From Archive Broadcasting to Web-based Archive Services: Examples from INA

Bruno Bachimont
Institut National de l’Audiovisuel
Université de Technologie de Compiègne
Some words about INA
Key facts

The 1974 law concerning freedom of communication, creates seven audiovisual societies from ORTF:

INA is an industrial and commercial public institution, founded on January, the 6th, in 1975
INA position: present

- Preservation and valorization of the French audiovisual memory
  - Radio and TV
  - Soon (!), Media part of the Web
  - Every content that is temporal and that is streamed (hertzian, cable, web, etc.).

- Two main modalities:
  - Audiovisual legal deposit:
    - All broadcasted programs (public or private), with a French production;
  - Archiving the public broadcasting channels:
    - Radio France, France Télévision.
Two types of exploitation

• Legal deposit: a *national* heritage
  – Accessed by researchers to found their scientific inquiry;
  – Tools provided by INA to help researchers consult AV materials.
  ➔ Contents do *not* belong to INA, nor to State, but to the Nation: cannot be sold and re-exploited.

• Archiving Public channels: a *public* heritage
  – Sold for re-exploitation to producers, journalists, channels, etc.
  ➔ Contents belong to INA (and many rightholders): they can be sold and re-exploited.
Some numbers:

- **Professional Archives:**
  - 1,1 million hours of radio and TV programmes;
    - 575,000 hours / Radio;
    - 535,000 hours / TV
    - + ~ 50,000 hours / year

- **Legal Deposit:**
  - 930,000 hours (NB: professional archives are partially included in legal deposit).
    - 430,000 hours / TV
    - 500,000 hours / radio
    - + 500,000 hours / year

- 2,5 millions documents covering 113km of shelf space;
  - 8km / year

- 133 years for watching or listening all archives;

- 220,000 hours digitized and accessible online (B2B).
INA situation: future

- Making accessible the French Audiovisual Heritage to main public and end users:
  - Technologically possible: end users have equipment required to receive and treat contents
  - Culturally necessary:
    - demand for public memory;
    - TV is now an old media: more and more programs have never been watched by contemporaneous people;
    - INA has a monopole on this kind of programs.
From an economical point of view

• Making money with archives:
  – Selling programs or excerpts for re-broadcasting:
    • Small increase in the future because:
      – Broadcasting is already 24h/day and 7d/week;
      – People capacity to watch TV is not extensible (3h30/day !)
    • New incomers are not rich enough to invest in the INA high quality archives
  – Selling programs or excerpts for downloading:
    • From program guides (duration constraints) to catalogs (organisational constraints);

• Conclusions:
  – Slow market increase for broadcasted excerpts.
  – Necessity and opportunity to consider new markets offered by content distribution.
Our perspective

1. Structuring
   • archived contents into *collections* composed of contents and excerpts selected according to their meaning.

2. Documenting
   • Collections in order to make intelligible and usable selected contents.

3. Repurposing
   • Using documentation to edit contents according to
     • the delivering device and format,
     • the targeted audience.
Multiple perspectives
New business, but new practices
Digital archive distribution

Distributing archives on line is not only a new market, it is a new business and a new professional practice.

- Nature of indexing is changing;
- Status of content is no longer the same;
- The archival stance should be adapted.
From index to metadata

• Traditionally:
  – Index:
    • To point the document relevant to the aimed use;
  – Purpose:
    • To retrieve documents as they are;

• Current trend:
  – Metadata:
    • Information that makes content useful and usable
  – Main property:
    • Can point to any arbitrary part of content;
  – Purpose:
    • To select a segment to reuse it in a new context.
From document to resource

• What is called “document” is from now on a “container” from which one can extract segments.
• Those segments are resources whose meaning is determined by use conditions and exploitation context.
• Initial coherence of the source document is no longer what constitutes content and defines it.
From retrieving to editing

- What motivates indexing is information retrieval:
  - Find out documents that contain and express the desired information.

- What motivates metadata is to select resources in order to create new contents.
  - The resource value is no longer a consequence of its original context (production context) but a consequence of its exploitation context.
From indexing to repurposing

• Objective:
  – Find out resources to create new contents

• Constraints:
  – Segments are decontextualized from their original context;
  – Need for recontextualizing them in their exploitation context.
    • Specific rendering device;
    • Coherence with other segments of the final created content.

Creating new information and new content to integrate selected resources.
Several perspectives

• The Genealogical stance:
  – Resources are selected to be enriched and situated in their original context;
  – The editing process has to publish the archival knowledge..

• The Amnesic stance:
  – Selected resources are reused without any consideration of their original meaning.
  – The editing process is a new creation using archives as stock shots.
Histoire du journal télévisé

1960 : le modèle de la radio

JT, RTF, 24.11.1960

Claude Dargent semble un journaliste particulièrement décontracté en faisant écouter une correspondance d'Alger sur la tentative d'insurrection menée par Lagailarde et Ortiz. C'est qu'il vit et pense son travail sur le modèle de l'information parlée de la radio : l'image du direct est encore très rare et l'attitude même du journaliste n'est pas pensée pour capter l'attention du téléspectateur.
Main issues

• Technical issue:
  – Dealing with content repurposing on a big scale:
    • Mass editing
    • Need for industrial solutions to manage huge amounts of data as well as complex process

• Cultural issue:
  – Dealing with content repurposing in a “faithful” way:
    • Genealogical editing
    • Need for technical solutions to manage complex knowledge representations and document structures.
Industrial Aspect
Repurposing

Génération/validation de métadonnées

Sélection/validation

Extraction de métadonnées
(structure, index, régions d’intérêt)

Codage scalable

Sélection du contenu
- extraits de vidéo
- crop
- résolution
- métadonnées

Edition
- édition Web
- création de résumés

Edition basée sur l’application finale

Metadonnées + vidéo scalable

Interaction utilisateur

Production

Vidéo
Métadonnées

Enrichissement

WEB

Mobile PDA

DVD

PlayOut
Stream line the operation matching workflow

Handle complex content management

Aggregate Metadata for rich media repurposing

Facilitate operation with task based action

Include Full text capability

Show content according search / task

Allow video segmenting and trimming

Speed-up production with pre-editing
Research aspects
Using what we know
And handling content
Conclusion

• Opportunity coming from the shift to content distribution.
• Paradigm shift in documentation and archiving;
• Needs for tools to represent content and meaning in order to repurpose it in a genealogical fashion

A question for the Semantic Web?