et puis aussi parce que l'on va collecter systématiquement le site ou le blog associé du membre dans lequel le plus souvent, les vidéos qu'il a postées viennent s'inscrire et définissent tout un paratexte des vidéos qui parfois sont incompréhensibles. Par exemple, dans le blog d'un candidat à une élection, tout un discours « légende » la vidéo.

Nous archivons des membres, mais lesquels ? Il nous semble qu'il est souvent aisé de faire la part des membres qui s'adressent à un public large et de ceux qui s'adressent à un public restreint (famille, amis ou même personne). Les membres qui sont l'émanation d'une personne morale, d'une entreprise, d'un parti politique, d'un label musical, d'un producteur de cinéma, etc., sont les plus simples. Nous sommes dans une logique de continuité évidente avec ce que nous collectons sur supports depuis de nombreuses années. Pour ceux-là, pas de choix à opérer, ils sont collectés dès qu'ils sont repérés.

La nouveauté de l'exercice est de sélectionner parmi les membres purement individuels, ceux qui sont dans une logique de production, d'autoproduction certes, mais de production quand même, et de les faire ressortir de la masse des vidéos personnelles ordinaires, du simple film du copain tourné avec un téléphone et posté immédiatement.

Schématiquement, les vidéos assimilables à des publications par destination se rangent sur trois axes :

- les vidéos de type documentarisant, reporters amateurs alimentant des sites d'actualité collaboratifs, films touristiques, etc. ;
- les vidéos de type militant, remontage d'images d'actualités à des fins politiques, parodies, etc. ;
- et les vidéos de jeunes créateurs, courts-métrages de fiction, films d'animation créés par des jeunes sortis de leur école d'infographie, de cinéma ou d'audiovisuel.

La page d'accueil du site Dailymotion assure régulièrement une visibilité à ces membres en les faisant remonter à la surface, en les mettant en avant. Il est donc intéressant de noter régulièrement ce qui s'y passe et reste ensuite à partir à la pêche dans les profondeurs de la plate-forme pour compléter cette première offre.

À côté du critère membre, nous avons une approche thématique seulement réduite à la portion congrue, sauf dans le cas de projet de collectes thématiques.

Un autre mode d'approche possible serait de collecter les 15 vidéos les plus vues du mois écoulé. Cela nous permet de garder un historique des phénomènes de forte audience sur une courte durée. On appelle cela communément sur Internet, le *buzz*, « Casse-toi pauvre con ! » remporte plusieurs dizaines de milliers de consultations pendant un week-end, c'est un moment de la vie sociale de la France. À ce stade, le risque est aussi de donner une vision faussée de la réalité de la plate-forme en tant que média dans son ensemble. Si l'on fait une sélection majoritairement axée sur des critères de profil de membre, on va donner une vision fautive. Je me réfère à la déclaration récente et parfaitement éloquentes du patron de *YouTube*, très auto-ironique : « *YouTube, ce n'est pas juste un chien qui fait du skate-board, non, c'est aussi un écureuil qui fait du ski nautique* ». Le complément de l'échantillonnage raisonné que je viens de décrire est donc de collecter périodiquement sur une journée aléatoirement choisie toutes les vidéos apparues sur Dailymotion pendant cette journée pour replacer les contenus que nous avons sélectionnés sur des profils de membres essentiellement dans leur contexte, qui est le flux quotidien des vidéos ordinaires.

Sur la base de ces critères, de l'été 2007 à aujourd'hui, trois grandes campagnes de collecte ont été menées sur Dailymotion par la BnF. Ces campagnes ont permis de collecter de 50 000 vidéos à 80 000 vidéos à chaque fois.

Une archive n'a bien sûr d'intérêt que si elle est accessible. Depuis avril 2008, les archives de l'Internet sont consultables en salle de lecture de la BnF au niveau recherche. Sous une forme encore

inaccomplie, le lecteur peut naviguer dans l'espace du Web à un moment donné ou dans les états d'un même site à travers le temps. Dans un proche avenir, il sera aussi aidé sur l'interface de consultation par plusieurs *corpus* éditorialisés (un seul aujourd'hui, en bas de l'écran), réalisés par des bibliothécaires qui ont fait les sélections et qui ont une connaissance de la manière dont l'archive a été constituée, et doit nécessairement être explicitée pour que les chercheurs futurs puissent en tirer parti au maximum.

Je conçois qu'il peut être encore difficile de toucher du doigt l'intérêt patrimonial d'une telle collecte. Quantité de vidéos collectées depuis deux ans sont encore en ligne et nous ne savons pas quand elles seront susceptibles de disparaître du moment qu'elles sont là légalement. Nous manquons d'expérience sur le temps que cela va prendre pour que ces contenus s'évaporent ou ne s'évaporent pas.

J'espère pouvoir emporter votre conviction en vous montrant des cas concrets de documents qui n'existent déjà plus en ligne et que nous avons archivés. Ces trois types de disparition ont des causes tout à fait différentes qui permettent d'en saisir l'intérêt.

Premier exemple : des vidéos disparaissent suite à un incident technique. Le site *Cinematicfilm* a disparu du jour au lendemain. La personne qui maintenait ce site a indiqué sur son site personnel qu'il y avait eu un *crash disc*, un accident technique qui fait que le site n'a plus été accessible. Cette situation perdure maintenant depuis un an, ce qui me laisse à penser que ce site ne refera peut-être pas son apparition. Peut-être que cela n'intéresse plus la personne qui l'a édité, peut-être qu'elle a perdu les vidéos. Ce site, *Cinematicfilm*, est un site de jeunes vidéastes plasticiens qui ont été parmi les premiers à constituer un catalogue de vidéos d'artistes mis en ligne sur Internet. Ce site a disparu, mais nous avons une archive incomplète (les deux tiers des vidéos) à travers l'archivage réalisé à l'automne 2005.

Deuxième exemple : la disparition par choix. Le membre décide à un moment de retirer toutes les vidéos qu'il a en ligne. Ce membre de Dailymotion, que vous identifiez aisément, le Front national de la Jeunesse, après la campagne présidentielle 2007, a retiré ses vidéos et a détruit son profil. Ces vidéos étaient des vidéos de communication politique mises en ligne à un moment donné, et retirées ensuite. Ce document sera très intéressant pour les historiens dans le futur puisqu'il manifeste la politique d'image du parti politique à un moment donné, etc.

Troisième exemple : la disparition pour des raisons juridiques. Ici, nous avons le membre CGT Valéo, représentant un syndicat qui a été engagé dans une lutte importante contre la fermeture d'une usine. Il est toujours présent avec la majorité de ses vidéos sur la plate-forme Dailymotion, mais trois vidéos ont disparu du lot sur les 19 vidéos d'origine. Il n'y en a plus que 16 aujourd'hui. Vous avez deux exemples en bas des trois vidéos en question. Michel Sardou et Jean Ferrat sont des vidéos qui posaient des problèmes de droits sur la bande son, et qui, néanmoins pour l'historien, sont parfaitement passionnantes. Nous sommes heureux de les avoir archivées puisqu'il s'agit d'un bon exemple contemporain de la vieille pratique de réutilisation de la culture musicale dans les manifestations ouvrières.

D'autres cas existent. Par exemple, la disparition d'un blog alors que les vidéos sont toujours en ligne. Le blog, dans lequel s'inscrivaient les vidéos, était le paratexte qui lui donnait son sens. Les vidéos sont encore là, mais une partie du sens est perdue si le blog disparaît, d'où l'intérêt de l'archiver aussi.

Nous n'avons pas le recul suffisant pour prévoir ce qui disparaîtra et à quel rythme. On ne peut pas exclure des disparitions massives, comme on peut imaginer qu'il n'y ait pas de disparitions massives du tout. Je lisais très récemment que Google se fait du souci, parce qu'il n'arrive pas à gagner de l'argent avec *YouTube* qu'il a acheté fort cher et certains analystes n'excluaient pas que Google ferme YouTube. Nous avons très vite pris l'habitude de croire que YouTube fait partie pour toujours de notre existence, eh bien non, peut-être pas.

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