

Archimages07

Entre l'offre et la demande : l'entreprise patrimoniale

22.23.24 octobre 2007

DE NOUVEAUX PRODUITS POUR DE NOUVEAUX PUBLICS

De la documentation à l'édition ? Évolutions des produits, des usages et nouveaux métiers

Table ronde animée par :

Isabelle GIANNATTASIO,

Conservateur Général des bibliothèques, Directrice du Département de l'Audiovisuel à la Bibliothèque nationale de France.

Avec la participation de :

Emmanuelle BERMES,

Responsable fonctionnelle de la bibliothèque numérique de la BnF,

Anne-Marie GRAPTON,

Administratrice de la base de données des Archives françaises du film du Centre national de la cinématographie,

Thierry ROLLAND,

Directeur de l'atelier des Archives,

Catherine TEILLOU,

Responsable de la Vidéotheque et de la Phototheque au sein de la Direction des Archives Audiovisuelles de l'Ina

Isabelle GIANNATTASIO

Le thème de cette table ronde porte sur le métier de documentaliste. Je suis très heureuse qu'on en parle parce que c'est finalement un peu les acteurs de l'ombre. Être un acteur de l'ombre, c'est aussi être un acteur de pouvoir. Ce qui s'est dit hier, mais encore ce matin, ou tout récemment par Marie-Claire AMBLARD, nous interroge sur ce pouvoir des documentalistes. Dans ce nouvel environnement de diffusion planétaire des informations sur les fonds et de parties des collections, y a-t-il un bouleversement du métier traditionnel de la documentation ? Est-ce encore une évolution ou une révolution ? Si on reprend le schéma classique de la documentation, qui est le centre de la relation et même de l'équation entre les documents et les usagers, assiste-t-on donc à une adaptation ou une révolution ?

Autour de moi, représentant autant leurs institutions que leurs compétences et leurs expériences :

- Emmanuelle BERMES de la Bibliothèque nationale de France, responsable fonctionnelle de la bibliothèque numérique de la BnF.
- Anne-Marie GRAPTON pour le Centre national de la cinématographie. Elle administre la base des données des Archives françaises du Film.

- Thierry ROLLAND de l'atelier des Archives. Il a sa propre entreprise de documentation, et c'est très rassurant de le voir ici, car cela veut dire qu'il y a une rentabilité de la documentation et des archives, visible par la création d'une entreprise privée.
- Catherine TEILLOU de l'Ina, responsable de la vidéothèque, et vit en direct ces évolutions du métier de documentaliste.

Dans un premier temps, je vais passer la parole aux uns et aux autres, sur la problématique des fonds : les connaître, les décrire. Pour aborder ces sujets, je leur demanderai quels sont les fonds qu'ils ont à traiter, comment ils les traitent, et quelles évolutions ils ont pu constater ces dernières années dans cette perspective de mise en ligne.

Anne-Marie GRAPTON

Comme le disait Isabelle GIANNATTASIO, je représente les Archives françaises du Film et du CNC. Je suis également administratrice de la base de données des archives LISE. Je ne peux témoigner que de mon expérience. Mes « maux » n'engagent que moi. Quand Isabelle GIANNATTASIO m'a invité à participer, je me suis dit, ayant été précédemment dans le public, que l'on a toujours l'impression de parler du même sujet dans ces rencontres, et donc que j'ignorais ce dont j'allais parler. Je me suis dit finalement que rien n'avancait vraiment. Et pourtant, c'est l'intérêt de ces rencontres de parler de cela.

Les choses avancent donc lentement, pas aussi vite que la mondialisation, mais cela convient à mon rythme. Je ne veux surtout pas disparaître dans l'enfer de l'économie. Je voudrais seulement rappeler ce qui me semble une véritable évolution sémantique dont on a tous conscience, mais qu'il est utile de rappeler : on ne parle plus de documents aujourd'hui mais de ressources. On ne parle plus de conservation mais de gestion de contenus. Et on ne parle presque plus d'usagers, à part les vieilles documentalistes dont je suis, mais d'interactions, bien que cela ne dise pas trop où cela interagit, pourquoi et comment. Il s'agit beaucoup plus aujourd'hui pour les archives comme les nôtres de créer la demande plus que de l'écouter ; d'être une entreprise et non plus une institution culturelle ; et également d'aller au-delà de la conservation physique. Nous nous disons qu'enfin nous allons nous attaquer au sens, mais la réalité est un peu plus déprimante aux archives. Je vais rapidement vous décrire ce qu'est ma réalité aux archives.

Parfois avec les fonds, on touche le fond... Aux archives, il y a beaucoup de ressources qui ne sont presque que des boîtes. Dans le vocabulaire, on appelle cela des archives mouvementées. On mouvemente beaucoup de boîtes, ce qui est un petit peu compliqué. Pour vous parler un petit peu de la culture de cette archive, pour que vous compreniez, je vous dirais que nous sommes loin de la bibliothèque numérique aujourd'hui. Je ne travaille pas dans un palace patrimonial, mais dans un ancien fort militaire à Bois d'Arcy. En 1954, lorsqu'on a interdit les films sur support nitrate, on a fait le constat que l'on avait plus de dix tonnes à classer. On a commencé par cela, et nous n'avons pas terminé aujourd'hui. Nous avons réussi à sauver à peu près 15 000 œuvres sur support nitrate, grâce au plan de sauvegarde des films anciens. Tout cela pour vous dire que la logistique a longtemps primé à Bois d'Arcy. Mouvementer des boîtes, gérer des flux : des contenants, mais toujours pas tellement de contenus. Mais cela avance : aujourd'hui, on continue donc le catalogage. Ce mot paraît assez ringard : être catalogueur n'est pas forcément très valorisant. Il y a une expression utilisée par les anglais que je trouve intéressante : « Cataloging unlocks the collection », soit « Cataloguer déverrouille les collections ».

Malgré tout, quand on catalogue, on crée des substituts condensés – c'est encore les anglais qui emploient cette expression – pour exprimer des ressources complexes de l'image animée et aussi de la culture. Encore aujourd'hui, comment peut-on faire une fiche pour épuiser la description de ces images animées qui sont presque toutes des œuvres d'art, et du patrimoine culturel ? Le catalogage est vital, même si ce n'est pas le plus plaisant à faire. Aux archives, nous faisons encore beaucoup de fiches, et nous avons une culture de la fiche. Aujourd'hui, nous utilisons une fiche très complète qui

réunit à peu près 117 champs. On se dit qu'avec autant de champs, on va nécessairement pouvoir décrire ce qu'on manipulait avant. Pourtant, on parle encore aux archives de notre fonds de 100 000 œuvres et 1 million de boîtes.

On a donc vraiment du mal à s'attaquer au sens. Nous n'avons pas aujourd'hui de véritables cartographies du fonds, parce que l'on a encore du mal aujourd'hui à oser parler de collections. On ne sait pas de quel monde parlent ces films et quel monde ils montrent. Les archives sont encore jeunes puisqu'elles sont nées en 1979. Nous avons donc encore de l'espoir. Elles ont une identité relativement récente : leur appellation précédente était « service des archives du film », puis s'est rajouté « du dépôt légal » ; enfin on a quitté « service » mais on est toujours « archives ». On a développé une base de données récente puisqu'elle date de 2004 pour sa mise en production. On a aussi ouvert très récemment un site WEB en 2006, au moment même de l'ouverture de l'antenne de consultation à la BnF. Je dirais donc que nous sommes encore dans une phase archaïque, mais l'archaïsme est intéressant parce qu'il implique de la proximité. Nous sommes proches des gens, des films et des consultants.

On a aujourd'hui traité le nitrate parce que c'était l'urgence, qu'il fallait sauver ce patrimoine, c'est-à-dire le transférer, le conserver physiquement. Mais il reste aujourd'hui tout le post-nitrate, qui est un passif énorme. Je ne pourrais même pas vous donner un chiffre, mais j'évalue cela à la moitié du fonds. L'outil dont on dispose est une base de données assez classique, un petit peu rigide. Aujourd'hui, on ne fait pas d'ECM (Environment Content Management) aux archives, donc pas de gestion de contenus, mais on espère y arriver. Nous avons donc beaucoup de nouveaux problèmes à gérer : le syndrome du vinaigre ; le support numérique, avec l'arrivée du cinéma numérique. Je dirais que l'on a une archive qui est « apprenante » sur le plan du traitement documentaire. Notre base actuelle comporte des données riches et pauvres : nous conservons pour nous les données pauvres et mettons les riches sur Internet, même si elles ne sont pas nécessairement exhaustives et demandent parfois des explications. Les données manquent parfois de contexte, et n'ont pas toujours de sources. Ce sont de vraies difficultés. En plus, lorsque nous avons la chance d'avoir identifié les ayants droit, il y a une autre difficulté mais qu'on évoquera peut-être un petit peu plus tard.

Numériser semble la panacée : on se dit que l'on va tout numériser. Mais il faut d'abord savoir pour qui et comment on va numériser. Enfin, comment montrer quand on n'a pas la propriété des œuvres ? C'est encore un nouveau problème.

Isabelle GIANNATTASIO

Merci beaucoup. Cela reflète bien la question de l'histoire des fonds et de la nature des collections. Je vais passer la parole à Catherine TEILLOU, car l'Ina est une masse énorme à la fois de documents et de documentaires.

Catherine TEILLOU

Je pense que l'on vous a déjà parlé de l'Ina entre hier et aujourd'hui, donc je vais essayer d'être succincte. Nous avons un peu les mêmes problèmes que le CNC, mais peut-être à un degré inférieur. Nous sommes une institution plus ancienne, et nous avons commencé plus tôt. Je dirais que la particularité de l'Ina est de gérer deux fonds différents :

- D'une part, les archives professionnelles issues des diffuseurs publics.
- D'autre part, les archives légales car nous sommes dépositaires du dépôt légal de la radio-télévision.

Pour donner une idée de la volumétrie, les archives professionnelles représentaient à peu près 700 000 heures ; le dépôt légal représentait un petit plus de 700 000 heures. Mais l'accroissement annuel des archives professionnelles était à peu près de 25 000 heures par an, alors que

l'accroissement du dépôt légal est de 350 000 heures par an. Cette volumétrie assez considérable est une des raisons – avec l'obsolescence des matériels ou la dégradation des supports physiques – qui ont fait que, dans les années 2000, l'Ina s'est lancé dans un vaste plan de numérisation des collections et de captation du flux entrant.

Nous avons aussi des strates d'indexation puisque la télévision a commencé à peu près après la guerre. On commence vraiment à conserver à partir de 1947, pour nos premières émissions ou extraits d'émissions. Nous avons des catalogues manuels qui commencent à apparaître dans les années 1950. On a une première informatisation, en semi-conversationnel avec des micro-fiches en 1975. Et puis une informatisation qui ressemble plus à ce que l'on connaît aujourd'hui dans les années 1985. Tout cela a des strates, et les contrôles de cohérence que l'on peut faire aujourd'hui n'ont rien de semblable avec les contrôles que l'on pouvait avoir à ces moments-là. Nous nous sommes aussi lancés dans des politiques d'inventaire dès les années 1980-1985, que ce soit des inventaires physiques ou des inventaires documentaires. Sur le plan national, les inventaires sont terminés. Ce n'est pas tout à fait vrai sur le plan régional. Toutefois, un inventaire implique que tout soit catalogué, mais pas que tout soit indexé. De même que, lorsque l'on dit « numériser », nous n'avons que 430 000 œuvres numérisées, soit un peu plus de la moitié. Ainsi, ces œuvres ont été transférées sur un nouveau support, mais cela ne veut pas dire qu'elles ont été restaurées.

Isabelle GIANNATTASIO

Merci. Je vais tout de suite passer la parole à Emmanuelle BERMES qui, en tant qu'experte, pourra nous dire comment on attaque ces masses documentaires. On a parlé des strates documentaires à l'Ina. Dans une maison comme la BnF, qui a des catalogues et des collections plus anciens de différents médias, c'est aussi un véritable casse-tête. Peut-on et comment attaquer ces masses documentaires ?

Emmanuelle BERMES

Je parlerai de mon expérience de la bibliothèque numérique, de son expansion appuyée sur un fonds effectivement assez massif. Nous nous positionnons aujourd'hui à un moment très important pour cette bibliothèque numérique. C'est un moment d'évolution à la fois de la façon dont la bibliothèque se constitue, de la façon dont elle se présente à son public, et de la façon dont elle se construit, c'est-à-dire des métiers qui sous-tendent cette bibliothèque numérique.

Gallica, la bibliothèque numérique de la BnF, est effectivement un double héritage. La numérisation a commencé dans les années 1990 à la BnF, et Gallica a aujourd'hui dix ans, puisque la première version du site était sortie en 1997. Dix ans est un âge avancé pour un site WEB. Cela veut dire, pour notre site dont la dernière version date de 2000, sept ans sans avoir connu d'évolutions majeures. Il y a un questionnement autour de cet héritage. C'est aussi un héritage d'apprentissage, puisque cette numérisation s'est construite pendant plus de dix ans au fur et à mesure, en apprenant son propre métier.

L'héritage de Gallica repose également sur l'héritage de la bibliothèque, ses collections très riches et très diverses, mais aussi sur son catalogue. Une des richesses sur lesquelles repose Gallica aujourd'hui est l'ensemble des descriptions qui existent déjà, même si, dans le catalogue de la BnF comme d'autres exemples cités, on a aussi des effets de strates dus aux différentes conversions qui ont amené à des données riches ou pauvres, pour reprendre l'expression précédemment citée.

Depuis quelques années, on se situe à un moment-clé de cette évolution et de cet apprentissage, puisqu'on connaît aujourd'hui un grand changement d'échelle dans la numérisation. On passe de quelque chose qui était véritablement encyclopédique, mais dans une dimension humaine, à quelque chose qui s'approche beaucoup plus du traitement des masses de la bibliothèque. Pour vous donner simplement un chiffre très peu significatif, mais donnant une bonne image de ce que représente ce

changement d'échelle, on passe d'une numérisation qui touchait de 5 000 à 6 000 documents par an, à une numérisation qui va toucher 100 000 documents par an. Cette numérisation qui s'accroît impose un questionnement sur la masse de documents et de descriptions qui sont à notre disposition et qui vont servir de socle à cette numérisation.

Cette nouvelle numérisation de masse va se construire sur les collections de la bibliothèque, donc aussi sur l'état des collections. Et on va retrouver des problèmes logistiques dans la manipulation des objets, puisqu'on est dans le numérique mais pas dans le virtuel. Avant d'obtenir cette bibliothèque numérique, on manipule les objets, boîtes, etc., de la même façon que pour le CNC précédemment. Ce sont des questionnements logistiques un petit peu nouveau dans le domaine du numérique. Il y a aussi des questionnements sur le type de description et de traitement que l'on va être capable de faire sur ces masses qui sont, nécessairement, moins détaillés et moins précis que ce que l'on pouvait faire lorsque l'on traitait des quantités de documents de moindre importance. Nous allons de plus en plus nous tourner vers une réflexion qui va porter sur l'automatisation et sur les outils qui sont à notre disposition aujourd'hui pour rentrer à l'intérieur des documents. Ce qui ne veut pas dire que l'on va se passer de la description des documents bibliographiques traditionnelle, toujours aussi essentielle, mais elle va être complétée, en quelque sorte, par des technologies qui vont venir apporter des descriptions nouvelles, ou qui vont rendre les documents auto-descriptifs. Par exemple, la reconnaissance automatique de caractère, ou OCR, permettra d'indexer en plein texte tous les mots des documents. Bien sûr, nous avons cette possibilité parce que l'on est sûr du texte. Dès lors que l'on commence à s'intéresser à d'autres types de documents que sont les images, les documents audiovisuels et les documents sonores, cette possibilité n'existe plus, du moins à ma connaissance. Dans l'état actuel des technologies, nous n'en sommes pas à faire de la transcription automatique des documents.

Du point de vue de l'évolution du métier, c'est tout de même très intéressant de voir l'apprentissage progressif que l'on va avoir sur la façon dont ces technologies nous obligent à regarder différemment nos documents. Par exemple, à les regarder en se demandant si tel document va être compréhensible pour une machine ou pas. Cela questionne la part de l'humain dans les points d'accès, dans les descriptions que l'on va devoir faire de ce document, et quelle est la part que l'on peut raisonnablement et pragmatiquement confier ou déléguer à une machine.

Isabelle GIANNATTASIO

Merci Emmanuelle. Je vais passer la parole à Thierry ROLLAND, car il sera dans cette affaire à la fois expert et chef d'entreprise. Il pourra dire si ces technologies nous apportent aussi quelque chose. Peut-on capitaliser et utiliser aussi ces technologies pour nous aider à produire l'information documentaire ?

Thierry ROLLAND

Je vais d'abord vous décevoir parce que je ne suis pas un banquier qui investit massivement dans les archives, mais simplement un historien de formation qui a passé neuf ans comme documentaliste à l'Ina, puis qui a été documentaliste free-lance pendant deux ans. J'ai également eu la joie de diriger pendant dix ans les archives Pathé. Fort de cette passion, j'ai choisi malgré tout de continuer dans le monde des archives, ce qui est apparu pour beaucoup comme une folie ou une obstination. J'essaye de vivre de cette passion, et d'en vivre un petit peu. Mais je ne suis pas encore coté en bourse...

La collection de l'atelier des Archives représente la confiance obtenue auprès d'un certain nombre de personnes qui nous ont confié leurs archives :

- Les archives anglaises British Movietone. Nous avons eu beaucoup de chances, car il s'agissait d'archives non seulement décrites mais entièrement numérisées. Le travail pour nous en a été simplifié.

- On a ensuite obtenu un accord de commercialisation avec l'AFP. Nous avons très vite été imités par l'Ina, donc on est en co-mandat avec l'Ina sur l'AFP, mais nous essayons tout de même de nous en sortir.
- Nous avons également une collection appelée « Paris Première », qui est une des seules archives de chaîne qui a produit beaucoup de documents et les a décrits pendant des années.
- D'autres collections sur lesquelles on se retrouve en contact avec le CNC, en particulier la collection de Benoît LEVY ; et une formidable collection qui s'appelle « La France en marche ». Ce sont effectivement des collections qui ont émergé grâce au travail du CNC, des Archives françaises du Film, qui ont fait ce travail de restauration et ont réglé ensuite la question des ayants droit. Nous sommes ensuite passés derrière pour faire connaître et commercialiser ces documents.

Nos principaux problèmes lors de la création de la société étaient que le « marché » n'était pas ouvert à tous les clients potentiels. Pendant des années, j'ai fait du B2B sans le savoir. Pour ceux qui n'ont pas compris ce que c'est, il s'agit du « Business to Business ». Dix ans après avoir été chez Pathé, quelqu'un m'a dit que je faisais du B2B et pas du B2C, ce dernier signifiant « Business to Consumer », donc le marché grand public. Je ne le savais pas, mais je faisais effectivement du B2B car la possibilité de toucher un grand public sans perdre d'argent n'est pas totalement évidente. Cela va s'améliorer, avec de nouvelles solutions à trouver. C'est un peu le but de ce colloque.

Je me suis demandé ce qu'il fallait faire comme base de données pour attirer la clientèle des professionnels. J'ai encore eu beaucoup de chance car je venais de réaliser un projet avec Pathé où nous avons dépensé 500 000 € à 600 000 € pour créer une base de données et la numérisation. Il se trouve que les choses ont évolué : des technologies comme MySQL permettent d'avoir le talent de Vincent PROST et de sa société Opsomai. Vincent PROST est quelqu'un qui vient de l'Ina et qui a fait énormément de choses pour les systèmes d'information pour l'Ina. Nous avons conçu une base qui s'appelle Opsomai, qui fut ensuite adoptée également par le CNDP et par d'autres institutions. Il s'agit de quelque chose de très évolutif qui permet la numérisation et surtout l'intégration des données. C'est ainsi qu'on a réussi à importer une dizaine de bases de données complètes qui étaient dans des formats aussi divers que Phraséa. Cela fonctionne aujourd'hui sous la forme d'une base unifiée de 100 000 documents, dont beaucoup sont numérisés.

Je donnerai quelques points importants à travailler face à la numérisation. Les collections sont pour moi un fil rouge. A l'époque de Pathé, la BnF nous a permis de faire quelque chose avec deux collections que sont *Les chroniques de France* et *Pathé revue*. *Chroniques de France* concerne des films en 16mm pour le Ministère des Affaires étrangères avec, tous les mois, trois sujets sur la France qui se faisait. C'est une collection extraordinaire dont vous pouvez voir une partie à l'exposition Giacometti. La BnF nous a dit que cette collection les intéressait. Ils ont donné un peu d'argent, et Pathé a fait des efforts pour rassembler la collection, régler les problèmes juridiques, et les transferts ont pu se faire. Voilà comment on a réussi à constituer une collection. La même chose s'est également passée avec *Pathé revue* qui était aussi une collection incomplète sur laquelle on a réussi à faire un vrai travail de description en suivant ce fil rouge, en cherchant les journaux de programme, en cherchant les catalogues. Je pense que c'est très important d'avoir ces niveaux et ces catalogages. D'ailleurs, encore aujourd'hui, il est plus facile de faire les imports de données parce qu'il y a des zones qui sont bien déterminées.

Juste un petit point technique à ce sujet : il y a en fait trois niveaux quand on numérise.

- *Numérisation par sujets* : il est difficile de numériser un document de trois heures. Cela marche à peu près aujourd'hui, techniquement parlant. Pour certaines choses, cela me semble indispensable. Par exemple, quand vous avez une chanson, vous n'allez pas la découper. C'est en général important de garder l'intégrité du sujet.
- *Numérisation par séquences* : ce que permet de faire la numérisation, c'est également un découpage en séquences.

- *Numérisation par plans* : troisième niveau, qui est celui du « plan », très présent dans la publicité. Cela sert pour les gens qui font de la publicité et qui veulent un plan d'une loutre qui sort de l'eau et met sa tête de droite à gauche. Même si c'est très bien, cela a parfois des limites.

Je vais vous en citer un exemple. Lorsque Marilyn MONROE débarque en Corée, elle est accueillie par les soldats. C'est une image magnifique mais qui dure dix secondes. J'ai été très content de découvrir à Washington un film muet de 30 minutes de Marilyn MONROE en Corée, filmé par différentes unités de l'armée. Il y a même une production de Serge VIALLET de l'Ina qui a utilisé ces très beaux documents. Jusqu'au jour où je découvre que ce film existe en couleurs. Le problème est que c'est très mal découpé : on a un plan où Marilyn a un bras en l'air, un autre où elle a un bras autrement, etc. C'est découpé et il n'y a aucun historique de ce que c'est, comment ils l'ont eu, etc. C'est absolument inutilisable pour un documentaire d'avoir des plans de 3 s. Pour cela, je pense qu'il est nécessaire, même si on découpe par plan, de garder et de pouvoir remonter au niveau supérieur et avoir l'historique de la chose.

Isabelle GIANNATTASIO

Merci beaucoup Thierry ROLLAND. Je vais repasser la parole à Anne-Marie GRAPTON parce que je pense que vous êtes en plein dans ces problématiques aux Archives françaises du Film. Cette question de l'œuvre, de la séquence et du plan ; cette question des différents matériels et du contexte, comment vous les posez-vous ?

Anne-Marie GRAPTON

Nous n'arrivons pas à tout faire, donc nous avons fait des choix, notamment de ne pas aborder le séquentiel. Ce choix est fait aussi en fonction de la mission de cette archive qui n'est pas, en résumé, de vendre, mais de donner accès aux œuvres. Ainsi, nous ne les découpons pas, même si cela nous arrangerait souvent, quand on voit par exemple la longueur des génériques et la difficulté de saisir un générique de manière exhaustif pour l'équipe de l'inventaire. Cela peut tourner au cauchemar, surtout quand on regarde le générique à la loupe. Nous ne découpons pas, mais nous traitons l'unité documentaire : l'œuvre, le film dans son entier.

En revanche, cela devient de plus en plus compliqué car cataloguer n'est pas indexer. Mais là aussi, un thésaurus trop long, trop hiérarchique, est trop difficile à maîtriser et à faire évoluer. Nous nous sommes lancés dans le « libre ». Sauf que l'outil ne permet pas aujourd'hui de traiter pleinement toutes les questions de synonymie, etc. Et puis nous nous sommes dit que les mots étaient insuffisants, donc nous avons rajouté des phrases, des résumés pour décrire ces œuvres. Ces résumés sont aussi faits en langage libre. Je passe beaucoup de temps à valider des résumés en langage libre, car je pose beaucoup de questions qui nous permettent d'avoir des vrais débats. Nous sommes vraiment dans cette problématique. Nous avons des objets qui ne sont pas que des objets mais aussi des œuvres. Pour prendre un exemple précis de corpus ou de genre qui peut être difficile, il y a les films idéologiques, de propagande. Comment aborde-t-on le message idéologique dans un résumé « informatif » ? On ne peut certainement pas être objectif.

Nous avons aux archives tous les films de toutes les époques, depuis la naissance du cinéma, y compris des films amateurs, des films pornographiques, des films avec des opérations médicales d'amputations, des films pédagogiques, beaucoup de fictions. Cela devient difficile de faire des résumés objectifs de la fiction, et c'est encore pire pour l'indexation. On s'arrête donc au résumé et on ne fait pas d'indexation pour la fiction.

Nous avons fait ce choix car la grande valeur ajoutée des archives, si je peux m'exprimer ainsi, est de conserver infiniment des documentaires inconnus ou méconnus, notamment des années 1950 et bien avant, si le mot documentaire est adapté ou s'il faut l'étiqueter « non-fiction ». Nous avons de la

fiction, de la non-fiction, et puis du docu-fiction. Je sais qu'il y a de grands débats à l'Ina qui ne nous aident pas, parce que nous voudrions des réponses pratiques pour savoir ce qu'est la docu-fiction dans la forme.

Nous avons justement commencé à travailler sur le corpus des films restaurés, pour lesquels nous avons donc une copie de visionnage, et pour lesquels nous avons aussi résolu les problèmes de droits, puisqu'on signe des conventions avec les ayants droit identifiés. Nous avons commencé un travail de thématization sur ce corpus, à la suite de l'Ina, ce qui permet de voyager dans les fonds, de naviguer dans les collections, de faire d'autres découvertes. Nous ne sommes pas descendus jusqu'à quinze niveaux de thématization sinon nous retombions dans le cauchemar du thésaurus, donc nous sommes arrêtés à deux niveaux. Mais si les documents semblent trop classifiés, cela reste relativement ouvert. C'est un angle d'attaque intéressant quand on fait du langage libre partout.

En réaction à ce qu'a dit Thierry ROLLAND, je dirais qu'aujourd'hui ce sont beaucoup plus les usagers qui nous font connaître les collections parce qu'ils s'attaquent à des sujets, et font ainsi émerger des sujets. Cela entraîne des réactions et des actions de notre part, en sortant des boîtes, en y trouvant parfois des problèmes de conservation, d'inventaire, de catalogage ou d'indexation. Nous prenons conscience de l'intérêt d'un sujet, jusqu'à envisager sa programmation. Ce sont de plus en plus les chercheurs, puisque nous ne recevons que très ponctuellement du grand public. Je voudrais aussi remercier toutes les équipes des archives, pas seulement les catalogueurs, mais les gens qui font des prêts, de la médiation culturelle, des permanences pour les consultations, et les stockistes qui continuent de mouvoir des boîtes. On les appelle toujours « les agents », mais ils sont en fait beaucoup plus des acteurs.

Isabelle GIANNATTASIO

Je vais passer la parole à Catherine TEILLOU, car je n'ose plus vous demander quelle est votre unité documentaire ou votre choix. La présentation de Marie-Claire AMBLARD entre inamedia et ina.fr a montré l'amplitude du public visé. En tant que médiateur, comment faites-vous ?

Catherine TEILLOU

C'est vrai que, jusqu'à ce que nous numérisions, nous avons fait des descriptions des émissions de façon très précise. Il fallait que la fiche documentaire rende compte du contenu, et il fallait que ce soit suffisamment précis parce que le recours au visionnage n'était pas quelque chose de nécessairement facile. Nous allions donc jusqu'à une description des séquences et du plan par plan. À partir du moment où nous avons numérisé, nous nous sommes questionnés sur ce que cela changeait à nos pratiques documentaires. Nous avons conclu que nous allions toujours avoir besoin de connaître le contenu, mais que nous pouvions nous exonérer de la description des séquences et du plan par plan. Nous avons fait ce pari qui s'est révélé juste, c'est-à-dire que notre indexation s'arrête à une description de contenu.

Pour ceux qui étaient là l'année précédente, il y a avait une intervention de Bruno BACHIMONT qui m'avait beaucoup intéressée parce qu'il avait dit que le numérique était comme une bombe à fragmentation : cela permet de prendre une intégrale et de faire autant d'extraits que l'on veut, que l'on va pouvoir manipuler à volonté, et réutiliser dans des contextes très différents. Je trouve un peu extrême ce qu'il disait par rapport à l'Ina, mais c'est assez vrai. Maintenant, nous faisons des extraits. Pour nous, un extrait est quelque chose qui est cohérent en termes de droits et de contenus, mais nous faisons malgré tout des extraits. Un extrait est par exemple une chanson, un sketch, ce sont toutes les petites phrases dans les discours des hommes politiques, ce sont les moments forts d'une interview, etc. Si vous prenez par exemple des conférences du Général DE GAULLE, il y a toujours une phrase mythique dans son discours. Savoir dans quel discours c'était et à quel moment, ce n'est pas toujours évident. Mais, à partir du moment où cela a été repéré, c'est-à-dire que l'on a été capable

de poser des time codes, où l'on a été capable d'en faire un objet documentaire indexé, on peut faire des recherches et le retrouver directement. C'est un peu la nouveauté.

Nous avons mis en valeur notre fonds, en faisant des corpus documentaires à base d'extraits et d'intégrales sur des thèmes ou des personnalités. La numérisation nous a aussi permis de faire des cartographies du fonds.

Isabelle GIANNATTASIO

Sur la question de la granularité ou de l'accès aux données grâce aux méta-données, Thierry ROLLAND voulait réagir.

Thierry ROLLAND

Effectivement, je souhaitais réagir. J'ai passé la semaine dernière à chercher quelque chose de très simple : des femmes tondues. Mais je ne cherchais pas dans des sujets d'actualité, mais dans des chutes et dans des rushes, dans tout ce qui a été filmé par les Alliés à partir du débarquement en Normandie. Je peux dire que cela n'a pas été facile, car il fallait d'une part chercher « shave » dans la base anglaise ; on tombait sur des « femme tondu(e) » avec un « e », des « femmes tondues » avec un « es ». C'est un exemple qui peut montrer l'importance de bien gérer son vocabulaire. La recherche semblait évidente, mais cela ne l'a pas été par les personnes qui ont pu indexer ces choses. Je me suis aussi retrouvé à chercher des tongs, des espadrilles, des bateaux Riva. C'est typiquement le genre de demandes qu'on peut avoir aujourd'hui pour une émission comme *Capital*. C'est bien de numériser car cela permet de voir, mais le problème est qu'on a toujours autant de documents à voir. Malheureusement, il n'y a pas de solution miracle. Même du côté de l'indexation et de la reconnaissance automatique, il reste du bruit documentaire.

Je voulais rectifier une erreur : ce n'est pas le CNDP, mais le CNRS-audiovisuel qui a Opsomai.

Isabelle GIANNATTASIO

Merci. Je vais passer la parole à Emmanuelle BERMES qui va nous parler de bruit et de silence, de granularité, de toutes ces notions documentaires.

Emmanuelle BERMES

Effectivement, la granularité est une notion qui nous est très chère dans le contexte de la numérisation. Elle est un peu sur les mêmes problématiques dont nous avons déjà parlé : dans les bibliothèques, et en particulier dans tout ce qui tourne autour du livre, il y a une facilité de gestion de la granularité qui s'appuie sur l'objet matériel. On a l'objet livre entre les mains qui constitue, *a priori*, de façon plus ou moins évidente, le niveau auquel on s'intéresse et qu'on a traditionnellement décrit dans les catalogues. Le numérique explose évidemment tout cela, à la fois par le haut et par le bas.

- *Par le haut* : C'est l'idée de collection qui est très importante, comme le soulignait tout à l'heure Thierry ROLLAND. C'est-à-dire l'idée qu'au-dessus du document se trouve une réalité, et que cette réalité est nécessaire pour rendre le corpus visible. Cela peut-être un point commun entre tous les documents, une indexation thématique, ou quelque chose de similaire. En tout cas, le numérique nous fait perdre quelque chose que l'on avait dans le monde physique ou la bibliothèque physique : la possibilité d'appréhender la collection en la regardant. On peut prendre l'exemple d'un usager qui arriverait dans une bibliothèque où il y aurait du libre accès, et qui aurait la possibilité de regarder tout ce qu'il y a dans la collection. On a trop souvent perdu ce niveau avec le numérique. On essaye de le reconstituer en utilisant des classifications, des thèmes, qui vont permettre de reconstituer virtuellement ce niveau du rayonnement tout en ayant un certain respect de la façon dont les utilisateurs ont

envie de voir la collection. On ne va pas leur imposer un seul cadre de classement, mais leur permettre de choisir l'angle qui les intéresse, puisque le numérique donne la possibilité extraordinaire de classer un même document à plusieurs endroits, ce qui n'était pas possible dans le monde physique.

- *Par le bas* : Pour ce qui est de la granularité à l'intérieur même du document, nous sommes un peu confrontés au même problème que lorsqu'on s'occupe de livres, et indépendamment du fait que l'on dispose éventuellement d'une version automatiquement transcrite en textes. C'est le fait que nous n'avons pas les moyens de rentrer dans le document en termes de description. Nous l'avons fait un petit peu en traitant par exemple les tables des matières, qui sont un bon outil parce qu'elles représentent déjà un découpage du livre en unités significatives. Mais, jusqu'à un certain point, on n'ira jamais plus bas. Parfois, pour des matériaux comme par exemple la presse du XIXe siècle, que nous sommes en train de numériser, nous n'avons pas l'équivalent de ces découpages, de ces structures internes des documents. Là, la piste est peut-être de se laisser guider par les usagers comme le suggérait Mme GRAPTON tout à l'heure. Il faut alors leur donner la possibilité et les outils pour faire sortir dans les collections et à l'intérieur des documents, les choses que nous n'aurions pas vues parce que ce n'est pas notre mission, ou parce que nous n'avons pas la possibilité d'entrer à ce point-là dans le détail. Ainsi, ils pourraient créer leur propre parcours, à la fois au-dessus des documents (par des regroupements ou des parcours de lecture), et aussi à l'intérieur (par de l'indexation, des commentaires). Cela peut être des choses plus ou moins structurées, de type indexation par les utilisateurs par des systèmes de tags ou d'étiquettes, en somme par l'ajout de mots-clés ; ou que ce soit des choses simplement en langage naturel (résumés, commentaires), qui peuvent aussi être traitées par des moteurs de recherche aujourd'hui. L'intérêt est de permettre aux usagers de créer du matériau textuel que nous allons pouvoir exploiter pour créer de nouveaux points d'accès pour ces documents.

Thierry ROLLAND

Je suis désolé, mais c'est une idée qui me choque énormément. Je pense que si l'indexation est faite sous la forme de tags, on ne fait plus la distinction entre l'Internet de Wikipédia, où l'on suppose que ce soit des gens à peu près experts, et une indexation ouverte à tous. Aujourd'hui, nous savons aussi à quel point il y a des manipulations dans Wikipédia. Que l'Internet permette à des gens spécialistes de telle ou telle question de rajouter des éléments, mais que ce soit comme dans DailyMotion ou YouTube, que les simples lecteurs fassent les mots-clés, cela me choque profondément. Ce n'est pas simplement parce que je suis conservateur, documentaliste, etc., c'est parce que je crois au modèle de l'encyclopédie, qui me semble n'avoir rien à avoir avec l'indexation par le public, qui tend plus au règne du n'importe quoi.

Isabelle GIANNATTASIO

Alors, Catherine TEILLOU, est-ce que le documentaliste devient un modérateur ou un validateur ?

Catherine TEILLOU

Là on touche au cœur du problème. Je vais donner un exemple. J'étais la semaine dernière à la FIAF où il y a eu une présentation très rapide d'une télévision hollandaise qui fait faire ses indexations par des tags inscrits par les utilisateurs. Ce n'est pas tout le monde, seulement un certain nombre d'utilisateurs. Je n'ai pas été voir ce que cela donne, donc je ne peux pas vous en dire plus, je sais simplement que cela commence à être fait.

La question que cela me pose me renvoie à un exemple concret. Nous avons à l'Ina, dans mon service, 1,2 million de photo sur les débuts de la télévision et de beaucoup de radios, sur de la

technique. Et il y a des photos où l'on ne sait pas reconnaître les personnes qui sont dessus, parfois on ne sait pas ce que c'est. Même en télévision, il y a des documents muets anciens qui nous inspirent peu. On se demande alors si on a une possibilité de demander aux internautes de nous aider, dans ce cas en reconnaissant des personnes. Je pense qu'à partir du moment où c'est publié par une institution comme la BnF ou l'Ina qui est publique, cela pose le problème de la fiabilité des informations. A partir du moment où l'on va mettre quelque chose sur notre site, il faut que les informations soient justes. Une des évolutions que je perçois pour le métier de documentaliste, c'est celle de la position de modérateur. En effet, si on ouvre ce genre de choses, il va falloir vérifier la fiabilité de ces informations.

Isabelle GIANNATTASIO

C'est une évolution de nos métiers.

Catherine TEILLOU

Oui, mais les métiers évoluent énormément en ce moment. C'est vrai que le rôle du documentaliste a évolué énormément sur une décade. Je ne vais pas dire qu'il a radicalement changé parce que c'est toujours de la documentation, c'est toujours mettre à disposition des utilisateurs des images. Mais il évolue évidemment.

Isabelle GIANNATTASIO

Je voudrais reprendre l'expression qu'a eue Emmanuelle BERMES tout à l'heure pour dire que « le numérique explose vers le haut », et parler de la question des collections. Pourriez-vous nous parler des thématisations, cette démarche de l'Ina qui a fait école ? Thématiser, est-ce constituer une collection virtuelle de documents qui peuvent éventuellement appartenir à plusieurs collections ?

Catherine TEILLOU

Oui, c'est tout à fait cela. Je vous ai parlé de notre volumétrie. C'est vrai que si vous arrivez dans les bases de l'Ina et que vous tapez François MITTERRAND, Jacques CHIRAC, Tour de France, ou des choses comme cela, vous allez avoir tellement de réponses, que cela aurait été impossible d'ouvrir les bases. Nous devons absolument guider nos utilisateurs. Nous l'avons fait dans deux directions :

- Nous avons essayé de structurer les fonds en proposant une sorte de cartographie. Nous avons pris les collections, en prenant la première et la dernière émission d'une collection, ainsi que quelques émissions marquantes, pour essayer de dire ce que nous avons dans notre fonds sur telle thématique. On peut donner une idée de l'amplitude et de la durée.
- Sur les élections présidentielles, nous avons essayé de proposer à nos utilisateurs une sélection organisée et cohérente. Soit c'est d'une façon thématique, soit c'est chronologique. Si c'est une personnalité, c'est bien sûr la personnalité telle qu'elle a été traitée à la télévision. Ce n'est pas nécessairement tous les aspects d'une personnalité si tous ces aspects n'ont pas été traités. Nous essayons de traiter aussi bien la vie publique que la vie privée, et de faire une sorte de « Best Of ». C'est-à-dire que si vous êtes pressés, vous verrez ce qui nous semble à nous, dans notre sélection, une vingtaine de sujets, d'émissions ou d'extraits qui vous donneront une idée de la personne, du thème ou de l'événement.

C'est véritablement un guide. Vous n'êtes pas obligé de suivre cette entrée, et cela ne vous empêche pas après d'aller dans l'intégralité du fonds.

Isabelle GIANNATTASIO

Merci. Est-ce que les participants de la table ronde pensent avoir énoncé tous les stratagèmes que nous pouvons avoir pour faire connaître nos fonds et les proposer ?

Thierry ROLLAND

D'abord, ce qu'il ne faut pas faire. Il y a beaucoup de gens qui arrivent aujourd'hui dans des centres d'archives en étant un peu en dehors du métier. Ils ont l'idée très simple de faire comme dans toute entreprise, d'éliminer les 90 % de la « marchandise » qui ne se vend pas. C'est une démarche qui est non seulement condamnable, au sens historique, mais également idiote. C'est ce que l'on va éviter. Après, nous avons le problème des grosses collections. Lorsqu'on tape « Guerre d'Algérie » et qu'apparaissent 30 000 notices, il est tout à fait important de faire ces espèces de compilations.

Il y a d'autres choses qui permettent d'aller vers le public par le média classique des émissions. Je voudrais parler de trois choses que j'ai vues faire pour thématiser et reformater des archives.

- 1) *La cassette anniversaire* : c'est quelque chose d'assez classique et qui fonctionne toujours. C'est ce que vous offrez à votre grand-père : par exemple, les actualités de 1938. À mon avis, c'est plus offert que regardé. Car je pense que les époques qui nous parlent le plus sont celles de notre jeunesse, pas celles de notre naissance. Offrir des choses sur lesquelles nous avons irrémédiablement une amnésie ne me semble pas pertinent, mais semble fonctionner.
- 2) *Une éphéméride* : c'est encore une chose très classique que nous avons faite chez Pathé pour France 5. Qu'est-il arrivé le 23 octobre 1954, 1968, 1972, etc. ? C'est aussi un produit des archives qui va fonctionner, notamment sur les sites des journaux.
- 3) *La répétition des choses* : c'est quelque chose que je trouve plus intéressant. Cette répétition du sujet sur le timbre anti-tuberculose, traité tous les ans de manière similaire mais avec de légères évolutions, me semble quelque chose de très intéressant. J'ai eu la chance de travailler sur une émission qui s'appelait *L'ABC d'hier*, et dans laquelle nous reprenons, en six minutes, par ordre alphabétique, trois sujets qui nous permettent de remonter des choses. Nous avons fait : A comme Année nouvelle ; B comme Blum ; C comme Chien ; D comme Dirigeable ; et on a décliné cela.

Je pense qu'il y a plein de choses à créer qui permettent de rendre attrayantes les archives. Maintenant, je pense qu'il faut plutôt faire des guides-mémoires, car on aime bien cette espèce de nostalgie lorsqu'elle est encadrée. Il faut déjà supposer que l'on a gardé une notion de collection, ainsi qu'une chose qui semble toute bête : l'intégrité des documents. Chez Pathé, il y a eu pendant très longtemps un fonds qui s'appelait « France Actualités », qui est le pendant de « La France en marche ». Ce fonds fut longtemps commercialisé muet. Heureusement, le son a été préservé et transféré. Mais je crois qu'il est important d'avoir une notion d'intégrité de collection.

Isabelle GIANNATTASIO

Cela veut également dire qu'il faut conserver les conservateurs, puisque cette mémoire des fonds est aussi indispensable. Peut-on arriver à capitaliser ces connaissances, à faire des bases de connaissances ?

Anne-Marie GRAPTON

Je voudrais réagir à ce qu'a dit Thierry ROLLAND, pour dire que l'on a beaucoup de films sur la tuberculose, et notamment sur les timbres. On s'interroge sur l'indexation pour savoir si on met « prévention » ou « prophylaxie ».

Je voudrais saluer la granularité de votre approche. On n'a pas toujours la possibilité d'entrer dans ce détail et cette finesse. Je mentionnerais une approche qui rejoint la bibliothèque numérique parce que cela veut dire proposer pour le plus grand nombre, afin que quelqu'un trouve quelque chose pour lui. Aujourd'hui, je participe à un groupe de travail européen pour un projet de normes. Bien sûr, il y a une crainte devant l'arrivée de nouvelles règles se rajoutant aux difficultés en interne. Ce projet de normes européennes porte sur l'identification des œuvres cinématographiques et l'interopérabilité des bases de données. Ce projet est intéressant mais un peu compliqué, parce que c'est aussi un système, un mandat confié par la commission européenne, qui est parfois très loin de la réalité du terrain. Il n'y a pas toujours une compréhension des besoins. Il est intéressant aussi de se confronter aux autres pays, alors que nous sommes rarement confrontés à d'autres pratiques, d'autres usages et d'autres questionnements. Nous pouvons nous rendre compte que nous avons tous les mêmes problèmes ou les mêmes vœux. Il y a deux notions importantes dans ce projet : identité des œuvres, c'est-à-dire pouvoir distinguer une œuvre et son nom parmi la masse de documents ; et aussi, comme nous ne pensons pas tous pareils, l'idée n'est plus d'avoir une base partagée, comme cela a été une expérience et l'est encore par exemple pour la base LISE des Archives françaises du Film, qui est partagée avec la Cinémathèque française et plusieurs cinémathèques régionales. La base partagée est déjà une première rencontre. Mais dans l'approche européenne, l'idée est d'être dans la diversité et pour le bruit, et l'identité dans le bruit. On ne va pas refaire tous ces réservoirs, tous ces catalogues, toutes ces descriptions, tout ce travail d'humain. Et je comprends les questions d'Emmanuelle sur la façon d'automatiser ce travail. L'idée est d'essayer d'harmoniser les pratiques descriptives. C'est ce qu'on fait aussi avec, par exemple, la Fédération Internationale des Archives de Film (FIAF), qui est en train de réviser ses règles de catalogage. L'harmonisation passera par les méta-données et un modèle conceptuel de références. Cela nous permettra de communiquer entre nous et, dans un futur que je souhaite proche, de permettre à n'importe quel internaute d'accéder à la ressource qu'il veut voir.

Isabelle GIANNATTASIO

Merci de nous avoir mis en garde contre la Tour de Babel et d'avoir suggéré plutôt un portail.

Suivi éditorial : Loraine Tambrun – Chargée de mission pour le patrimoine cinématographique / INP.