# Projet 2 : Le corpus vidéo du CEM

Présentation

Rappel : le fonctionnement de la cellule audio vidéo

Mission : coréalisation de montage, post projet de films doc, diffusion et formation

Motivation : Rendre accessible et préserver l'intelligibilité des données donc faire les migrations des formats.

Corpus : 153 documents vidéo de plus de 200h depuis 2001. Depuis 2012, sauvegarde d'une sélection de rush.

Diffusion : Cnrs images, archives.org, viméo. Tout est en ligne sur le site (spip) du Centre Edgar Morin. Les vidéos sont embarquées depuis archives.org.

Quand il s'agit d’un Master l’accès est libre sauf pour quelques cas, par exemple certaines vidéos de Daniel Friedman (éditées par les éditions Montparnasse et donc restrictions dans la diffusion).

Sauvegarde actuelle : Copie pour le dépôt légal à la BNF en format MP4 en résolution standard. Pas de sauvegarde physique. Seulement un master et un site miroir mais dans la même pièce, donc situation précaire.

Information de descriptions :

* le numéro de master, la durée,
* la date de production,
* la version,
* les chercheurs auteurs/réalisateurs, montage
* un petit résumé
* mots clefs pour les films réalisés après de 2012 (archive.org).

RQ : Chaque titre est numéroté : un titre peut contenir plusieurs vidéos. Ces vidéos peuvent avoir plusieurs auteurs... un titre peut par exemple être une série d'entretiens sur une thématique précise. Chaque vidéo a un master (format sans compression et résolution d’origine).

Dans la fiche web du titre, les vidéos sont identifiées sous forme de liste, sous le même numéro (ex. : L’archive des festivals panafricains : http://www.iiac.cnrs.fr/CentreEdgarMorin/spip.php?article683)

Rosine Lheureux remarque que ce choix ne permet pas facilement de repérer le volume en termes de vidéos disponible et propose de numéroter chaque vidéo de série – 080-1, 080-2, etc.

Etape du travail

1. Etablir un protocole de compression.
2. Vérifier la validité du format, par FACILE du Centre informatique de l'enseignement supérieur (CINES)
3. Décrire les vidéos. Arghyro souligne que ce n’est pas son domaine de compétence et qu’elle est actuellement seule en poste à la cellule audio-vidéo. C’est pour cela que le Centre Edgar Morin/IIAC a demandé du soutien au consortium.
4. Enfin, il est envisagé de déposer, les données sur Nakala. Il y a des champs à remplir et cela nécessite une collaboration avec les chercheurs, notamment pour la description du contenu.

Pour le stockage sur Nakala il n’y a pas restrictions en volume.

Arghyro évoque ensuite les choix de formats vidéo pour l’encodage et la diffusion : (les conteneurs : fichiers de type archive contenant d'autres fichiers en général des flux vidéo et/ou audio... Les formats de conteneurs vidéo les plus répandus sont AVI, MOV, MPG, MKV/MKV (Matroska), OGM, FLV, 3GP,...

Les codecs : programmes qui (de)compressent les flux vidéo et/ou audio rassemblés dans les conteneurs 🡪 MPEG-2**,** pour diffuser du contenu SD par DVD, TNT, et MPEG-4, famille de normes dont l’usage est particulièrement adapté au web et aux périphériques mobiles.

Elle réalise au moins deux versions, l’une pour la conservation, l’autre pour la diffusion.

Le master est toujours sauvegardé dans le format dans lequel il a été monté et ensuite il est décliné dans les formats requis par les différents canaux de diffusion.

Le format du Master peut différer, étant donné que chaque chercheur filme avec ses moyens (SD, HD, etc…), Arghyro se charge du montage dans la meilleure qualité possible et de la conformation.

Etablir un cahier des charges.

La question du catalogage et niveau de description

Après discussions, il semble qu’il faut favoriser :

* Une description à l’unité. C’est-à-dire non pas une fiche par master mais une fiche documents audiovisuels (les films).
* Puis établir des collections de documents audiovisuels.
* On peut donc partir des numéros de master et affiner le catalogage ensuite par la création collection pour les masters qui regroupent plusieurs films. Mais il est important de donner un identifiant unique à chaque document audiovisuel.
* Exemple pour la mise en place d’un système de cotation/identification( par ex : pour le master 80 qui possède x films, créer un identifiant 80-1 pour le premier films, 80-2 pour le second, etc…) . Pour ce travail, Arghyro pourra être accompagnée par des archivistes et ou documentalistes du comité technique, qui rappelons-le, n’est ni archiviste, ni documentaliste.

Pour la question de la description des documents audiovisuels.

* On peut déjà s’appuyer sur la liste des champs de description proposée par Nakala et sélectionner ceux qui sont les plus pertinents pour la description d’un document audiovisuel.
* Aurélie Hess propose aussi de partir des 15 champs de description du Dublin Core en les adoptant au document audiovisuel. (cf. dans la présentation Plozcorpus, le protocole de description utilisé pour les ressources audio)
* Objectif : Etablir un modèle de fiche type avec les champs de descriptions qui auront été sélectionnés. Là aussi, il est proposé que le comité technique vient accompagner ce travail. Le prototype pourrait ensuite être soumis à l’expertise de l’ensemble du comité technique pour discussion et validation.

Une fois le protocole établi : créer les notices de descriptions par unité (une fiche par films)

A ce niveau, l’enveloppe budgétaire que le CS du consortium a mis à disposition du Centre Edgar Morin pour ce projet pourra être utilisé pour le recrutement d’un stagiaire ou d’un contractuel.

Une question se pose :

Réaliser les fiches directement sur Nakala ou en amont et faire une saisie sur Nakala par la suite? La deuxième solution semble être privilégiée, car permettra de disposer de données et métadonnées réutilisables et potentiellement moissonable par d’autres services que ceux d’Huma-Num.

Avantage et inconvénients : Nakala et moissonnage

Serge Wolikow rappelle qu’il est envisagé d’adresser à l’équipe d’huma-num, de manière collective, une série de questions partagées par plusieurs des membres du consortium au sujet de Nakala. On pourrait leur demander par exemple s’ils peuvent ou non implémenter des fichiers Excel ou des extractions Drupal dans Nakala.

Hédi Maazaoui rappelle brièvement les fonctions de Nakala ne fait pas de l'archivage pérenne, ce n'est pas parce que le format est valide et stocké qu'il est archivé pour 30 ans. Néanmoins c’est un premier pas pour l’archivage des données. Nakala génère des entrepôts OAI PMH qui sont ensuite accessibles directement depuis ISIDORE, mais c'est surtout une interface de saisie de notices normalisées.

Serge Wolikow se demande si Nakala est le plus à même de recevoir les films d'auteurs achevés.

Ne devrait-on pas envisager de se rapprocher de Canal U et des plateformes audiovisuelles du Réseau De MSH ? Canal U en cours de restructuration tend à avoir vocation à gérer tous les films de chercheurs et enseignants. La vocation de Nakala c'est déjà d’archiver les archives vidéos/audio qui ont servi au travail intermédiaires de la recherche, mais pas forcément des œuvres terminés.

Si l’on travaille avec Nakala, il faudrait donc aussi avoir un moyen de valoriser les ressources archivées avec Isidore déjà, mais aussi avec d’autres catalogues et plateformes plus spécialisés dans les archives audiovisuelles. Ce qui conforte l'idée de travailler en amont car plus le travail est fait plus les fichiers seront facilement susceptibles d'être éditorialiser dans différents lieux.

Ressources et compléments :

Arghyro Paouri cite quelques ressources :

- Écrire un cahier des charges de numérisation de collections sonores, audiovisuelles et filmiques, BNF :

<http://www.culturecommunication.gouv.fr/Politiques-ministerielles/Recherche-Enseignement-superieur-Technologies/Innovation-numerique/La-numerisation-en-pratique/Recommandations-techniques>

- Guide méthodologique pour le choix de formats numériques pérennes dans un contexte de données orales et visuelles, TGE ADONIS :

<http://www.tge-adonis.fr/ressources/guides>

-Logiciel d’annotation de films, Ligne de Temps de l’IRI :

<http://www.iri.centrepompidou.fr/fr/atelier.html>

Une autre ressource évoquée : le séminaire archives audiovisuelles mis en place par le consortium ArcMC. Serge rappelle que la dernière séance-atelier devrait servir à poser un programme pour l’établissement de préconisation a minima pour traiter des grands types de documents audiovisuels de la recherche. Christian Dury souligne que les équipes de recherche ont des usages différenciés des documents audiovisuels, pour des coopérations avec la télévision, à des fins de recherches, pédagogiques.

Aurélie Hess évoque Omeka qui est un logiciel libre de gestion de bibliothèques numériques ( cf. une présentation par une équipe de recherche. <http://alambic.hypotheses.org/4924>) . Elle l’utilise. C’est une bibliothèque (qui est d’office en Dublin core). C'est plus simple que drupal, c'est au départ pour faire des bibliothèques numériques et ensuite par ajouts de plugins, on peut permettre de rendre les données moissonables.

Christian Dury évoque la rénovation du site des ressources audiovisuelles de l’ISH de de Lyon : site 25image/SHS (**25images**.**ish**-lyon.cnrs.fr/). C’est un site Drupal, qui reprend les 500 heures videos.

Les descriptions sont réalisés mais pour l’instant un peu bloqué par l'OAI PMH, qui n’est pas encore développé sur le site, mais là aussi des plugins peuvent permettre de le mettre en place.

Réutilisation des données et Citabilité

Aurélie Hess évoque le travail de réutilisation des documents audiovisuels. Ils peuvent devenir des sources pour l'historien. C’est donc nécessaire de bien les documenter. Mais un problème se pose, il n’existe pas aujourd'hui vraiment de système de citation des sources audiovisuelles. Comment faire pour les citer et les faire reconnaitre comme sources.

Sébastien Jacquot parle de son travail autour de la question de la citabilité qu’il a mené dans le cadre du portail FANA. ( cf. CR de la Séance 5 archives audiovisuelles).