

PROJET CAMPUS-AAR

UNE PLATEFORME COMMUNAUTAIRE DESTINÉE À
L'INDEXATION/ANALYSE ET LA PUBLICATION D'ARCHIVES AUDIOVISUELLES

Francis Lemaitre

Pôle Service et Infrastructures Numériques
Fondation Maison des Sciences de l'Homme
lemaitre@msh-paris.fr

Séminaire ArcMC

23 Juin 2016

MSH Paris Nord, Saint-Denis

1. Vue d'ensemble
2. Ressources métalinguistiques
3. Analyse
4. Publication

- **Projet ANR**
 - 2014-2016
 - CONTINT / *Digital Humanities*

- **Partenaires**
 - ESCoM-AAR (FMSH, coordinateur)
 - Armadillo
 - INA (équipe CAP)
 - CCSD (CNRS)

- Le démonstrateur d'une plateforme numérique
 - Travail **scientifique et éditorial** à partir d'archives audiovisuelles
 - A partir des principaux entrepôts de l'enseignement supérieur et de la recherche : **AAR, CANAL-U, MediHAL**
 - Basée sur les technologies du **web sémantique** (RDF/OWL)
 - **Hébergé au CCSD** (CNRS)

- Une version améliorée du Studio ASA
 - Ce projet constitue le prolongement des travaux du laboratoire **ESCoM-AAR**, dirigé par Peter Stockinger et hébergé à la FMSH jusqu'à fin 2015.
 - Il s'appuie notamment sur les résultats des projets ANR SAPHIR (2006-2009) et ASA-SHS (2009-2011) et de la **plateforme ASA** : <http://www.semioscape.fr>

- **BUT** : Adapter la description de médias à différents besoins
 - Description pilotée par des **ontologies** -> dynamique
 - « **Ressources métalinguistiques** »
 - « **Analyse** » = description de médias structurée et contextuelle
 - Plusieurs niveaux de description : corpus, média, segment, ...
 - Analyse du contenu, du discours, visuelle, sonore, narrative, ...
 - Pilotée par des **ontologies de domaine**
 - Exploitation des analyses
 - Portails thématiques -> « **publications automatiques** »
 - Editorialisation -> « **publications auteur** »
 - Recherche **sémantique** de contenu

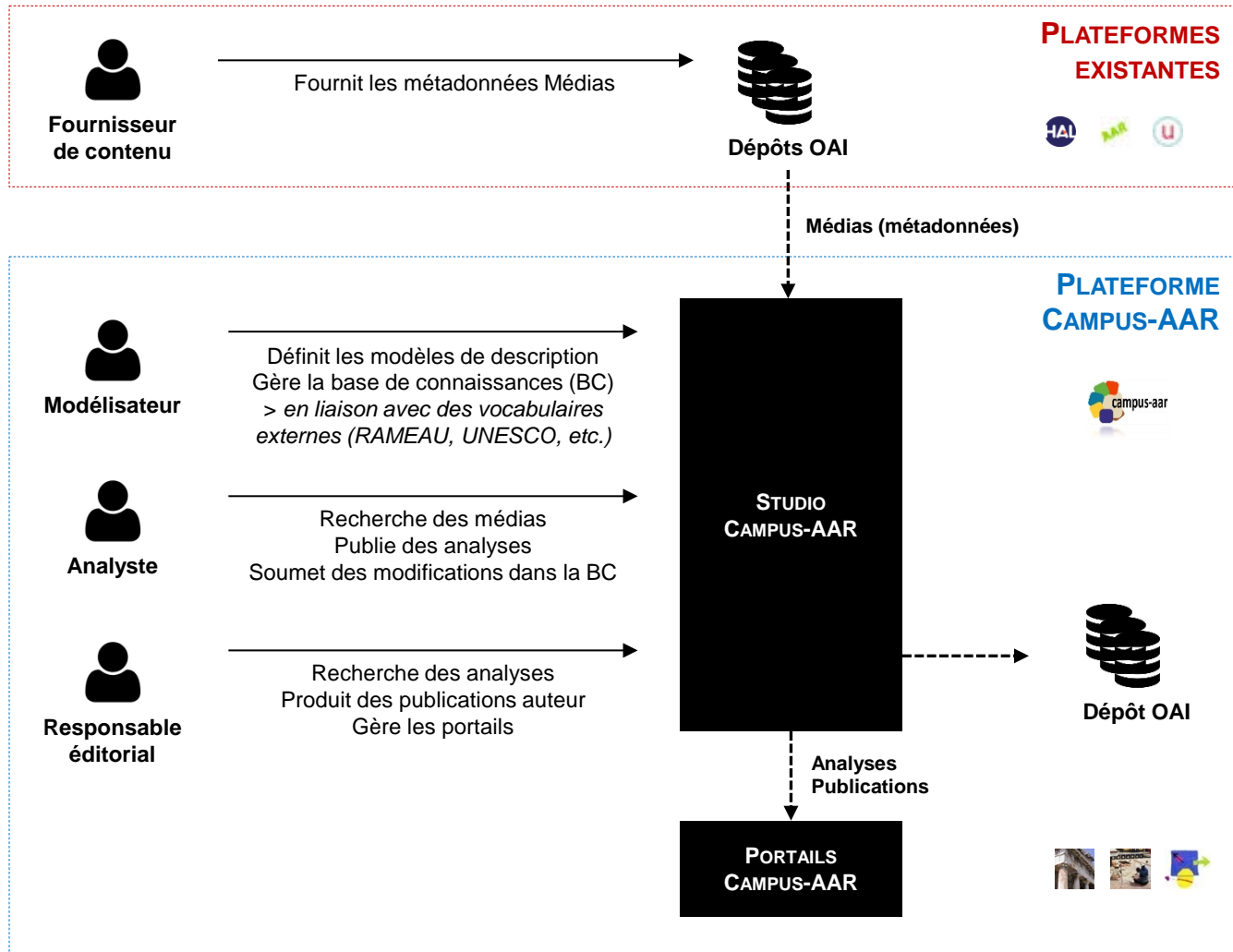
- 3 domaines développés dans le cadre du projet (ESCoM-AAR)
 - **AGORA** : « Patrimoine audiovisuel de la recherche en SHS »
 - **ARC** : « A la Rencontre des Cultures »
 - **AHM** : « Archives Histoire des Mathématiques »
- + **INA** (prototypé)

	Corpus	Disciplines	Contexte	Acteurs
AGORA	AAR	Sciences Humaines et Sociales	Patrimonial	Divers
ARC	INALCO	Linguistique, anthropologie, art, ...	Enseignement	Étudiants, enseignants
AHM	F2DS	Mathématiques	Recherche	Chercheurs
INA	Émissions TV	Information, Communication	Professionnel	Documentalistes

Vue d'ensemble (assez grossière) des domaines

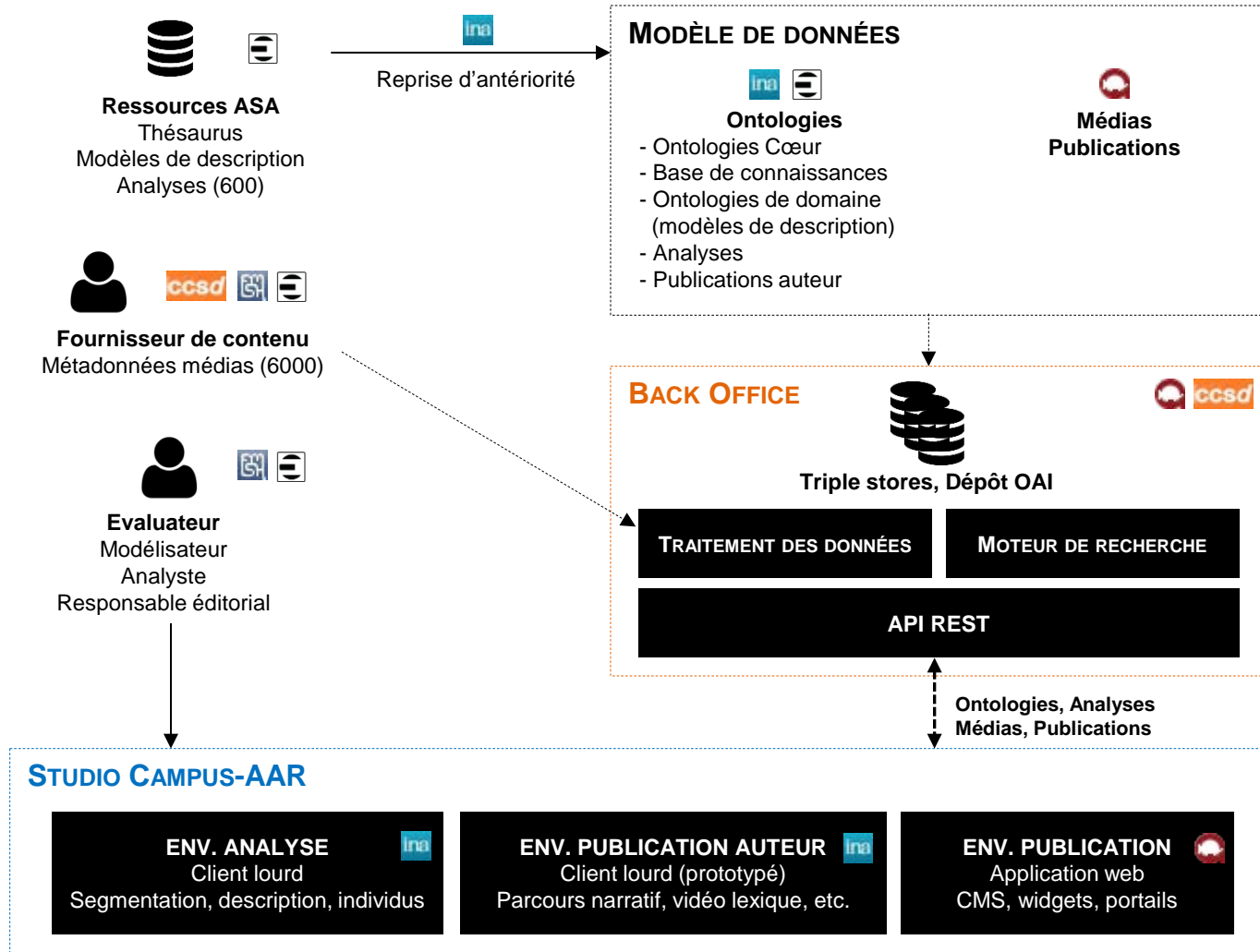
PROJET CAMPUS-AAR

VUE D'ENSEMBLE



Rôles et interactions

VUE D'ENSEMBLE



Architecture

RESSOURCES MÉTALINGUISTIQUES (INA ET P. STOCKINGER)

- Ontologies de base

- **Ontologie cœur** (INA)

- Modèle de données : corpus, media, segment, strate, analyse
- Vocabulaire ontologique de base

- **Thésaurus** (entités nommées)

- Thésaurus ASA
- Thésaurus SKOS importés : UNESCO, TheSoz (GESIS), RAMEAU

- Ontologies de domaine

- Structuration de l'analyse : strates thématiques, visuelles, sonores, ...
- **Modèles de description** : graphes de description (des sujets notamment)

- Ontologies de publication

- **Modèles de publication auteur** : vidéo-lexique, parcours narratif, ...

-> Éditeur d'ontologies en développement (sinon : Protégé)

- Interopérabilité

- Depuis des sources externes

- Dépôts OAI (médias)
- **SKOS** (thésaurus)
- Web de données : dbPedia, freebase, wikidata

- Interne

- **Entités nommées partagées**

- Vers des sources externes

- Dépôt **OAI** de CAMPUS-AAR (analyses/publications)
- A priori sans limitation technique, à développer au cas par cas

ANALYSE (STUDIO CAMPUS-AAR)

- Multi-strates
 - Plusieurs segmentations selon le type d'analyse
 - Typées : description adaptée au type de strate
 - **Strates média, thématique, visuelle, sonore, rhétorique**
- Objets d'analyse
 - Média (métadonnées importées depuis OAI)
 - **Segment**
 - **Corpus** (de médias)
 - Analyse
- Analyse
 - Dynamique : formulaires suivant des modèles ontologiques
 - Structurée et contextuelle : **WHAT-WHEN-WHERE-WHO-HOW**

-> **TOUT CECI PARAMETRABLE DEPUIS LES ONTOLOGIES**

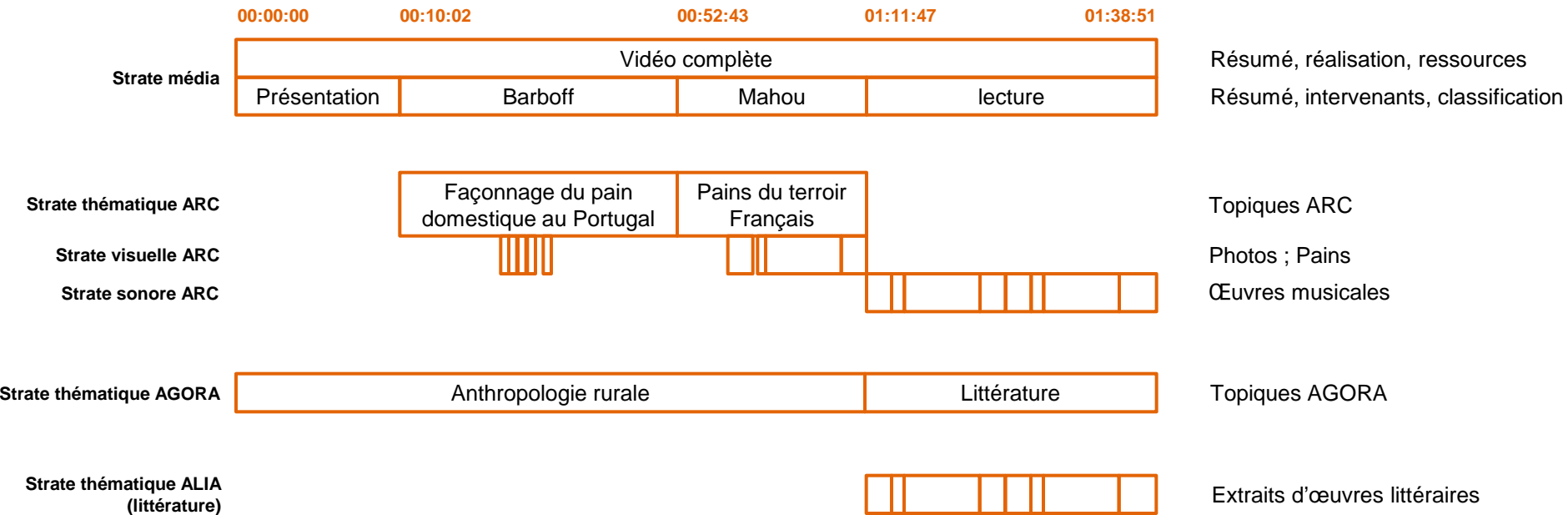
ANALYSE : EXEMPLE

● Exemple : « Le pain dans tous ses états »



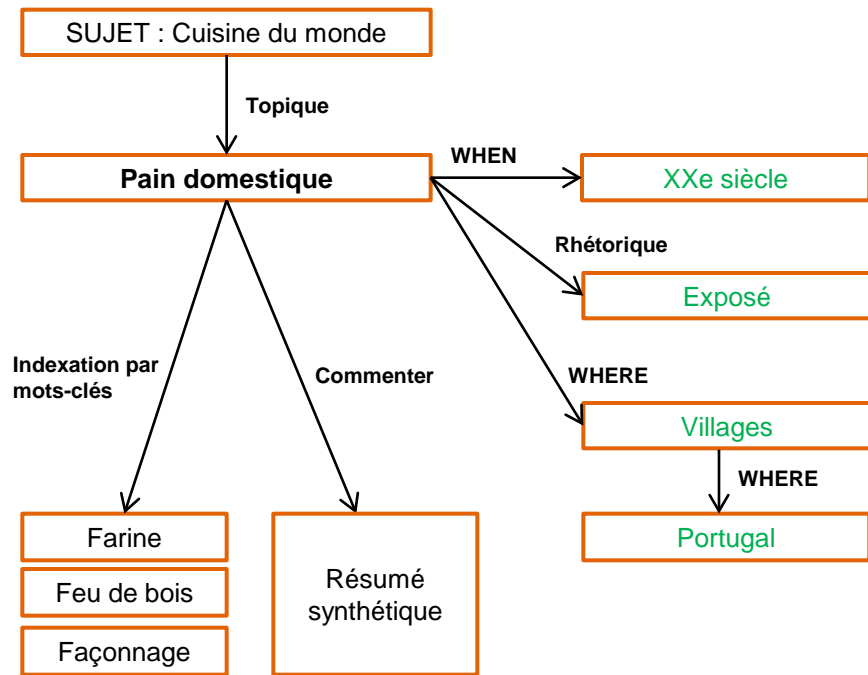
- <http://www.archivesaudiovisuelles.fr/FR/video.asp?id=1396&ress=4366&video=79105&format=68> (FMSH-ESCoM)

-> Les exemples qui suivent sont fictifs et seulement illustratifs !

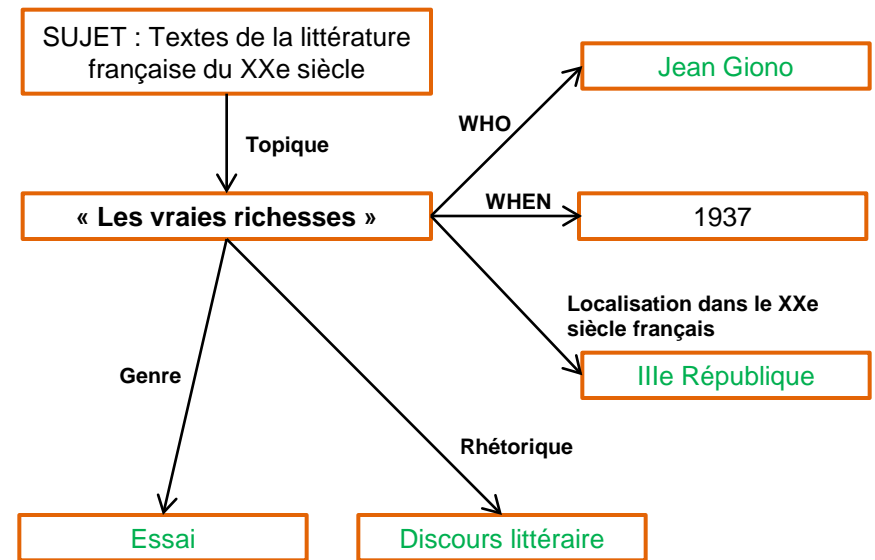


ANALYSE : EXEMPLE

Topique « Fabrication du pain domestique au Portugal »



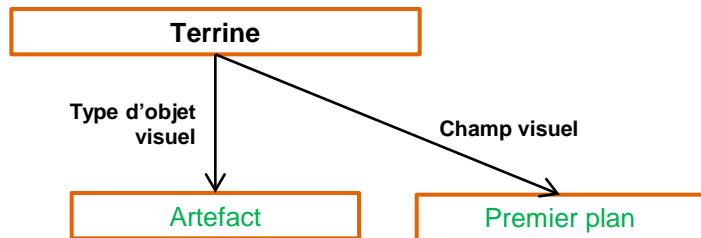
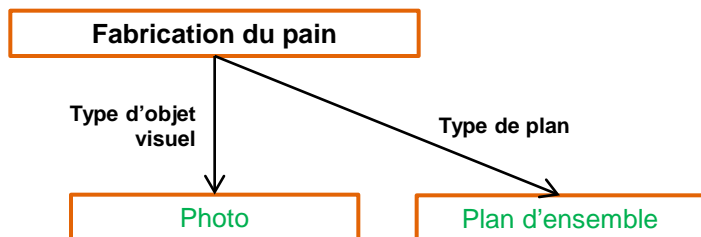
Topique « œuvre littéraire »



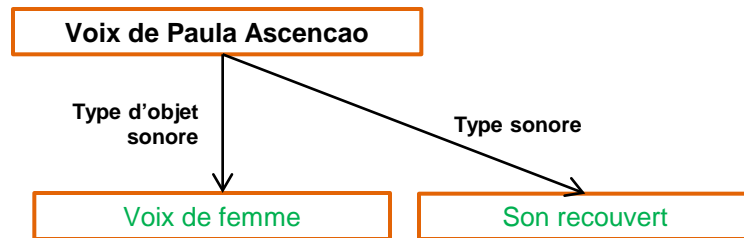
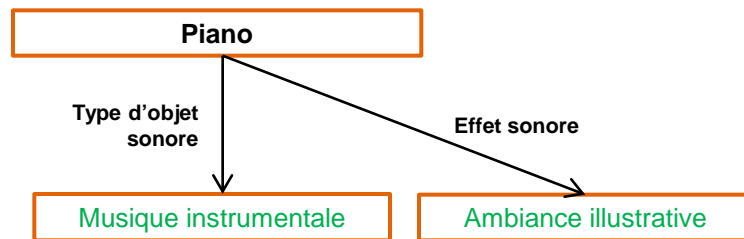
En vert : entités nommées partagées

ANALYSE : EXEMPLE

Analyse visuelle



Analyse sonore



En vert : entités nommées partagées

PUBLICATION (STUDIO ET CMS CAMPUS-AAR)

● Publications auteur

- Structures et règles définis par les **modèles** (ontologies)
 - Vidéo-lexique, parcours narratif, ...
- Intégration de contenus grâce à des **requêtes sémantiques**
- Travail **éditorial** : ajout de rubriques, de textes, d'images, ...
- **Exportables** sur les portails CAMPUS-AAR (ou autres)

● Publications automatiques

- **Portails** dédiés à chaque domaine (incluant un CMS)
- **Accès** au contenu (média ou segment) selon différents filtres
 - sujets (**topiques**), lexique (thésaurus), **carte géographique**, **frise temporelle**, contributeurs, plans visuels, plans sonores, ...
 - **Dossiers audiovisuels** : sélection de médias (auteur)
 - **Widgets** exportables dans toute page d'un portail

PUBLICATION : EXEMPLE

- Publication auteur : « Fabrication du pain domestique au Portugal »
 - Intro : Segment dont la topique est « pain domestique » ET localisée au Portugal ET avec la rhétorique « Définition »
 - Exemple: Segments dont la topique est « pain domestique » ET ayant un plan visuel contenant le mot « pain »
 - Exposé : Segments ayant pour topiques « pain » et « Portugal », avec pour rhétorique « exposé » et localisation temporelle > 1850
 - Etc.
- Accès automatiques
 - Par sujets : « Cuisine du monde » > Tous les segments ayant pour topique « Pain domestique »
 - Par lieu géographique > tous les segments dont une topique est localisée au Portugal
 - Lexique : tous les segments dont la topique est « Jean Giono »
 - Par objet sonore : tous les segments avec de la musique instrumentale
 - Etc.

 [Démon Studio Campus-AAR](#) (INA, 5 mn)

 [Démon CMS Campus-AAR](#) (Armadillo, 5 mn)

- Blog *Hypothèses* : <http://campusaar.hypotheses.org>
- Portail Campus-AAR : <http://campus-aar.armadillolab.fr>
- Prototypes
 - AGORA : <http://campus-aar.armadillolab.fr/agora>
 - ARC : <http://campus-aar.armadillolab.fr/arc>
 - AHM : <http://campus-aar.armadillolab.fr/ahm>

- Studio ASA : <http://www.semioscape.fr>
- AAR : <http://www.archivesaudiovisuelles.fr>
- Canal-U : <https://www.canal-u.tv>
- MediHAL : <https://medihal.archives-ouvertes.fr>

Francis Lemaitre

Pôle Service et Infrastructures Numériques
Fondation Maison des Sciences de l'Homme
lemaitre@msh-paris.fr