Audio/Video Digital Libraries:

designing, searching for documents, and generating metadata

Giuseppe Amato
Claudio Gennaro
Pasquale Savino
ISTI-CNR

ECHO
European CHronicles On-line

DLib Center
Digital Library Competence Center

The goal

Provide a theorical and experimental background on the techniques and the methodologies for the organization, creation, and management of an Audio/Video Digital Library.

Tutorial program

- Introduction to Audio/Video Digital Libraries
- How to build an Audio/Video Digital Library
- A practical example: the creation of a documentary film Digital Library
- Metadata models
- Automatic indexing of A/V documents
- Manual indexing of A/V documents
- The ECHO metadata editor

Audio/Video Digital Libraries

An introduction

Pasquale Savino
ISTI-CNR

European CHronicles On-line

Digital Library Competence Center

Outline

- What is a Digital Library?
- Characteristics of an Audio/Video DL
- Applications of Audio/Video DLs
- Types of data managed
- The characteristics of digital Audio and Video
- The main functions
- Automatic and manual indexing
- Retrieval functionality
- Logical architecture of a video DL
- User's categories
- Overview of existing systems

Outline [Part 1]

- What is a Digital Library?
- Characteristics of an Audio/Video DL
- Applications of Audio/Video DLs
- Types of data managed
- The characteristics of digital Audio and Video
- The main functions
- Automatic and manual indexing
- Retrieval functionality
- Logical architecture of a video DL
- User's categories
- Overview of existing systems

What is a Digital Library?

Definition

A Digital Library is an organized collection of digital objects, including text, images, audio, video and services for its access and retrieval, as well as for selection, organization and maintenance of the collection.

The digital objects

- In general, a Digital Library may contain not only text documents but also ….
  - Video
  - Audio
  - 3D objects
  - Virtual-reality worlds
  - ……..

Key library services

- Access and retrieval
  - Catalogs
  - References
  - Indexes
- Preservation
- Management
  - Access control
  - Data sharing
  - Management of collaboration
    - E.g. collaborative filtering, cataloging, …..
  - ……

The importance of video

- Video can be considered today the primary information and communication channel, due to
  - Richness in information contained
  - Appeal
- Video libraries will become essential in many application fields
  - Personal information
  - Distance learning
  - Telemedicine
  - ……

Video characteristics

- High video production vs print production
  - TV stations produce 50 Million hours of video per year (25,000 TB)
  - Newspapers and periodicals produce less than 200 TB of data per year
- Storage and transmission problems
  - Video is usually compressed
- Richness in content
  - Difficulties in automatic extraction of content description

Services of A/V Digital Libraries

- Digital Video Libraries are more complex than traditional DLs; they require the integration of several specialized technologies
- They offer the same services of text digital libraries
- Specific characteristics of Indexing and retrieval services
  - Indexing based on the integration of different technologies for the automatic feature extraction
  - Integration of manual and automatic indexing
  - Retrieval based on different video features

Characteristics of an Audio/Video DL

The need of A/V DLs

- Nowadays, video is present in many situations
  - TV broadcasting
  - Professional applications, such as medicine, journalism, advertising, education, training, surveillance, etc.
  - Movies
  - Historical videos
  - Personal videos
- The combination of audio and video is a very powerful communication channel
  - approximately 50% of what is seen and heard simultaneously is retained

Advantages of A/V DLs

- Most of the video material produced is used only once, due to the difficulty to archive it, to protect and to retrieve.
- A large video library of distributed and network searchable videos would enable
  - Preservation of precious and expensive material
  - Reduction of production costs for new videos, through the reuse of existing material
  - Diffusion of knowledge

In general, it will enable the access to information that could have been lost.

A/V vs traditional DLs [1/2]

- Library creation
  - Traditional DLs, contain text documents
    - Library creation requires automatic acquisition of text, extraction of document content, and indexing
    - This process is well known and many different techniques have been developed
  - Video is extremely rich in “content” but ….
    - the indexing of video content is difficult, expensive, and extremely dependent from the user and the application
    - A possible approach consists in an appropriate integration of automatic content extraction (e.g. speech recognition, image analysis, etc.) and manual indexing

A/V vs traditional DLs [2/2]

- Library exploration
  - Traditional DLs, contain text documents
    - Library exploration requires simple interfaces to formulate queries on free text and document metadata.
  - Video libraries should permit
    - To formulate queries on many different “dimensions”
      - Text, as extracted from speech and captions
      - Images extracted as key frames
      - Motion information
      - Other features automatically extracted
      - Metadata provided manually

Applications of Audio/Video DL

Who may use A/V DLs?

- We consider four main categories
  - Large companies
    - Large corporations that may use Digital Video for their internal business, for advertising, promotion, etc.
  - Media and entertainment
    - The most traditional area. Video is one of the key assets.
  - Education
    - Video recording of courses
    - Video used as course material
  - Others
    - Health and medicine
    - Government
    - Surveillance
    - Etc.

Large companies

- Audio/video digital libraries are used for
  - Sales
  - Product launches
  - Marketing
  - Relation with investors
  - Product design (acquisition and analysis of customer's needs)
  - Support for online sales
  - Video archives for internal use
  - Special services for customers, such as web access to specialized video archives, e.g.
    - News
    - Economic information
    - Products
    - Materials
    - Etc.

Media & Entertainment [1/3]

- Broadcasting companies
  - Many broadcasters are creating and distributing video programs on the web. A video archive is very helpful to them to add a new service for accessing old video material.
  - Examples:
    - ABC News
    - Mediaset
    - RAI
      - Archive of old programs
      - Archive of daily programs
      - Additional material w.r.t. tv programs

Media & Entertainment [2/3]

- Video archives
  - Many national and private organizations own old video material. The digitalization and archiving of this material is beneficial for content owners (for example, they can promote the use of their material) and for users belonging to different categories: e.g. professional users (that need the material to produce their video programs) or researchers or general public.
  - Examples:
    - Istituto Luce

Media & Entertainment [3/3]

- Movie production companies
  - Many large movie production companies own a large amount of video material, composed of the films and of related material, such as cuts not used in the final film version, interview, video trials, etc. This material is very helpful for many purposes, from the production of DVD version of the film up to the critical study of the video. Providing access to the general public of this material is also a powerful promotion and advertising channel.
  - Examples:
    - MGM
    - 20th Century Fox

Education

- Digital video used for different purposes
  - Promotion and advertising
    - Online preview of training content
  - Store and distribute the video courses
    - Remote access of the courses
    - Keep track of classroom discussion
  - Used as course material
    - Delivery of video clips to students, either online or in the classroom
    - From remote sites provide students and teachers with on-demand, searchable access to whole programs and video clips
    - Free search and access to the video library con be used by students to find answers to specific questions, to study in depth some topics, etc.
  - Production of new courses
    - Improve the course production procedures, allowing teachers and producers to remotely access the video library
  - Examples:
    - Princeton University
    - Harward Business School
    - University of Arizona

Other Applications [1/2]

- Health and medicine
  - Health and social care info to the general public
  - Information to physicians for special purpose medical procedures
  - Training

Other Applications [2/2]

- Government
  - Enhancement of the governmental decision making process, by recording and archiving of public meetings and discussion.
- Surveillance
  - A large amount of video is produced for surveillance purposes.
    - Required automatic video analysis
    - Archiving for successive search

Outline [part 2]

- What is a Digital Library?
- Characteristics of an Audio/Video DL
- Applications of Audio/Video DLs
- Types of data managed
- The characteristics of digital Audio and Video
- The main functions
- Automatic and manual indexing
- Retrieval functionality
- Logical architecture of a video DL
- User's categories
- Overview of existing systems

The characteristics of Digital video

European CHronicles On-line

Digital Library Competence Center

Types of data managed

- A digital video is composed of a sequence of frames plus possibly an audio track.
- In general, it is possible to view an audio/video document from different perspectives
  - The audio part can be separated into
    - Speech
    - Sound
  - Sequence of frames (video shot and sequence)
  - Single frames as images
- From all of them is possible to extract information that can be used for indexing and retrieval purposes

Digital video characteristics

- Sequence of frames with a certain frame rate
  - NTSC 30 frames/sec, PAL 25 f/s, HDTV 60 f/s
  - Minimal change between frames
- Single frames resolution
  - 768 x 576 PAL, 720 x 480 NTSC
- Uncompressed video requires high storage space and bandwidth
  - For example, one second of uncompressed PAL video requires

    768 x 576 x 16 x25 ~ 172 MByte

Digital video storage and transmission [1/3]

- The high storage requirements of video imposes the adoption of compression techniques.
- High compression rates are possible with video signals, due to the following reasons:
  - Spatial correlation: correlation among neighboring pixels
  - Temporal correlation: correlation among pixels in different frames
  - A significant part of video data is not perceived

Digital video storage and transmission [2/3]

- Compression can be divided in two broad categories
  - **Lossless compression**, that allows one to compress decompress video without any degradation
    - Lossless compression provides low compression factors
    - An example of lossless compression is MJEPG, where each frame is compressed using the JPEG format
    - Examples of lossless coding techniques are run-length coding, Huffman coding

Digital video storage and transmission [3/3]

- **Lossy compression**, where the complete cycle of compression and decompression introduces some degradation of the original video
  - Lossy compression allows to obtain high compression factors
  - Examples are the MPEG compression family (MPEG1, MPEG2)
  - Example of lossy coding is DPCM
    - DPCM compares adjacent pixels and stores only their difference

MPEG

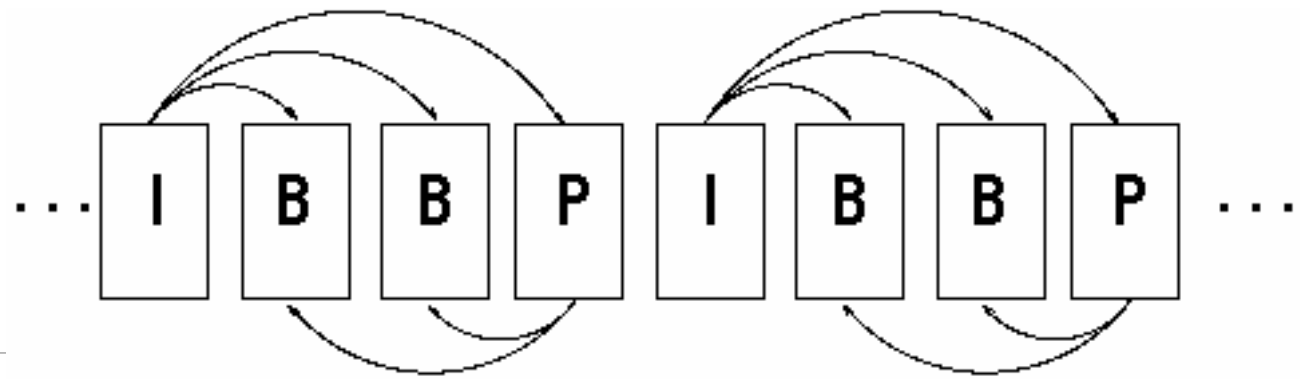
- MPEG (Moving Pictures Experts Groups)
  - MPEG1 has a bit-rate up to 1.5Mb/sec
    - Designed for storage and retrieval of VHS quality video on CD-ROM
  - MPEG2 Designed for broadcast video quality
    - Bit rate: 2Mbps or higher
    - Used for DVD, cable TV, etc.
  - MPEG4 is object-based, multi stream
    - Variable bit-rates, from <64 kbps, up to 4Mbps and more (in the future)

MPEG-1 [1/2]

- Compression based on intra-frame and inter-frame encoding
- Intra-frame coding
  - Each frame is subject to compression
  - Uses DCT compression schema
- Inter-frame coding
  - Exploits temporal redundancy
  - Predictive coding
    - current picture is modeled as a transformation of picture at some previous time
  - Interpolative coding
    - Uses past and future pictures for reference

MPEG-1 [2/2]

- MPEG uses three types of frame coding
  - I frames: intra-frame coding
    - Moderate compression
    - Access points for random access
  - P frames: predictive-coded frames
    - Coded with reference to I or P frames
  - B frames: bi-directionally predictive coded
    - Coded using previous/next I and P frames
    - High compression



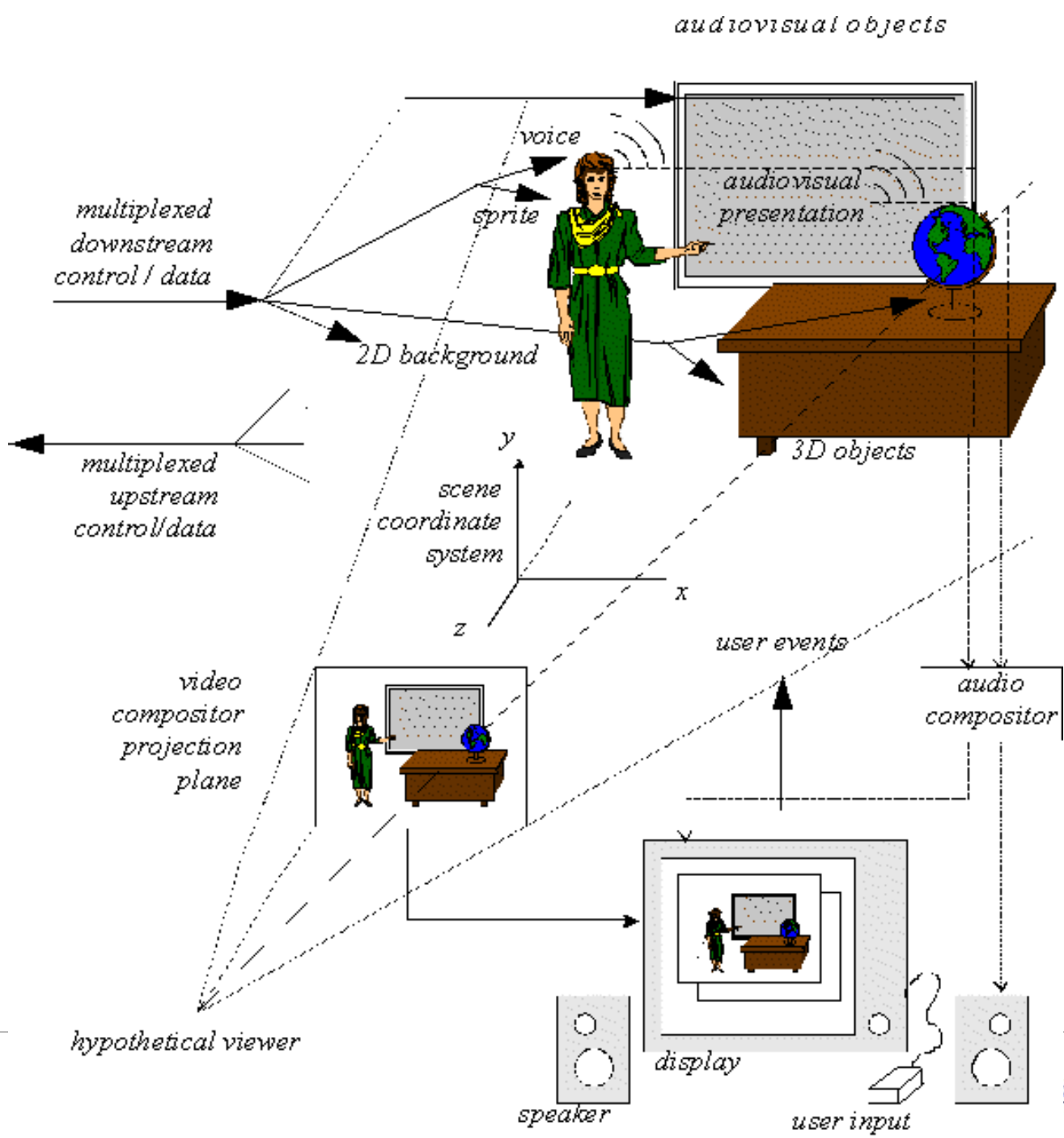
MPEG-4 [1/4]

- Scalability of bit rate vs quality
- Better Audio/Video compression than MPEG-1
- Content based coding
- Support for efficient streaming

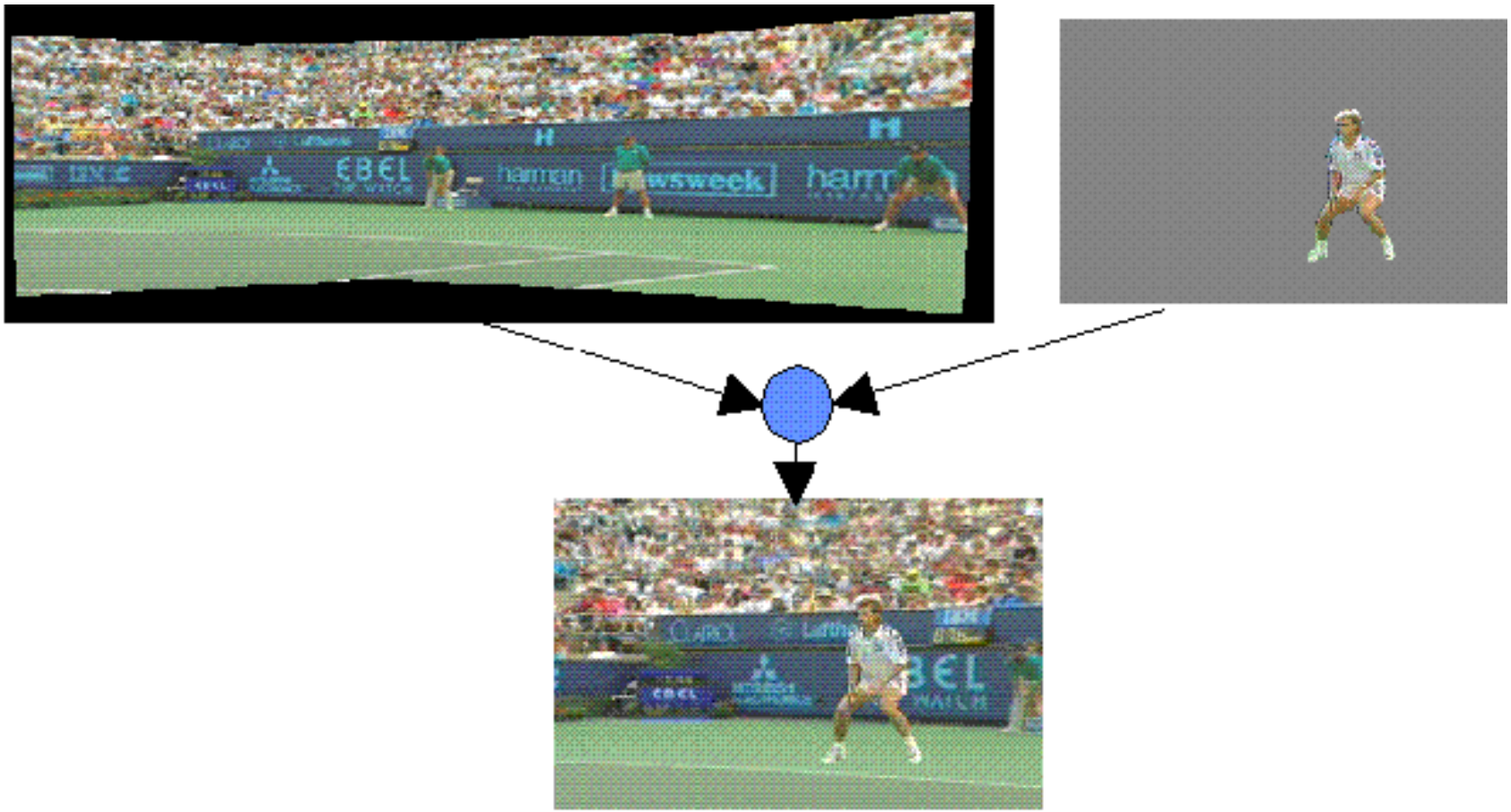
MPEG-4 [2/4]

- Content based coding
  - Reusability of object coding
  - Adaptation (different coding for different objects)
  - High quality for interesting parts
  - Possibility of scene composition
    - Integration of natural and synthetic content
    - Tele-presence

MPEG-4 [3/4]


audiovisual objects
voice
sprite
audiovisual presentation
multiplexed downstream control / data
2D background
3D objects
multiplexed upstream control/data
y
scene coordinate system
x
z
user events
video compositor projection plane
audio compositor
hypothetical viewer
display
speaker
user input

MPEG-4 [4/4]



Digital Video representation

- Video is composed of a sequence of **frames**
- Video is separated into **shots**
  - A shot is a sequence of frames separated by a transition
  - Transitions between shots are given by
    - Camera break
    - Dissolve
    - Wipe
    - Fade-in, fade-out
- A video can be separated into **sequences**, that are semantically meaningful groups of shots, possibly non consecutive

The main operations of an A/V Digital Library

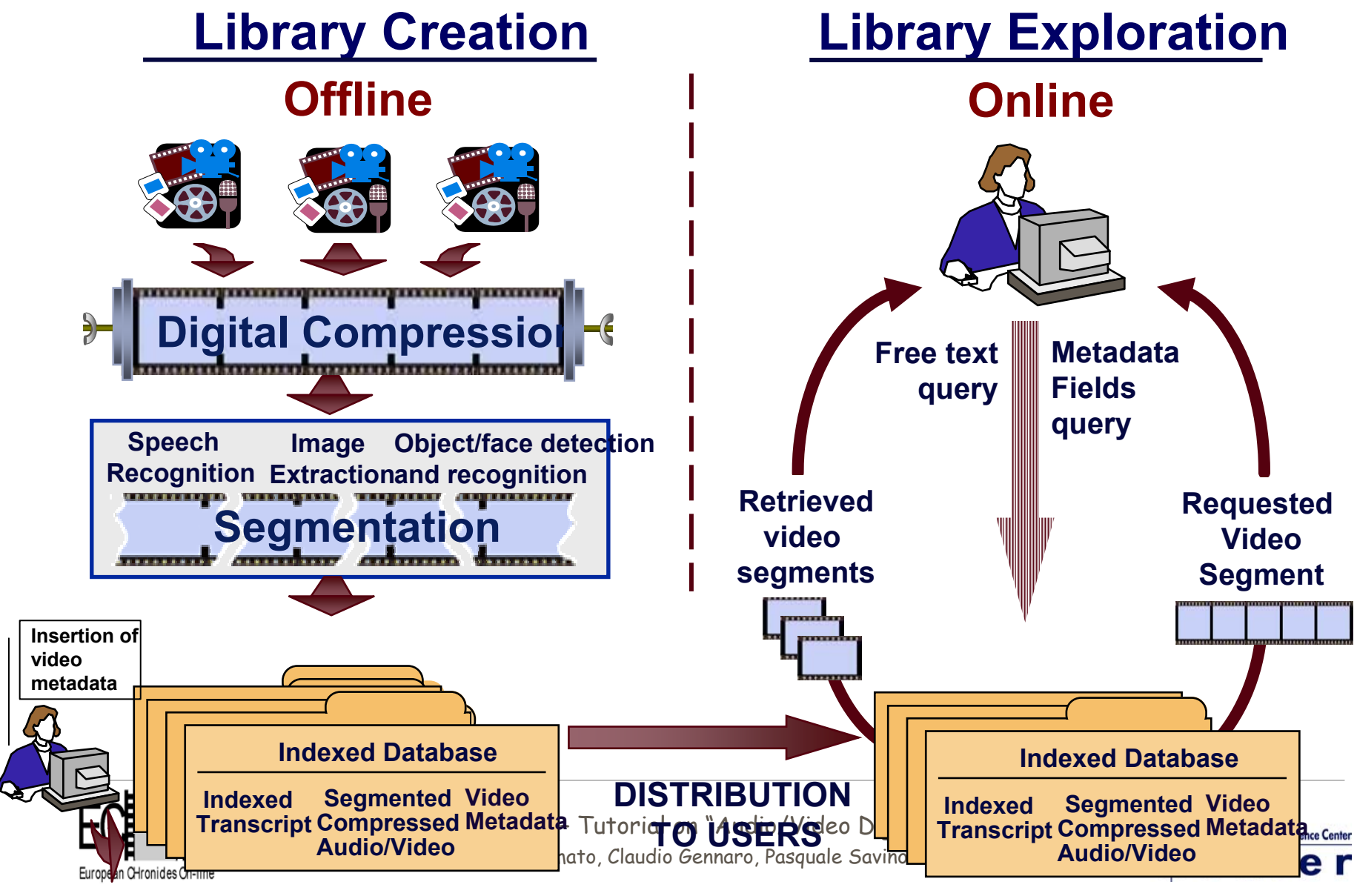
European CHronicles On-line

Digital Library Competence Center

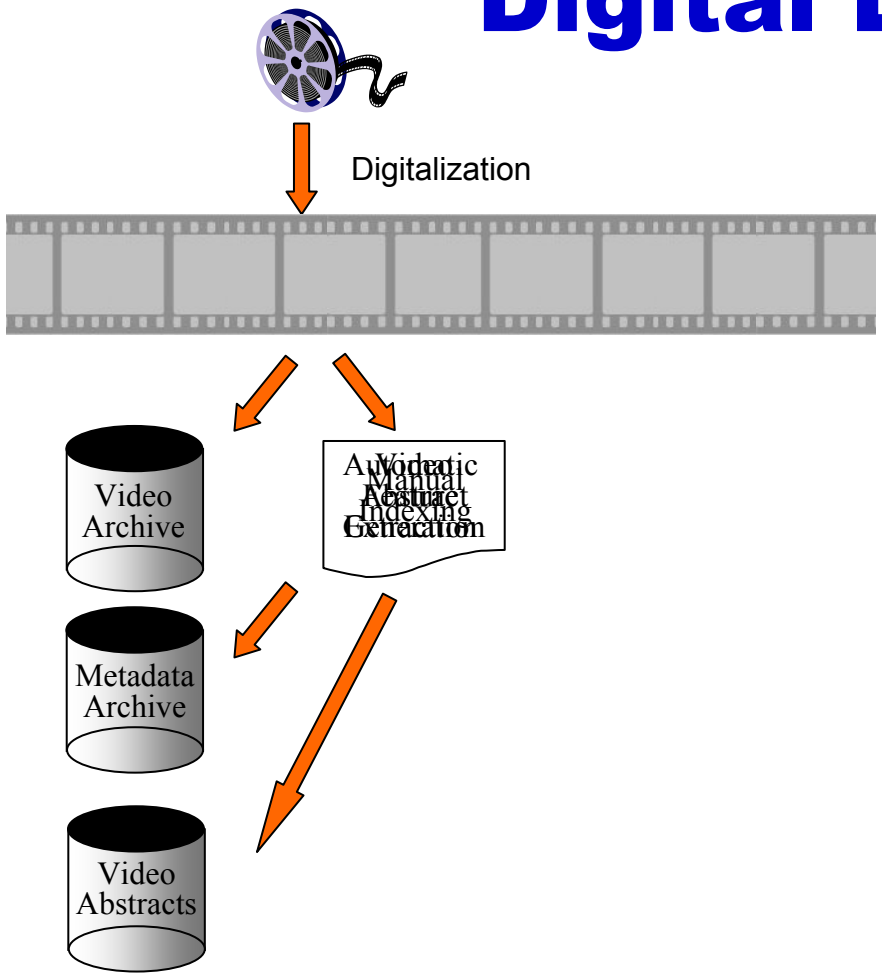
Operations of an A/V Digital Library

- Video archiving and indexing
- Video storage
- Content-based search
- Video access (visualization and copy)

Summary of all phases & operations



Data flow of the Digital Library creation



Digital video

Archival of raw video

Archival of automatically extracted features

Archival of metadata inserted manually

Archival of Video Abstracts

Outline [Part 3]

- What is a Digital Library?
- Characteristics of an Audio/Video DL
- Applications of Audio/Video DLs
- Types of data managed
- The characteristics of digital Audio and Video
- The main functions
- Automatic and manual indexing
- Retrieval functionality
- Logical architecture of a video DL
- User's categories
- Overview of existing systems

Automatic and manual indexing of Audio/Video documents

What is the purpose of video indexing

- The indexing process provides a “description” of video content that can be used to support the retrieval process
- Three main categories of video descriptions
  - Keywords describing the entire video
  - Visual properties
  - Semantic information

Automatic vs manual indexing

- The goal is to provide a completely automatic indexing
  - Fast
  - Reliable (user independent, error reduction)
- In many cases this is not possible
  - Complexity of the task (e.g. semantic interpretation of a shot content)
  - Information is not available in the video (e.g. creation date, place where the movie was recorded)

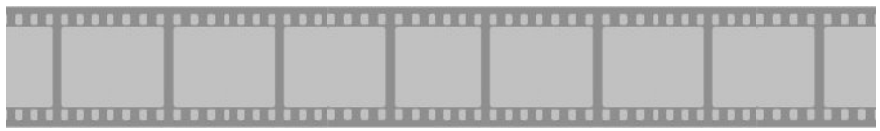
Information that cannot be extracted automatically

- Background information, e.g.
  - Creation date
  - Author
  - Names of the actors
  - Ecc.
- Semantic information
  - Relations among different shots
  - Interpretation of the meaning of a shot
  - Interpretation of the meaning of a frame
- All this type of information must be provided manually, possibly by using a specific tool

Information that can be extracted automatically

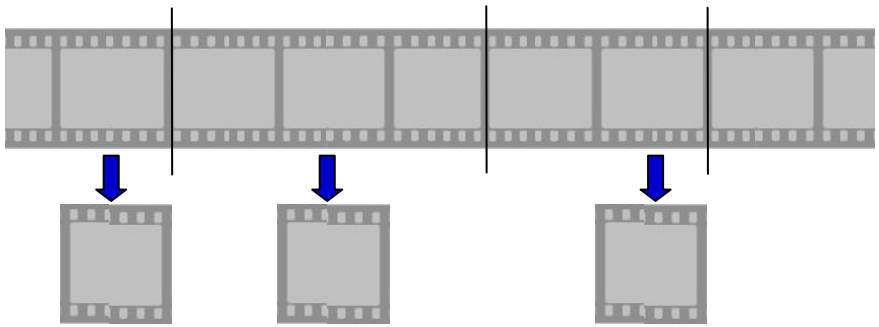
- Features that can be extracted from the entire video,
  - e.g. frame rate, resolution, b&w or color video, etc.
- Features that are associated to the audio part
  - e.g. the transcript of the speech
- Features that can be extracted from each shot
  - e.g. object track, camera movement, recognition of specific objects, recognition of faces, text captions, key frames
- Features that can be extracted from each frame
  - these are typical image features, such as color distribution, texture, object's shapes, etc.

Automatic feature extraction



Digital Video

Segmentation



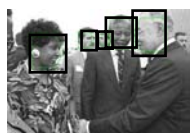
Key frame extraction

Speech extraction and recognition

Xyz xxx abc vvmm skks ….. ….. …

Object recognition

Face detection and recognition



Video segmentation [1/2]

- Segmentation is needed in order to identify the index units for video content. These units are *generic clips* which correspond to individual camera shots.
- A *generic clip*, which is the basic indexing unit, is defined as a single uninterrupted camera shot.
- Video partitioning consists in detecting boundaries between consecutive camera shots.
- The type of transitions between shots are
  - camera break (the simplest to be detected), dissolve, wipe, fade-in, fade-out

Video segmentation [2/2]

- Detection of camera break
  - **Pair-wise Pixel comparison** Given two consecutive frames, corresponding pixels are compared and the number of pixels changed is determined. For monochromatic images, a pixel is judged as changed if the difference between its intensity values in the two frames exceeds a given threshold T.
  - **Histogram comparison** This method uses a comparison of some feature of the images. For example it may use the histogram of intensity levels. The principle behind this approach is that two frames having an unchanging background will show little difference in their respective histograms. This method is less sensitive to object motion because it ignores the spatial changes in a frame.
  - **Motion continuity** Motion can be represented quantitatively by assigning a field of motion vectors to the pixels of an image. The motion vectors are computed by dividing each frame into blocks and determining where each block is located in the successive frame. A correlation between the two frames can be computed; a low correlation between two consecutive frames is interpreted as a camera break.

Image indexing

- Image indexing is performed on key frames
- Image indexing is difficult, since the concept of image similarity is not precise
- Two different indexing approaches
  - Based on image visual features
    - Color, texture, object's shape, etc
  - Based on a semantic information

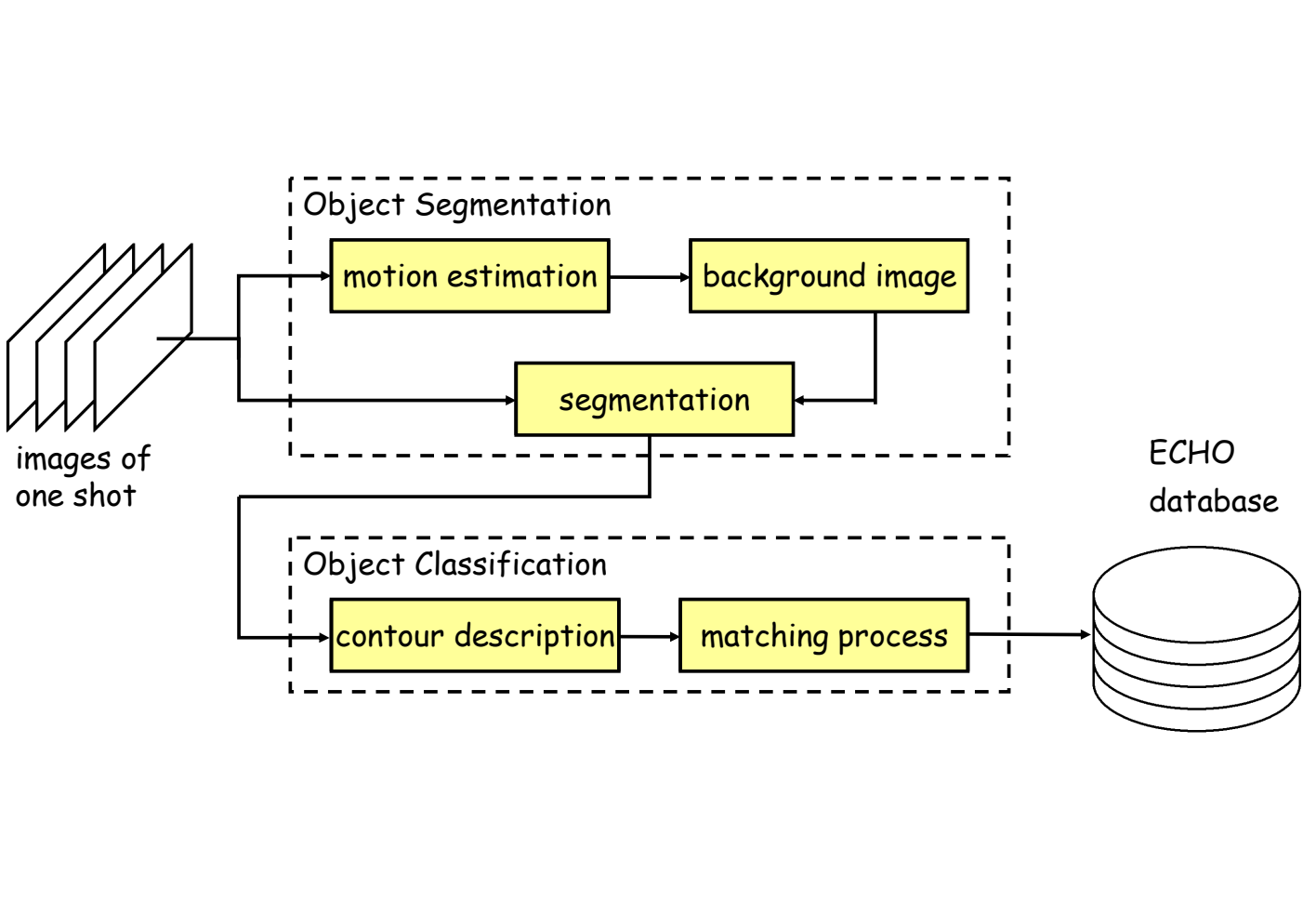
Speech recognition

- The purpose of speech recognition is the generation of a transcript to be used as a support for retrieval
- Main functionality required
  - Speaker independent
  - Multiple languages
  - Operating also with low quality audio
- Does not require perfect recognition
  - Retrieval quality is acceptable for W.E.R. up to 30-40%

Object detection and recognition

- The system for moving-object recognition consists of two components, a *segmentation* module and a *classification* module.
- For each shot in the video, a background panorama image is constructed. The foreground objects in this background image are removed by means of temporal filtering (median).
- The object is segmented by comparing each frame of the video to the background image.

Phases of Object detection and recognition



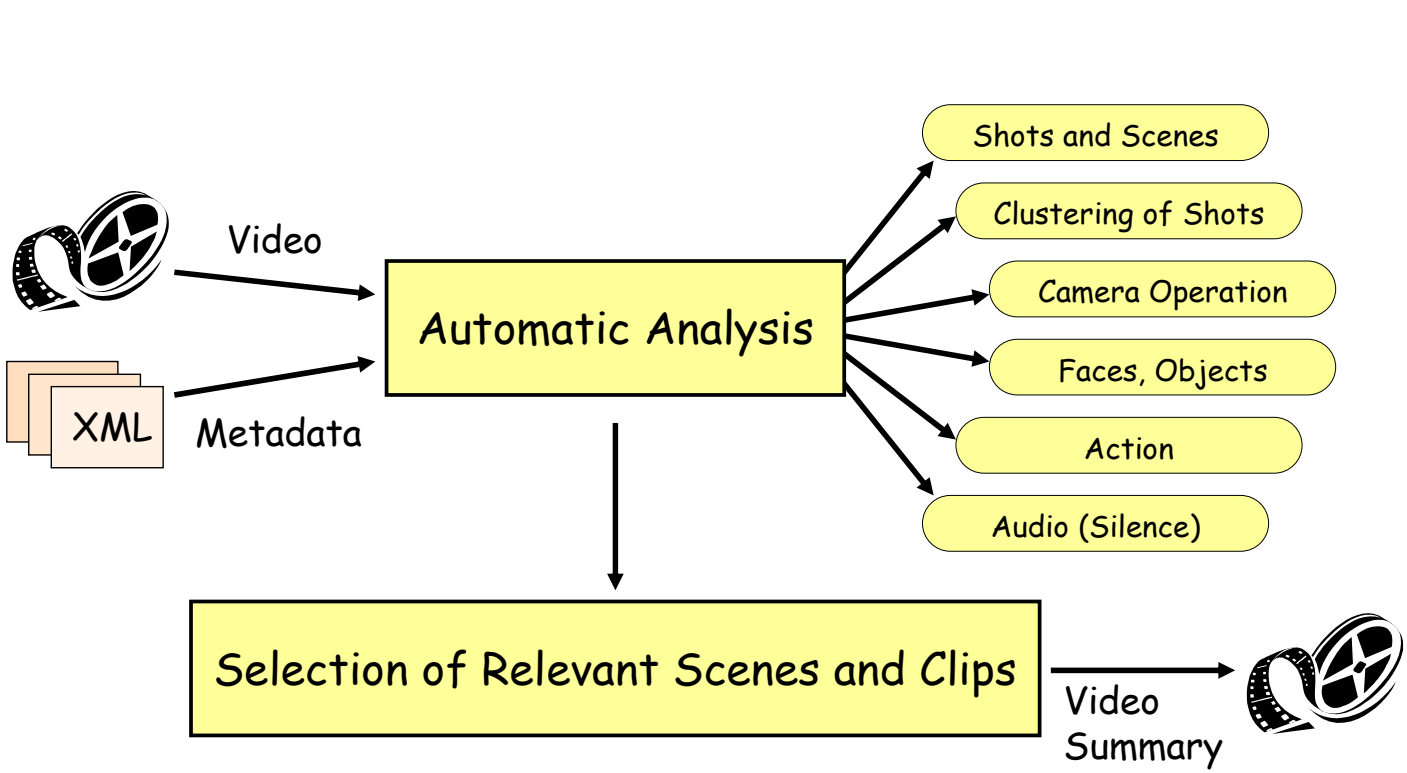
Example: Cars



Video abstract generation [1/2]

- A video abstract is a part of a much longer video, which preserves the essential message of the original video.
- A video abstract does not change the presentation medium.
- The users get a quick overview of a much longer video.
- The video abstracting application will:
  - Select relevant clips
  - Order these clips
  - Define a transition between two clips
  - Modify the audio track

Video abstract generation [2/2]



Representation of video content

- Metadata are used to represent the video content in order
  - To support video retrieval and navigation, as well as video management and processing
- Simple attribute values can be used as metadata to represent the video content (e.g. Dublin Core)
- or complex representations can be used to describe content information extracted from videos (e.g. MPEG-7)
- Metadata can also provide a description of video structure

Retrieval functionality

European CHronicles On-line

Digital Library Competence Center

Retrieval functionality

- Retrieval is based on queries expressed on metadata values.
- Both automatically extracted metadata, as well as metadata associated manually to the video can be used.
- The user may not distinguish between these metadata types; system behavior may be different

Type of queries

- Queries can be expressed on
  - Metadata associated to the entire video
    - E.g. ***find b&w videos produced before II world war by Istituto Luce***
  - Metadata associated to video shots
    - E.g. ***find a shot where the audio transcript contains the words "Attentato Banca Nazionale dell'Agricoltura"***
  - Metadata associated to single frames
    - E.g. ***find a video that contains a frame similar to this image [the image is provided as an example]***
  - Any combination of the previous cases

Retrieval characteristics

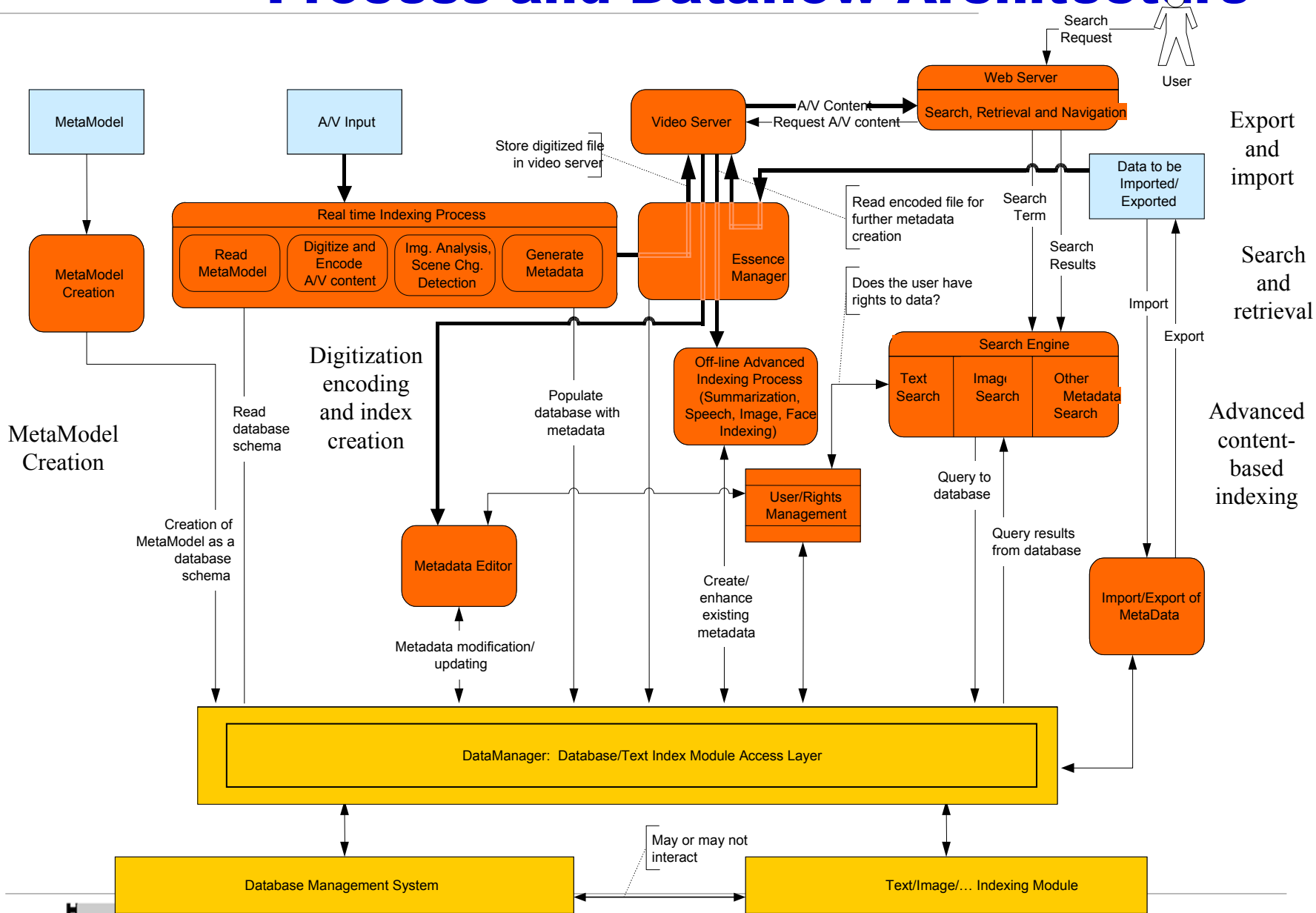
- Retrieval is based on an approximate match between the query and the retrieved videos. This is mainly the case when imprecise query elements are used (e.g. free text, images)
- Retrieved videos are returned to the user in decreasing relevance order, possibly indicating the degree of relevance of the retrieved items.
- Due to the imprecision of the method (i.e. some of the retrieved items are not relevant for the user and some relevant items are not retrieved), it is helpful to have a query refinement and a relevance feedback mechanism.

Logical architecture of a Video Digital Library

Process and Data Flow

- MetaModel creation
- Digitization, encoding and index creation
- Advanced content-based indexing
- User rights management and access control
- Search and retrieval
- Export and import

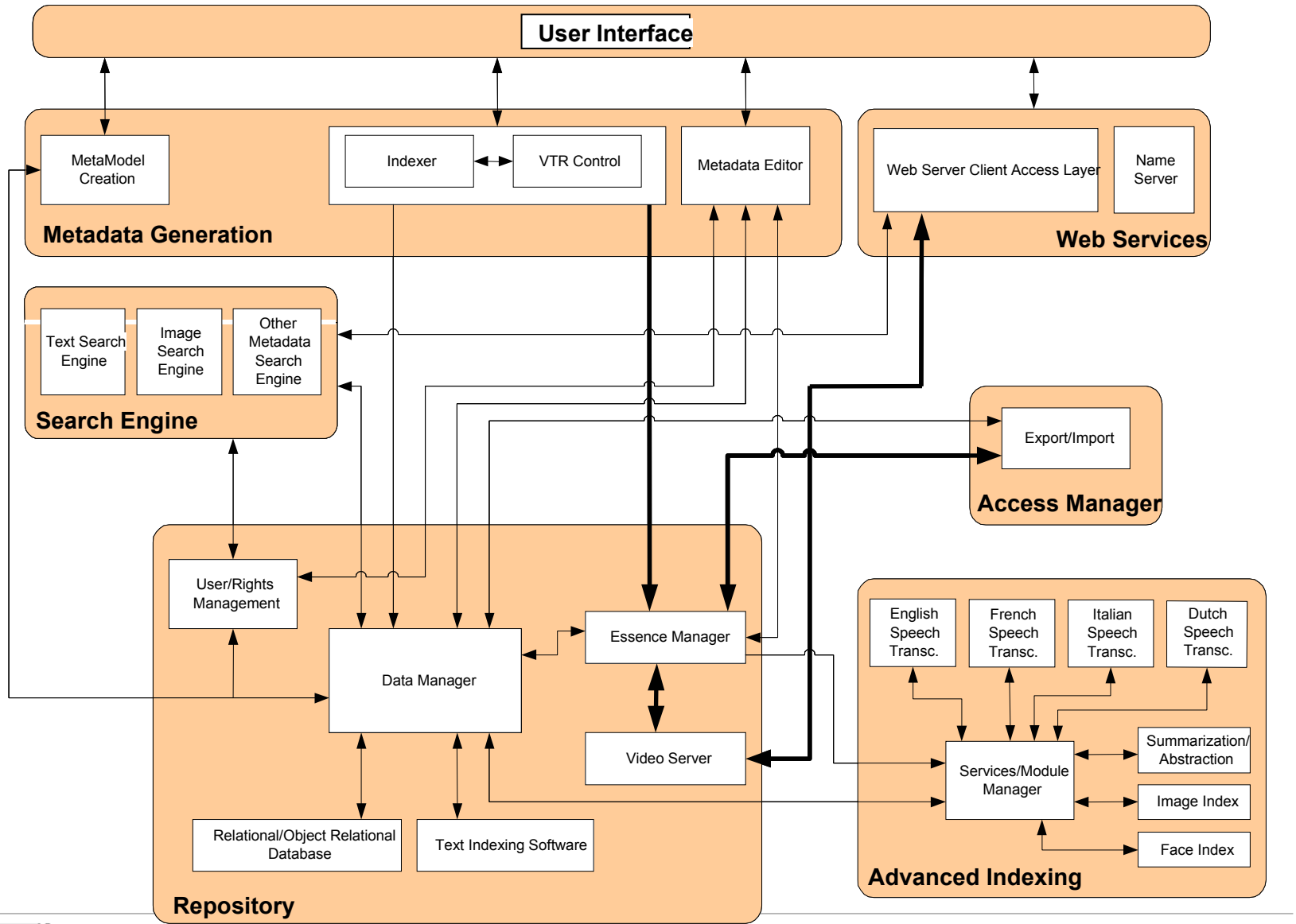
Process and Dataflow Architecture



Functional Decomposition

- User interfaces
- Repository
- Metadata generation and manipulation
- Search engine
- Web-based client access layer
- Access Manager
- Advanced Indexing

System Architecture



Outline [Part 4]

- What is a Digital Library?
- Characteristics of an Audio/Video DL
- Applications of Audio/Video DLs
- Types of data managed
- The characteristics of digital Audio and Video
- The main functions
- Automatic and manual indexing
- Retrieval functionality
- Logical architecture of a video DL
- User's categories
- Overview of existing systems

User's categories

User's categories

Three main user's categories

- Administrator
  - Manages the entire system
- Cataloguer
  - Manages the ingestion of new video material and indexes it
- Information seeker
  - Searches videos in the Digital Library.
  - There are different types of seekers.

Administrator

The Digital Library administrator manages the procedures to

- Control the access to the system
- Manage system security
- Manage backup and recovery
- Manage billing and accounting

Cataloguer

The cataloguer is responsible for all procedures needed to ingest and index new video material.

- Procedures to ingest new videos
- Procedures to associate and revise metadata of existing videos

Depending on the system characteristics, this operation can be automatic or it can heavily require user intervention.

Information seeker

Users searching for information in the Digital Library. Operations may depend on the application.

- Naïve users
  - Typical search is performed on the Web
  - Users access the archive for cultural interest, learning, etc.
- Professional users
  - Production of video documentaries
  - Production of learning material, etc.

Overview of existing systems

Existing Digital Library Systems

- Many DL systems (e.g. Greenstone) manage video as an unstructured data type.
- Indexing is based on metadata associated to the entire video.
- Simple retrieval support is based on metadata associated to the video

Existing DL Systems (cont.)

- More advanced video archiving and retrieval systems (e.g. Virage, Informedia) use part of video content to support retrieval.
- Indexing is mainly automatic
- Other systems (e.g. ECHO) also offer typical DL services combined with powerful indexing and retrieval capabilities.
- Indexing is partly automatic and partly manual
- We will review the functionality of four different DL systems
  - Greenstone
  - Virage
  - Informedia
  - ECHO

Greenstone

- Greenstone is a Digital Library software from the New Zealand Digital Library Project at the University of Waikato (www.mkp.com/DL)
- Greenstone provides services to
  - Build digital library collections
  - Deliver the information to the users

Greenstone

- Types of data managed
  - Information is stored in collections that are composed of documents
  - Documents in a variety of formats is accepted, and converted into a standard XML form for indexing
  - Collections may contain text, pictures, audio, and video.
  - Non textual material is either linked into the textual documents or accompanied by textual descriptions to allow full-text searching and browsing

Greenstone

- Building a digital collection with Greenstone
  - Collections are created by specifying the source material
  - During building, indexes for browsing and searching are constructed
  - The metadata used for indexing are either provided in specific files or automatically extracted by using specific programs
    - Dublin Core is used for defining metadata types
  - Adding new material to an existing collection, requires to rebuild the indexes

Greenstone

- Searching in Greenstone
  - Support of free text search on the entire textual content of the library
  - Support of boolean and ranked queries
  - Search can be restricted to document components (e.g. titles, paragraphs, etc.)
  - Different tools provided to the user to simplify the search, e. g.
    - Query history
    - Stemming
    - Phrase search
  - Presentation of results in decreasing relevance order, with a short abstract of the document

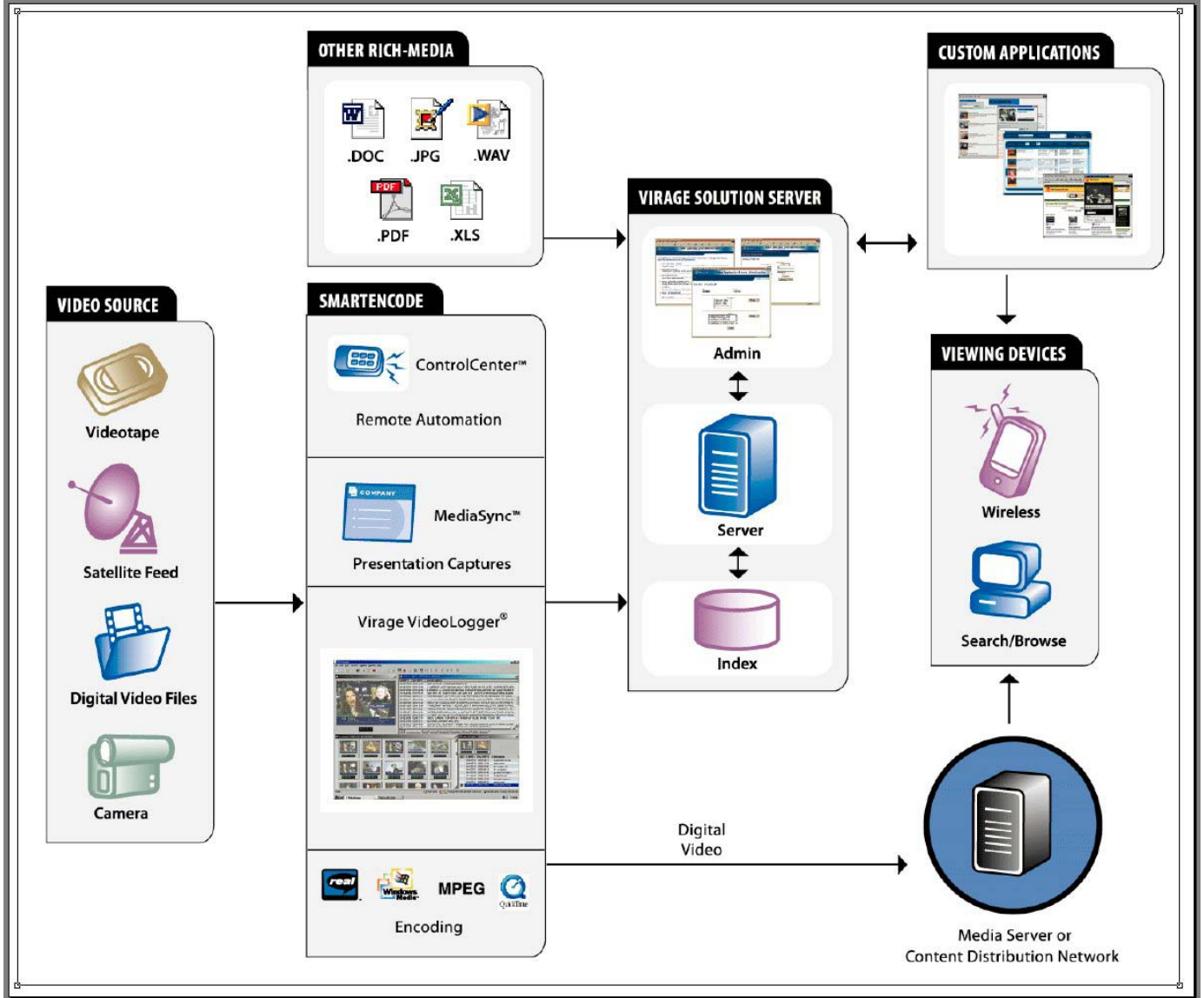
Greenstone

- Browsing in Greenstone
  - Browsing is the unsystematic access to documents in the collection
  - Metadata associated to documents are used to guide the user, by supporting different browsing activities
  - Some examples
    - Browsing alphabetical lists
    - Browsing by date
    - Browsing by subject

Virage

- Virage (http://www.virage.com) is a provider of video and rich media software
- The Virage Rich-Media Application platform enables
  - Management, distribution and dynamic publishing of streming video
- It is composed of two main modules
  - Smart Encode
    - Encoding of analog video
    - Digital video indexing
    - It is composed of the Control Center and the Video Logger
  - Virage solution server
    - Application server that dynamically generates HTML pages based on queries to the video index.

Virage



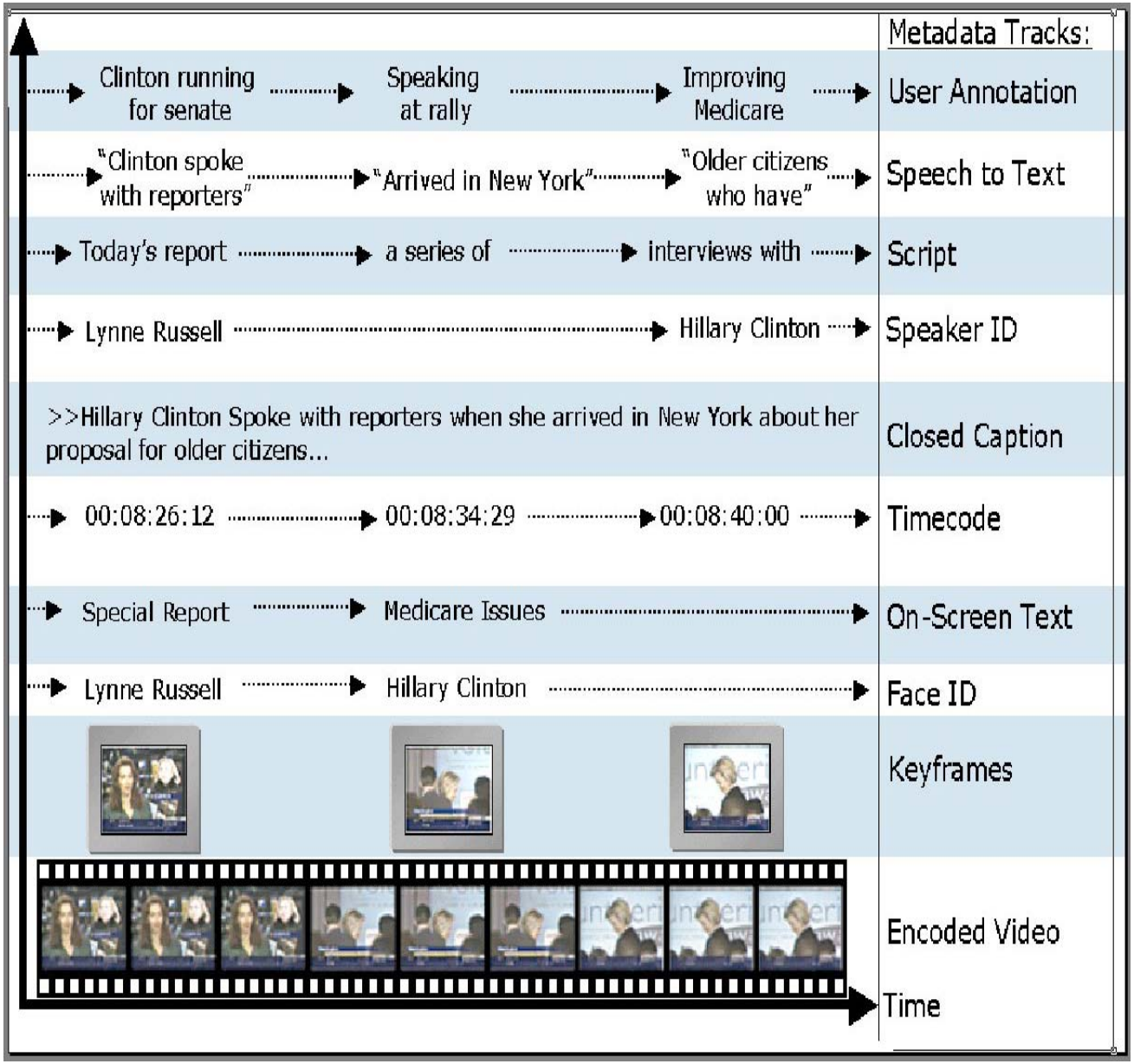
The Virage Application Platform

Virage

- The Control Center
  - This is a workflow application that centralizes the management of the Smart Encode process.
- The Video Logger
  - Encoding into different formats (AVI, Real Video, Windows Media, QuickTime, MPEG)
  - Text extraction of closed caption and teletext
  - Time-based keyframing

Virage

- The Video Logger (cont.)
  - Media Analysis plug-ins
    - Speech recognition
    - Speaker identification
    - Audio Classification
    - Face recognition
    - On-screen text recognition
  - Data export for different DBs
  - Support of user annotations
  - User defined video and clip labels



Virage

Video is indexed by using different tracks, which are time-synchronized

The Virage metadata architecture is extensible, through the use of the VideoLogger SDK

Virage

- Virage Solution Server
  - Web-based sw platform to **search, share, manage,** and **store** video and other multimedia objects
- Main features
  - Content sharing and elaboration
  - Administration
  - User management and authentication
  - Advanced search
  - Support of different data types

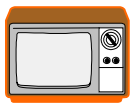
Informedia

- Informedia (http://www.informedia.cs.cmu.edu/) is a research effort coordinated by Carnegie Mellon University.
- The project aims to achieve machine understanding of video and film media, including all aspects of search, retrieval, visualization and summarization in both contemporaneous and archival content collections.

Informedia System Overview

Library Creation

Offline


Video

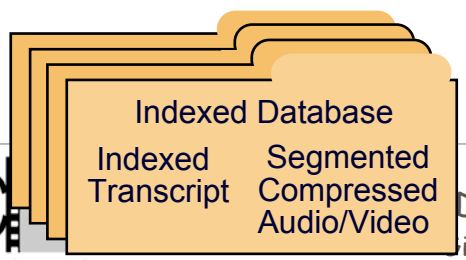

Audio


Text

Digital Compression

Speech Recognition | Image Extraction | Natural Language Interpretation

Segmentation



DISTRIBUTION TO USERS

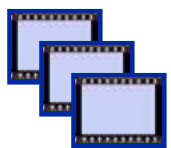
Library Exploration

Online

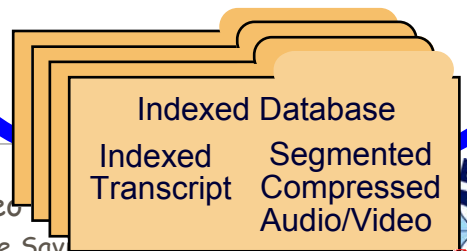
Spoken Natural Language Query

Semantic-Expansion

Story Choices



Requested Segment



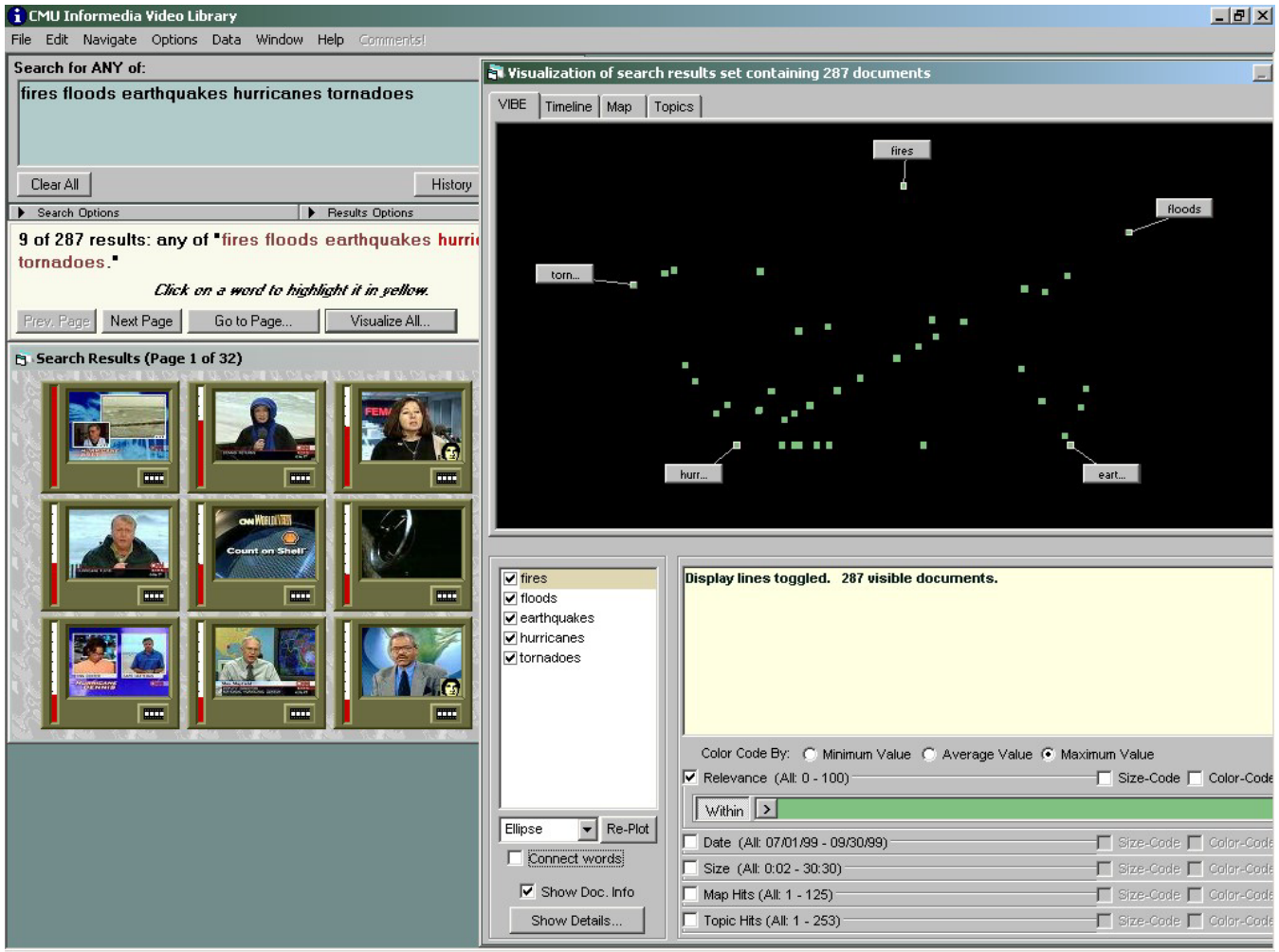
Informedia

- Automatic indexing through the integration of diverse technologies
  - Speech understanding for automatically derived transcripts
  - Face, text and object recognition
  - Key frame extraction and indexing
  - Geocoding
  - Topic assignement
- Automatic abstract generation

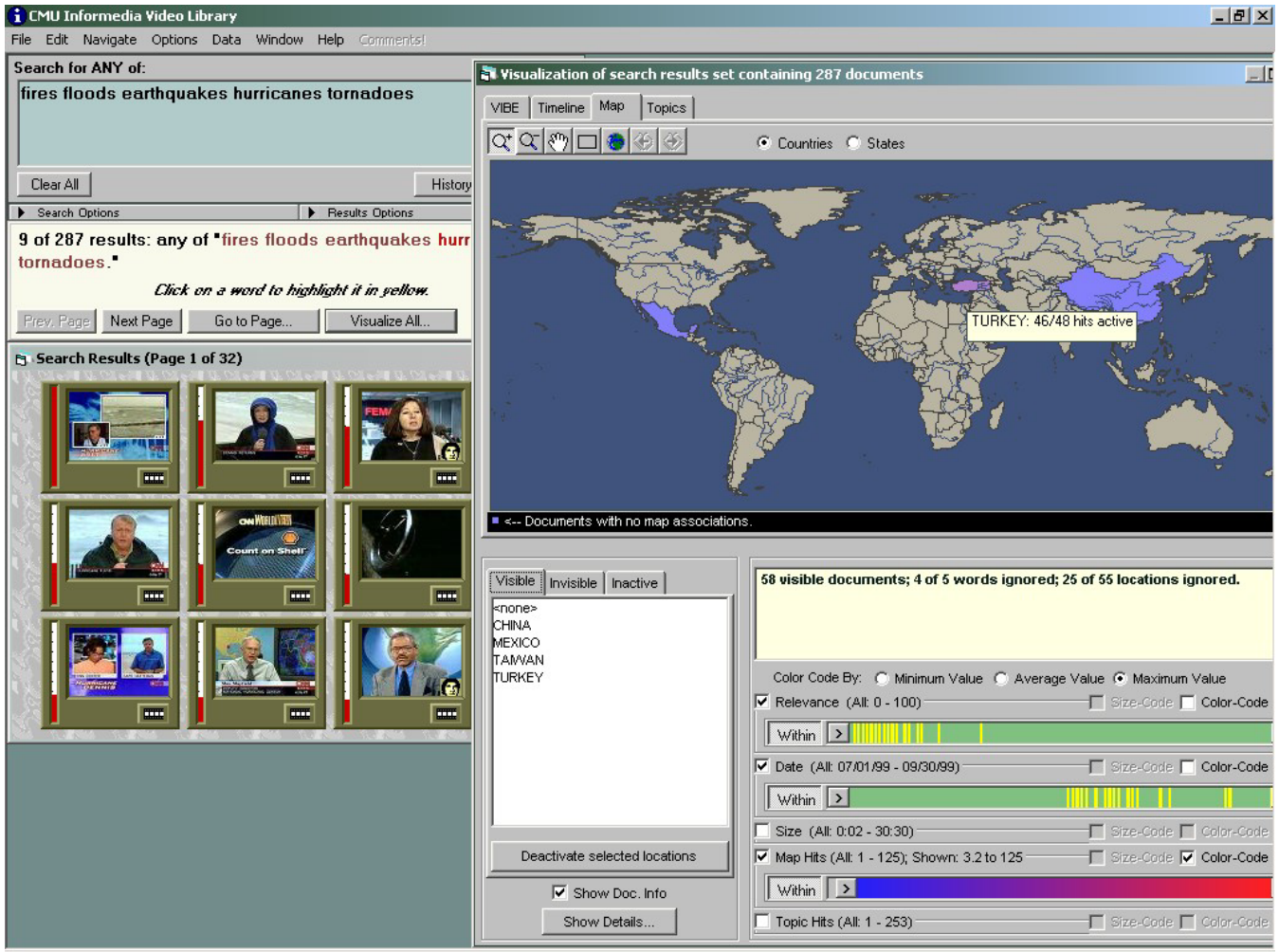
Informedia

- Retrieval based on
  - Free text
  - Image similarity
  - Face and object similarity
- Multiple presentation styles of query results
- Use of geographical information

Informedia – an example



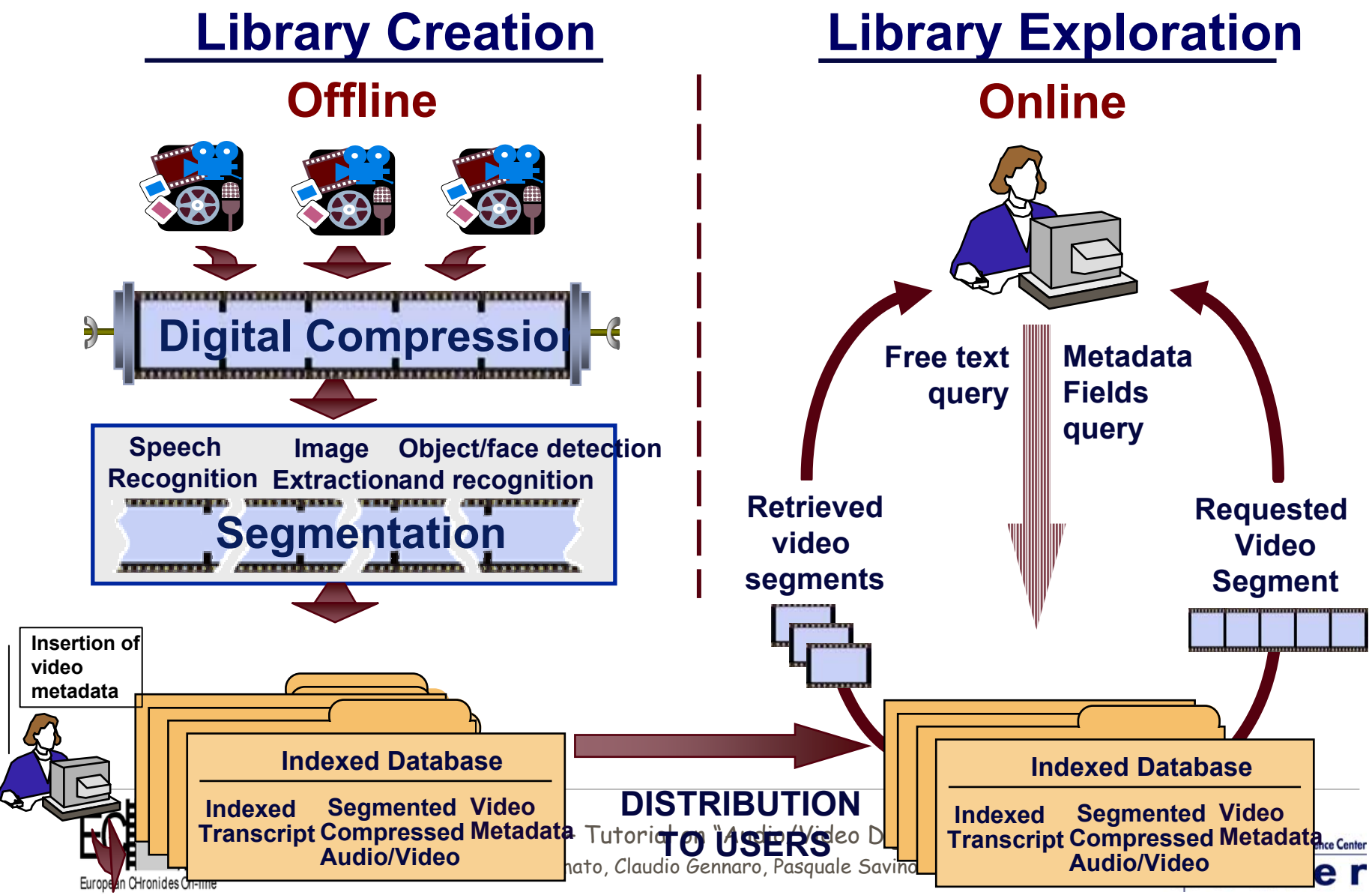
Informedia – an example



ECHO (European Chronicles On line)

- ECHO (http://pc-erato2.iei.pi.cnr.it/echo/) is a Project funded by the European Union under the 5th FP
- The project started in February 2000 and was completed in March 2003
- ECHO aimed at building a Digital Library system for old documentary films, at creating an experimental DL, and at experimenting its use in real application settings
- Partners of ECHO were research institutions, software developing companies, content providers

ECHO System Overview



Key system functionality

- Main system functionality
  - Software infrastructure for audio visual digital libraries
  - Powerful metadata model for Audio Video documentaries
  - Web-based access to large collections in multiple languages
  - Automatic speech recognition (for multiple languages) of old documentary material
  - Simple Cross-language retrieval based on a multi-lingual thesaurus
- Advanced system functionality
  - Cross-language retrieval on audio transcripts
  - Object detection and recognition
  - Face detection and recognition
  - Similarity retrieval of key frames
  - Automatic creation of video summaries

Main characteristics of the ECHO metadata model

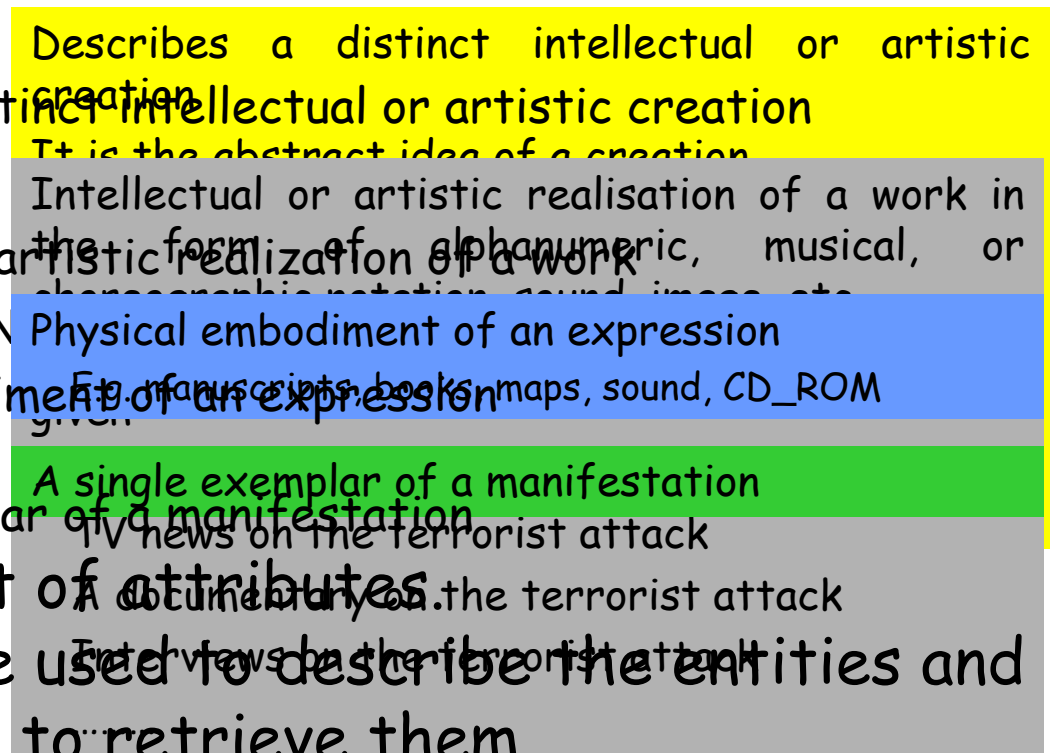
- Supports a multi-layer and hierarchical description of audio-video documents
  - Description of different aspects of the same document
- The model can be adapted to specific application needs
- Describes metadata that are automatically extracted as well as metadata manually extracted
- Multi-lingual support

Metadata model

Extends the IFLA-FRBR model

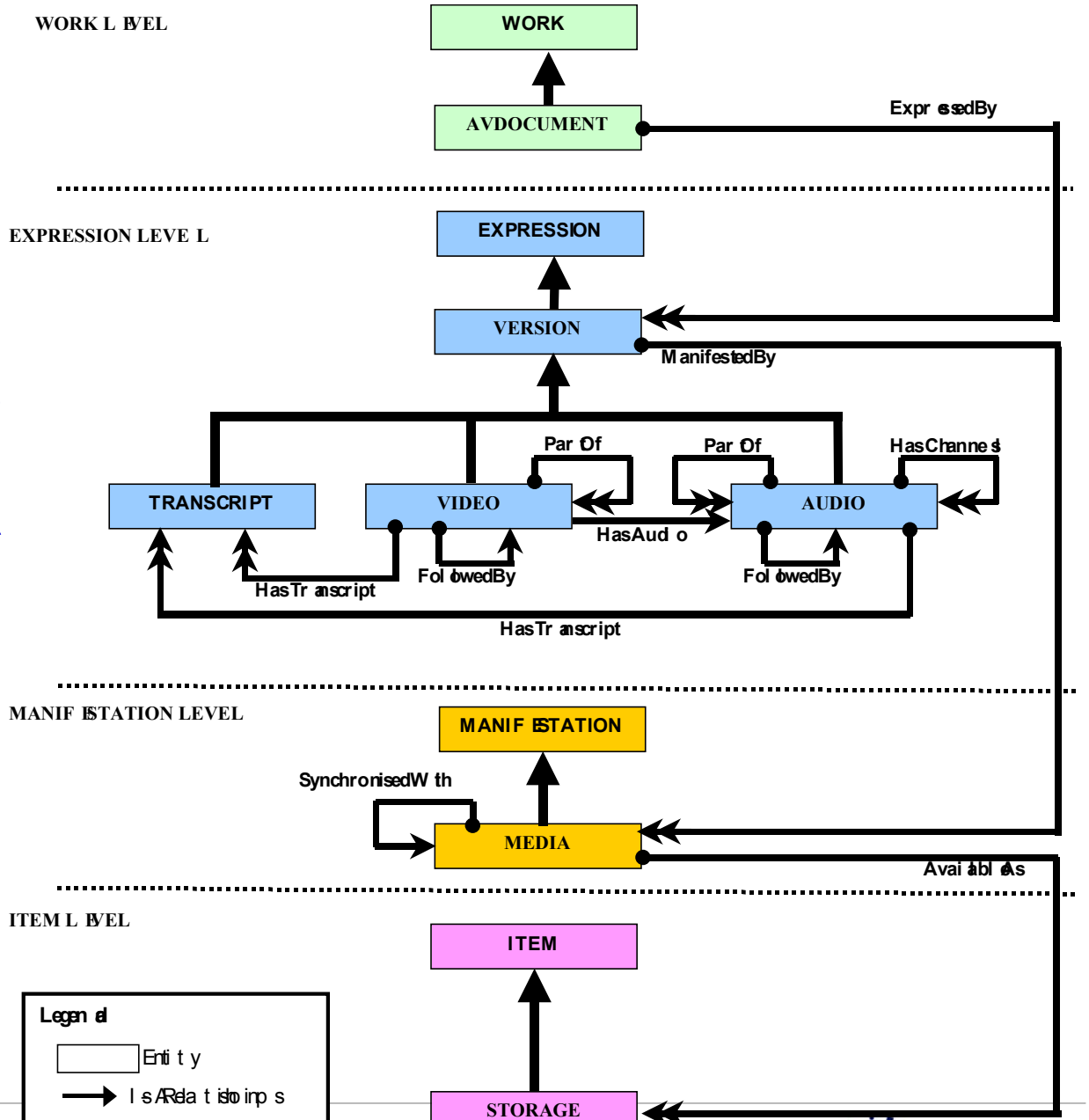
Four entities used to describe different aspect of a resource:

- WORK
  Describes a distinct intellectual or artistic creation
- EXPRESSION
  Intellectual or artistic realization of a work
- MANIFESTATION
  Physical embodiment of an expression
- ITEM
  A single exemplar of a manifestation

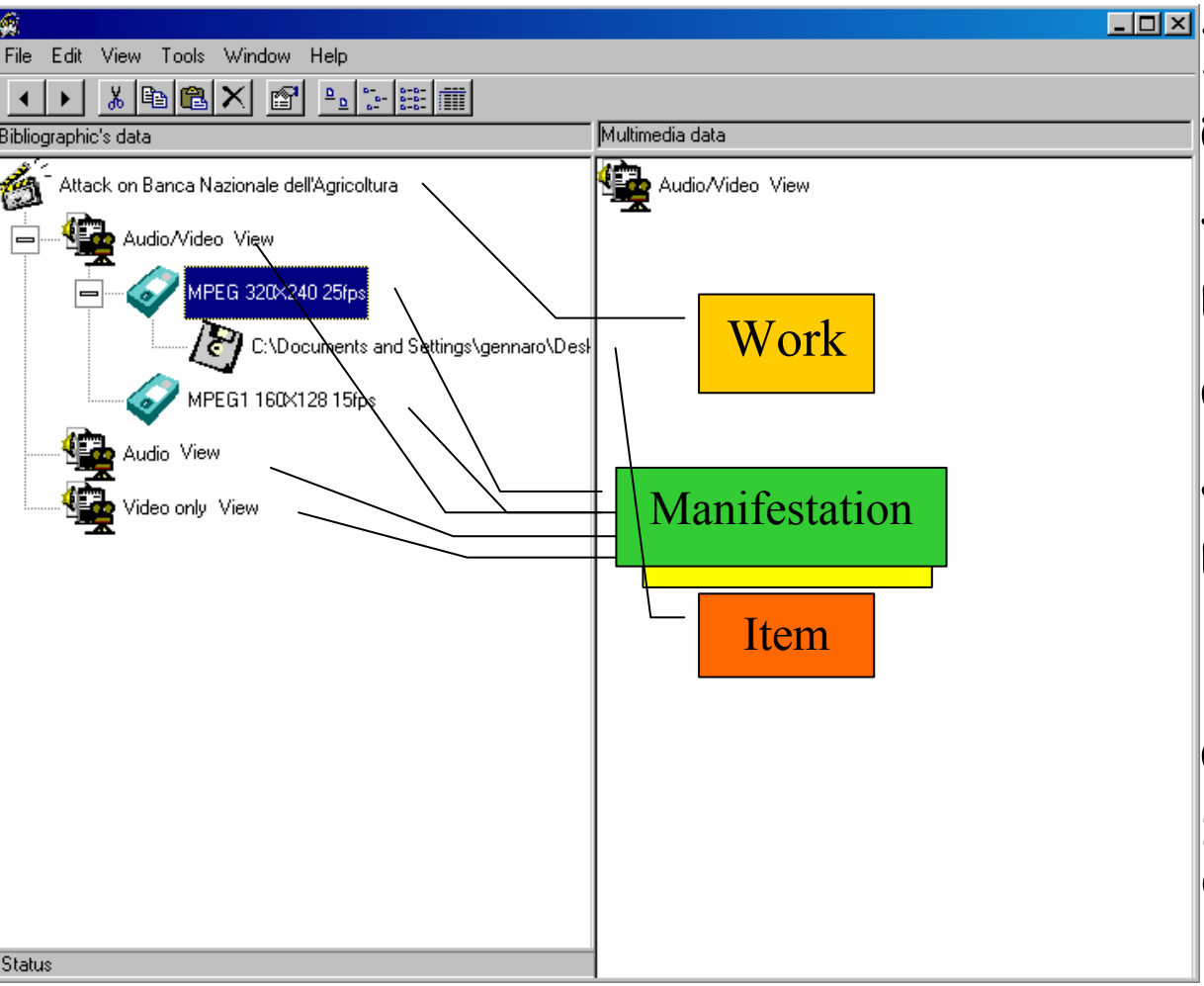


Each entity has a set of attributes.
These attributes are used to describe the entities and to formulate queries to retrieve them.

The ECHO metadata model



Metadata editor

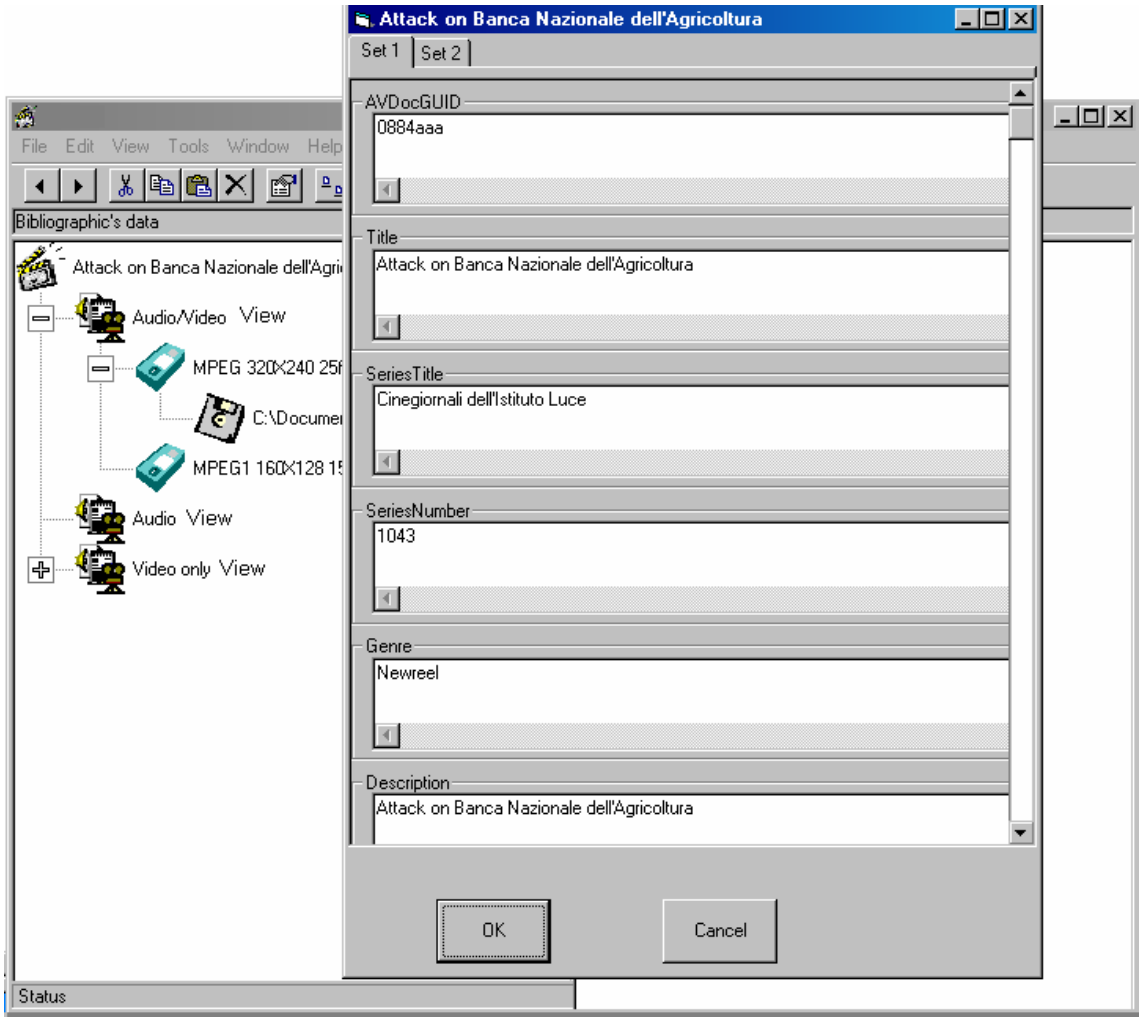


The editor supports the browsing of the structure of the Bibliographic metadata

Here, it shows the EXPRESSIONS that compose a WORK

Similarly, we can list all ITEMS of each MANIFESTATION of each EXPRESSION

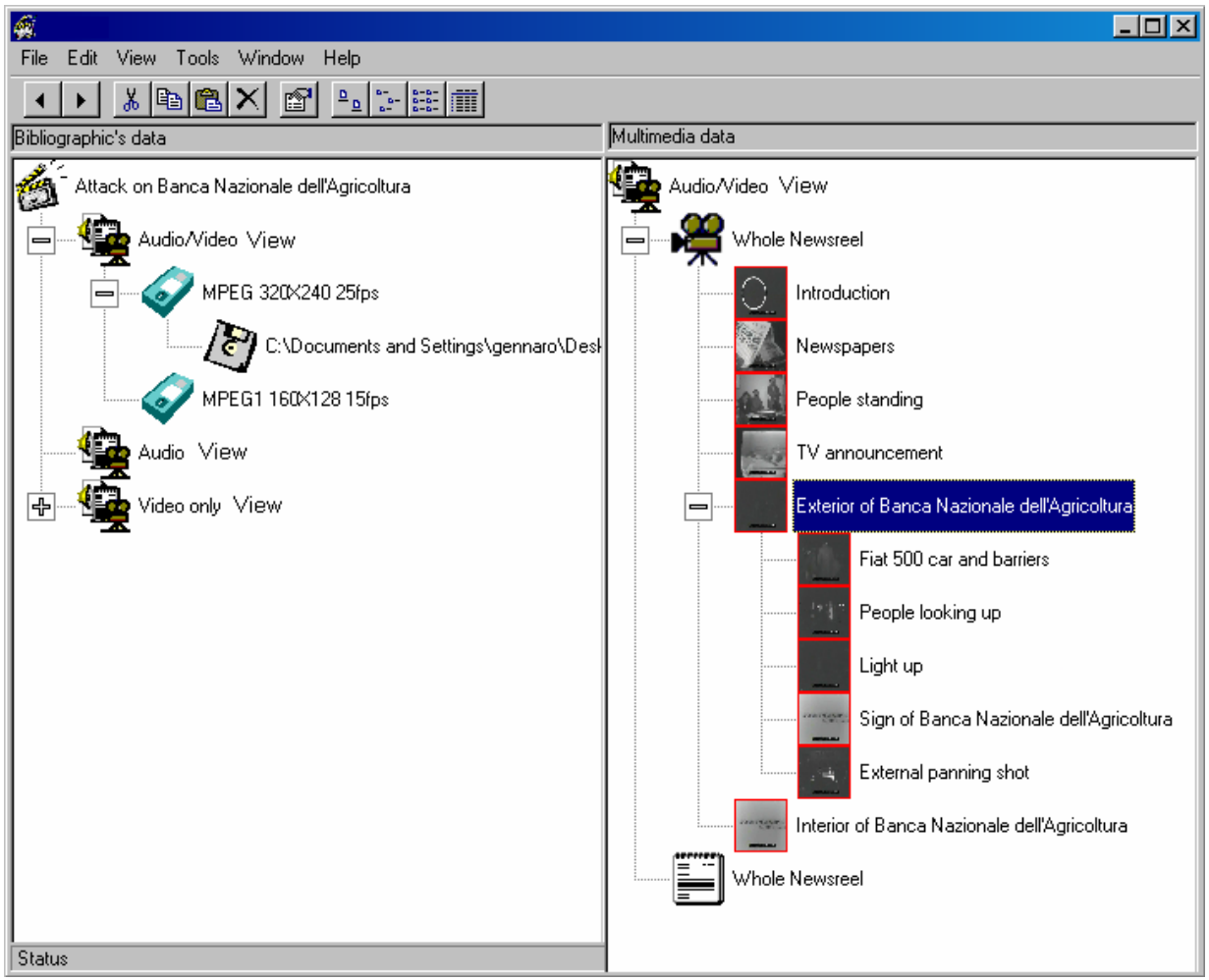
Metadata editor



At any time we can modify the attributes of each element.

Here, we are modifying the WORK attributes

Metadata editor

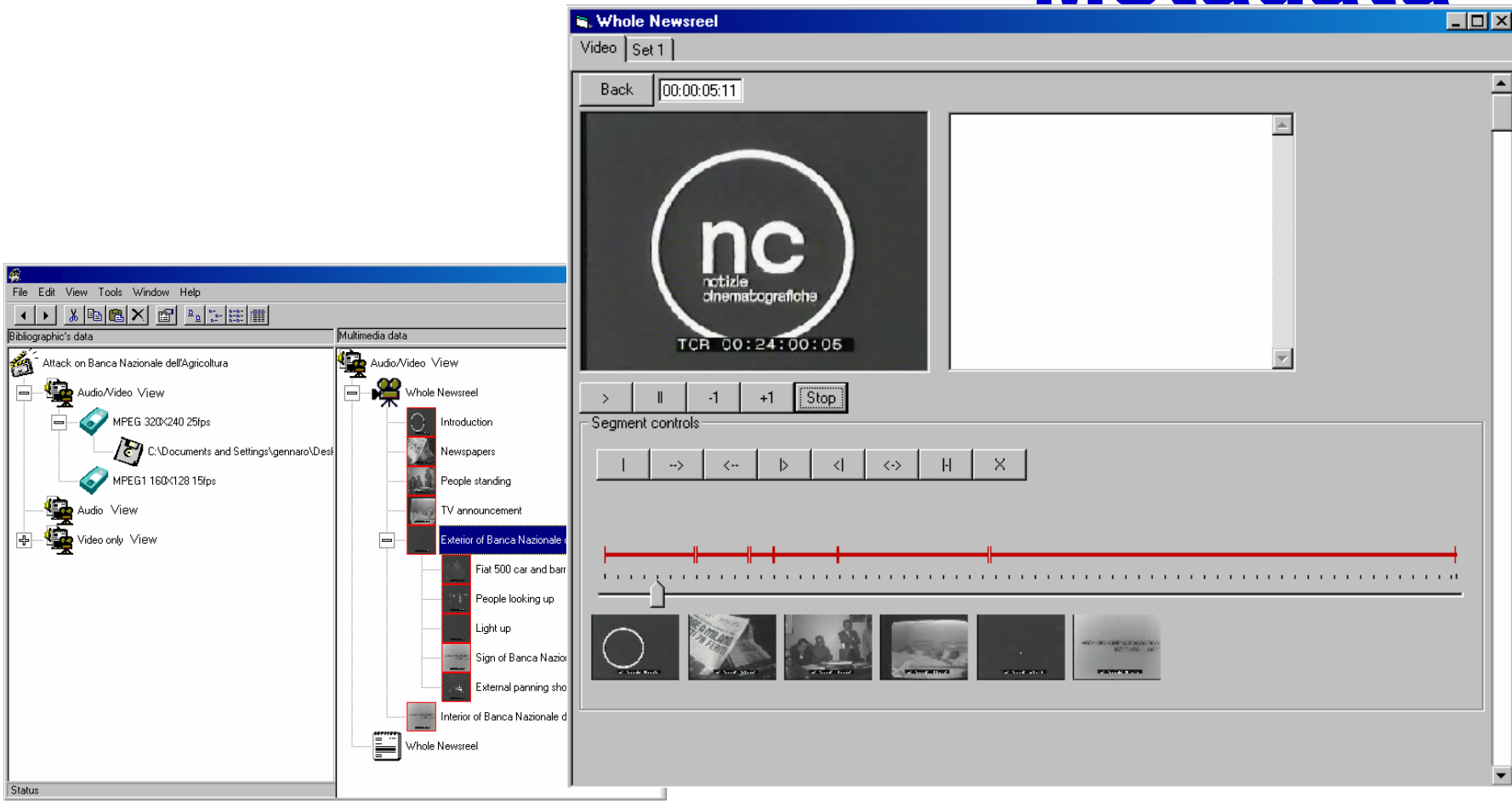


Given an EXPRESSION we can modify its bibliorafic metadata as well as its structure.

An expression can be composed of SUB-EXPRESSIONS

at different levels of detail

Metadata

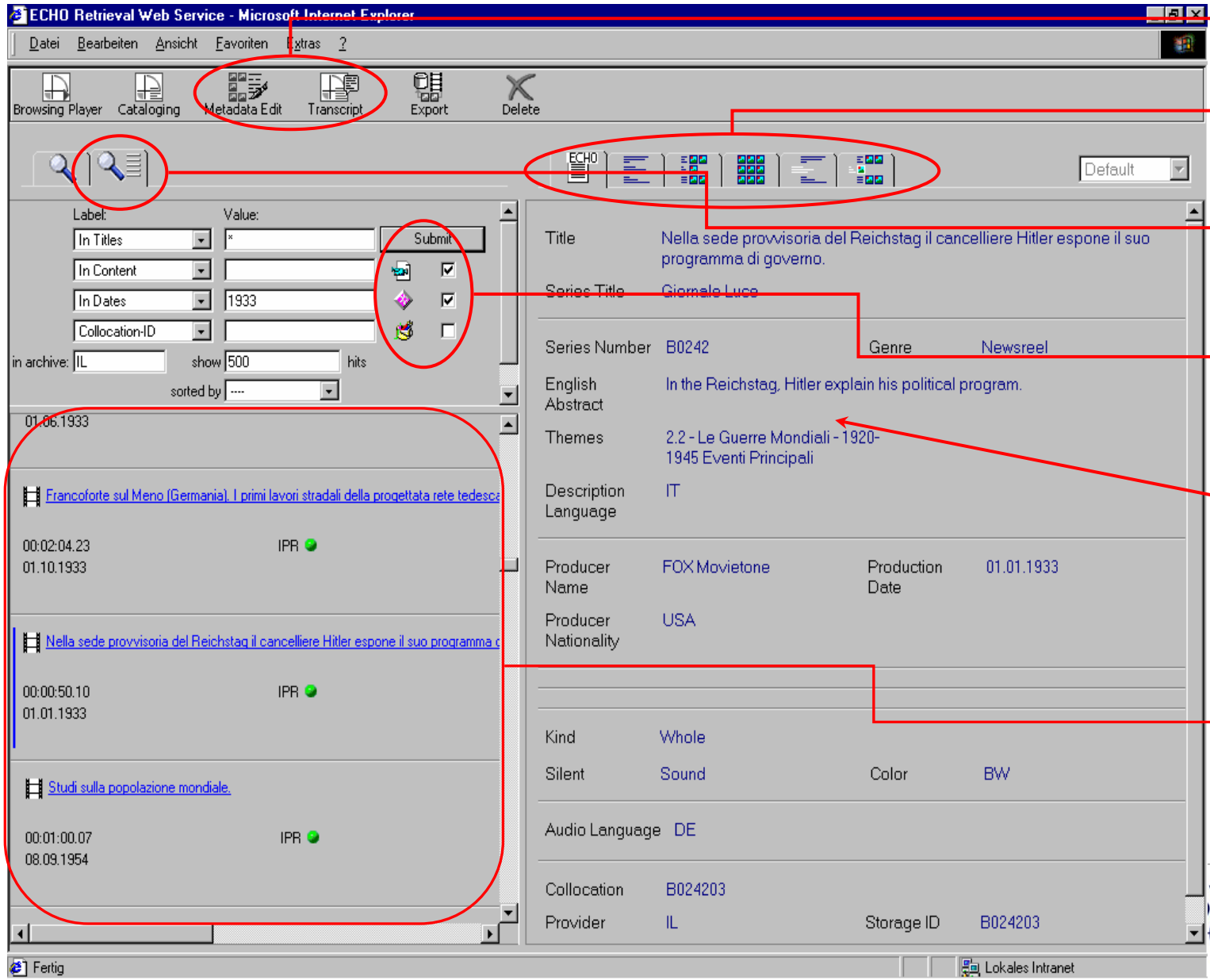


Here, we are modifying the EXPRESSION components

Retrieval functionality

- Free text to search in the transcript of the soundtrack
- Sample images for similarity searching on the film frames
- Keywords to search in the metadata
- Support of cross language retrieval

Retrieval Interface



Integration of Clients and Services in GUI

Views on the material

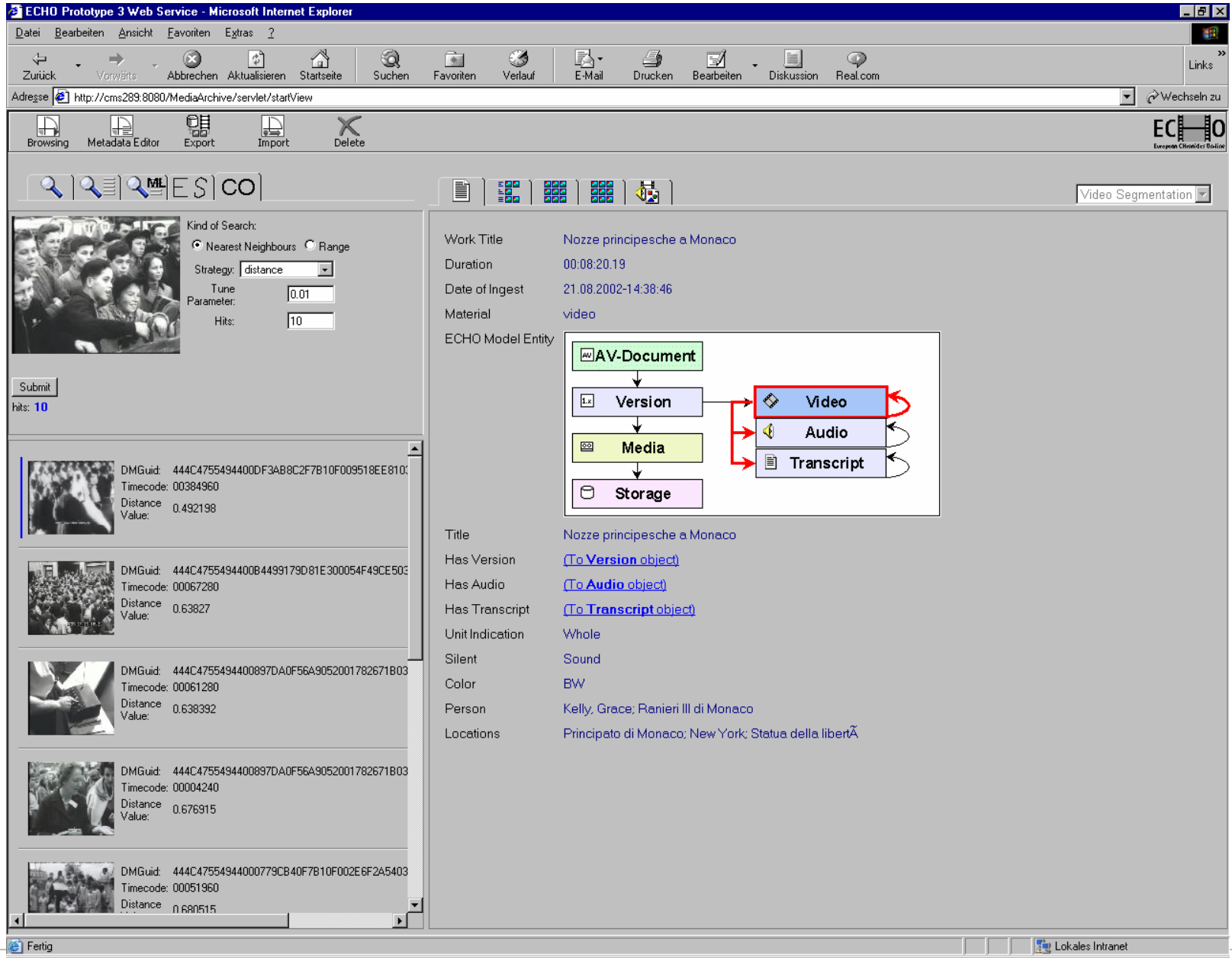
Attribute Search Retrieval interface

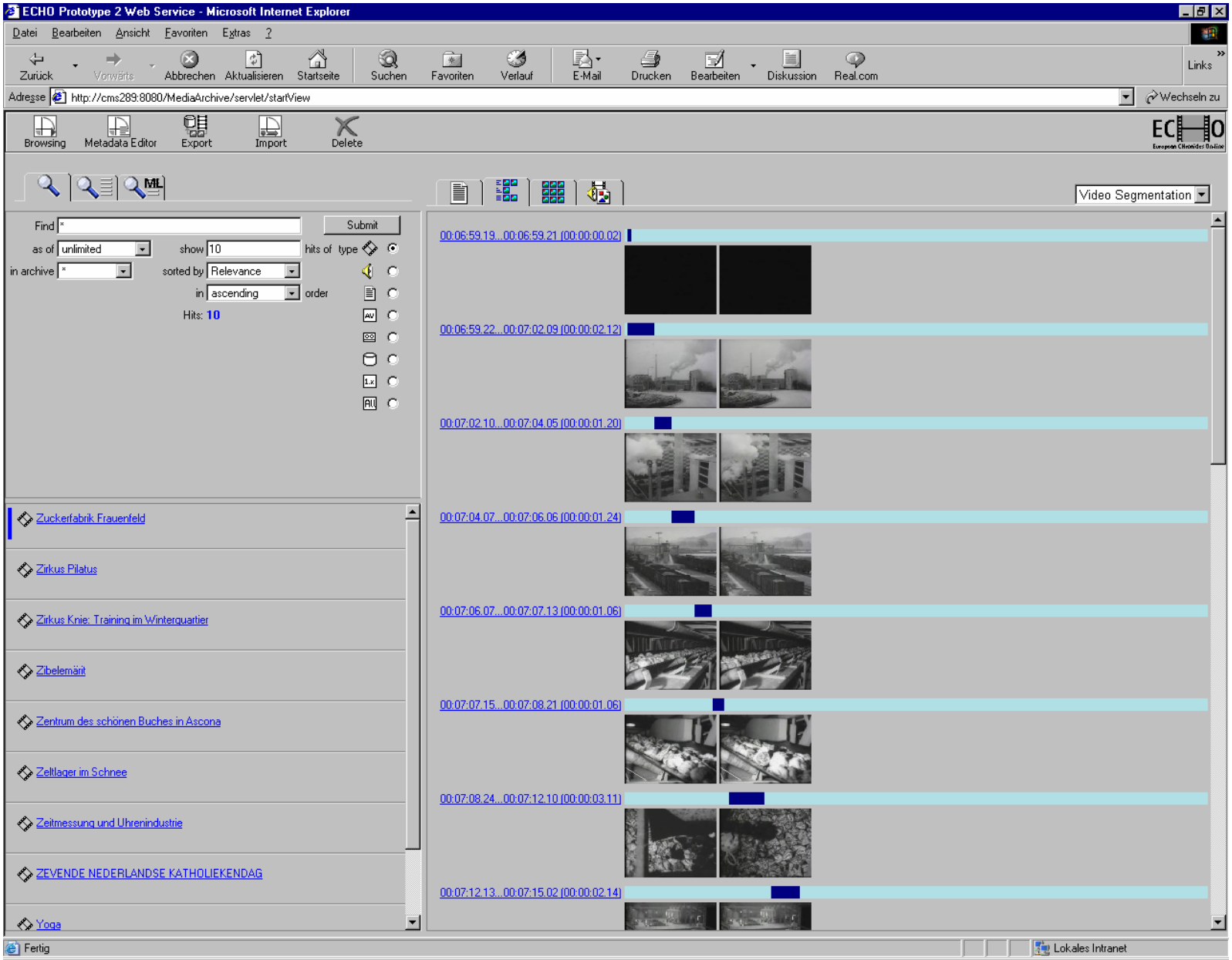
Search for category level of material (Work, Expression, Manifestation, Item)

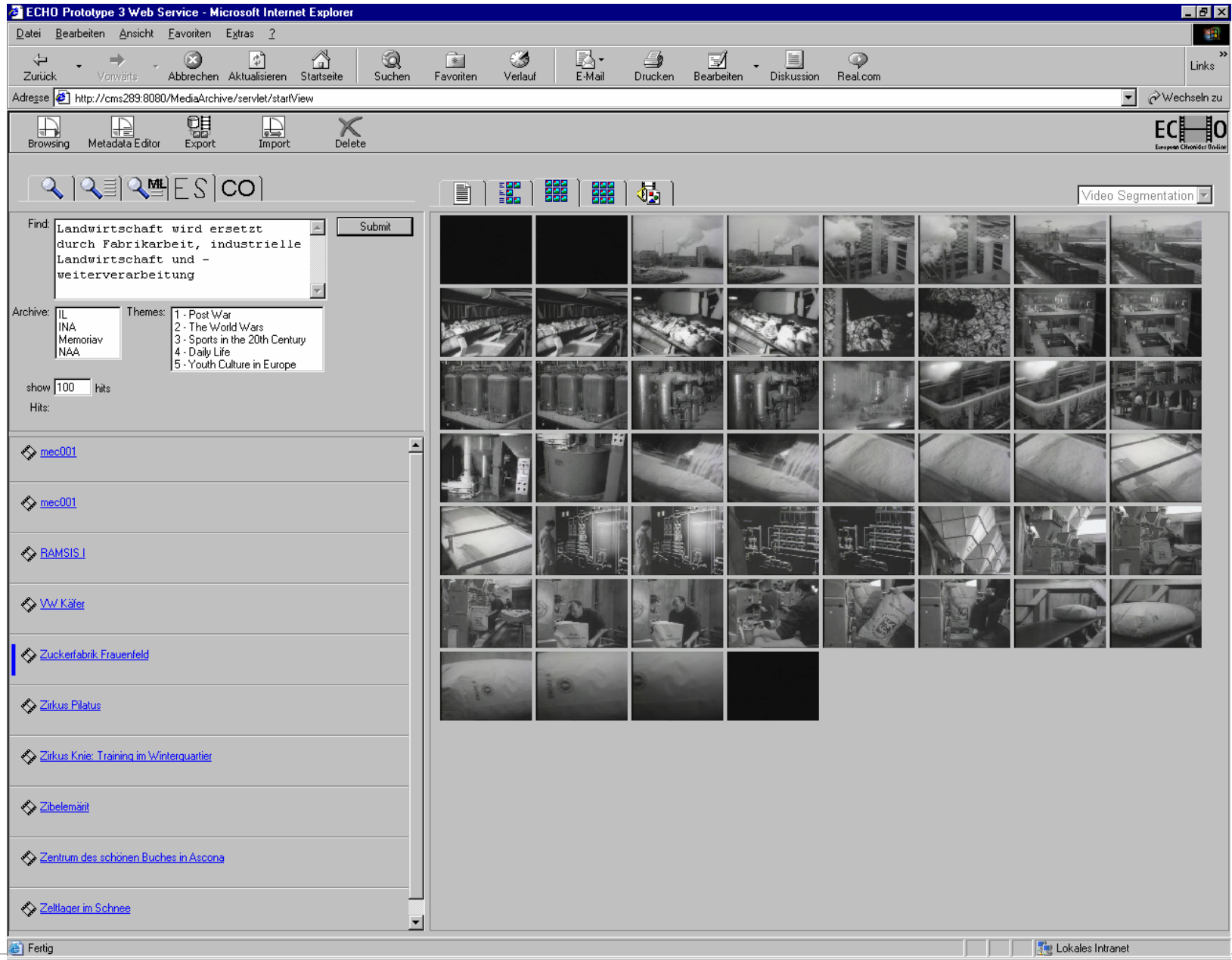
Detailed view on an item corresponding to the full ECHO data model (structures, links, ...)

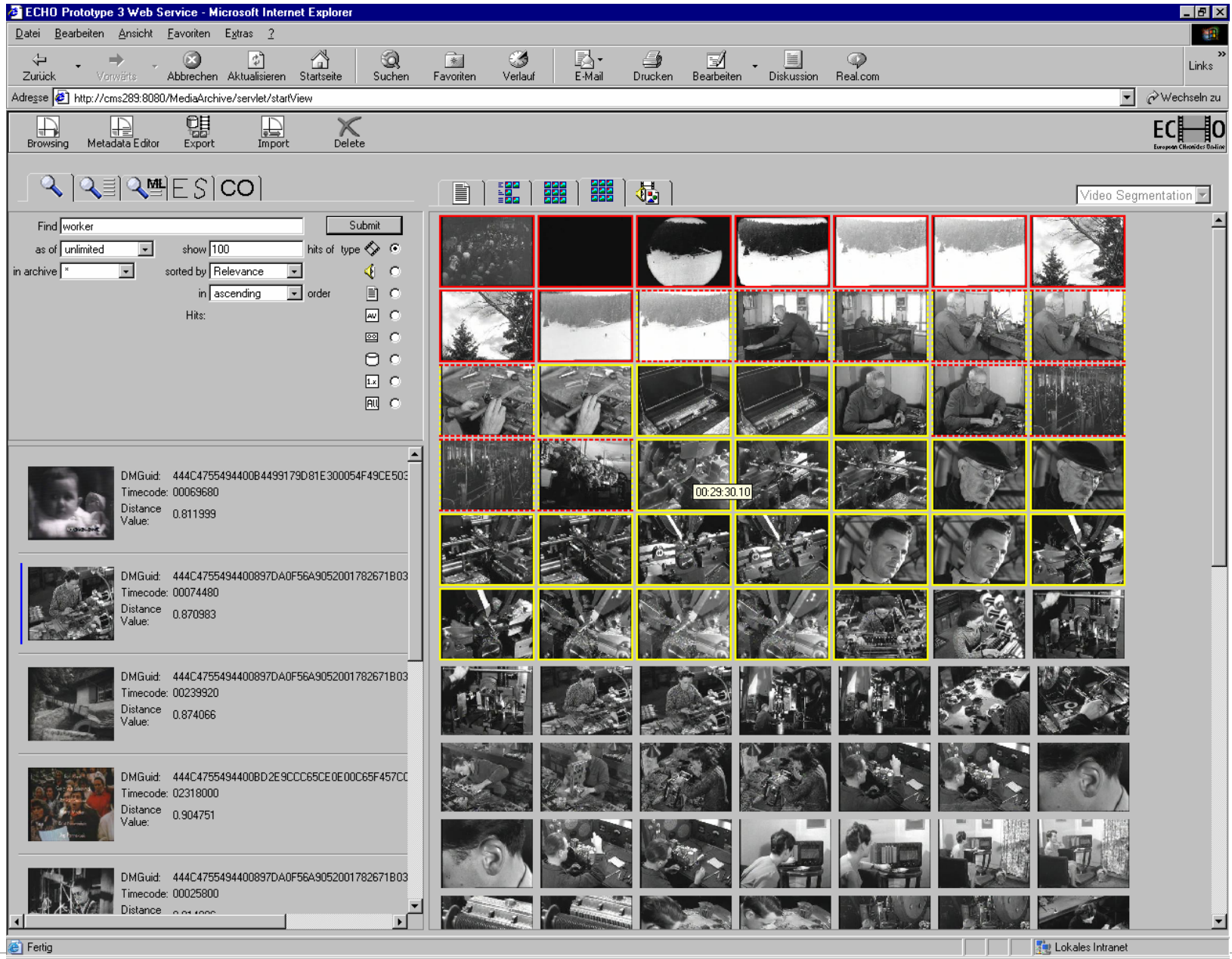
List of retrieved items

Digital Library Competence Center


ECHO Prototype 3 Web Service - Microsoft Internet Explorer
Datei Bearbeiten Ansicht Favoriten Extras ?
Zurück Vorwärts Abbrechen Aktualisieren Startseite Suchen Favoriten Verlauf E-Mail Drucken Bearbeiten Diskussion Real.com Links
Adresse http://cms289:8080/MediaArchive/servlet/startView Wechseln zu
Browsing Metadata Editor Export Import Delete
ECHO
ES CO
Video Segmentation
Kind of Search:
Nearest Neighbours Range
Strategy: distance
Tune Parameter: 0.01
Hits: 10
Submit
hits: 10
DMGuid: 444C4755494400DF3AB8C2F7B10F009518EE810
Timecode: 00384960
Distance Value: 0.492198
DMGuid: 444C4755494400B4499179D81E300054F49CE503
Timecode: 00067280
Distance Value: 0.63827
DMGuid: 444C4755494400897DA0F56A9052001782671B03
Timecode: 00061280
Distance Value: 0.638392
DMGuid: 444C4755494400897DA0F56A9052001782671B03
Timecode: 00004240
Distance Value: 0.676915
DMGuid: 444C4755494400779CB40F7B10F002E6F2A5403
Timecode: 00051960
Distance
0.680515
Work Title Nozze principesche a Monaco
Duration 00:08:20.19
Date of Ingest 21.08.2002-14:38:46
Material video
ECHO Model Entity
AV-Document
Version
Media
Storage
Video
Audio
Transcript
Title Nozze principesche a Monaco
Has Version (To Version object)
Has Audio (To Audio object)
Has Transcript (To Transcript object)
Unit Indication Whole
Silent Sound
Color BW
Person Kelly, Grace; Ranieri III di Monaco
Locations Principato di Monaco; New York; Statua della libertÃ
Fertig
Lokales Intranet


ECHO Prototype 2 Web Service - Microsoft Internet Explorer
Datei Bearbeiten Ansicht Favoriten Extras ?
Zurück Vorwärts Abbrechen Aktualisieren Startseite Suchen Favoriten Verlauf E-Mail Drucken Bearbeiten Diskussion Real.com Links
Adresse http://cms289:8080/MediaArchive/servlet/startView
Wechseln zu
Browsing Metadata Editor Export Import Delete
ECHO
Video Segmentation
Find *
Submit
as of unlimited
show 10
hits of type
in archive *
sorted by Relevance
in ascending order
Hits: 10
Zuckerfabrik Frauenfeld
Zirkus Pilatus
Zirkus Knie: Training im Winterquartier
Zibelemärit
Zentrum des schönen Buches in Ascona
Zeltlager im Schnee
Zeitmessung und Uhrenindustrie
ZEVENDE NEDERLANDSE KATHOLIEKENDAG
Yoga
00:06:59.19...00:06:59.21 (00:00:00.02)
00:06:59.22...00:07:02.09 (00:00:02.12)
00:07:02.10...00:07:04.05 (00:00:01.20)
00:07:04.07...00:07:06.06 (00:00:01.24)
00:07:06.07...00:07:07.13 (00:00:01.06)
00:07:07.15...00:07:08.21 (00:00:01.06)
00:07:08.24...00:07:12.10 (00:00:03.11)
00:07:12.13...00:07:15.02 (00:00:02.14)
Fertig
Lokales Intranet


ECHO Prototype 3 Web Service - Microsoft Internet Explorer
Datei Bearbeiten Ansicht Favoriten Extras ?
Zurück Vorwärts Abbrechen Aktualisieren Startseite Suchen Favoriten Verlauf E-Mail Drucken Bearbeiten Diskussion Real.com Links
Adresse http://cms289:8080/MediaArchive/servlet/startView Wechseln zu
Browsing Metadata Editor Export Import Delete
ES CO
Video Segmentation
Find: Landwirtschaft wird ersetzt durch Fabrikarbeit, industrielle Landwirtschaft und -weiterverarbeitung
Submit
Archive: IL INA Memoriav NAA
Themes: 1 - Post War 2 - The World Wars 3 - Sports in the 20th Century 4 - Daily Life 5 - Youth Culture in Europe
show 100 hits
Hits:
mec001
mec001
RAMSIS I
VW Käfer
Zuckerfabrik Frauenfeld
Zirkus Pilatus
Zirkus Knie: Training im Winterquartier
Zibelemärit
Zentrum des schönen Buches in Ascona
Zeltlager im Schnee
Fertig
Lokales Intranet


ECHO Prototype 3 Web Service - Microsoft Internet Explorer
Datei Bearbeiten Ansicht Favoriten Extras ?
Zurück Vorwärts Abbrechen Aktualisieren Startseite Suchen Favoriten Verlauf E-Mail Drucken Bearbeiten Diskussion Real.com Links
Adresse http://cms289:8080/MediaArchive/servlet/startView Wechseln zu
Browsing Metadata Editor Export Import Delete
ECHO European CHronicles On-line
Video Segmentation
Find worker Submit
as of unlimited show 100 hits of type
in archive * sorted by Relevance
in ascending order
Hits:
DMGuid: 444C4755494400B4499179D81E300054F49CE503
Timecode: 00069680
Distance Value: 0.811999
DMGuid: 444C4755494400897DA0F56A9052001782671B03
Timecode: 00074480
Distance Value: 0.870983
DMGuid: 444C4755494400897DA0F56A9052001782671B03
Timecode: 00239920
Distance Value: 0.874066
DMGuid: 444C4755494400BD2E9CCC65CE0E00C65F457CC
Timecode: 02318000
Distance Value: 0.904751
DMGuid: 444C4755494400897DA0F56A9052001782671B03
Timecode: 00025800
Distance
00:29:30.10
Fertig
Lokales Intranet

The ECHO A/V documents

- Composed of 200 hours of video documentaries
  - From four collections of National Archives of video documentaries, 50 hours per archive
  - Interesting for the user communities
  - 'National' footage' 'European' angle
  - Interrelated themes
  - Timespan 1920-present (Focus on 1920-1960 period)
  - Structure: thematically/chronological.
  - Selected documents belonging to 5 themes, each one subdivided into subthemes

The 5 Main Themes

1 Post-War
2 The World Wars
3 Sports in the 20th Century
4 Daily life
5 (Youth) Culture in Europe

How to design and build an audio/video digital library

Pasquale Savino
ISTI-CNR

ECHO
European CHronicles On-line

DLib Center
Digital Library Competence Center

Outline

- Preliminary analysis
  - User needs analysis
  - Analysis of the video material
  - Analysis of the needed functions
  - Selection of an appropriate digital library system
- Design
  - Selection of relevant metadata
  - How to organize the video data
- Digital library creation
  - Video ingestion
  - Video analysis
  - Indexing

User needs analysis

- Select a representative group of users for your application environment
- Define a procedure for acquiring their needs
  - On-line questionnaire
  - Interview (free style)
  - Interview (with a questionnaire)
- Collect user requirements
- Analyze user requirements
  - Determine different categories of requirements (e.g. mandatory, desirable, not necessary)

Analysis of the video material

- Format of available video
  - Analog/Digital
  - Type of format
- Quality of video
  - Resolution, frames/sec, etc.
  - B&W or color
- Quality of audio
- Depending on the type of indexing and retrieval that will be performed, it could be useful to perform an analysis of the types of content of the video material
  - Presence of faces, titles, closed caption, etc.

Analysis of main functionality needed and of a DL system

- Prepare a functionality requirement document, based on the user requirements analysis
- Analyze the functionality of existing systems and compare them with the required functionality

Selection of relevant metadata

- Determine the precise set of metadata needed to describe the video material
- Select the metadata that can be automatically extracted and those that need user intervention
- Determine if there is the need of thesauri

How to organize the video data

- Subdivision of video material into collections and sub-collections, according to their characteristics, type of retrieval required, etc.
- Physical distribution of video material into different archives
  - Centralized vs. distributed organization
  - In a distributed organization define how data are distributed, if data replication is allowed

Digital Library creation

- Digitalization of the video material (depends on the available format)
- Video ingestion
  - The video material is stored in the DL.
- Video analysis
  - Video segmentation into shots, scenes, key frame extraction
  - Generation of video summaries
- Automatic indexing
  - Speech recognition
  - Key frames indexing
  - Face, text, objects, etc. recognition
- Manual association of metadata

A practical example

The creation of the ECHO Digital Library

Pasquale Savino

ISTI-CNR

ECHO
European CHronicles On-line

DLib Center
Digital Library Competence Center

Outline

- Preliminary analysis
  - The results of the ECHO User needs analysis
  - The characteristics of ECHO video material
  - The main functionality needed
  - The ECHO DL system vs other systems
- Design
  - The selection of relevant metadata in ECHO
  - The organization of the ECHO video data
- The ECHO Digital library creation
  - Video ingestion
  - Video analysis
  - Indexing

User's requirement collection and analysis

User needs assessment

60 users subdivided in different cathegories

Educational Environment

- Teachers (film/television/new media)
- Students (history, film)
- Scientific researchers
- Documentary makers

Film & Entertainment Industry

- News editors/journalists

Cultural Heritage Institutions

- Keepers/custodians
- Educational department
- Exhibitioners

Audiovisual Industry

- Product developers
- Application designers
- Researchers

AV Archives

- Archivists
- Sales managers
- Researchers

Over 250 questions divided into 8 categories, linked to several functionalities

- Data entry management
- Indexing
- Retrieval
- Output

- Export & re-use
- Billing and account
- Usage in general
- Content

User needs

1. Data entry management

Manually added metadata as well as (semi) automatic metadata extraction from digital film information are distinct features of ECHO.

2. Interface & related databases

Full Web-based interface; the possibility to search in multiple archives of one or more countries in different languages.

3. Administration

ECHO will have a price and billing mechanism in order to take care of acces, control, authentication, privacy and accounting.

User needs (cont.)

4. Cross linguality

To facilitate full text retrieval and automatic keyword extraction, the original speech of the ECHO content will be converted into text via Dutch, French, German and Italian speech recognizers.

5. Retrieval

Echo will provide content-based searching and retrieval.
As the content is conveyed in both narrative and image form, collaborative interactions of technologies will be adopted for satisfactory recall and precision.

6. Presentation results

ECHO will provide several presentations, either at the document level or at the level of the search result.

User needs (cont.)

7.Visual Abstract

The required automatically created (moving) visual summary should capture content and structure of the underlying documentary film for intuitive understanding and give a good overview of the entire content.

8. Browse Copy

The ECHO content should be made available via computer networks and the Internet. The original historical films will be converted into browse copies.

9. Storyboard

In order to get a quick overview of the visual content of a retrieved programme and to be able to start the browse copy at any given point, a storyboard should be generated of each item in ECHO.

User needs (cont.)

10.Cross Document Viewing

The visualization and summarization of the content across all stories in a result set is a desired viewing functionality that helps the user understand the chronological, geographical, and/or subject context of the retrieved content.

11. Re-use of the content

Functionalities that facilitate the profit and non-profit use of the ECHO content, either the browse copy or a high res copy of the original material, which is expected in several user communities, active in public and commercial broadcasting, historical studies, teaching and the creation of new AV-products.

The characteristics of ECHO video material

Physical characteristics

- MPEG-1 format
- Duration from 1 minute to 1 hour
- Mainly black&white video
- Low level audio quality
- Presence of speech
- Presence of text
  - Titles
  - Name of authors

The content

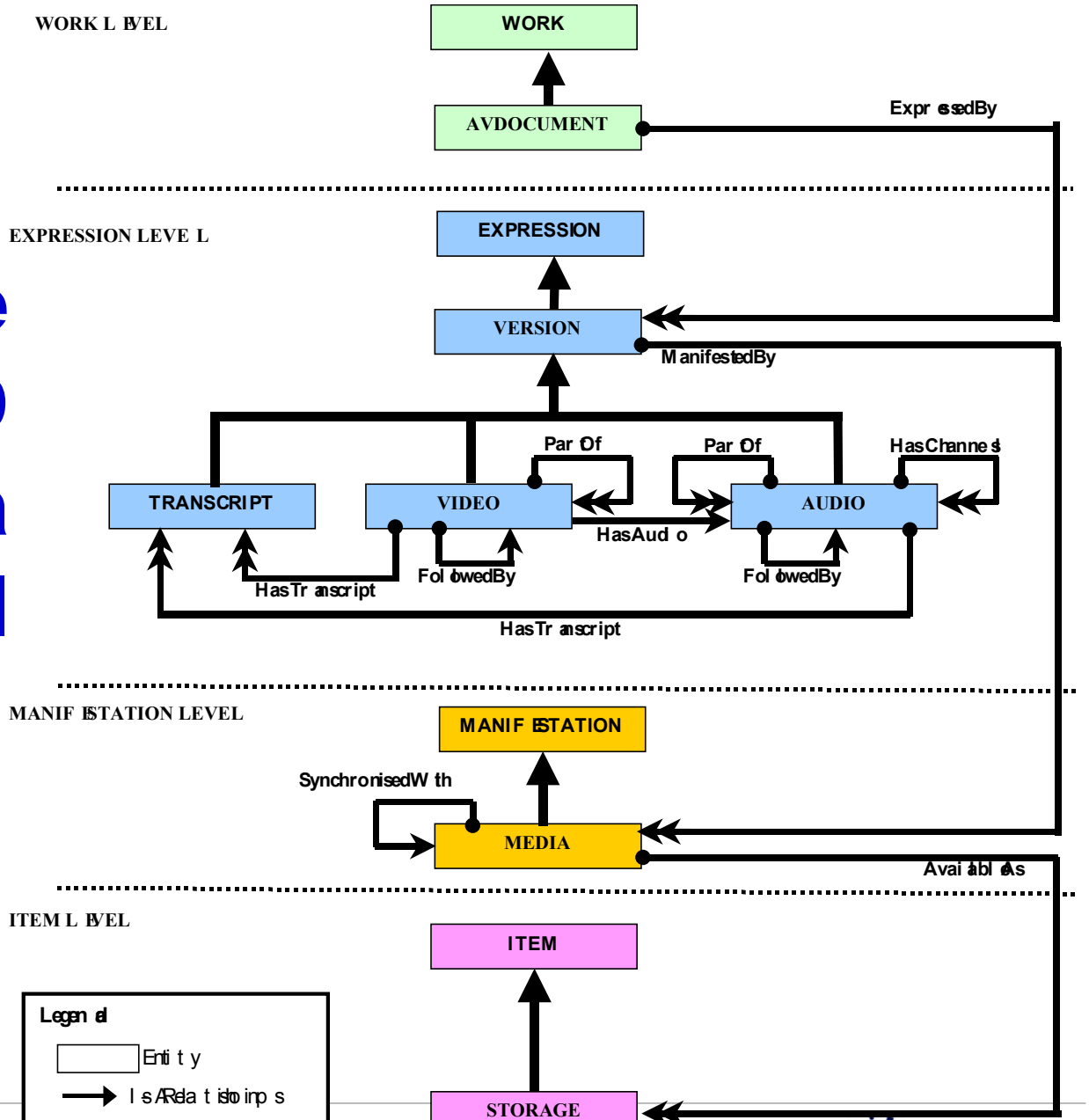
- Historical documentaries produced from the 20ties until 70ties
- Videos can be grouped into different themes
- Speech and text in four languages
  - Italian, French, Dutch, and German
- Some videos have associated multilingual sheets with speech transcripts
- Most of the video has speech descriptions
- Most of the videos have existing metadata

Selection of relevant metadata

Main functionality

- Traditional audio-visual access funct.
  - by the name of the producer
  - by the series title
  - by the “tape”identifier, …
- Advanced audio-video access funct.
  - by key-frames
  - by features
  - by visual abstract
  - by words in the transcript, ...

The ECHO metadata model



Metadata examples

- Work
  - Title
  - SeriesTitle
  - EventDate
  - Director
  - Censorship
  - Description
- Expression
  - Edition, Duration
  - The video component of the expression has
    - Kind, SubtitleLanguage, ImageDescription, VideoAbstract, KeyFrame, Faces, Objects, CameraMovement, etc.
  - The audio component has
    - Kind, AudioLanguage, Frequency, Type, etc.
  - The transcript component has
    - Transcript, SpeakerID, Gender, SpeakerLanguage, etc.

Metadata examples (cont.)

- Manifestation
  - Format,
  - size, etc.
- Item
  - Collocation
  - Provider
  - StorageID
  - PublicAccess, etc.

How the video data are organized

The 5 Main Themes

1 Post-War
2 The World Wars
3 Sports in the 20th Century
4 Daily life
5 (Youth) Culture in Europe

Theme 1 Post-War

- European Communities
- Continuing Life in the City
- Emigration Movements
- Rebuilding the Military Forces
- Cold War and International Relationships
- (Changes in) Society

Theme 2 The World Wars

- Aftermath World War
- 1920-1945 Major Events
- 1920-1945 Propagand
- 1920-1945 International Relationships
- 1920-1945 Socio-economic Factors
- 1939-1945 The Development of the Second World War

Still from the Istituto Luce Collection: "Cinema is the strongest weapon"

Theme 3 Sports in the 20th Century

- Sociological Developments
- National Sports
- Mass-Events
- European Contests
- Sponsoring
- Vandalism

Theme 4 Daily Life

- Work and Leisure
- The European Family
- Education
- Food and Drink
- Sickness and Health

Still from the INA Collection: A Bicycle Taxi

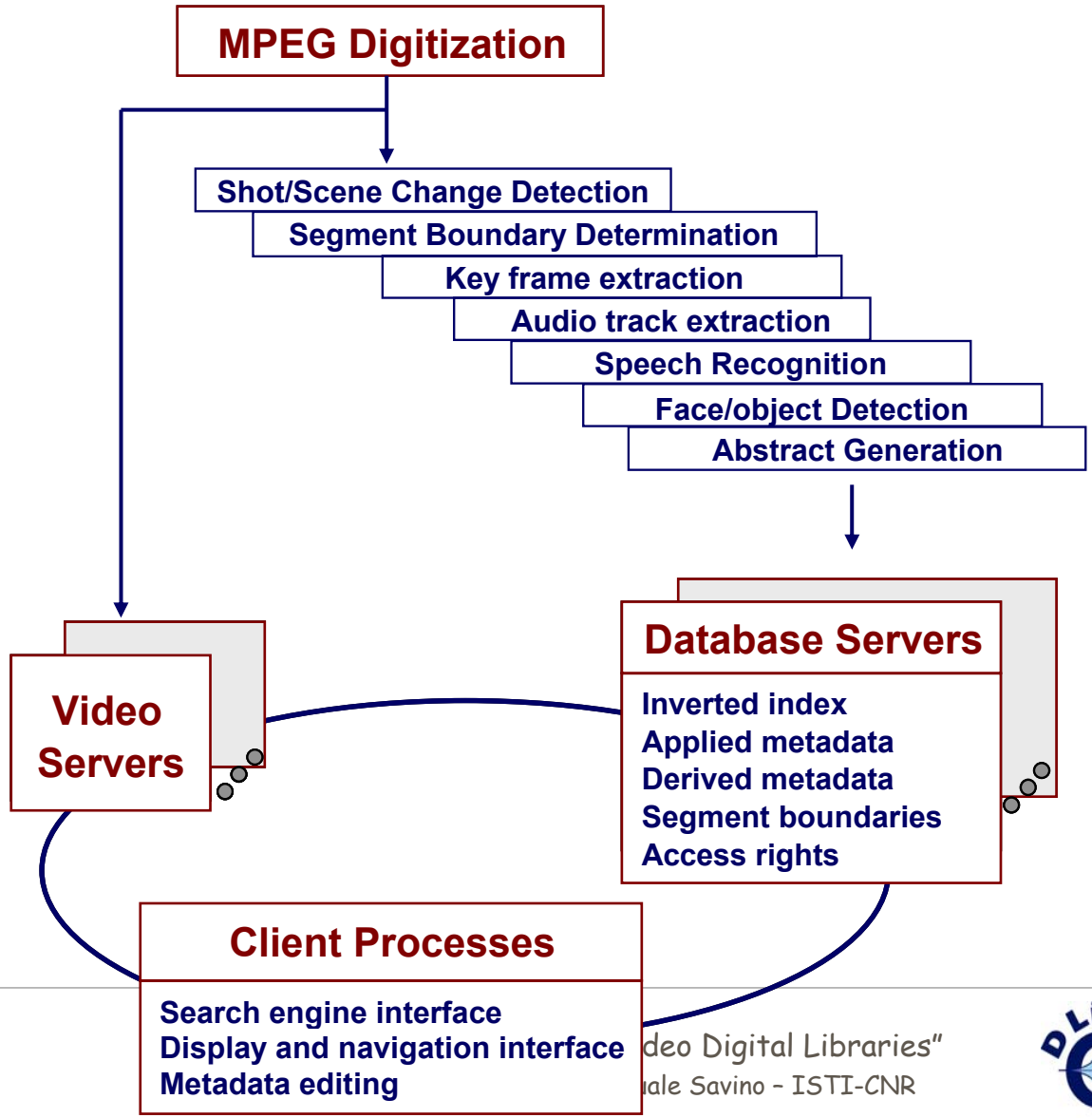
Theme 5 (Youth) Culture in Europe

- Fashion, Clothing, Lifestyle
- Student Revolts
- Sexual Revolution
- Television and Radio
- The Arts

Still from the NAA (Smalfilm) Collection: "Boy"

The ECHO Digital Library creation

Library creation overview



Video Ingestion

- ECHO has tools for the ingestion of multiple video documents
- It is also possible to store single video documents
- During ingestion an identifier is associated with each video document, and a first set of metadata fields are created (e.g. name, creation date, etc.)

Video analysis

- During the ingestion, each video is segmented into shots, and key frames are extracted from each shot
- The audio track is extracted in order to be used during the automatic speech recognition
- The video summary of each video is created as a separate process (elapsed time ~ 10 time video duration)

Indexing

- Automatic indexing
  - All these procedures are performed independently. At the end each procedure updates the metadata of the video
  - The audio track is sent to specialized modules for speech recognition (Italian, French, Dutch)
  - In case of written transcripts of the speech, an OCR is performed
  - Key frames indexing
    - Extraction of features representing color distribution and texture
  - Face, text, objects, etc. recognition
- Manual association of metadata
  - The Metadata Editor is used to modify all metadata fields of the video

Outline

- What is a Digital Library?
- Characteristics of an Audio/Video DL
- Applications of Audio/Video DLs
- Types of data managed
- The characteristics of digital Audio and Video
- The main functions
- Automatic and manual indexing
- Retrieval functionality
- Logical architecture of a video DL
- User's categories
- Overview of existing systems

Tutorial program

- Introduction to Audio/Video Digital Libraries
- How to build an Audio/Video Digital Library
- A practical example: the creation of a documentary film Digital Library
- Metadata models
- Automatic indexing of A/V documents
- Manual indexing of A/V documents
- The ECHO metadata editor

References

- What is a Digital Library?
- Characteristics of an Audio/Video DL
  - Ian H. Witten, David Bainbridge, “How to build a Digital Library”, Morgan Kaufmann Publishers, 2003
  - William Y. Arms, “Digital Libraries”, The MIT Press, 2001
  - Michael Lesk, “Practical Digital Libraries”, Morgan Kaufmann Publishers, 1997
  - Gary Cleveland, “Digital Libraries: Definitions, Issues and Challenges”, IFLANet, March 1998,
- Applications of Audio/Video DLs
  - Virage, Inc. Case studies, http://www.virage.com/customers/case_studies/
  - Sonic Foundry, partners, http://www.mediasite.com/partners/partnes.asp
- The characteristics of digital Audio and Video
  - V.S. Subrahmanian, “Principles of multimedia database systems”, Morgan Kaufman Pub., 1998
  - Borko Furht, Stephen W. Smollar, Hongjiang Zhang, “Video and image processing in multimedia systems”, Boston Kluwer Academic Pub. 1995
  - Home page of the Moving Picture Experts Group (MPEG): http://mpeg.telecomitalialab.com/
- Overview of existing systems
  - Ian H. Witten, David Bainbridge, “How to build a Digital Library”, Morgan Kaufmann Publishers, 2003
  - Greenstone software: http://www.greenstone.org/english/home.html
  - Virage home page: http://www.virage.com/
  - Informedia home page: http://www.informedia.cs.cmu.edu/
  - ECHO home page: http://pc-erato2.iei.pi.cnr.it/echo/
  - ECHO final report: http://pc-erato2.iei.pi.cnr.it/echo/documents/public/Final%20Report.pdf
- How to design and build an audio/video digital library
  - Ian H. Witten, David Bainbridge, “How to build a Digital Library”, Morgan Kaufmann Publishers, 2003
  - Sun Microsystems, “The Digital Library Toolkit”, January 2003, 3rd edition
  - Annemieke de Jong, Johan Oomen, Pasquale Savino, Hanneke Smulders, Paola Venerosi, “ECHO: User Requirements Report”, June 2000