

# ESCoM-AAR 2014: Rapport d'activités de l'Equipe Sémiotique Cognitive et nouveaux Médias - Archives Audiovisuelles de la Recherche.

Peter Stockinger

► **To cite this version:**

Peter Stockinger. ESCoM-AAR 2014: Rapport d'activités de l'Equipe Sémiotique Cognitive et nouveaux Médias - Archives Audiovisuelles de la Recherche.. [Rapport de recherche] ESCoM-AAR (Equipe Sémiotique Cognitive et Nouveaux Médias - Archives Audiovisuelles de la Recherche). 2014. hal-01330114

HAL Id: hal-01330114

<https://hal-inalco.archives-ouvertes.fr/hal-01330114>

Submitted on 10 Jun 2016

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



# ESCoM-AAR

## 2014

The screenshot shows the homepage of the Archives Audiovisuelles de la Recherche (AAR). At the top, there is a banner with a world map and the text "Archives Audiovisuelles de la Recherche". Below this is a navigation menu with links: Accueil, Les AAR, Equipe, Comité scientifique, Documents, Dons & Mécénat, Votre avis, Nous contacter, Liens. There is also a search bar and a "Recherche avancée" button. The main content area is divided into several sections:

- La vidéothèque AAR**: Includes links for "Domaines et disciplines", "Les collections", "Vidéos en version originale", "Les auteurs", and "Tout le fonds audiovisuel".
- Les éditions AAR**: Includes links for "Tous les dossiers", "Dossiers par discipline", "Dossiers pédagogiques", "Profil et niveau de connaissance", "Les dossiers bilingues", and "Les encyclopédies AAR".
- Actualités**: Features a "Newsletter" sign-up form, a section for "Vidéos de la Semaine" with a video titled "Le féminisme en mouvements", and a section for "LES ACTUS" with a video titled "L'après-sujet, séance du 16 novembre 2012".
- LES ARCHIVES AUDIOVISUELLES DE LA RECHERCHE**: A section highlighting film resources from various AAR portals.

At the bottom, there is an "ATTENTION" notice: "Aucun tournage n'est assuré par l'ESCoM et le programme de financement recherche de ses membres. Vidéo".

30/10/2014

Rapport d'activités de l'Equipe Sémiotique  
Cognitive et nouveaux Médias - Archives  
Audiovisuelles de la Recherche

## Sommaire

1. LE PROGRAMME ESCOM-AAR .....	2
1.1. Présentation synthétique.....	2
1.2. Composition de l'équipe et conseil scientifique.....	2
2. ACTIVITÉS DE R&D.....	3
2.1. Axe 1 : AAR – Archives Audiovisuelles de la Recherche : <i>Archivage numérique et diffusion de patrimoines scientifiques et culturels</i> .....	3
2.1.1. Le fonds AAR .....	4
2.1.2. Les publications de la banque des données audiovisuelles des AAR.....	4
2.1.3. Depuis mai 2013 – une « nouvelle génération » de portails AAR .....	7
2.1.4. Réappropriation de corpus audiovisuels et republications – l'ingénierie du « repurposing » .....	9
2.2. Axe 2 : Studio ASA (Atelier Sémiotique Audiovisuel) : <i>Plateforme de travail scientifique et éditorial des AAR</i> .....	10
2.2.1. L'environnement AAR.....	11
2.2.2. L'environnement ASA .....	12
2.2.3. Le Studio ASA (Atelier Sémiotique Audiovisuel) .....	13
2.3. Axe 3 : Campus AAR : <i>Transfert et partage des connaissances et technologies pour l'archivage et la diffusion de patrimoines scientifiques et culturels</i> .....	14
2.3.1. Les objectifs du projet Campus AAR.....	15
2.3.2. Partenariats consacrés à la mise en place d'archives audiovisuelles .....	16
2.4. Axe 4 : e-Sémiotique : <i>Recherche et développement sur les nouveaux écosystèmes de l'information et de la connaissance</i> .....	16
2.4.1. Le « semiotic turn » dans les archives numériques .....	17
2.4.2. Méthodologie Projet Archives numériques et remédiation : Procédures, documentation et outils	20
2.4.3. Eco-systèmes d'information, acteurs sociaux et patrimoines de connaissance.....	22
3. RÉFÉRENCES BIBLIOGRAPHIQUES .....	24

# 1. LE PROGRAMME ESCOM-AAR

## 1.1. Présentation synthétique

L'ESCoM-AAR est un programme de recherche, hébergé à la Fondation Maison des Sciences de l'Homme (FMSH) à Paris et créé en 1990 par Peter Stockinger, professeur à l'Institut National des Langues et Civilisations Orientales (INALCO) avec l'aide de Charles Morazé, co-fondateur de la FMSH, Clemens Heller (alors administrateur de la FMSH) et François de Labriolle (alors président de l'INALCO).

Les activités de recherche et de développement de l'ESCoM-AAR portent aujourd'hui sur la problématique de la réalisation d'archives numériques et de l'utilisation/réappropriation des archives numériques dans des contextes sociaux spécifiques tels que l'enseignement, la recherche, la conservation et la valorisation de patrimoines scientifiques et culturels.

En 2000, Peter Stockinger a conçu et mis en place avec les chercheurs et ingénieurs de l'ESCoM et d'un réseau scientifique national et international, le programme « Archives Audiovisuelles de la Recherche » (AAR) – programme qui forme, aujourd'hui, une plateforme à la fois scientifique et technique pour la production, la description/indexation, la diffusion et la valorisation des connaissances en SHS à l'aide des technologies du numérique.

Une grande partie des réalisations du programme ESCoM-AAR est consultable en ligne :

1. Le portail central du programme AAR qui propose un accès à plus de 6800 heures de vidéos documentant plus que 1100 manifestations ou événements scientifiques : <http://www.archivesaudiovisuelles.fr/> .
2. Le portail Semioscape – la plateforme en ligne de publication du programme AAR qui offre le téléchargement du Studio ASA développé par l'ESCoM-AAR et ses partenaires. Le Studio ASA est une suite de logiciels spécialisés pour l'analyse de corpus audiovisuels et leur publication sous forme d'archives personnalisées : <http://www.semioscape.fr/> .
3. Le site de l'ESCoM-AAR qui présente d'une manière détaillée toutes les activités de l'équipe ainsi que, plus particulièrement, toutes les publications depuis 1996 (monographies, ouvrages collectifs, articles scientifiques, conférences, cours, ...) : <http://aar.hypotheses.org/> .
4. Le carnet de recherche du projet ASA-SHS (Atelier de Sémiotique Audiovisuelle pour les SHS) qui décrit *in fine* les enjeux scientifiques (sémiotiques), technologiques et pratiques des archives numériques pour les « digital humanities » : <http://asashs.hypotheses.org/> .

## 1.2. Composition de l'équipe et conseil scientifique

**1) Directeur:** Peter Stockinger - P.U. INaLCO.

**2) Programme AAR (Archives Audiovisuelles de la Recherche):**

1. Elisabeth de Pablo (FMSH): responsable éditorial du Programme AAR.
2. Francis Lemaitre (FMSH): responsable informatique du programme AAR.
3. Neli Dobрева (FMSH): communication, valorisation et production de contenus.
4. Gaël Vilpoix (FMSH): administration de l'infrastructure informatique.

**3) Projet ANR Campus AAR :**

1. Valérie Legrand (FMSH - INALCO): Recherche et développement.

2. Dominique Flament (CNRS): Archives Histoire des Mathématiques (AHM).
3. Gabrielle Laumonier (FMSH - INALCO): suivi administratif du projet Campus AAR.

#### 4) Chercheurs associés :

1. Véronique Alexandre Journeau (AIDIAA-Association internationale de dialogues artistiques, Paris-Sorbonne, OMF et Creops): Archives Langarts - audiovisuelles, arts et esthétique.
2. Tarek Ouerfelli (Université de la Manouba à Tunis, Institut Supérieur de Documentation): Archives audiovisuelles numériques.

#### 5) Comité Scientifique:

1. Barbara Glowczewski (directeur de recherche; Collège de France-LAS/CNRS).
2. Florence Descamps (EPHE).
3. Monica Schpun (CRBC - EHESS).
4. Richard Pottier (Université de Paris V).
5. Steffen Lalande (INA).
6. Patrick Courounet (INA).
7. Laurence Favier (professeur des universités ; GERIICO - Université de Lille 3).
8. Widad Mustafa (professeur des universités ; GERIICO – Université de Lille 3).

## 2. ACTIVITÉS DE R&D

Les activités de l'ESCoM-AAR s'organisent autour de quatre axes complémentaires :

- 1) **Axe 1** : AAR – Archives Audiovisuelles de la Recherche : Archivage numérique et diffusion de patrimoines scientifiques et culturels ;
- 2) **Axe 2** : Studio ASA (Atelier de Sémiotique Audiovisuelle) : Plateforme de travail scientifique et éditorial des Archives Audiovisuelles de la Recherche ;
- 3) **Axe 3** : Campus AAR : Transfert et partage des connaissances et technologies pour l'archivage et la diffusion de patrimoines scientifiques et culturels ;
- 4) **Axe 4** : e-Sémiotique : Activités de R&D consacrées aux nouveaux écosystèmes de l'information et des connaissances.

### 2.1. Axe 1 : AAR – Archives Audiovisuelles de la Recherche : Archivage numérique et diffusion de patrimoines scientifiques et culturels

Les Archives Audiovisuelles de la Recherche (AAR) forment un *programme de recherche appliquée* de l'ESCoM dans le domaine des archives/bibliothèques numériques pour la recherche scientifique, l'enseignement et la valorisation de patrimoines scientifiques et culturels. Il a été élaboré en 2000, dans le cadre d'un projet de R&D français s'intitulant OPALES (coordinateur : INA) et mis en place concrètement à partir du mois d'avril 2002. Deux objectifs sous-tendent les AAR :

- a) *Constituer et préserver* des corpus audiovisuels numériques qui documentent le patrimoine scientifique dans les différentes « disciplines » en SHS.

b) *Publier et diffuser* en ligne une partie des corpus archivés sous forme d'événements (conférences, séminaires, entretiens, recherches de terrain, ...) et de collections (« dossiers ») d'événements scientifiques en français et dans d'autres langues.

### 2.1.1. Le fonds AAR

Entre 2002 et 2009, l'ESCoM-AAR a déployé une intense activité de production de données audiovisuelles documentant la très grande richesse du patrimoine scientifique et culturel des SHS. Depuis 2009 les activités de production ont été complètement externalisées pour des raisons budgétaires internes à la FMSH. Voici quelques caractéristiques du fonds des AAR tel qu'il se présente aujourd'hui, en 2014 :

#### 1) Types de données :

- a) *Données vidéos* : enregistrements bruts, enregistrements traités a posteriori et montés par l'ESCoM-AAR, documents filmiques édités (reportages, documentaires, ...).
- b) *Données visuelles statiques* : photographies brutes, photographies traitées à posteriori ; photographies d'archives de chercheurs.
- c) *Données textuelles* : (pré-)publications des contributeurs, CV, bibliographies, prépublications, « working papers ».
- d) *Données structurées* : une base de données contenant des informations sur les auteurs du fonds AAR (environ 3100 personnes), sur les abonnés à la newsletter AAR (plus que 10.000 abonnés) et sur quelques 500 visiteurs/utilisateurs des AAR (questionnaires en ligne réalisés entre 2008 et 2010).
- e) *Données à caractère juridique* : contrats de mise en ligne (environ 3200 contrats signés et classés).

**2) Volume du fonds** (vidéos seulement) : environ 7000 heures de vidéos produites et/ou coproduites par l'ESCoM et un réseau national et international de partenaires; quelques 5800 heures de ressources AV en accès libre (« open access ») sur le portail général du programme AAR; des milliers d'heures de ressources AV également librement accessibles sur une série de portails plus spécialisés (cf. ci-après).

**3) Formats des fichiers vidéo** : Les formats de stream utilisés actuellement sont *mpeg4*, et – sporadiquement – *wmv* (Microsoft) et *flash* (Adobe).

**4) Auteurs**: *communauté internationale* d'environ 3100 chercheurs-enseignants, artistes, journalistes, hommes/femmes politiques, mais aussi témoins (victimes politiques, réfugiés, descendants de colons européens, représentants d'anciens métiers, ...) – communauté originaire de quelques 75 pays du monde.

5) Le fonds AAR par **langue** : une très grande partie des vidéos est monolingue ; une petite partie est monolingue assorti d'un résumé en une ou plusieurs autres langues (exemple *français + résumé en russe ; français + résumé en hébreu ; anglais + résumé en japonais ; ...*) ; une grande partie des vidéos est en français (environ 65%) ; vidéos également produites en anglais (plus de 500 heures), espagnol (plus de 400 heures), allemand, chinois, japonais, russe, arabe, etc.

### 2.1.2. Les publications de la banque des données audiovisuelles des AAR

Le fonds des AAR est publié et diffusé en ligne sous forme de 3 genres de base différents :

1. Le genre principal est l'**événement AAR** (« événement scientifique », « événement culturel », ...) : chaque événement filmé est publié sous forme d'un site d'événement ou encore d'un site événementiel. Sur le portail central du programme AAR, on peut accéder, aujourd'hui, à environ 670 sites d'événements ; l'ensemble des portails du programme AAR compte environ 1030 sites événementiels.

2. Les **collections des AAR** : « *Collection* » veut dire ici *ensemble cohérent* de sites d'événements. Un ensemble de sites d'événements constitue un *corpus publié* (enligne) d'événements. Les AAR comportent, aujourd'hui, *40 grands corpus* d'événements audiovisuels et quelques *180 corpus plus circonscrits*.

Exemple :

Le corpus *Sociologie* comporte, aujourd'hui, 160 sites d'événements (représentant quelques 960 heures de vidéo en ligne).

Ce corpus comprend, à son tour, 46 corpus spécialisés comme, par exemple, le corpus spécialisé *Sociologie urbaine* (composé de 8 sites événementiels), le corpus spécialisé *Sociologie de la culture* (composé de 27 sites événementiels) ou encore le corpus spécialisé *Sociologie des migrations* (composé de 11 sites événementiels), ...

Un corpus AAR est composé d'enregistrements d'entretiens (avec chercheurs, experts, journalistes, témoins, ...), de séminaires et/ou de colloques (tables rondes, journées d'étude, ...) mais aussi de documentaires AV, d'enregistrements de terrain documentant le thème/le sujet, d'enregistrements d'événements culturels (expositions, concerts, ...), de prises de vue de « scène de vie ».

(figure 1 : le portail central des AAR)

3. Les **portails des AAR** (portail général du programme AAR; portails à thème). Un portail est un point d'accès central aux événements et/ou collections d'événements. Les AAR sont composés : d'un portail central<sup>1</sup> (cf. figure 1) en huit langues et d'une série de *portails spécialisés* (i.e. de portails

<sup>1</sup> <http://www.archivesaudiovisuelles.fr/FR/>

proposant à leur tour des accès communs et/ou spécifiques à une sélection du fonds audiovisuel de la banque des données ouvertes AAR):

- Portail AVERROES – La médiathèque France-Maghreb<sup>2</sup> (cf. figure 2) ;
- Portail PCM – Peuples et Cultures du Monde<sup>3</sup>,
- Portail MDD – Mondialisation et Développement Durable<sup>4</sup>,
- Portail DLC – Diversité Culturelle et Linguistique<sup>5</sup>,
- Portail SCC – Semiotica, Cultura e Comunicazione<sup>6</sup> (coopération avec l'Università di Roma – Sapienza), ...



(figure 2 : le portail spécialisé Averroès – la médiathèque France-Maghreb)

Le portail général AAR ainsi que les portails spécialisés proposent un ensemble d'accès standards qui permettent d'explorer la richesse du patrimoine audiovisuel diffusé gratuitement :

- accès par « nouveautés » : publications les plus récentes; les « grands titres », ... ;
- accès par domaines et disciplines ;
- accès par types d'événements ;
- accès par langue (version originale des vidéos) ;
- accès par auteurs ;
- accès à des dossiers hypermédias spécialisés issus d'un processus de republication de corpus audiovisuels ; ...

<sup>2</sup> <http://www.france-maghreb.fr/FR/>

<sup>3</sup> <http://www.culturalheritage.fr/FR/>

<sup>4</sup> <http://www.evolutiondurable.fr/FR/>

<sup>5</sup> <http://www.languescultures.fr/FR/>

<sup>6</sup> <http://www.archiviosemiotica.eu/IT/>



### 2.1.3. Depuis mai 2013 – une « nouvelle génération » de portails AAR

Grace à une série de projets de R&D européens (6<sup>ème</sup> et 7<sup>ème</sup> PCRD) et français (ANR) dans lequel le programme ESCoM-AAR a été impliqué en tant que partenaire ou coordinateur, une *nouvelle génération de portails audiovisuels* a pu voir le jour depuis **mai 2013**.

Les principales caractéristiques de cette nouvelle génération de portails sont les suivantes :

- Accès sémantique à des *corpora* audiovisuels choisis (accès par *thesaurus*, accès par *thèmes* ou *sujets de discours*, accès par *localisations géographiques* et *temporelles*, accès aux *références culturelles*, accès aux corpus par genre de discours ; accès aux corpus par profil d'usage ; etc.).
- Descriptions et indexations riches et détaillées des vidéos pour des usages dans l'enseignement et la recherche.
- Republications « auteurs » totalement intégrées dans le processus de gestion éditorial d'un portail sous forme de dossiers thématiques, de parcours narratifs, de vidéo-lexiques, etc. (cf. ci-après, section 2.1.1.4).
- Customisation de l'interface physique et graphique ainsi que des accès aux corpus audiovisuels.
- Intégration des médias sociaux dans chaque portail et chaque « page vidéo » (i.e. interface d'exploration d'une ressource vidéo).

The screenshot shows the website 'Les mondes de l'Amérique latine' with a navigation menu and two video thumbnails. The first video is titled '6. Carlos HERRERA : Naissance du parti socialiste argentin' and the second is '12. Carlos HERRERA : L'industrialisation et l'immigration intérieure dans les années 1930'. Both videos are in French and feature Carlos Herrera.

(figure 3 : le portail spécialisé AMSUR – Les mondes de l'Amérique Latine)

Ces améliorations significatives ont été rendues possibles grâce à l'environnement de travail « unifié » Studio ASA de l'ESCoM (cf., ci-après, l'Axe 2).

La figure 3 montre l'exemple de la page d'accueil du portail AMSUR<sup>7</sup> donnant accès à quelques 380 heures de vidéos finement analysées et indexées. On peut y identifier trois barres de menu supérieures :

1. La barre supérieure proposant l'accès à des rubriques habituelles d'un site/portail ;
2. la barre du milieu « Publications » donne accès aux publications/republications spécialisées (dossiers thématiques, dossiers pédagogiques, etc.) ;
3. la barre inférieure propose des accès par sujets, thesaurus, localisation géographique, contributeurs, collections filmiques, genres de discours, etc.



(figure 4 : La « page vidéo » sur le portail AMSUR permet l'exploration systématique d'une ressource vidéo)

Enfin, la figure 4 montre encore l'exemple de l'interface (adaptable aux métadonnées descriptives) de l'exploration d'une ressource audiovisuelle (d'une vidéo en entière, d'un segment, de vidéos similaires à une vidéo donnée, ...) :

- exploration d'une ressource audiovisuelle séquence par séquence ;
- présentation de l'ensemble de la ressource audiovisuelle ;
- présentation d'une séquence sélectionnée par le lecteur/visiteur du portail ;
- propositions de ressources documentaires (filmiques, photographiques, textuelles, ...) supplémentaires pour mieux comprendre un sujet, une séquence ;
- traductions (synthétiques, libres, adaptées au public, ...) d'une séquence ;
- propositions dans quel cadre et comment utiliser une séquence ;

<sup>7</sup> <http://www.amsur.msh-paris.fr/>

- propositions d'explorer des ressources audiovisuelles similaires (par genre, par thème, par mot-clé, par auteur, ...);
- mentions légales (relatives à la ressource vidéo, à l'analyse produite, ...).

#### 2.1.4. Réappropriation de corpus audiovisuels et republications – l'ingénierie du « repurposing »

Le portail AMSUR donne accès à un *corpus audiovisuel* qui provient de la *banque de données ouvertes AAR* et est consacré à l'histoire et aux cultures de l'Amérique Latine. Ce corpus constitue donc *une partie du fonds commun* de la banque des données des AAR (environ 380 heures sur une totalité de 7000 heures). Il est également accessible via le *portail général des AAR* mais uniquement par le biais des accès généraux du portail (qui « valent » pour toutes les vidéos composant la banque des données AAR). Sur le portail AMSUR, le corpus en question se présente sous une forme *profondément retravaillé selon les objectifs éditoriaux propres au portail AMSUR*.

The screenshot shows the AMSUR website interface. At the top, there is a navigation bar with links like 'Accueil', 'Les AAR', 'Equipe', 'Comité scientifique', 'Documents', 'Dons & Mécénat', 'Votre avis', 'Nous contacter', and 'Liens'. Below this is a search bar and a language selection menu. The main content area is titled 'La musique ouïgoure' and includes a 'Portefeuille de lecture pour un cours de civilisation ouïgoure portant sur la musique, niveau master'. There are tabs for 'Accueil', 'Passages', 'Thématiques', and 'Lexique'. The central video player area displays the title 'Un regard ethnopsociologique sur les musiques chinoise et ouïgoure', its duration (0h 56mn 02s 62), and language (Français). It also features a list of 'Intervenant(s)', including Sabine Trebinjac, and a detailed biography of her. On the left side, there are sections for 'Dossiers hypermédias' and 'Dossiers pédagogiques', each with multiple entries in different languages.

(figure 5 : Accès à une série de republications d'un entretien avec l'ethnomusicologue Sabine Trebinjac du CNRS : republication sous forme de plusieurs dossiers bilingues et pédagogiques)

Un *corpus audiovisuel* « profondément retravaillé » veut dire que les vidéos qui font partie du corpus sont :

1. **localisées** dans le fonds (AAR) et **sélectionnées** selon leur pertinence en référence à la politique éditoriale du portail ;
2. **redécoupées** en segments pertinents (selon la politique éditoriale propre au portail) ;
3. **reclassées** à l'aide d'un métalangage (i.e. d'une *ontologie de domaine*) propre au portail ;
4. **re-décrites** et **re-indexées** à l'aide d'une bibliothèque de modèles de description propre au portail ;

5. **re-contextualisées** (i.e. mises en relation avec d'autres données selon la politique éditoriale du portail) ;
6. « **mélangées** » avec d'autres données afin de donner lieu à de nouvelles publications.

Ces six étapes composent la problématique du « *repurposing* » (aussi : *réappropriation, réécriture, republication, re-documentarisation, ...*) des données numériques archivées. Elles font aujourd'hui *intégralement* partie du travail d'archivage numérique des patrimoines culturels et scientifiques par le programme ESCoM-AAR.

Les genres de republications expérimentés à l'ESCoM-AAR depuis 2007 sont les suivants :

- Les **vidéo-lexiques** : Corpus de *segments assez brefs* de vidéos qui illustrent un terme, une notion, un domaine, ... *Exemples* : Le vidéo-lexique « *Peuples du monde* »<sup>8</sup> ; Le vidéo-lexique « *Introduction à la Sémiotique* » (en italien, français et anglais)<sup>9</sup>.
- Les **dossiers bilingues** : Republication d'une vidéo (ou d'une partie d'une vidéo) dans une *langue-cible différente* de la langue-source. *Exemples* : Republication (en arabe, chinois, anglais, italien, ...) d'un *entretien en français avec Sabine Trebinjac (CNRS) consacré à la musique ouïgoure*<sup>10</sup> (figure 5).
- Les **dossiers thématiques** : analyse et explication des grandes *thématiques* dans une prestation (entretien, séminaire, ...) ; les vidéos sont *totalemt délinéarisées!* *Exemples*: Republication – en français et arabe - d'une conférence de Joëlle Allouche-Benayoun (02/04/2008) sur *La place de la femme dans le judaïsme* (passages thématiques – encyclopédie – lexique)<sup>11</sup>.

## 2.2. Axe 2 : Studio ASA (Atelier Sémiotique Audiovisuel) : Plateforme de travail scientifique et éditorial des AAR

L'objectif du programme AAR lancé en 2002 par l'ESCoM n'étant pas seulement de constituer un projet d'archivage numérique du patrimoine scientifique des SHS, une partie importante de ses activités de R&D depuis 2004 a été consacrée au développement d'une *plateforme scientifique et technique* pour la production et la gestion d'archives audiovisuelles ainsi que la description et la publication du fonds faisant partie d'une archive. Cette plateforme est aujourd'hui constituée de 2 environnements de travail :

1. L'**environnement AAR**, développé entre 2004 et 2009, utilisé pour le portail des Archives Audiovisuelles de la Recherche<sup>12</sup>.
2. L'**environnement ASA**, développé dans le cadre du projet ASA-SHS à partir de 2009<sup>13</sup>, utilisé pour les portails **Semioscape**<sup>14</sup>.

<sup>8</sup> [http://www.culturalheritage.fr/FR/\\_encyclokeywords.asp](http://www.culturalheritage.fr/FR/_encyclokeywords.asp)

<sup>9</sup> [http://www.archiviosemiotica.eu/IT/\\_encyclo\\_knowledges.html](http://www.archiviosemiotica.eu/IT/_encyclo_knowledges.html)

<sup>10</sup> [http://www.archivesaudiovisuelles.fr/1051\\_fr/](http://www.archivesaudiovisuelles.fr/1051_fr/)

<sup>11</sup> [http://www.archivesaudiovisuelles.fr/1405\\_Allouche-Benayoun\\_fr/](http://www.archivesaudiovisuelles.fr/1405_Allouche-Benayoun_fr/)

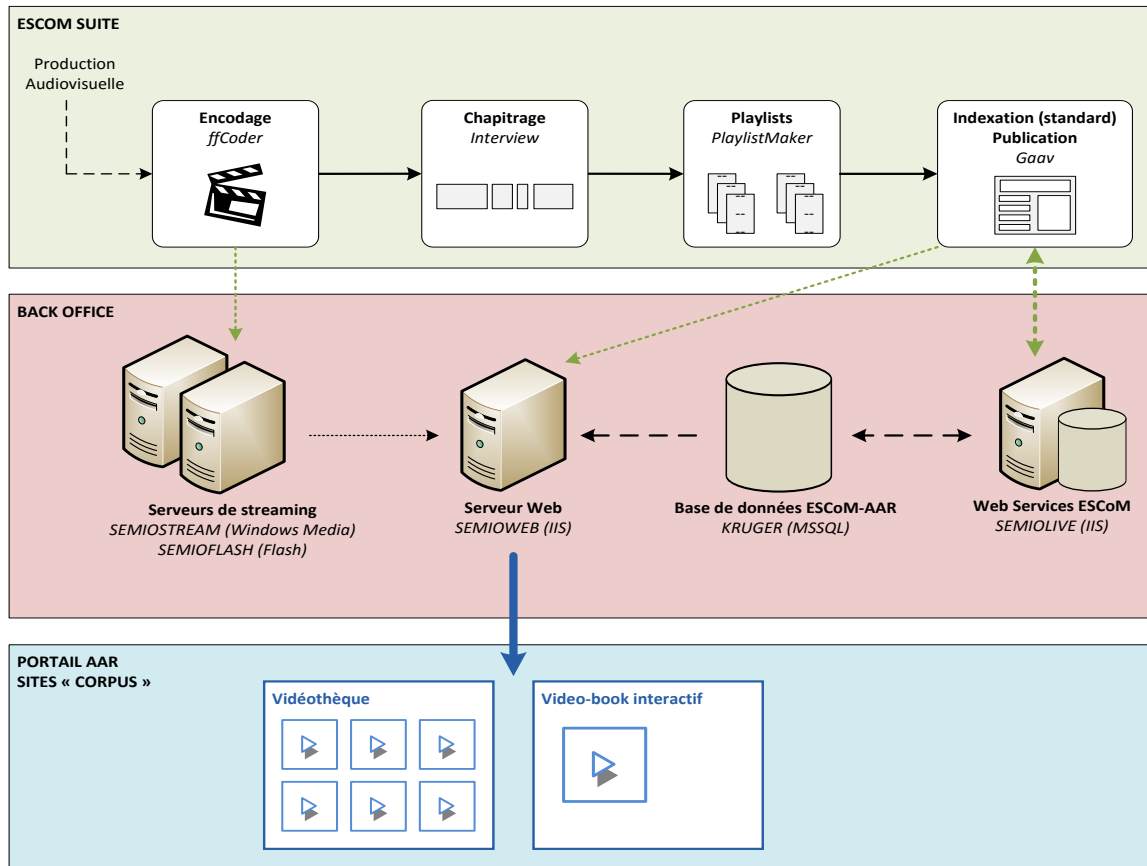
<sup>12</sup> <http://www.archivesaudiovisuelles.fr>

<sup>13</sup> Site officiel du projet ANR « ASA-SHS » : <http://www.asa-shs.fr/>; blog scientifique : <http://asashs.hypotheses.org/>

<sup>14</sup> <http://www.semioscape.fr>

### 2.2.1. L'environnement AAR

L'environnement AAR est celui toujours utilisé pour la gestion et la publication du corpus audiovisuel des Archives Audiovisuelles de la Recherche sur le portail AAR. La figure 6 représente schématiquement cet environnement.



(figure 6 : L'environnement AAR)

Le processus de travail suivi par les utilisateurs est défini par une suite de tâches successives réalisées à l'aide d'outils appelés *ESCoM Suite*. Ces outils communiquent avec les serveurs du back office par Internet (http ou ftp), et sont donc utilisables à distance.

1. Chaque vidéo est d'abord encodée avec *ESCoM ffCoder*, en 2 formats Windows Media (100 et 1400 kbps) et 2 formats Flash (16 et 700 kbps). Les fichiers ainsi produits sont ensuite copiés par ftp sur les serveurs de streaming, respectivement *SEMIOFLASH* (Red 5) et *SEMIOSTREAM* (Windows Media Server).
2. Chaque vidéo est ensuite chapitrée virtuellement avec *ESCoM-INA Interview*. Ce chapitrage est ensuite converti en playlists asx à l'aide de l'outil *ESCoM PlaylistMaker* (outil maintenant intégré dans *ESCoM-INA Interview*).
3. Enfin, le résultat de ce travail est importé dans *ESCoM Gaav*. Cet outil sert à gérer les publications des vidéos sous forme d'événements (ou de « mini-sites »). Outre le chapitrage des vidéos, l'auteur édite dans *Gaav* les informations relatives à l'événement qui constitue l'objet des documents audiovisuels (par informations signalétiques, présentation, intervenants, thèmes, pages additionnelles, ressources supplémentaires, etc.).
4. Ces informations sont enregistrées dans la base de données *ESCoM-AAR*, via le serveur de services web *SEMIOLIVE*. Elles sont ensuite traitées par le serveur web *SEMIOWEB* qui produit des publications automatiques de chaque événement sous forme de mini-site et d'accès aux contenus

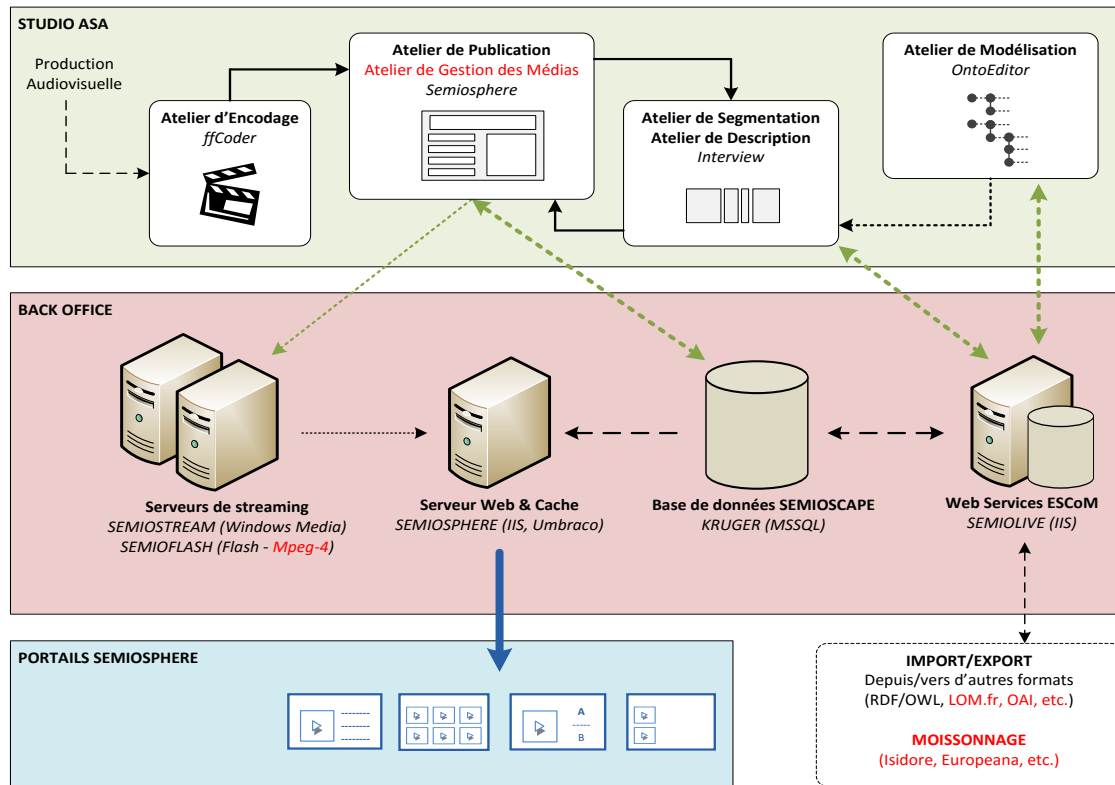
sous forme de vidéothèque. Cette technologie est celle utilisée par le portail AAR, ainsi que les sites « corpus ». En outre, SEMIOLIVE héberge des services supplémentaires tels que l'envoi des newsletters, la mise à jour des flux RSS, etc.

A noter que cet environnement a permis également de générer des publications spécialisées utilisant des ontologies et méta-descriptions – ces fonctionnalités ne sont pas abordées ici car elles ne sont plus utilisées, remplacées et améliorées dans l'environnement ASA décrit ci-après.

### 2.2.2. L'environnement ASA

L'environnement ASA est celui utilisé pour la description et la publication sémiotiques du corpus audiovisuel des Archives Audiovisuelles de la Recherche sur les portails *Semioscape*<sup>15</sup>. La figure 7 représente schématiquement cet environnement (les éléments en rouge sont en cours de développement et devraient être déployés en décembre 2013).

Le processus de travail suivi par les utilisateurs est défini par une suite de tâches successives réalisées à l'aide d'outils appelés Studio ASA. Ces outils communiquent avec les serveurs du back office par Internet (http ou ftp), et sont donc utilisables à distance.



(Figure 7 : L'environnement ASA)

<sup>15</sup> <http://www.semioscape.fr>

1. Chaque vidéo est d'abord encodée en plusieurs formats avec ESCoM ffCoder.
2. Chaque vidéo sera ensuite enregistrée dans la base de données *SEMIOSCAPE* à partir de l'interface *Umbraco* de SEMIOSCAPE, et chacun des fichiers compressés automatiquement uploadés sur les serveurs de streaming. Seront ainsi stockées les informations générales de la vidéo (titre, date, lieu, auteurs, etc.), ainsi que les urls de chaque format disponible pour la vidéo.
3. Les vidéos sont segmentées puis décrites avec ESCoM-INA Interview, en utilisant les modèles de description et méta-ressources préalablement développés avec *ESCoM OntoEditor* (ces méta-descriptions et méta-ressources sont enregistrées dans la base SEMIOSCAPE, via le serveur de services web SEMIOLIVE).
4. Ces données sont traitées par le serveur web SEMIOSCAPE qui produit des publications automatiques de chaque méta-description et des accès aux contenus sous forme de vidéothèque (accès par sujet, géolocalisation, type de collection, participant, etc.). Cette technologie est celle utilisée par tous les portails Semioscape.
5. L'interface Umbraco de SEMIOSCAPE permet également aux auteurs de publier ces méta-descriptions sur leur portail sous forme de dossiers vidéo (thématiques, bilingues, pédagogiques, etc.). En outre, des fonctionnalités CMS leur permet de personnaliser leur portail, définir et mettre à jour leurs rubriques, articles, actualités, page d'accueil, etc.

A noter que le serveur SEMIOSCAPE joue également le rôle de serveur de cache. Chaque nuit, les données de la base de données sont transformées en objets et stockées dans une base NoSQL (Redis Server). De plus, après leur premier chargement, chaque page est mise en cache (dans une base MongoDB) : c'est le cache qui est restitué lors des chargements suivants (et les temps de réponse sont ainsi très courts).

Ce cache est effacé chaque nuit (sinon les modifications apportées par les auteurs ne seraient pas visibles). Pour en limiter les effets, le serveur SEMIOLIVE simule chaque nuit des requêtes sur les pages les plus importantes (pages d'accueil et accès des vidéothèques) afin de générer leur cache.

### 2.2.3. Le Studio ASA (Atelier Sémiotique Audiovisuel)

Le Studio ASA est la plateforme de travail de l'ESCoM (son « lab ») pour instrumenter : a) un projet d'archivage numérique stricto sensu ; b) et son appropriation/exploitation par un « digital humanist », un « digital knowledge professionnel » dans le cadre, par exemple, d'un projet de remédiation ou republication d'un corpus de ressources audiovisuelles.

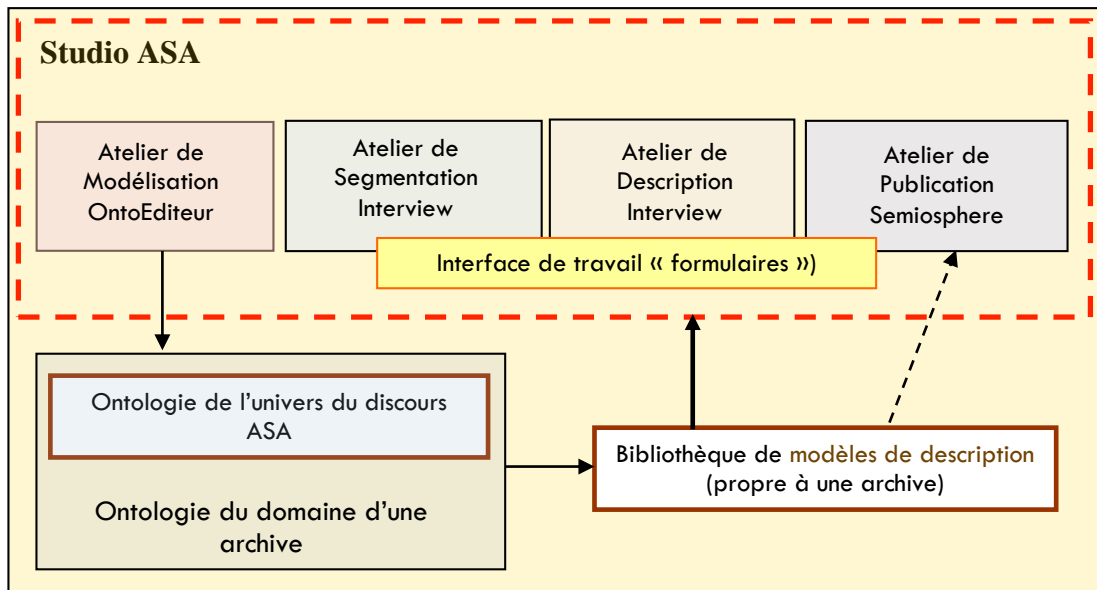
La figure 8 montre schématiquement le Studio ASA qui est composé de quatre ateliers :

1. Atelier de segmentation d'une ressource audiovisuelle ;
2. Atelier d'analyse d'une ressource audiovisuelle ;
3. Atelier de publication d'une ressource audiovisuelle ; et
4. Atelier de modélisation de l'univers du discours d'une archive audiovisuelle.

Le métalangage de description, lui, comprend des éléments métalinguistiques plus spécialisés nécessaires pour spécifier la bibliothèque de modèles de description propre à l'univers du discours d'une archive ou d'un corpus textuel comme, par exemple :

- un vocabulaire conceptuel (une ontologie) spécifiant les **objets d'analyse** qui font partie de l'univers du discours ASA ;
- un vocabulaire conceptuel (une ontologie) spécifiant les **procédures et activités d'analyse** de l'univers du discours ASA ;

- une bibliothèque de **schémas d'indexation** définissant les différents types d'indexation dont peut être pourvue une procédure ou activité de description ;
- un **thesaurus** réunissant des expressions normalisées à usage répété ;
- des **ontologies de domaines** dont chacune correspond à l'univers du discours d'une archive.



(figure 8. Le Studio ASA – représentation graphique)

Conception et développement de cet environnement ont été rendus possibles depuis 2006 grâce à plusieurs projets de R&D européens et français importants dont, plus particulièrement, le projet ASA-SHS<sup>16</sup> financé par l'ANR entre 2009 et 2011. Le Studio ASA fait également l'objet de nombreuses publications depuis 2010. Ensemble, avec les recherches autour du programme (scientifique) AAR, il est extensivement traité dans deux ouvrages collectifs réalisés fin 2011 et publiés (en français) par les Editions Hermes et (en anglais) par les éditions Wiley<sup>17,18</sup> à New York. Enfin, une monographie scientifique, publiée en 2012, est consacrée exclusivement au métalangage de description de corpus de textes (audiovisuels). Cette monographie a été de nouveau publiée chez Hermes (pour l'édition française) et chez Wiley<sup>19</sup> (pour l'édition anglaise).

### 2.3. Axe 3 : Campus AAR : Transfert et partage des connaissances et technologies pour l'archivage et la diffusion de patrimoines scientifiques et culturels

Un nombre de plus en plus important de personnes (chercheurs, enseignants, professionnels, ...) et d'institutions (universités, laboratoires de recherche, associations, musées, ...) cherchent à préserver, à partager et à valoriser leurs patrimoines intellectuels à l'aide des médias numériques (textes, images fixes,

<sup>16</sup> Site officiel du projet ASA-SHS : <http://www.asa-shs.fr/>; carnet de recherche : <http://asashs.hypotheses.org/>

<sup>17</sup> <http://public.j.eblib.com/EBLPublic/PublicView.do?ptilID=1124646>

<sup>18</sup> <http://books.google.fr/books?id=fUY16MW6rmMC&printsec=frontcover&hl=fr#v=onepage&q&f=false>

<sup>19</sup> <http://www.iste.co.uk/index.php?p=a&ACTION=View&id=494>



vidéos, ...) sous forme de corpus ou de collections thématiquement organisés et de publications adaptées à des usages spécifiques.

Campus AAR est un projet de R&D qui a comme objectif de mettre en place une plateforme « communautaire » pourvue de toutes les technologies, outils, méthodes et procédures indispensables pour pouvoir mener à terme un tel projet. Financé par l'ANR (Agence Nationale de la Recherche) dans la cadre de son programme CONTINT, le projet Campus AAR débute officiellement en janvier 2014.

Campus AAR est coordonné par l'ESCoM-AAR de la FMSH. Le consortium est composé par l'ESCoM-AAR, l'INA (Institut National de la Recherche), le CNRS-CCSD (Centre National de la Recherche Scientifique – Centre de Communication Scientifique Directe/HAL), le MESR-CERIMES (Ministère de l'Enseignement Supérieur et de la Recherche – CERIMES/CANAL U) ainsi que par la société Armadillo.

### 2.3.1. Les objectifs du projet Campus AAR

Plus concrètement parlant, les partenaires du consortium du projet Campus AAR se proposent de coordonner leurs efforts de recherche et de développement autour des **5 axes** majeurs suivants :

1. **Axe 1** : Réalisation d'un démonstrateur (d'un prototype opérationnel) d'une *plateforme logicielle* (appelée *Studio Campus AAR*) offrant la possibilité aux acteurs (individuels ou collectifs) du monde des SHS de réaliser et de gérer d'une manière autonome leurs *archives personnelles* de ressources audiovisuelles.
2. **Axe 2** : Mise à la disposition à tout utilisateur de la plateforme *Studio Campus AAR*, des *ressources conceptuelles* et *cognitives communes*, *prêtes à l'emploi* pour un projet d'archive personnelle :
  - 2.1) ontologie de base (= générique) pour les SHS, thesaurus commun et modèles de base de description/indexation de ressources AV ;
  - 2.2) interface de travail composée de « *formulaires* » *interactifs personnalisables* pour la constitution, description et (re-)publication de corpus AV ;
  - 2.3) possibilité d'utiliser des standards communs tels que OAI, DC ou LOM, de langages et de thesaurus usuels (ex. : RAMEAU) afin d'améliorer l'utilisabilité d'une archive en s'appuyant sur l'*interopérabilité* des systèmes d'organisation des connaissances grâce à la mise en œuvre de SKOS.
3. **Axe 3** : Mise à disposition à tout utilisateur de la plateforme *Studio Campus AAR* d'une riche bibliothèque de *templates* pour la publication (automatique et/ou « auteur ») de corpus audiovisuels : *templates* de portails AV personnalisés ; *templates* de publications (semi-automatiques) de vidéo-dossiers ; *templates* pour différents genres d'accès aux corpus AV.
4. **Axe 4** : Mise en place du (méta-)portail *Campus AAR* qui offrira:
  - 4.1) un accès central à tous les autres portails de la même technologie et une recherche transversale simplifiée basé sur un niveau de catalogage commun intégrant le standard OAI ;
  - 4.2) un accès pérenne aux *applications du Studio Campus AAR* et les *ressources partageables* pour tout utilisateur identifié et habilité.
5. **Axe 5** : Travail d'expérimentation en étroite concertation avec les acteurs impliqués dans les *digital humanities* en France :
  - 5.1) les grands programmes institutionnels français : côté *enseignement supérieur*, les UNT dont l'UOH, le Cerimes et Canal-U ; côté *recherche*, le CCSD du CNRS et les archives ouvertes de la recherche (*MédiHal*, *Hal-Vidéo*), ISIDORE, le TGIR Huma-Num ; côté *gouvernance* : Alliance Athéna, Campus Condorcet, PRES HeSam, UNPIDF, réseau des MSH, ... ;
  - 5.2) les acteurs individuels (enseignants, chercheurs, doctorants, ...) et acteurs collectifs (laboratoires de recherche, ...) engagés dans des projets d'archives AV en SHS.

Campus AAR s'appuiera plus particulièrement d'une part sur l'expertise de l'ESCoM-AAR en matière d'archivage numérique (cf. Axe I) ainsi que sur le *Studio ASA* et le *portail Semioscape*.

### 2.3.2. Partenariats consacrés à la mise en place d'archives audiovisuelles

En attendant le démarrage officiel du projet Campus AAR, planifié pour début janvier 2014, l'ESCoM-AAR a d'ores et déjà initié un ensemble de partenariats portant sur des projets concrets d'archivage numérique de patrimoines audiovisuels et de leur exploitation sous forme de publications enrichies et adaptées à des usages spécifiques :

- 1) Coopération avec l'Université François-Rabelais de Tours portant sur la création d'une archive audiovisuelle consacrée au patrimoine ouvrier de Vierzon<sup>20</sup>.
- 2) Coopération avec le laboratoire CRLA (Centre de Recherches Latino-Américaines) du CNRS et de l'Université de Poitiers autour du portail audiovisuel AMSUR<sup>21</sup> consacré au patrimoine culturel et historique de l'Amérique Latine.
- 3) Coopération avec les Archives Henri Poincaré (UMR 7117 - CNRS/Université de Lorraine) et le Centre d'Histoire des Sciences et des Philosophies Arabes et Médiévales (équipe de SPHERE, unité mixte de recherche du CNRS et de l'enseignement supérieur (UMR 7219), associée à l'Université Paris 7 Diderot) portant sur la mise en place d'une archive audiovisuelle en histoire des mathématiques et des géométries.
- 4) Coopération avec l'Association des Amis du Pays de Civaux portant sur la création d'une archive audiovisuelle consacrée au patrimoine rural et culturel de la commune de Civaux, département de Vienne<sup>22</sup>.

Du point de vue de l'ESCoM-AAR, tous ces projets de coopération poursuivent les objectifs suivants :

- 1) préparer d'une manière aussi efficace que possible le démarrage du projet Campus AAR lui-même ;
- 2) vérifier et généraliser les procédures d'archivage numérique mises en place par l'ESCoM-AAR depuis 2002 ;
- 3) vérifier la robustesse du métalangage ASA élaboré par l'ESCoM-AAR depuis 2006 ;
- 4) expérimenter des bibliothèques de modèles génériques de description facilement adaptables à des archives audiovisuelles les plus variées.

## 2.4. Axe 4 : e-Sémiotique : Recherche et développement sur les nouveaux écosystèmes de l'information et de la connaissance

Les activités consacrées aux trois premiers axes ont toujours été guidées par une vision scientifique spécifique de la problématique des archives (ou bibliothèques) numériques ainsi que de l'organisation, de la circulation et de l'utilisation de connaissances dans les médias (médias numériques, mais aussi médias au sens plus classiques). Cette vision est exprimée par le slogan du *tournant sémiotique* (en anglais, on parle plutôt du « *linguistic turn* ») dans la gestion des connaissances. Voici, d'une manière succincte, les 4 thèmes scientifiques autour desquels s'organisent les activités et projets de l'ESCoM-AAR :

---

<sup>20</sup> Prototype du portail « Mémoire ouvrière de Vierzon » : <http://www.memoirevierzon.msh-paris.fr/>

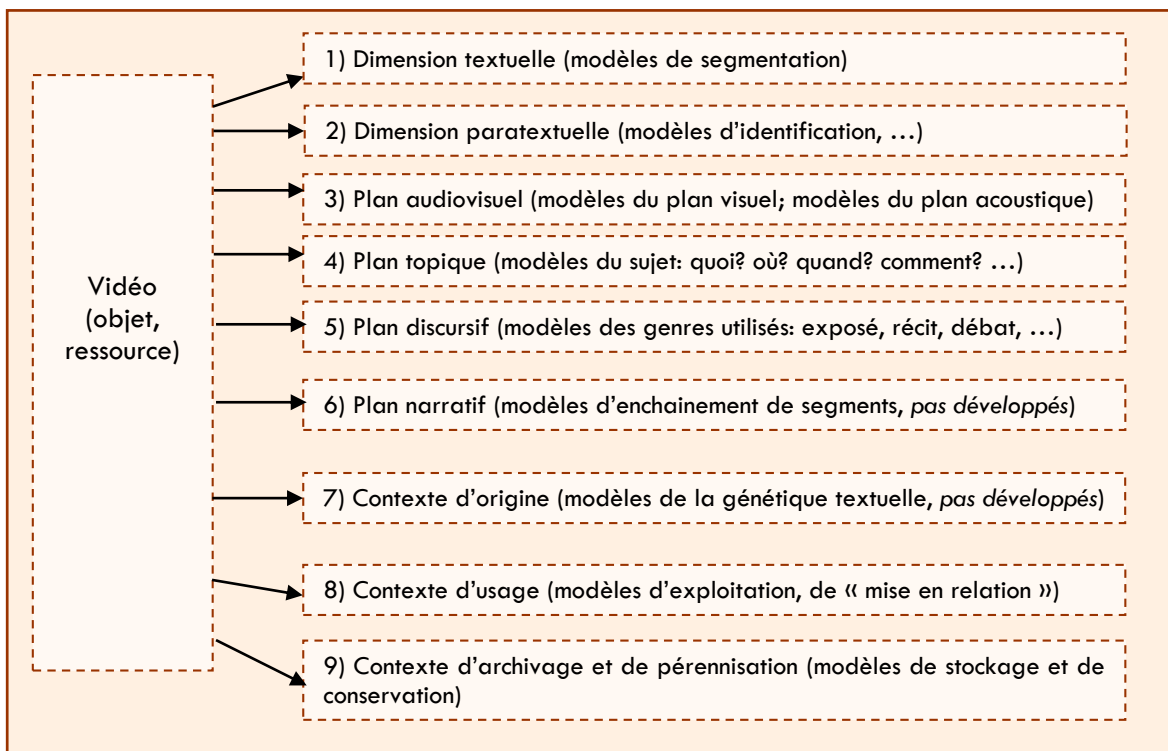
<sup>21</sup> <http://www.amsur.msh-paris.fr/>

<sup>22</sup> Prototype du portail « Mémoire de Civaux » : <http://www.memoirecivaux.msh-paris.fr/>

- 1) Le « semiotic turn » des archives numériques.
- 2) Méthodologie *Projet Archives numériques et remédiation*.
- 3) Ecosystèmes d'information, acteurs sociaux et patrimoines de connaissances.

#### 2.4.1. Le « semiotic turn » dans les archives numériques

Les activités de recherche et de développement de l'ESCoM-AAR consacrées aux archives numériques adoptent une approche résolument *textuelle* et *discursive* des archives (approche popularisée sous l'expression « *linguistic turn* » dans la recherche sur les archives. Défini d'une manière très générale, on suppose que les données (textuelles et autres) archivées forment, ensemble, un **univers de discours** qui exprime une **culture**, une conception culturelle (des savoirs, des visions, des doctrines, des valeurs, etc.) du « domaine », du « secteur de la réalité » documenté par les dites données archivées.



(Figure 9 Les différentes dimensions structurales de la donnée audiovisuelle prise en compte par les modèles de description composant l'interface de travail du Studio ASA)

Cette approche a été initiée, entre autre, par des savants comme Michel Foucault, Michel de Certeau ou Jacques Derrida. Récemment, le « *linguistic turn* » dans les archives a été discutée et approfondie par Louise Craven et al<sup>23</sup> ainsi que d'une manière plus générale par Trevor Owens en traitant la question du statut d'une donnée numérique dans les *digital humanities* comme suit:

“As a species of human-made artifact, we can think of *data sets* as having the *same characteristics as texts*. Data is created for an audience. Humanists can, and should interpret data as an authored work and the intentions of the author are worth consideration and exploration.”<sup>24</sup>

<sup>23</sup> Cf. Louise Craven (éd). *What are Archives ? Cultural and theoretical perspectives*; Ashgate Publishing 2008

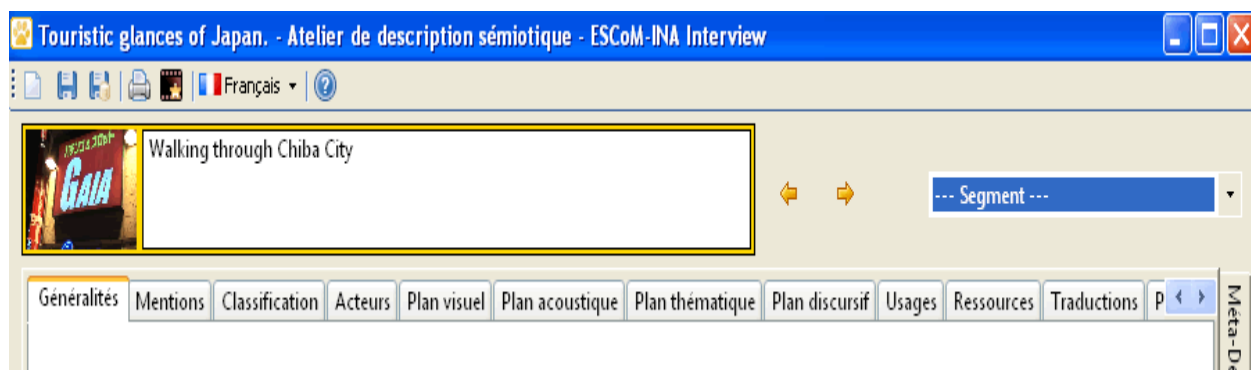
<sup>24</sup> in : *Journal of Digital Humanities* (1/1 2011).

Le “semiotic turn” dans les archives numériques ne reste pas, dans le cadre des activités de R&D de l’ESCoM-AAR, une simple notion de principe, un simple postulat.

Il est très concrètement implémenté dans le cœur même de la plateforme scientifique et technique du Studio ASA et expérimenté à travers des activités d’analyses et de (re)publications de corpus audiovisuels variés faisant partie des portails « nouvelle génération » (cf. Axe 1) comme, par exemple :

- le portail AGORA consacré au patrimoine audiovisuel de la recherche en SHS<sup>25</sup> ;
- le portail AMSUR consacré à l’Amérique Latine<sup>26</sup> ;
- le portail ALIA - Atelier Littéraire d’Ici et d’Ailleurs - consacré au patrimoine littéraire<sup>27</sup> ;
- le portail PCIA - Patrimoine Immatériel Andin - développé dans le cadre d’une thèse en anthropologie<sup>28</sup> ;
- le portail ADA – Atelier des Arkéonautes consacré au patrimoine scientifique archéologique<sup>29</sup> ;
- le portail MVC - Mémoire de Civaux consacré à la mémoire historique et culturelle du village Civaux<sup>30</sup> ;
- ainsi que de toute une série de portails actuellement en cours de développement.

Le « cœur » du Studio ASA est composé d’un ensemble de *modèles de description* qui représentent « l’univers du discours » d’une archive au sens des AAR. Les modèles de description sont élaborés à l’aide d’un *métalangage de description* (= métalangage ASA) et reposent sur une représentation structurale de l’objet « vidéo » (document vidéo).



(Figure 10. Accès aux modèles de description d’une ressource audiovisuelle via l’interface de l’Atelier de Description du Studio ASA)

Le Studio ASA, pour chaque dimension, chaque plan (cf. figure 9) spécifie et met à la disposition de l’utilisateur soit une classe de *modèles uniques* (i.e. valide pour toutes les archives), soit une classe de *modèles multiples* (i.e. adaptés aux spécificités d’une archive particulière et pouvant varier d’une archive à une autre).

<sup>25</sup> Cf. <http://www.agora.msh-paris.fr/>

<sup>26</sup> Cf. <http://www.amsur.msh-paris.fr/>

<sup>27</sup> Cf. <http://www.alia.msh-paris.fr/>

<sup>28</sup> Cf. <http://www.pcia.msh-paris.fr/>

<sup>29</sup> Cf. <http://www.ada.msh-paris.fr/>

<sup>30</sup> Cf. <http://www.memoirecivaux.msh-paris.fr/>

Un modèle de description sert à définir la structure et le contenu des métadonnées relatives à une donnée audiovisuelle. Les *métadonnées* « servent » à classer les données audiovisuelles dans une archive (= AIP au sens de OAI) ; à exporter les données (= SIP au sens de l'OAI) vers d'autres archives.

Il s'agit, par exemple, du dépôt légal du fonds audiovisuel et de l'ensemble des données du programme AAR à la Bibliothèque Nationale de France (BnF), du dépôt pérenne des données dans les archives HAL et HAL-Vidéo du CNRS-CCSD; de l'agrégation et de la diffusion plus large des ressources audiovisuelles du programme AAR via, par exemple, ISIDORE, UOH, Europeana, etc.

La figure 10 montre comment les modèles de description sont intégrés dans l'interface de l'Atelier de Description du Studio ASA et rendues accessibles à l'analyste travaillant sur une ressource audiovisuelle ou un corpus de ressources audiovisuelles.

Les **éléments du métalangage ASA** composant les modèles de description sont :

1. une hiérarchie de concepts relatifs aux *objets d'analyse des archives au sens des AAR* ;
2. une hiérarchie de concepts relatifs aux *procédures de description et d'indexation* d'un texte audiovisuel ;
3. des relations entre les concepts permettant d'associer les concepts entre eux pour en faire :
  - 3.1. des schémas et séquences de description et d'indexation qui forment à leur tour
  - 3.2. les modèles de description (composant l'interface de travail du logiciel Interview)
4. un thesaurus à facettes permettant l'utilisation de *valeurs prédéfinies* (des « descripteurs ») dans des contextes sémantiques variés.

La gestion du métalangage ASA, de ses différents éléments ainsi que des modèles de description communs à toutes les archives ou spécifiques à telle ou telle archive, se fait via l'outil *OntoEditeur*, un éditeur xml qui fait partie de l'Atelier de Modélisation du Studio ASA. *OntoEditeur* permet, plus particulièrement :

- la gestion du métalangage ASA (corrections, modifications, enrichissements, ...),
- l'élaboration des *modèles de description* formant la bibliothèque des modèles de description propres à une archive audiovisuelle (exemple: AMSUR; AGORA; PCIA),
- l'export du métalangage et des modèles de description en format rdf/owl et, bientôt SKOS (« passerelle » entre ASA et OAI, Rameau, MODS, MARC 21, EAD, etc).

Les principaux enjeux de cet axe de recherche sont les six suivants:

- 1) Rendre «massivement» interopérable le métalangage ASA avec les principaux standards, normes, langage d'indexation, thesaurus, ... utilisés dans le domaine des bibliothèques/archives numériques tel que OAI, RAMEAU, DCC, UDC, LCC, MODS, MOTBIS, MARC 21, EAD, LOM, ...) en utilisant les possibilités offertes par SKOS.
- 2) Produire des versions multilingues du métalangage ASA via un système de pré-translation automatique suivi d'une phase de validation par des traducteurs humains (actuellement, un premier prototype est un train d'être développé et testé à l'ESCoM-AAR sur des couples de langue « Français – Allemand » ; « Français – Anglais » ; « Français – Vietnamien » ; « Français – Arabe »).
- 3) Produire une bibliothèque de modèles de base de description pouvant être utilisés tels quels par tout acteur désirant créer une archive audiovisuelle mais ne possédant pas les ressources (humaines, financières, ...) nécessaires pour procéder à une véritable modélisation conceptuelle et cognitive du domaine de référence de son archive.
- 4) Expliciter davantage les contraintes sémantiques entre les différents éléments composant le métalangage ASA, les implémenter et les utiliser dans une construction « guidée (par l'application) » de nouveaux modèles de description.

- 5) Investiguer certaines dimensions de l'entité textuelle au sens sémiotique pas encore systématiquement explorées/prises en compte comme, par exemple, le plan acoustique d'une ressource audiovisuelle, le plan rhétorique et discursif ou encore la partie « usabilité » d'une ressource audiovisuelle.
- 6) Améliorer la gestion et le partage des index générés par les analystes sous forme, par exemple, d'affichage dynamique dans les interfaces des modèles de description.

#### 2.4.2. Méthodologie *Projet Archives numériques et remédiation* : Procédures, documentation et outils

Depuis 2002, l'ESCoM-AAR a investi beaucoup d'efforts dans la définition, l'explicitation et la documentation d'une approche méthodologique générale pour la production, description/publication, exploitation d'archives numériques (audiovisuelles). Aujourd'hui, cette méthodologie est parfaitement opérationnelle, totalement respectée dans le processus de travail quotidien de l'équipe. Elle sera mise à la disposition de toutes les parties prenantes dans le cadre du projet Campus AAR (cf. ci-dessus : Axe 3).

L'ESCoM-AAR distingue **deux étapes** centrales dans le travail avec et autour des archives numériques :

- 1<sup>ère</sup> étape : L'« **archivage numérique** » à proprement parler
- 2<sup>ème</sup> étape : L'**étape du « repurposing »** (*exploitation, appropriation, transformation qualitative, ...*) d'une donnée archivée.

**1<sup>ère</sup> étape : L'« archivage numérique ».** Le modèle organisationnel développé par l'ESCoM-AAR comprend 6 phases centrales caractérisant un projet d'archivage numérique au sens des AAR ...

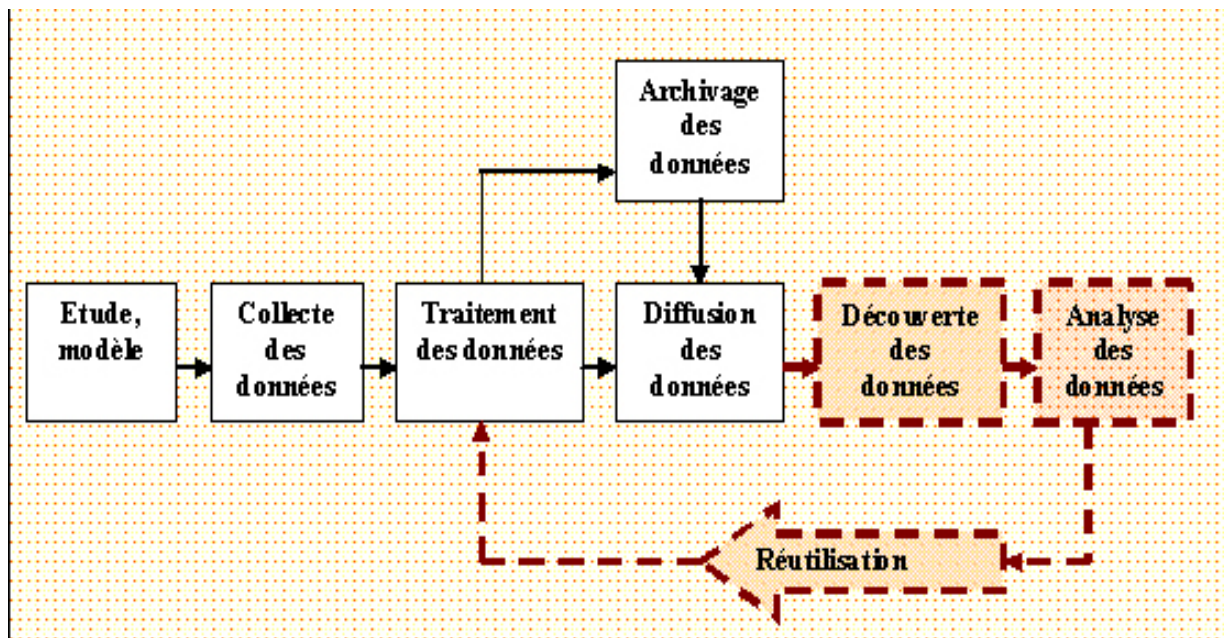
1. 1<sup>ère</sup> phase: Activités préparatrices en amont de la *documentation d'un « terrain » (lato sensu)*.
2. 2<sup>ème</sup> phase: Activités d'enregistrement du *terrain* (lato sensu), de collecte des données et de la constitution d'un corpus de documentation du terrain.
3. 3<sup>ème</sup> phase: Activités de préparation technique et conceptuelle (« métadonnées de base ») des corpus à verser dans les archives et versement des corpus (= les SIP de l'OAIS).
4. 4<sup>ème</sup> phase: Classement, description et indexation des données à archiver (= les AIP de l'OAIS).
5. 5<sup>ème</sup> phase: Edition et diffusion des corpus audiovisuels traités et analysés préalablement (= les DIP de l'OAIS).
6. 6<sup>ème</sup> phase: Activités de clôture du travail sur un corpus archivé et diffusé (dépôt légal; pérennisation; sauvegardes, ...).

Chacune des six phases se différencie en une série d'activités spécialisées, connaît des procédures explicites et documentées et est *instrumentée* (= outils + métadonnées; cf. ci-après le *Studio ASA*). Le modèle a été réalisée – *dans la mesure du possible* – par rapport à et à l'aide des deux *normes/standards techniques de référence* suivants:

- **OAIS** (Open Archival Information System; *norme ISO 14721*): Modèle conceptuel pour la production, gestion, diffusion et pérennisation de données numériques ;
- **PAIMAS** (*Producer-Archive Interface Methodology Abstract Standard*) – PAIS : Guide identifiant le cadre général de la constitution d'une archive numérique: phases, activités, paquets d'information OAIS.

**2<sup>ème</sup> étape : L'étape du « repurposing »** (*exploitation, appropriation, transformation qualitative, remédiation*) d'une donnée archivée. Cette deuxième étape concerne le travail, l'expertise du chercheur, de l'enseignant, du spécialiste, ... qui transforme une donnée audiovisuelle en une ressource intellectuelle à proprement parler (pour le chercheur, l'enseignant, ...).

Le modèle d'archivage numérique *stricto sensu* développé par l'ESCoM-AAR ainsi que les deux normes/standards de référence OAIS et PAIMAS ne formalisent que la première étape du travail d'un « digital humanist » (ou d'un « digital knowledge professionnel », ...), c'est-à-dire : ils ne formalisent que le travail de la numérisation des données documentant un domaine et leur archivage. Mais ils ne tiennent pas réellement compte de la deuxième étape, i.e. de celle de la transformation qualitative d'une donnée archivée par un acteur, de son appropriation par ce dernier ...

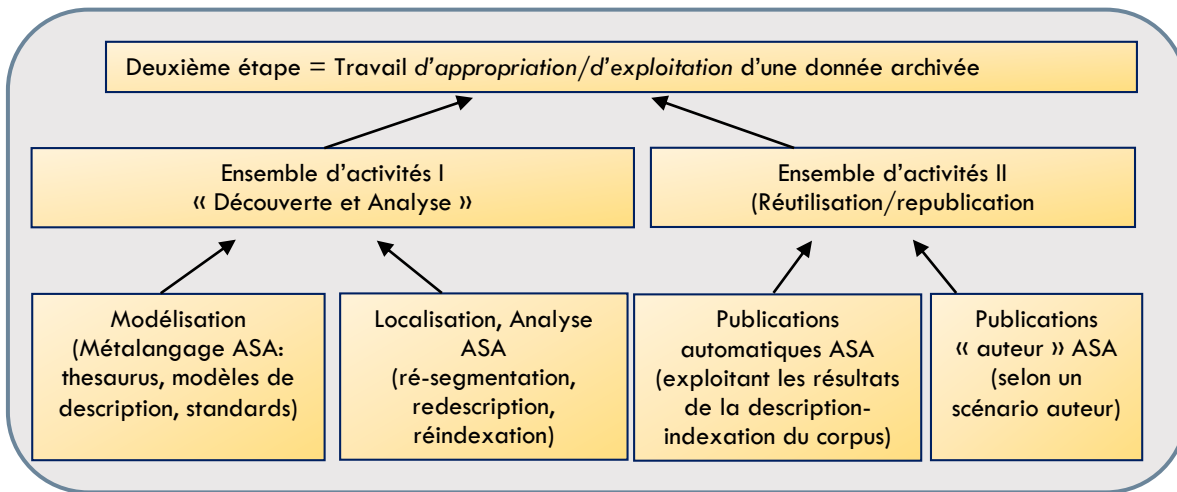


(figure 11 : Le « combined life cycle model » d'une donnée audiovisuelle dans le cadre des archives du programme ESCoM-AAR)

Une représentation plus complète des cycles de vie d'une donnée est fournie par la DDI (Data Documentation Initiative) et le CESSDA (Council of European Social Science Data Archives) sous le sigle du « combined life cycle model » (figure 11). Le combined life cycle model d'une donnée "... incorporates either direct dissemination to users or dissemination through data archives and recognizes that data can be reprocessed at later points in its life cycle, creating an iterative process. (...) This means that the life cycle is no longer linear but has become circular. We viewed repurposing of data as being a secondary use of the data from a study". (DDI 2009, p. 8).

La figure 12 représente schématiquement les principaux éléments qui sont à l'œuvre dans le processus du *repurposing* d'une ressource numérique (d'un corpus de ressources numériques) dans le cadre du programme ESCoM-AAR. Ce processus comprend, comme décrit ci-dessus (cf. section 2.14), toute une série d'étapes précises dans le travail de l'analyste (le « digital humanist », l'enseignant, le chercheur, ...) visant la republication de ressources numériques déjà existantes afin de les adapter à des usages spécifiques.

Dans le cadre du nouveau projet Campus AAR et à l'aide des divers partenariats consacrés à la création d'archives audiovisuelles spécialisées (cf. ci-dessus, l'axe 3 – le projet Campus AAR), l'ESCoM-AAR traitera plus en avant ce processus avec comme but de proposer à toutes les parties prenantes un modèle opérationnel et « normalisé »



(figure 12 : Les activités « Découverte et analyse » et « (Re-)publication dans le « combined life cycle model d'une donnée audiovisuelle dans le cadre des AAR)

### 2.4.3. Eco-systèmes d'information, acteurs sociaux et patrimoines de connaissance

En reprenant le concept de la *gestion du hub d'informations* de Vincent Ducrey<sup>31</sup>, on peut considérer que chaque acteur social (personne, groupe, communauté, organisation, ...) est en quelque sorte « enveloppée » par une médiasphère via laquelle l'acteur produit, diffuse, partage, ... des informations (sur lui, sur ses « idées », ses activités, ...).

En même temps, la médiasphère forme l'« agora » sur laquelle l'acteur est à son tour l'objet d'informations. La médiasphère d'un acteur n'est pas seulement composé des médias « classiques » (presse, édition, radio, télévision, ...) mais elle englobe également les médias de l'événementiel, de la communication directe et interpersonnelle ainsi que, bien sûr, les médias numériques (dont, plus particulièrement, les dits médias sociaux).

La médiasphère n'est pas une « enveloppe » neutre comme le montre, par exemple, la médiasphère contemporaine de la communication scientifique globale imposée aux chercheurs (notamment en SHS) qui les désavantage au détriment d'intérêts économiques (cf. les débats actuels très âpres sur la bibliométrie comme outil d'évaluation des chercheurs en SHS). C'est une structure sociotechnique qui, d'une part s'impose à un acteur social comme une « donne » historique et culturelle et qui, d'autre part, se voit être (ré-)appropriée, adaptée à ses intérêts et besoins.

Par ailleurs, la médiasphère ne constitue qu'une dimension (ou qu'un une strate) **de l'écosystème d'informations** d'un **acteur social**. D'autres dimensions (strates) sont représentées d'un côté par les **textes** (*lato sensu*) qui y circulent, se transforment, se « métissent » et de l'autre côté par tout un univers de **messages** et de **signes** de toute sorte qui composent les textes et qui réfère à des **univers de discours** au sens de Michel Foucault, les univers de discours faisant partie ou exprimant des visions, des doctrines, des savoir-faire, des idéologies, des valeurs, des normes – bref, des **cultures**.

<sup>31</sup> Ducrey, Vincent, *Le Guide de l'Influence. Communication, Médias, Internet, Opinion*. Paris, Eyrolles 2010



L'ambition de l'ESCoM-AAR ici est triple :

- 1) produire une vision théorique de l'écosystème d'information qui caractérise un acteur social (en s'appuyant sur des études de cas concrets dans le domaine de la communication des organisations et de la marque, par exemple, mais aussi de la communication scientifique) ;
- 2) définir une approche méthodologique robuste pour :
  - i. expliciter, analyser de tels écosystèmes d'information plus ou moins propres à un acteur social,
  - ii. concevoir et réaliser des projets d'archivage numérique de données faisant partie de ces écosystèmes d'information
- 3) repositionner la problématique des archives numériques dans une perspective d'appropriation/de réappropriation de données numériques de toutes sortes, par l'acteur dans ses activités quotidiennes, professionnelles ou simplement personnelles.

Ce quatrième axe de recherche est d'abord de nature exploratrice, se basant sur l'enseignement de certains des membres de l'ESCoM-AAR dans le cadre du parcours spécialisé de niveau Master (première année et deuxième année) *CIM – Communication, Information et (nouveaux) Médias* à l'INALCO (Institut National des Langues et Civilisations Orientales), il fera l'objet de séminaires, conférences et tables rondes spécifiques organisés dès 2014.

A terme, cet axe doit permettre un repositionnement assez « radical » des archives numériques dans un contexte de production, de gestion et de transmission de patrimoines pour des acteurs sociaux tels que les administrations publiques, les entreprises, les collectivités territoriales et locales, les institutions culturelles, les ONG, les associations, etc. Il s'agit ici de comprendre, d'analyser et d'instrumenter la communication numérique des organisations à l'aide d'une approche résolument sémiotique.

### 3. RÉFÉRENCES BIBLIOGRAPHIQUES

#### 2013

STOCKINGER, P. « La problématique des valeurs culturelles » ; in : Biglari, Amir (éd), *Les valeurs en sémiotique*. Limoges, PULIM 2013 (sous presse)

#### 2012

de PABLO, E., « Développement d'un environnement de travail pour la gestion et l'exploitation d'archives audiovisuelles. Etude de cas : Les Archives Rencontre des Cultures ou comment maintenir des archives vivantes », actes de colloque 1<sup>er</sup> congrès de l'AFEa – sept 2011, paru 2012, en ligne : <http://afea2011.sciencesconf.org/8750>

DESLIS J., « La communication scientifique via des réseaux sociaux : Etude de cas Atelier des Arkéonautes », actes de colloque 1<sup>er</sup> congrès de l'AFEa – sept 2011, paru 2012, en ligne : <http://afea2011.sciencesconf.org/8752>

STOCKINGER, P. *Analyser l'univers du discours des archives audiovisuelles. Un métalangage de description du texte audiovisuel en sciences humaines et sociales*. Paris – Londres, Editions Hermes Science Publishing 2012 (trad. en anglais aux éditions John Wiley & Sons, NY)

#### 2011

CHEMOUNY, M., « Indexer la vidéo », in Archimag, n°241, février 2011

LEMAITRE, F. *Introduction to Audiovisual Archives*. STOCKINGER P., CHEMOUNY M., DE PABLO E., DESLIS J., GUERINET R., LEMAITRE F., SAKUNDABHAI P. ISTE-John Wiley, London, Décembre 2011.

LEMAITRE, F. *Digital Audiovisual Archives*. STOCKINGER P., CHEMOUNY M., DE PABLO E., DESLIS J., EYYUBOVA A., LEGRAND V., LEMAITRE F. ISTE-John Wiley, London, Décembre 2011.

LEMAITRE, F. *Les archives audiovisuelles : description, indexation et publication*. STOCKINGER P., CHEMOUNY M., DE PABLO E., DESLIS J., GUERINET R., LEMAITRE F., SAKUNDABHAI P. Hermès-Lavoisier, Paris, Sept. 2011.

LEMAITRE, F. *Nouveaux usages de corpus audiovisuels numériques*. STOCKINGER P., CHEMOUNY M., DE PABLO E., DESLIS J., EYYUBOVA A., LEGRAND V., LEMAITRE F. Hermès-Lavoisier, Paris, Septembre 2011.

STOCKINGER, P. (éd.) *Les archives audiovisuelles : description, indexation et publication*. Paris – Londres, Editions Hermes Science Publishing 2011 (trad. en anglais aux éditions John Wiley & Sons, NY)

STOCKINGER, P. (éd.) *Nouveaux usages des archives audiovisuelles numériques*. Paris – Londres, Editions Hermes Science Publishing 2011 (trad. en anglais aux éditions John Wiley & Sons, NY)

#### 2010

LEMAITRE, F. « Working Environment for Management Exploitation of Audiovisual Archives - ASA-SHS Project », *Digital Heritage*, p. 492-503. Springer Verlag, Berlin/London/New York, Novembre 2010.

STOCKINGER, P. « L'adaptation et la republication de ressources audiovisuelles ». In: HERMES. *Cognition – Communication – Politique*. N° 56, pp. 36-52

**2009**

STOCKINGER, P. Das Programm „Digitale Audiovisuelle Forschungsarchive“. In: M. Becker (ed.), *Archiv im Netz*. Innsbruck, Studien-Verlag/Skarabaeus, pp. 113 – 121

**2008**

LEMAITRE, F. « ARA Environment and Educational Use Cases », *Cross-Media and Personalized Learning Applications with Intelligent Content*, p.79-87. LAIC, Budapest/Sofia, Septembre 2008.

STOCKINGER, P. « Les archives audiovisuelles de la recherche: un programme de production, traitement, conservation, diffusion en ligne de patrimoines numériques audiovisuels en sciences humaines et sociales ». In: *Gazette des Archives* N° 212, année 2008 – 4, pp. 101 – 119

**2007**

STOCKINGER, P. « Des archives audiovisuelles monolingues sur un site multilingue ». In: HERMES. *Cognition – Communication – Politique*. N° 49 – 2007, pp. 69 – 77

STOCKINGER, P. “Repurposing digital audiovisual resources for user communities possessing different cultural and linguistic profiles”. In: *LADL 2007 Cross-Media and Personalized Learning Applications on top of Digital Libraries*. Budapest 2007, pp. 75 – 91

**2005**

STOCKINGER, P. *Les sites web*. Description, évaluation et production. Paris, Hermes Science Publications (220 p).

**2003**

STOCKINGER, P. *Le document audiovisuel. Description et exploitations pratiques*. Paris, Hermes Science Publications (240 p).

STOCKINGER, P. (eds) *Portails et collaboratoires pour l’enseignement et la recherche*. Paris, Hermes Science Publications (230 p).

LEMAITRE, F. *Portails et collaboratoires pour la recherche et l’enseignement*. LEMAITRE F., STOCKINGER P., BAROUNI K., FADILI H., GIRAUDY E., KROMPHOLTZ S., THIVOLLE J.C. Hermès-Lavoisier, Paris/London, Octobre 2003.

**2001**

STOCKINGER, P. *Traitement et contrôle de l’information*. Perspectives sémiotiques et textuelles. Paris, Hermes Science Publications (195 p).

STOCKINGER, P. (avec M. Knassmüller and K. Sandner), *Die Unternehmensleitbilder*. Eine thematische Systematik. Wien, Universität der Wirtschaftswissenschaften

**2000**

STOCKINGER, P. (avec Charlotte Nikitenko; eds) *La publication en ligne*. Paris, Hermes Science Publications (225 p).

**1999**

STOCKINGER, P. *Les nouveaux produits d’information. Conception et sémiotique du document*. Paris, Hermes Science Publications (370 p).