Indexation des fonds et flux radio- et télédiffusés

Usages en vigueur et perspectives autour de la refonte du système d'information documentaire de l'Institut national de l'audiovisuel

Malika Mekdad, Hélène Allioui, Eleonore Alquier

2ème séance de l'atelier
Archives audiovisuelles et recherche
3 mai 2016

ina

Introduction : l'Ina en quelques dates

- **1974-1985 :** A l'éclatement de l'ORTF, l'Ina hérite des missions liées aux archives, à la formation, à la recherche et à la production de création
- **1986-1995 :** Tout en poursuivant ses recherches sur les nouvelles images, l'Ina se voit confier une nouvelle mission : le dépôt légal de la radio et de la télévision
- **1996-2005 :** Au tournant du millénaire, l'Ina lance un plan de sauvegarde et de numérisation de masse visant à assurer la pérennité de ses fonds.
- **2006-… :** Après la numérisation, l'accès : l'Ina ouvre ses fonds au grand public, via le site ina.fr, dont la troisième version est en préparation

Introduction : l'Ina en quelques chiffres

128,5 millions d'euros de budget en 2015

980 collaborateurs (fin 2013)

- CONSERVER

12 000 000 d'heures de documents télé et radio

13 000 sites web médias, 1 200 000 photos

120 chaînes captées au titre du dépôt légal (TV+radio)

1 200 000 heures sauvegardées et numérisées

- VALORISER

1 million d'heures de programmes numérisés en ligne (inamediapro.com)

103 millions de vues en 2014 et 43 000 heures de programmes sur ina.fr

60 films en productions et coproductions par an

- TRANSMETTRE

3 000 professionnels formés chaque année, 560 stages de formation continue au catalogue Ina EXPERT, 14 formations initiales en BTS, Licence et Master Ina SUP

5 projets de Recherche audiovisuelle en France et à l'international

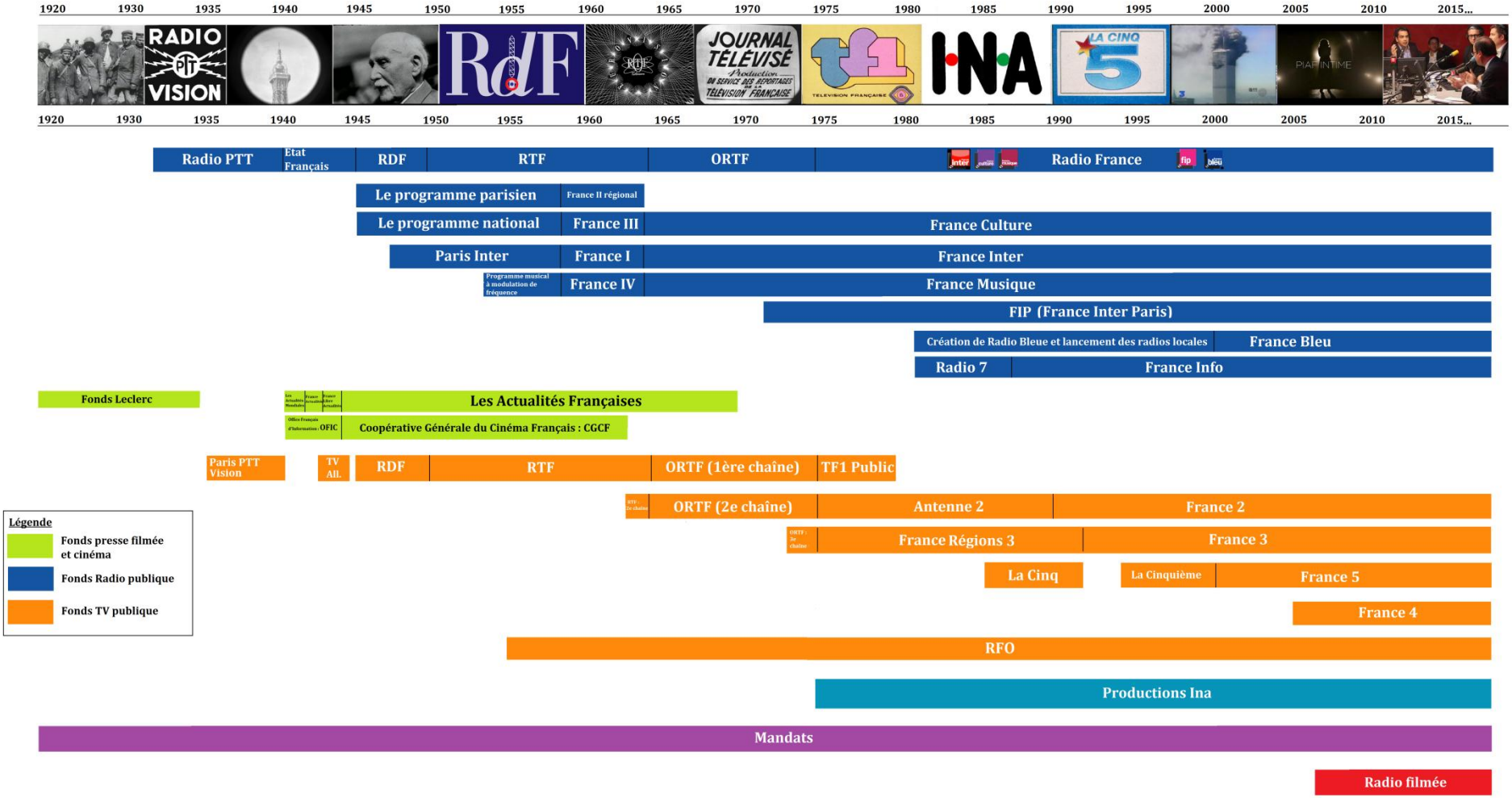
TRAITEMENT DES FONDS « HISTORIQUES » ET PROFESSIONNELS

75 ans de radio et 65 ans de télévision

Au total, + de 12 millions de notices documentaires et 1 million d'heures de programmes en ligne sur le site inamediapro au service des professionnels :

- Les émissions nationales et régionales diffusées par les chaînes françaises de télévision publique depuis 1949
- Les émissions produites et diffusées par les chaînes de Radio France
- La presse filmée des "Actualités françaises" (1939-1969)
- Les fonds privés de chaînes, producteurs ou réalisateurs acquis dans le cadre de mandats (TF1, AFP…)
- Les productions Ina depuis le Service de la Recherche de l'Ortf jusqu'au DPA.
- Le fonds photo : environ 1 200 000 clichés pris par les services photos de la RTF, l'ORTF, la SFP puis l'Ina depuis 1950
- Le fonds des images amateurs

Chronologie des fonds professionnels



Différentes formes de valorisation des fonds professionnels (1) :

Le traitement documentaire assure l'enrichissement des contenus documentaires des **collections** du fonds historiques radio et Tv et des mandats :

- Sélection des collections en collaboration avec les partenaires Ina (en fonction des grands projets de l'entreprise et de l'actualité
- Envoi en sauvegarde, restauration, numérisation des matériels
- Segmentation des fichiers numériques
- Analyse et indexation des images et des sons

Corpus multimédias et éphémérides : réalisation **d'ensembles documentaires** relatifs à un thème, une personnalité ou un évènement permettant de proposer aux clients de l'Ina une offre cartographiée et enrichie (563 corpus proposés sur le www.inamediapro.com et sur l'edatebook) :

- Recherche documentaire et sélection de documents radio, tv et photo
- Enrichissement documentaire des notices
- Création et indexation d'extraits et structuration du corpus

Différentes formes de valorisation des fonds professionnels (2) :

La préparation éditoriale : recherche, sélection, et indexation de contenus diversifiés pour :

- Pour le site grand public ina.fr
- les trailers mis en ligne sur inamediapro.com
- les fresques réalisées par le studio hypermédia
- et tous nouveaux modes de diffusion et de valorisation des contenus sur les réseaux sociaux, les sites partenaires, les plateformes de visionnage…

Garantir l'intégrité des bases de données : qualité des bases de données, actions de contrôle et de corrections des bases documentaires et matérielles :

- Contrôle et correction des données documentaires et matérielles du fonds historique
- Relations avec la hotline et les partenaires Ina : correction des données présentes sur les sites.
- Préparations des données documentaires avant import (exemple : RFO)

Base de données Totem

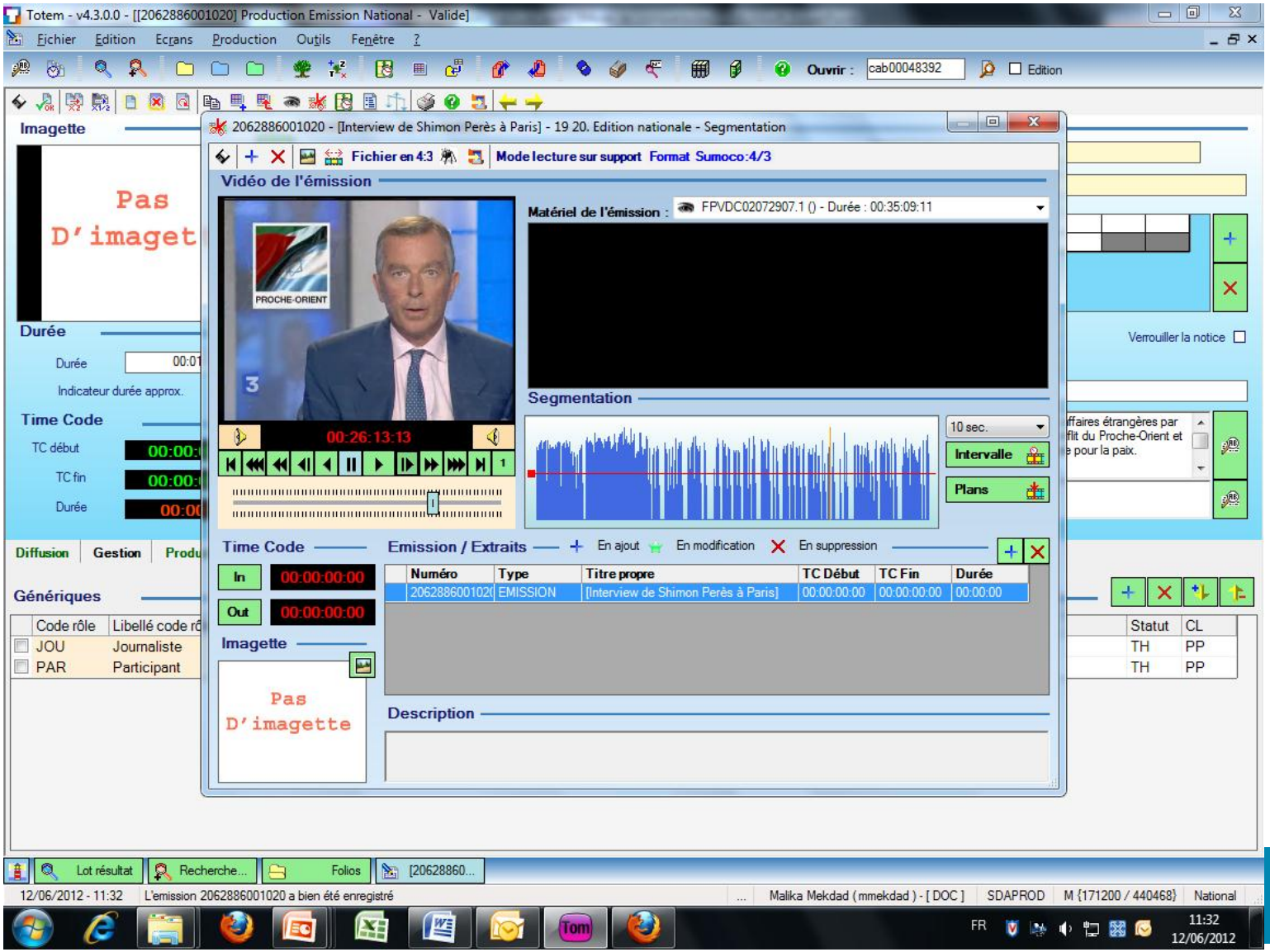
Développé en interne par les équipes de l'Ina pour les besoins spécifiques des Archives, Totem est une application qui réunit à la fois la production de données et la recherche documentaire. Totem pilote une base Oracle associée au moteur de recherche Exalead.

- Outil multimédia : vidéo, photo, radio
- Recherche documentaire tous médias
- Production documentaire :

  création de notices, enrichissement, indexation, segmentation de fichiers numériques
- Création d'extraits
- Constitution de corpus multimédia
- Export vers les sites ina.fr et inamédiapro

+ de 12 millions de notices documentaires et 1 million d'heures de programmes

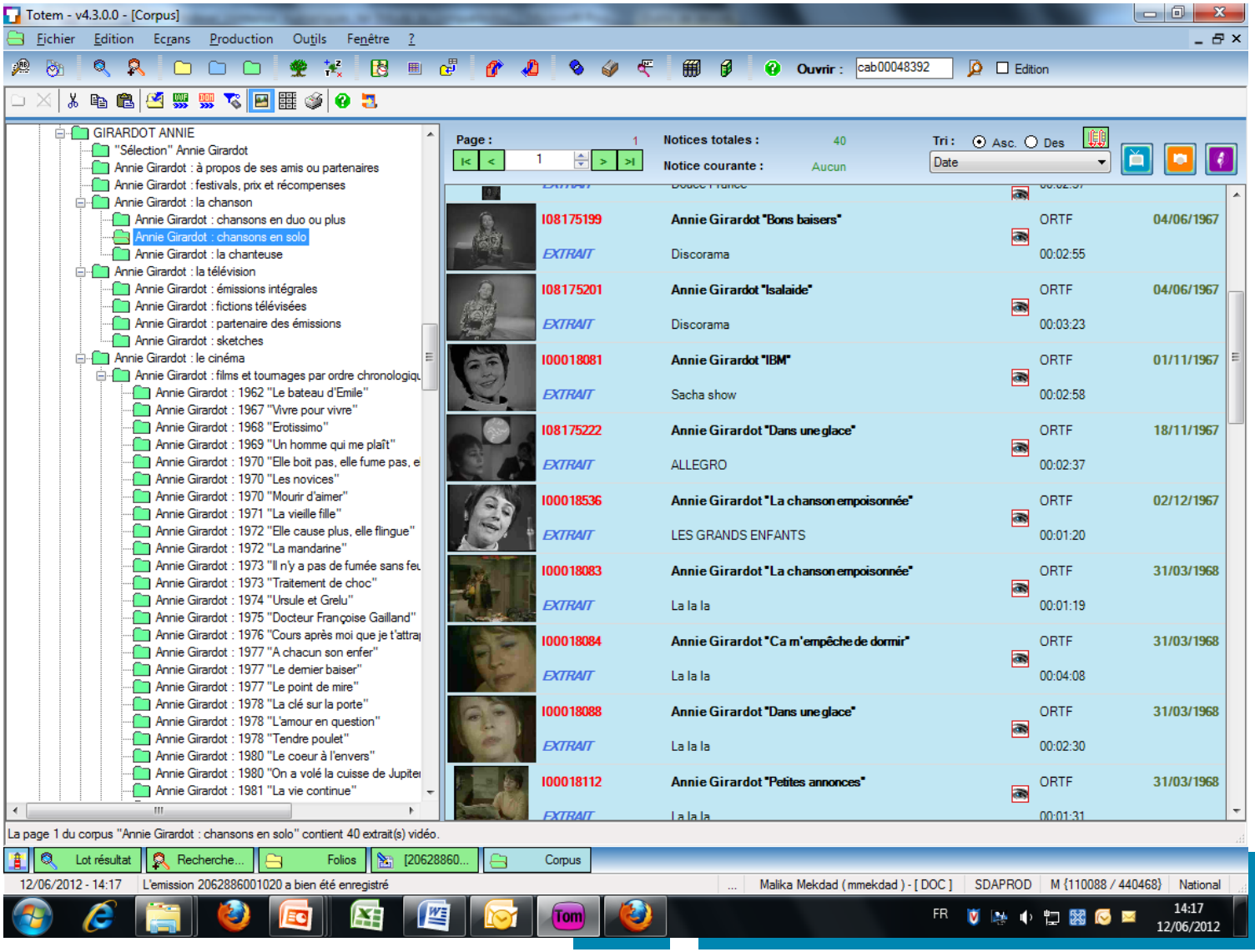
Un outil de visionnage, de segmentation et d'indexation



Exemple de notice documentaire (après reprise d'indexation)

Identifiant de la notice :	**CAF12005351**
Titre propre :	**Gérard Larrousse : pistard et routier**
Titre collection (Aff.) :	**Pourquoi pas l'exploit**
Diffusion (aff.) :	**03/01/1975 - type date: Diffusé -heure:17:30:03 - canal:1ere chaîne (ORTF) -ext.géo.:Nationale**
Durée :	**00:03:38**
Nature de production :	**Production propre**
Producteurs (Aff.) :	**Producteur ou co-producteur - Office national de radiodiffusion télévision française (ORTF) - Paris - 1974**
Genre :	**Magazine ; Reportage ;**
Thématique :	**Sports ;**
Générique (Aff. Lig.) :	**JOU Collaro, Stéphane ; PAR Larrousse, Gérard ;**
Descripteurs (Aff. Lig.) :	**DET: sport automobile ; DET: pilote automobile ;**
Résumé :	**Gérard LARROUSSE a totalisé 11 victoires à son actif cette année. C'est chez lui, en train d'entretenir son vélo en survêtement, que Stéphane COLLARO a rencontré le pilote. L'interview est entrecoupée de plans d'archive de courses en sport prototype, en formule 1 et lors de sa victoire aux 24 Heures du Mans avec Henri PESCAROLO.** Gérard LARROUSSE exprime son intérêt pour le vélo et le sport en général, parle de son parcours particulier qui l'a mené du rallye au circuit. Il évoque aussi son expérience en formule 1 lors du Grand prix de Belgique. Archives : course de sport-prototype 2 litres avec la voiture de Gérard LARROUSSE (écurie Renault), Henri PESCAROLO et Gérard LARROUSSE portés en triomphe lors de leur victoire aux 24 Heures du Mans 1974, images du Grand prix de Belgique de formule 1 à Nivelles avec la participation de Gérard LARROUSSE.
Notes :	**Alternances d'images couleur et noir et blanc**

La gestion des corpus dans l'outil Totem



TRAITEMENT DU FLUX : LE DÉPÔT LÉGAL

Le dépôt légal

Télévision : 100 chaînes

- Hertziennes
- Câbles et satellites, TNT
- FR3 régions

Le dépôt légal

RADIO : plus de 50 chaines

- Publiques
- Privées
- Locales de France bleu

01/01/1995

09/03/1999

01/01/2001

01/01/2001

01/12/2008

15/02/2012

07/04/2012

LE TRAITEMENT DOCUMENTAIRE DU FLUX RADIO ET TELEVISION

- SES CARACTERISTIQUES
- LE VOLUME DES NOTICES DOCUMENTAIRES
- LA CHAINE DOCUMENTAIRE

L'INDEXATION DE LA NOTICE DOCUMENTAIRE

Document / Synthèse FR5

Session Document Edition Choisir Impression Ecrans Rapprochement Aller à ?

Synthèse - Format page

IDENTIFICATION N° DL T 20150210 FR5 005 001

Chaine	Mardi	Début	JS	Fin	JS	Durée
FR5	10/02/2015	09:00:36		10:11:11		01:10:35:00

Tranche horaire
Collection: Les maternelles
Titre: Les maternelles : [émission du 10 février 2015]
Titre ELP: Les maternelles : [émission du 10 février 2015]
Id.Pro. Imedia 88419420
Réf. ext.

☒ Sélection DL.
☒ Sélection PRO
☐ Sélection THEMA
Repérage 09:00:36:00
Statut Première diffusion
Statut index. CATA
CC bar
☒ Validation CC;
DOC zan
☒ Validation DOC;
Type Desc. 31
Type ELP SSM
Fonds FR5PRO

Thématiques
Genres: Magazine
Typologie médiamétrie: EBL
Publics: Participation du public à

DESCRIPTION

Générique			Descripteurs	
AU	REA	RIMOND DE	DE	SOCIETE
GEN	PRE	VIGNALI JUL	DE	ECOLE PRIM.
GEN	PAR	RAWLINS AN	DE	COLLEGE
GEN	PAR	REVOL OLIV	DE	HARCELEME
GEN	RED	MALPHETTE	DE	DOCUMENTA
GENS	PRE	LE BRETON	DE	ECOLE MATE

Nat. Prod. 2
Société(s) de Production
France Télévisions
Neria Productions

Informations complémentaires...

Chapeau
Présenté par Julia VIGNALI, ce magazine consacré à la maternité et à l'éducation des enfants propose des chroniques, un dossier avec des invités et des reportages.

Résumé
UN MANIFESTE CONTRE LE HARCELEMENT SCOLAIRE
Entretien avec Andrea RAWLINS, réalisatrice pour l'agence Capa d'un documentaire sur le harcèlement scolaire diffusé dans le magazine Infrarouge.

Notes...
Notes : Diffusé en stéréo
NTI :
N.Gest. :

Oeuvres

Dispositif

Résumé Producteur
[Source iMedia]
En compagnie de chroniqueurs, Julia Vignali accueille témoins et experts qui viennent échanger et partager leurs expériences autour de la parentalité.

Sommaire
[Plateau invités : 1ère partie]
La minute puériculture : [habiller un bébé de 8 mois]
[Plateau invités : 2ème partie]
L'école maternelle : 3 ans pour devenir élève
[Plateau invités : 3ème partie]

CHAMPS TEXTUELS

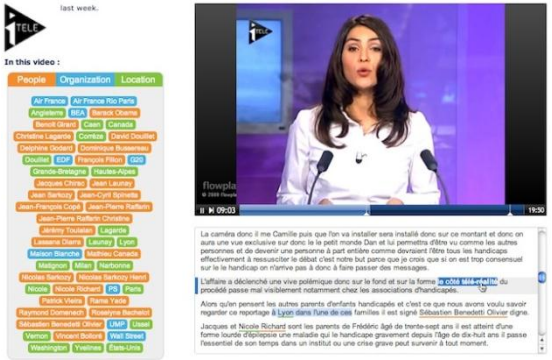
IMPORT S EXTERIEURS

DECOUPAGE EN SUJETS DE SOMMAIRE

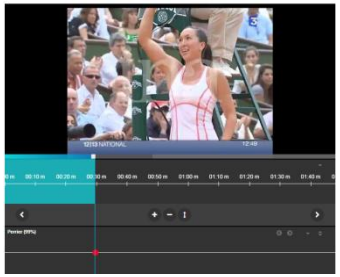
VERS UN SI « INNOVANT »

Industrialiser les techniques innovantes issues du département R&D

- Intégration des pratiques du big data / politique d'open data
- Transcription automatique avec
  - Extraction d'entités nommées
  - Détection de citations
  - Alignement du texte sur des time codes



- Connexion à des référentiels et sources extérieurs (dictionnaires d'œuvres, d'éphémérides, …)
- Analyse et reconnaissance de sons et d'images (visages, monuments, œuvres d'art, voix, …)



Merci de votre attention

ealquier@ina.fr

http://www.institut-national-audiovisuel.fr/en/home