



Multimedia Content Search: The view from the large Audiovisual Archive

Chorus Workshop
13 March 2007
Inria, Rocquencourt

Daniel Teruggi, Head of Research

dteruggi@ina.fr

Ina: Institut National de l'Audiovisuel



Institut national de l'audiovisuel



Missions:

- National Broadcast Archive, collections starting in 1933 for Radio and 1949 for Television
- Legal Deposit for Radio and Television since 1995
- Research and Experimentation (participation to major projects: PrestoSpace, Quaero, Infom@gic, Vitalas, Mesh, etc.)
- Archive based production
- Professional Training, Master in Multimedia Production and Digital archiving (starting september 07)

Status:

Public Industrial and Commercial
Enterprise



1959 - Tournage "Cinq Colonnes à la Une" pendant la Guerre d'Algérie
André Hugues - Photo INA



Professional Archives :

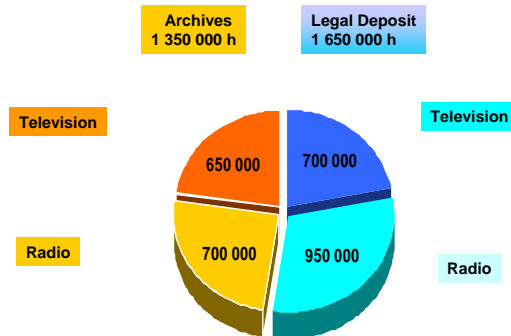
1 350 000 hours, 700 000 hours of radio and 650 000 hours of television
(Public broadcasters)

Ina owns the Production Rights, 30000 new hours every year. Any possible analogue format! Highly menaced collection, 300.000 hours already digitized and on-line

Radio and Television Legal Deposit :

1 650 000 hours, 950 000 hours of radio and 700 000 hours of television
(Public and Private)

51 Tv channels 17 radio channels, 300.000 new hours every year, digital media on shelf. No commercial use, access restricted for research only.



Ina makes this memory accessible



To professionals – through digitization and theme organisation

Since 2004 Ina proposes a unique service
www.inamedia.com



Largest on-line collection of digitized archives, more than 280 000 hours of Images and almost 3 million documents

*french and english versions
4000 registered users, 35% coming from outside France*



To scientists – to permit analysis and comprehension of the audiovisual domain

8 000 researchers are registered at the Inathèque premises at the National Library for access to the Legal Deposit



To the public – sharing the national memory

Since April 2006, through the website
www.ina.fr

Ina proposes an on-line access to 12000 hours of radio and television, for streaming or downloading. Specific contents dedicated to educational issues for secondary schools



Documentation is a major issue for Ina:

- Most of the video and film contents are already hand-documented, including Right management issues
- Many old audio documents are not documented
- An associated collection of photographs is undergoing a documentation and indexation process
- Documentation is strongly use oriented, description schemes were inherited from broadcast production
- Documentation is based on a huge, continuously updated, Thesaurus
- Ina is evolving from a mass Archive to an Archive structured in collections and themes



Documentation evolution in the recent years:

- Due to the large amount of documentalists (150), a certain fear has developed in the past concerning job security!
- Today there is an identified need for documentation assistance, in order to accelerate hand-documentation
- The passage from B2B to B2C has had strong implications in the documentation schemes and necessities
- The perception of documentation is changing, dynamic multimedia indexation will be a major issue in the future for professional and consumer uses



Major challenges for indexation:

- Indexation and documentation are never ending actions, they depend on uses and change objective through time
- Hand documentation is extremely precise and efficient for large collections, however it is not updatable and cannot deal with the amount of entering material
- Scalability is the major issue when searching for a content among millions, mainly when different contents are searched on the same document
- Organising contents in structured collections based on themes or other criteria
- Searching a specific content versus exploring a collection



Some essential questions:

- How to find sunsets?
- Is user annotation an issue for broadcast content holders?
- How to pre-process entering data (330.000 hours per year) without generating much noise?
- Content structuring and thematic clustering



Other issues for Audiovisual Archives:

- Storage issues for the future: grids, replication, long-term migrations, archive linking for content securisation (Web archiving)
- Ensuring the long term integrity of digital contents (security, integrity, tracability, persistency, etc.)
- User-generated documentation and their integration in documentation schemes and practices
- Documentation and Preservation models for complex contexts



Merci de votre attention

