D-Lib IST-2001-32587
Digital Library Competence Center

Audio/Video Digital Library – Report D5.1.1

Deliverable Type: REPORT
Number: D5.1.1
Nature: Public
Contractual Date of Delivery: month 13
Actual Date of Delivery :month 24
Task WP5.1

Name of responsible: Costantino Thanos, ISTI-CNR

Authors: Giuseppe Amato, Claudio Gennaro, Pasquale Savino, ISTI-CNR

contact info
pasquale.savino@isti.cnr.it

Abstract

This report details the organization of the course on Film Digital Library Environment and the manuals written for and distributed at the course itself. The complete documentation has been made available to the public through the D-Lib Center web site (http://dlibcenter.iei.pi.cnr.it/).

Executive Summary

The aim of this workpackage was to provide a theoretical and experimental background on the techniques and the methodologies for the organization, creation, and management of an Audio/Video digital library.

The D-Lib Center established, operated and managed an audio/video digital library environment. In particular, it made available the ECHO digital library system (http://pc-erato2.iei.pi.cnr.it/echo/) developed within a project funded by the IST Programme of the European Commission (5th FP). An overview of the ECHO system functionality is provided in Appendix A.

The ECHO system, used as a testbed during the course, provides a digital library service for historical films. It allows to index and retrieve the audio video material by using speech transcripts, video features automatically extracted from the video and metadata manually associated by the user. Metadata are described by using an audio-video metadata model based on the IFLA-FRBR standard.

The frontier of Digital Libraries consists in the possibility of managing multimedia documents other than pure textual information. In particular, due to the large amount of Audio/Video material that is available in a digital form and due to the importance of this material for many aspects - economic, environmental, health, cultural, social, etc. - of everyday life, the management of Audio/Video Digital Libraries is becoming of crucial importance.

Courses on audio/video digital libraries were organised and held at the D-Lib Center laboratory on the following dates:

20-21 January 2003
27-28 November 2003

Teaching Staff:
Pasquale Savino, ISTI- CNR
Giuseppe Amato, ISTI- CNR
Claudio Gennaro, ISTI-CNR

Technical Staff:
Carlo Carlesi, ISTI - CNR

1. Course overview

The Course had the aim to provide a theoretical and experimental background on the techniques and the methodologies for the organization, creation, and management of an Audio/Video Digital Library. It was organized in two full days, and it included face to face presentations, as well as training sessions. The course program was as follows:

1. Introduction to Audio/Video digital libraries
   - 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
2. How to design and build an Audio/Video Digital Library
   - Preliminary Analysis
     i. Analysis of video material
     ii. Analysis of needed functions
     iii. Selection of an appropriate digital library system
   - Design
     i. Selection of relevant metadata
     ii. How to organize the video data
   - Digital library creation
     i. Video ingestion
     ii. Video Analysis
     iii. Indexing

3. A practical example. The creation of a documentary film DL
    - This part illustrates the phases followed during the design and the creation of the ECHO digital library
4. Metadata Models for A/V Digital Libraries
    - Metadata
        i. Dublin core
        ii. MPEG-7
        iii. IFLA-FRBR / ECHO
    - Editing Metadata
    - Automatic indexing
        i. Text, speech, images, moving pictures
5. Manual indexing of A/V documents
    - Introduction
    - IBM MPEG7 editor tool (demo)
    - Tecmath eample
    - The echo model (restricted version)
    - The echo Metadata editor
    - Browsing the echo Metadata
        i. Demonstration of an A/V Digital Library
    - Echo Editor Architecture
    - Echo Editor demo
6. Demonstration of an A/V Digital Library
7. Training session on the use of an A/V Digital Library
    - The training session was based on a set of tasks that students had to accomplish, from simple free text search on the audio transcript, up to complex search based on the use of all functionality of the ECHO metadata model. Finally, the user has the possibility to use the system for free searches.
8. Training session on manual Indexing of A/V documents
    - Two metadata editors were used, the *VideoAnnEx* annotation tool (http://www.research.ibm.com/VideoAnnEx/download.html) and the ECHO metadata editor
9. Questionnaire
    - A questionnaire asking for the student’s interest and satisfaction on the course was distributed and filled.

The demonstration and the training activities were based on the ECHO prototype system and on the ECHO video collection. In particular, the hw/sw platform was composed of an ECHO server plus 5 workstations running the ECHO client, as well as the creation of the test video library composed of 50 hours of video selected among the complete ECHO collection of A/V documentaries.

The documentation distributed at the course and available for download is composed of: 1) Copy of the slides used during the course. 2) User manual of the ECHO metadata editor. 3) User manual of the ECHO retrieval module, 4) ECHO metadata model.

2. Tutorial

This section provides some more details on the different topics presented during the course.

2.1. Introduction to Audio/Video digital libraries

This part of the tutorial begins with a definition of what is a Digital Library and a comparison between traditional and A/V digital libraries. The differences and the similarities in terms of data types managed and services offered are illustrated..
Later on, the main application areas of A/V digital libraries are analyzed: in particular, the following four main categories have been considered:

- 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

This introduction continues with a detailed analysis of the characteristics of audio and video, the techniques used for video compression (MPEG1-2-4), the methods used for the representation of the video structure.

Later on, the main functions of an A/V digital library are presented:

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

The introduction continued with an overview of the techniques used for automatic indexing of audio/video, starting from the techniques used for video segmentation, automatic feature extraction (e.g. audio track, speech recognition, key frame extraction and analysis, object and face detection and recognition)

The retrieval functionality and the types of query managed in an A/V digital library are also presented.

Then, the logical architecture of a video DL is illustrated.

The introduction is completed with an overview of existing systems that can be used as A/V digital libraries. In particular, the description looked first at systems that are able to manage video only as an unstructured data type (such as Greenstone www.mkp.com/DL). These systems provide an indexing based on metadata associated to the entire video, while retrieval is based on these metadata. No use is made of the video content. More advanced video archiving and retrieval systems (e.g. Virage http://www.virage.com , Informedia http://www.informedia.cs.cmu.edu) use part of video content to support retrieval. Indexing is mainly automatic. Other systems (e.g. ECHO http://pc-erato2.iei.pi.cnr.it/echo/) also offer typical DL services combined with powerful indexing and retrieval capabilities. Indexing is partly automatic and partly manual.

2.2. How to design and build an Audio/Video Digital Library

This part of the course aimed at presenting the methodology that should be followed to design and to build an audio/video digital library. The methodology was based on a preliminary analysis of the characteristics of the video material (e.g. the format, the quality, etc.), followed by an analysis of the user needs. The user needs analysis should be conducted through an interview of the users in order to define precisely the functionality of the system needed to manage the videos. After this, it is possible to select the most appropriate digital library system.

At this stage, it is possible to initiate the design of the DL, first through a selection of the relevant metadata, and with the definition of the organization of the video material (e.g. physical placement of videos, organization in collections, etc.).

Finally, the real creation of video DL is performed, through an ingestion and indexing of the video material.

2.3. A practical example. The creation of a documentary film DL

This part illustrates the phases followed during the design and the creation of the ECHO digital library. The creation of the ECHO digital library was performed according to the guidelines outlined in the previous section.

2.4. Metadata Models for A/V Digital Libraries

We first provide an introduction to the concept of metadata, and an overview of most popular metadata models used for A/V DLs: Dublin core, and MPEG-7. Then we illustrate in detail the metadata model used in ECHO, which is based on an adapataion of the IFLA-FRBR model.

Then, the problem of A/V indexing is addressed, looking at the techniques for automatic indexing and manual indexing.

On automatic processing of video material, the following aspects have to be considered:

- Cut detection
- Visual features extraction
- Transcript generation
- Object recognition
- Face recognition
- Geospatial information
- Video abstract generation

Automatic indexing requires to support indexing of text, speech, images, moving pictures. All these aspects are illustrated in detail.

2.5. Manual indexing of A/V documents

This module of the course presents the methods used to perform the manual indexing of video material. This is strictly related to the metadata model adopted. We present three different approaches to manual indexing: the first one is based on the VideoAnnEx Tool developed by IBM, which supports the annotation of videos; the second one is based on the video metadata editor developed by Tecmath, which is based on a model that looks at videos at different strata; for each stratum it is possible to provide specific time dependent metadata. Finally, the metadata editor used in the ECHO prototype system, based on the ECHO metadata model is presented in detail.

2.6. Demonstration of an A/V Digital Library

A demonstration of the ECHO audio video DL was given.

2.7. Training session on the use of an A/V Digital Library

The training session was based on a set of tasks that students had to accomplish, from simple free text search on the audio transcript, up to complex search based on the use of all functionality of the ECHO metadata model. Finally, the user has the possibility to use the system for free searches. The tasks performed in the training session are listed in Appendix B.

2.8. Training session on manual Indexing of A/V documents

Two metadata editors were used, the *VideoAnnEx* annotation tool (http://www.research.ibm.com/VideoAnnEx/download.html) and the ECHO metadata editor

2.9. Questionnaire

A questionnaire asking for the student's interest and satisfaction on the course was distributed and filled.

3. Course organisation

Two courses have been organised; the first one was held on January 18-19, 2003 and the second on November 27-28, 2003; both courses were organised at ISTI-CNR, in Pisa. For both courses, we used the equipment of the D-Lib Center Lab, composed of a server running the ECHO prototype system, and five workstations used as clients, running the retrieval interface and the metadata editor. The course was based on part of the video material selected and indexed in the context of the ECHO project. This material was composed of 50 hours of documentaries provided by four European video archives (INA, NAA, Istituto Luce, and Memoriav). The material was dated from the 20ties to the 60ties of the 20th century. It belong to five different categories:

- Post-war
- The world wars
- Sports in the 20th Century
- Daily life
- (Youth) Culture in Europe

The course has been advertised in a large number of specialised mailing lists (see Appendix C) and through direct mailing contacts. A flyer (see Annex B) advertising the course has been distributed to all major international conferences in the field.

34 people registered to the course. Most of them were archivists; only few of them had a background in computer science, while more had a background in preservation of cultural heritage material. Most of them were Italians, even if there were 6 from other countries (Switzerland, Portugal, Cameroon, Spain, Germany, Brasil) .

4. Other activities

As outlined in the previous section, most of the course attendants were Italians. This is related to many reasons such as problems in following course held in a foreign language and cost of the participation. For this reason, and in order to extend the audience of the course, we prepared a one day tutorial based on the course material. This tutorial was selected and presented at the two main conferences on Digital Libraies, the IEEE-ACM joint Conference on Digital Libraries, held at Houston (USA) in June 2003. and the European Conference on Digital Libraries held in Trondheim in August 2003.

A. Appendix A

This appendix provides an overview of the ECHO project and of the results obtained in ECHO.

The ECHO project has been funded by the IST programme of the European Commission under the V Framework; the project began in February 2000 and will be completed in March 2003. Its objective has been to develop a digital library service for historical films owned by large national audio-visual archives. Actually being able to see and hear an account of a historical event, filmed in the original context, is very different from reading about it. The ECHO services allow users to search and access these documentary film collections and exploit the content for their own particular requirements, whether commercial, educational, leisure, or whatever.

The project has involved a number of European institutions holding or managing unique collections of documentary films, dating from the beginning of the century until the seventies: Istituto Luce, Italy; Institut Nationale Audiovisuel, France; Netherlands Audiovisual Archive; and Memoriav, Switzerland. Academic partners are ISTI-CNR (coordinator), CNRS-LIMSI, ITC-irst, University of Twente, University of Mannheim and Carnegie Mellon; industrial partners are Techmath and Eurospider Information Technology. These collections are of great value as they document different aspects of life in Europe (social, cultural, political, economic) during this period of time. The set of services implemented by ECHO thus provides users with access to significant portions of their cultural heritage, which would otherwise be almost inaccessible. In addition, users can compare the way in which an event or phenomenon is documented in their own country with how it is reported in others, or can investigate how different countries have documented a particular historical period of their life, etc.. This means that the ECHO services have to operate over linguistic, cultural and national boundaries, while respecting the requirements of international standards.

A1. System Functionality

The services to be provided by the system were defined on the basis of the results of a user needs analysis performed by the project. A main requirement was to support interoperability over distributed, heterogeneous digital collections and services. Achieving interoperability in the digital library setting is facilitated by conformance to an open architecture as well as agreement on items such as formats, data types and metadata conventions. These issues have already been addressed with varying degrees of success by digital libraries handling textual collections; the challenge in ECHO was to solve the numerous technical problems that up until now have obstacled the inclusion of audio-visual material in a searchable digital environment. The aim has been to make the film collections available and searcheable to as broad a range of users as possible. To achieve this goal, the following components were developed and included in the system.

A2. Audio/Video Metadata Model

When the project began, there were no well-defined metadata models for an adequate description of film data. A major effort of the project has been to define a suitable metadata model to represent the audiovisual contents of the archive. The model that has been implemented is an adaptation of the IFLA model, a general conceptual framework used to describe heterogeneous digital media resources. The model is composed of four levels describing different aspects of intellectual endeavour: work, expression, manifestation and item. The entities of the model are hierarchically ordered from the top level (work) to the bottom (item).

A3. Intelligent Access

The ECHO system assists the user during the indexing and retrieval of A/V documentaries. Semiautomatic indexing is supported: the system automatically extracts a number of metadata information such as the scenes composing the video, key-frames that describe each scene, image features describing each key frame, spoken dialog (automatically transformed into text through a speech recognition process), faces, specific objects. Later on, the user can complete the indexing by specifying metadata that cannot be automatically extracted. Search and retrieval via desktop computer and wide area networks is performed by expressing queries on the audio transcript, on the metadata, or by image similarity retrieval. Retrieved films or their abstracts are then presented to the user. By the collaborative interaction of image, speech and natural language understanding technology, the system compensates for problems of interpretation and search that arise when handling the error-full and ambiguous data sets.

A4. Multilingual User Interface

The ECHO film archives are made up of language dependent (speech, text) and independent (video) media. Thus although users querying over collections in different languages may not understand the spoken dialog, they can still identify useful documents (or parts of documents) via the images. This has facilitated the implementation of a relatively simple multilingual search interface that can still provide useful functionality. The approach adopted has been to implement on-line cross-language search tools based on the use of standard metadata formats and mechanisms that provide a mapping between controlled vocabularies agreed between the content providers. Access is provided by local site interfaces in the local languages but a common user interface in English is also maintained on the project Website for external access.

A5. Creating Visual Summarization

The project has developed techniques to produce visual summaries. The aim is to capture the content and structure of the underlying documentary film in a brief visual abstracting process. The summary consists of a sequence of moving images, much shorter than the original film, but preserving the essence of the original message. It should provide a good overview of the entire film documentary. The creation of visual summaries is based on the use of a number of video features, such as the different scenes recognized, faces, text, objects, and action scenes detected. After this initial analysis, the more relevant clips are determined and assembled in order to maintain the flow of the story. The abstract is usually set to 8\% of the length of the original video, but other values can be specified, depending on user and application needs. The video abstracting process is performed off-line, after video archiving, since it requires approximately 10 times the video duration.

A6. Security

In order to make a digital library of films possible, the copyright owners must be guaranteed that their property will be protected and that its use will be measured in order to provide them with appropriate compensation. ECHO thus includes mechanisms which support access control, authentication, security and privacy.

A7. Analyzing user’s needs

The list of new features included in ECHO has been defined by taking into account the functionality provided by existing commercial systems, including Informedia and Media Archive. In addition, the experience of ECHO content providers was taken into consideration. However, given the unique characteristics of ECHO's audio-visual material and the specific characteristics of potential users of ECHO, we concluded that it was important to collect user requirements for the ECHO system.

In order to collect the user requirements we have prepared an extensive questionnaire that has been submitted to approximately 15 users per participating country (giving a total of 58 users). The primary purpose of the questionnaire was to capture major functional aspects of ECHO, resulting in the user requirements list. Moreover, the questionnaire was constructed extensively enough to give as good an idea as possible on what the interviewees think about using a digital video archive like ECHO.

Three different user types were identified per target market or user community. Each identified user type was expected to have different interests in the ECHO system and to use ECHO in different ways. Furthermore, the user types identified had different levels and special focuses of experience in searching traditional film archives catalogues.

The selection of interviewees was very thorough; it represented the four national archives and the customers of these national archives. Experts on indexing video, on information retrieval, on film making and film history, and on scientific historical research were chosen because they were expected to provide interesting comments on the functionality of ECHO and because they were identified as potential future users of ECHO.

The interviewees themselves indicated their expected future use of ECHO in the questionnaire as follows

- research on historical film material
- browsing historical film material
- reuse of historical film material in new AV products
- data entry in ECHO
- reuse of historical film material for teaching

ECHO can be used for different purposes. The project team distinguished the following types of usage:

- Data entry for archiving
- Retrieval for reference searching, without interest in viewing the content
- Retrieval and viewing the content for the general public
- Professional retrieval and viewing the content for scientific study (history or film science)
- Professional retrieval, viewing and (non-)profit reuse of the content. (including general educational usage

Each type of usage refers to a specific goal for which someone might want to use ECHO.

To actually collect the user requirements, the questionnaire was submitted to users and specific support was provided during the activity.

Given that users could be split in two broad categories, the first including users who were professionally concerned with the database creation process and the second with those concerned with database consultation, two questionnaire were prepared.

The questionnaire was analysed using a numerical approach. The method of analysis was carried out in different steps.

1. Quantitative aspects: thorough analysis of the spreading of the percentages over the answers. The contact persons had the strong impression that the choice in answering *Yes, mandatory* or *Yes, desirable* depended more on the character of the interviewee than on the actual necessity of the functionality concerned. It was therefore decided to analyse these answers together.

2. Qualitative aspects: in the case of a rather equal spreading of the percentages over the answers or when a relatively large number of interviewees answered *No* or *Yes*, optional to a functionality

that is generally regarded as essential to a digital video-archive, the following was taken into account:

The macro-level of the user requirements could be obtained, while:

- the answers to the open questions that were added, to give as good an idea as possible on what the interviewees think about using a digital video archive like ECHO, were not always suitable for transformation into directly related user requirements
- those questions in the questionnaire which asked for a detail of the same major functional aspects of ECHO were analysed together
- the questions concerning the content (category 6) were left out of the analysis for this user requirements report. This category is analysed in Workpackage 2: Content Selection.
- only those questions referring to automatically created metadata were analysed. All other questions that might have an impact on the metadata model of ECHO, concerning manual formal and subject indexing, were not analysed for the general user requirements list of ECHO. The results concerning the manual formal and subject indexing will be discussed and decided upon in the context of Workpackage 3 on Metadata.

A8. The ECHO audio/video collection

The four participating National Archives house enormous collections of audiovisual material starting from the first experiments in the end of the nineteenth century till the present. The origin of the material is diverse and includes cinema newsreels, television programmes, commercials, amateur film etc. etc. These collections keep growing rapidly as audiovisual production increases.

The characteristics of the collections owned by four ECHO archives are briefly listed in the following:

- **Istituto Luce**.

The archive of the Istituto Luce conserves the vastest historical audio-visual memory of Italy, from the first film clips at the time of the Lumière Brothers to the documentaries of the 80's. History, entertainment, customs: a cross-section of their most recent past documented by various cinematographers and news agencies. The most significant part of this patrimony is available on-line. Seventy years of history have been catalogued and computerised and are today available to whoever feels the need to examine their past. The entire audio-visual patrimony of the Istituto Luce contains 3,500 hours of film to which must be added a million photographic negatives. These cover the period between 1916-1960. Within the cinematographic sector, about 3,000 hours have been digitised and computerised. The Istituto Luce archives films that are mainly documentaries and newsreels, mainly in black and white, and from 1932 with sound. The productions with sound were made between 1925 and 1977 for the big screen and not for television.

- **INA**

Created in 1975, INA is the legal guardian of the French audio-visual heritage of 3 millions of radio and television programmes of the public and private broadcasters in France.

The collections go back to 1949 for television programs, 1940 for the news (35mm cinema newsreels) and 1933 for radio archives.

The material include discs and tapes for radio assets and films (mainly 16mm), tapes and cassettes for television recordings (from 2 inches to ¾ inch, at various standards analogous or digital).Those archival materials are kept, restored, distributed for professional uses : re-broadcasting and new broadcasting productions.

Since 1995, the public service mission is extended, making INA, the radio and television Legal Deposit. This mission consists in managing the collections allowing scientific use for researchers and graduate students.

- **Memoriav**

For different reasons (high costs, federal tradition) it was not possible to create a national audiovisual archive in Switzerland. Therefore subsequent considerations led to the foundation of the Association for the preservation of the audiovisual heritage of Switzerland, known as Memoriav, at the end of 1995. The essential concept of Memoriav is creating and sponsoring a network of all resources in Switzerland which are already engaged in different fields in the task of rescuing and conserving audiovisual cultural heritage.

The most ambitious project of Memoriav is the digitisation of reports on Swiss politics in films and TV programs.

One important part of this project is the realisation of the accessibility to the Swiss cinema newsreels (Schweizer Filmwochenschau) and to the TV-News (1954-1989), which have been re-archived by a team of television documentalists.

- **Netherlands Institute for Sound and Vision**

The Dutch Audiovisual Archive was founded in 1996 when the four major keepers of Dutch audiovisual Cultural Heritage merged. The Netherlands Institute for Sound and Vision houses the material of the public broadcasters, documentary films, the National Music Depot, The Broadcast Museum, The Smalfilmmuseum, Commercials, a research department and other audiovisual material dating from 1898 to the present day.

In total the collection of the Netherlands Institute for Sound and Vision consists of more than 600.000 hours of radio, television, film and music. The usergroup of the Netherlands Institute for Sound and Vision is diverse and ranges from producers, publishers and scholars to families that visit the museum.

The Netherlands Institute for Sound and Vision searched for original documents such as Dutch Newsreels from the period 1920 to 1980 and for films and documentaries made for cinema as well as television broadcasts by Dutch directors. It's also interesting to show documentaries and items produced for current affairs programmes in the 60's – 90's, which refer back to certain developments in European society and to European main events in the twentieth century. In order to have a reality look at the European daily life the Netherlands Institute for Sound and Vision, in collaboration with the Filmarchive Smalfilmmuseum, will select a number of amateur films

Netherlands Institute for Sound and Vision is accessible on the internet: http://www.naa.nl (unfortunately only in Dutch) and http://www.smalfilmmuseum.nl .

Taking their catalogue as starting point, a set of themes that could be relevant for the project has been selected. The final theme framework consists of 5 themes and 28 subthemes. The ECHO digital collection is composed of 200 hours of documentaries belonging to all these themes and subthemes, taken from the four archives.

THEME 1 **Post War**	The second World War had a devastating effect on every aspect of the society and international affairs. It also revived the wish for a economic community. Apart from these matters directly linked to WW II, the ECHO collection will also include developments that indirectly originated from the war, like the consumer society and the emigration movements.
European Communities	Films documenting the founding of the various organisations prior to the EC. Successively: Marshall Plan, European Coal and Steel Community, European Economic Community, European Community.
Continuing Life in the City	After the war the cities and the infrastructure were rebuilt and the economic, environmental, cultural and social life were slowly normalised. See also 1.6 Society.
Emigration Movements	Films concerning various forms of emigration: Post War exodus, political (human rights), economical (from Third World to Europe), humanitarian (War, Third World famine), etc. Immigration also falls under this heading.
Rebuilding the Military Forces	Films on NATO, dissuasion and nuclear weapons, etc.

Cold War and International Relationships	Euro-missile crisis (1977-1984), Afghanistan (1973-1990), Cold War 1945-1963 (Berlin Crisis 1958-1962, Cuban Missile Crisis 1962), Nicaragua (1978-1990), Nuclear non-proliferation policy (1945-1991), etc.
(Changes in) Society	This subtheme concerns matters of labour, feminism, demography, poverty, consumer society, and technical inventions.

THEME 2 **The World Wars**	The World War section ECHO collection will include material from the 1920's to 1945.
Aftermath World War 1	The first World War had serious consequences on the developments of the already deteriorating international relations and, in particular, some events determined the somehow predicted economic and political crisis at the end of the 1930's.
1920-1945 Major Events	The fundamental events in those twenty years that influenced the political developments in the second world conflict. Some important topics: Mussolini at power, Stalin at power, Hitler at power, Civil war in Spain, Monaco Conference .
1920-1945 Propaganda	Provides material of how not only dictatorships but also democratic countries used audiovisual material to support their political beliefs among the people
1920-1945 International Relationships	The international relations between the most important European countries: official visits, diplomatic meetings, etc., selected to underline the new developments in the international alliances and the changes from the old system.
1920-1945 Socio-economic Factors	A documentation on the economic transformations which caused Europe's economic quick recovery through rearmament after the crisis at the end of the first World War. Some important topics: The economic crisis after WWI, The 1929 crisis in the USA and Europe, Rearmament.
1939-1945 The Development of the Second World War	Some selected events, such as: Occupation of Poland, Drôle de Guerre, Occupation of Paris, War in the USSR, Stalingrad, Fascist defeat, Normandy, Occupation of Berlin

THEME 3 **Sports in the 20th Century**	The importance of sports in European societies has developed during the 20th century. The goal is not only to show nice motion pictures of different kinds of sport but also to give an impression of the social and national meanings of sports in Europe.
Sociological Developments	Films showing the (changing) importance of sports for a society or for a social group like youth, women, working class etc.
National Sports	Films on sport with a large national identification, sometimes even for people who aren't interested in sports at all; often practised by a national team. E.g. ski-races in Switzerland.
Mass-Events	Films on sports events with a large audience like Olympic Games, World Championships or famous bike races.
European Contests	Films on European Championships in different sports.
Sponsoring	Films showing the growing importance of money in sports and the discussions about it.
Vandalism	Films on the ever growing violence around sports events.

THEME 4 **Daily Life**	Daily life is a broad theme, including matters like: work, education, food, drink and health matters.
Work and Leisure	Films dealing with the production of goods and services and the time available

	for ease and relaxation following these activities. Labour movements is also placed under this heading. (this subtheme includes religious feasts, folklore and holidays)
The European Family	Films on parents and their children in the 20th century. This is also a very broad subtheme, ranging from housing, marriage and funerals to more demographic matters.
Education	Films concerned with teaching, pedagogy, upbringing of children and government legislation will be assembled here.
Food and Drink	Films on the production, distribution and possible lack of food in Europe. European legal matters and legislations will also be discussed here.
Sickness and Health	Films dealing with various health matters, ranging from hospitals, the work of relief organisations to more scientific subjects.
THEME 5 (Youth) Culture in Europe	The theme Culture not only includes different art forms, like literature, film, painting. It also includes the system of socially acquired values, beliefs, and rules of conduct in any given society. This system is created by lifestyle, television and forms of personal expression.
Fashion, Clothing, Lifestyle	Films concerning the manner of living that reflects the people’s values and attitudes, expressed in their appearance.
Student Revolts	Revolts started by students against the ruling classes and their school board.
Sexual Revolution	The alteration of the right or good conduct concerning matters of sex and relationships.
Television and Radio	How the technique behind film and radio was invented and how it was introduced to the public. Films about television personalities, legendary broadcasts and the social impact of these modes of mass communication .
The Arts	This very broad subtheme is filled with various products of human creativity ranging from literature and cinema to painting, theatre and music.

A9. A metadata model for A/V documentaries

The ECHO audio/video metadata model extends the IFLA-FRBR model in order to support the description of audio visual documents and to cope with the specific requirements of audio/video digital libraries.
IFLA-FRBR suggests to describe resources using four different entities: (1) *Work* entity to describe the abstract idea of a resource, (2) *Expression* entity to describe different versions of the same work, (3) M*anifestation* entity to describe different physical supports where an expression can be embodied, (4) I*tem* entity to describe different copies of the same manifestation.

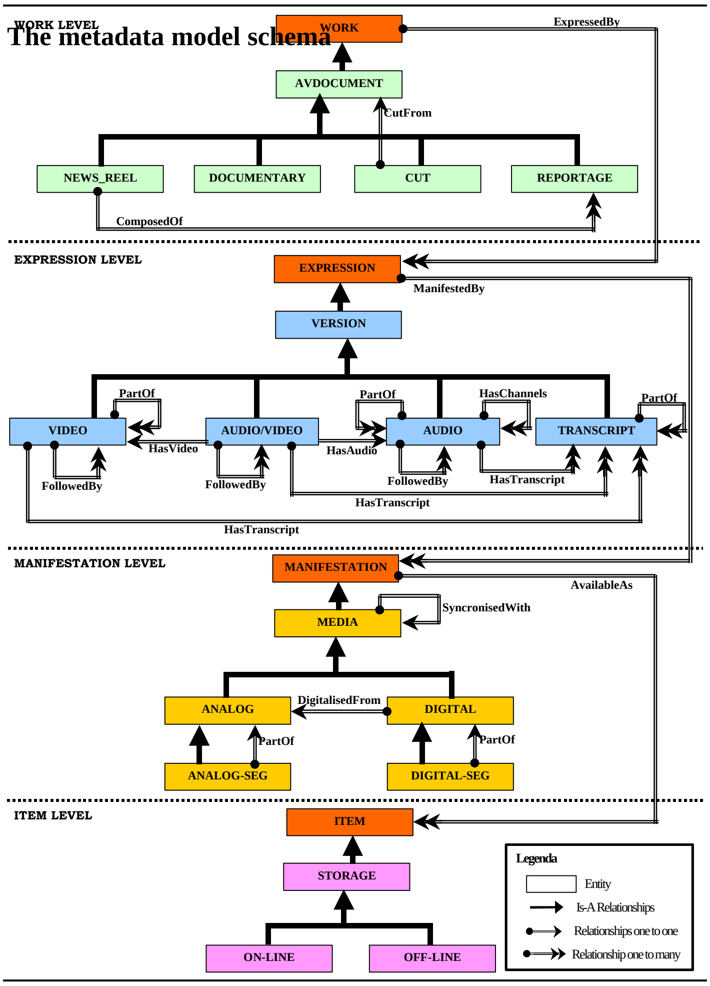
The ECHO metadata model includes new sub entities for each of the previous IFLA-FRBR entities to offer a more specialised mechanism to describe audio visual documents. The overall schema of the metadata model is sketched in the next Figure.

The Work entity was extended by defining the sub entity *AVDocument* that contain attributes specialised for describing abstract ideas of audio visual documents as, for example, Director, Event, Date, Person, Location, Description. Some of these attributes can be defined in such a way that cross-language search can be performed effectively. For instance, a controlled vocabulary can be used where terms are chosen in such a way that a translation exists in all supported languages.
A specific version of a work can have several forms. For instance, it can be a silent movie, an audio stream, a black and white movie, the Italian translated version, etc. To this aim, the *Expression* entity was extended by defining the *Version* entity, which contains specialised attributes like VersionTitle, Duration, etc. The *Version* entity was also extended by other entities

specialised to describe specific versions: the *Video* entity, corresponding to a silent video expression, *Audio/Video* entity, corresponding to an audio/video expression, *Audio* entity, corresponding to an audio only expression, and *Transcript* entity, corresponding to the word spoken in an audio stream.

It is also important to describe separately relevant portions of the entire expression as, for instance, scenes and shots. Scenes and shots of a video are represented by the *Video* entity itself (since portions are just videos). *PartOf* relationships relate composite videos and video components. The same technique is used to represents portions of Audio/Video, Audio and Transcript portions. Other appropriate relationships model the possible links among these expressions such as for example *FollowedBy, HasAudio, HasVideo, HasChannel* and *HasTranscript*. Using these entities and relationships, it is possible to separately describe audio streams and corresponding video streams. It is possible, for instance, to separately keep track of audio streams produced in different languages, associated with the same video stream, and their corresponding transcripts. It is also possible to describe the fact that the same sequence (for instance a video sequence) was used in different documents.

Expressions are associated with features that are extracted from them in order to support similarity retrieval. Typically, these features are extracted automatically by means of image, audio, and text



processing. Examples of features for visual documents are colour histograms, shapes, motion vectors, etc.

Audio video documents can be stored on analog and digital supports. Examples of analog supports are VHS and Betacam, while examples of digital supports are DVD and MPEG files. This was modelled by extending the *Manifestation* entity with the *Media* entity that was further refined by means of the *Digital* and *Analog* sub entities.

In order to relate expressions, corresponding to relevant portions of entire documents to their physical position in the corresponding manifestations, *Digital* and *Analog* sub entities were further refined by defining the sub entities *Analog-Sequence* and *Digital-Sequence*. A relation *SinchronisedWith*, included in the *Media* entity, permits to relate two manifestations that may be played synchronously as, for instance, a movie and its soundtrack, in case they are stored separately.

A library may have several copies of the same manifestation. For instance the same MPEG file may be stored on several video servers, or the library may have available several copies of the same VHS tape. To represent this situation, the *Item* entity was refined by the *Storage* entity, containing attributes for the right access management and ownership, which was further refined by *On-Line* and *Off-Line* entities. The first one represents and describes resources that can be accessed directly by a computer, the second describes resources that are stored off-line, for instance in a shelf of the library. In the first case, it is possible to describe that on-line resources may have different access speed, and the systems where they are stored have different workload. In the second case, it is possible to describe the quality of preservation of different copies.

In designing this metadata model great care was taken to the provision for interoperability. We are currently experimenting the mapping of this model into other metadata models to evaluate its interoperability level. Another aim in designing the model was to render it extensible. This is a key feature of any metadata model since it permits a successive expansion to cover new emerging needs.

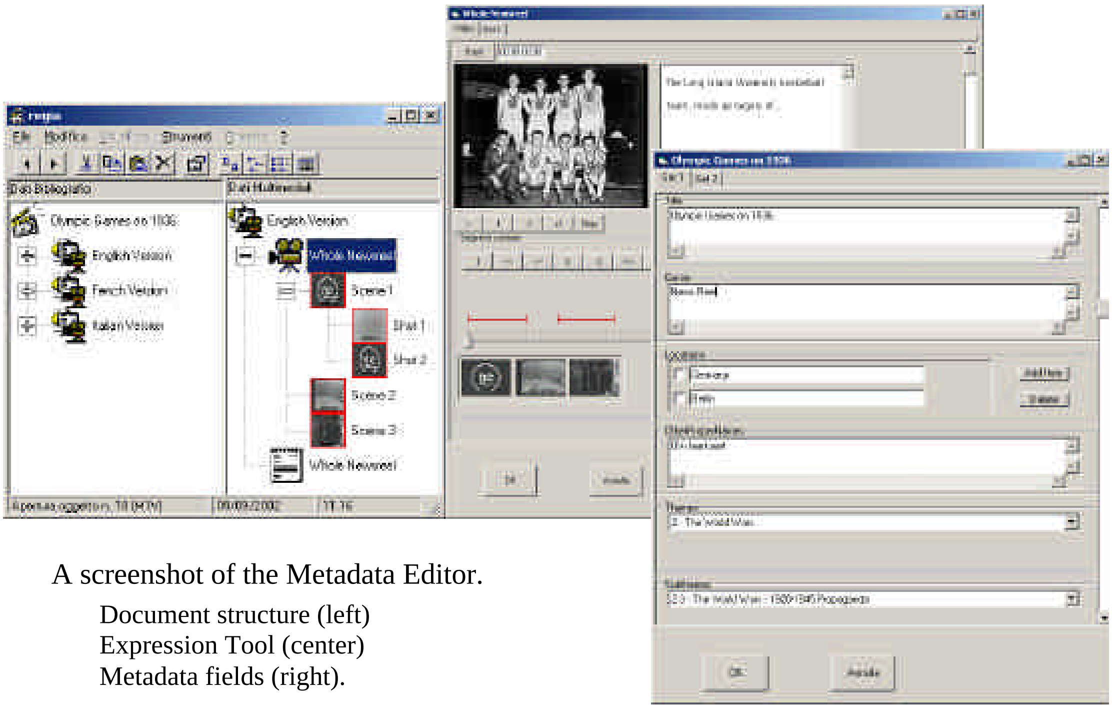
The current version of the model does not cover yet the typical requirements for an efficient “right approach”. Future work will inquire at what degree of detail of rights description a content information system should supply and how: by providing a right issue specialization, or by activating a systematic flux with a right management system.

A10. The Metadata Editor

The Metadata Editor is a client tool used by the cataloguer of the archive for editing the metadata automatically inserted by the indexer. The typical cataloguer workflow is the following:

- A new audiovisual document is digitalized or transformed from one digital format into another;
- The document is archived by the system in the Essence Manager;
- The document is processed for automatic indexing (extraction of scene cuts, speech recognition, etc.);

- When the automatic indexing has been completed, the user is informed by the system and the manual indexing can start;
- The user typically edits the textual description for typos or factual content, reviews or sets values of the metadata fields, adjusts the bounds of the document segments, removes unwonted segments and merges multiple documents. This phase is usually performed starting from the top level of the model (the AVDocument), and continuing by modifying/editing the lower-level objects connected to the AVDocument (i.e., Version, Media and Storage objects).



A screenshot of the Metadata Editor.
Document structure (left)
Expression Tool (center)
Metadata fields (right).

Since the metadata model is relatively complex, the design of the metadata editor is of primary importance.

The interface of the editor is designed in such a way that it is possible to browse the tree structure of an audio/video document. The Figure shows a screenshot of the interface: the window on left side displays a document like a folder navigation tool. On the top level of the tree, there is an icon representing an AVDocument object (the work of the “Olympic Games on 1936” in our example). Connected to the work object the editor presents all the entity instance of the document.

The navigation tool on the left side of the window shows only the main expressions belonging to the documents. The editor allows to browse a single Version one at a time by using a second frame on the right side of the window. In this way it is possible to see the possible Video, Audio and Transcript Versions (at least one of them must exist) of the document and, for each Version, to browse the video segmentations in scenes, shots, etc.

By clicking on the icon corresponding to a metadata object, it is possible to modify, in a separated window, the metadata fields of the object. A particular attention has been paid to the expression window design, i.e, the Expression Tool. The Figure gives an example of the Expression tool interface. Besides the textual fields, the Expression Tool allows the access to the metadata relative to the video segmentation, and allows one to modify them. More precisely, the user can view the video, and read the transcript. The window shows also an overview of the video segmentation, by means of a timeline tools (see the bottom of the Expression Tool window), which represent the video and a the list of the keyframes of the video segments (scene, shots, etc.). The timeline is subdivided into partitions that represent how the video was segmented. By selecting a segment (or the corresponding keyframe), the Expression Tool shows the video corresponding to the segment.

A11. The ECHO digital library system

All the functionalities of the ECHO have been integrated into a digital library system. This system bases upon the media archive® Enterprise Asset Management System. This is a fully distributed archiving system running under WindowsNT and 2000 and relying upon CORBA as middleware.
The base system provides a separate management of audiovisual assets (i.e. audio, video files) and the metadata related to them using SQL databases. Basic functionalities are the ingest and the indexing of a video clips and a full-text search upon the archive's metadata. The system provides client applications e.g. to import video clips and to annotate them. A web based interface provides search and browse functions to access the content.

Modules that have been developed in the ECHO project have been integrated seamlessly into the existing framework. The hierarchical ECHO metadata model has been integrated without any loss of the full-text search performance. Special import tools have been designed to import video assets and metadata into the system.
The object and the face detection modules are integrated as extension of the standard video analysis. The metadata of these are imported into the ECHO metadata model. The video abstracting module works as an extension to the video conversion service – it converts not the format but the length of the video clip. The audio transcripting in Dutch; French German or Italian works remotely using distant speech recognition engines.

Two extra search engine have been integrated
1. the cross-language retrieval provided by EIT and
2. the Image Retrieval Engine provided by CMU and IEI.

The web service has been extended to show all the extended information and to provide convenient navigation facilities. The web interface itself is multilingual it can be configured for 4 languages.

As a result of the project there are two prototypes available now:

- a commercial prototype with all standard functionality of the ECHO project, such as metadata management, video browsing etc. and
- an experimental prototype demonstrating extended features such as face detection and image search.

A12. Speech recognition

Large vocabulary continuous speech recognition is a key technology that can be used to enable content-based information access in audio and video archives. State-of-the-art speech recognition technology was used to produce automatic transcriptions of the historical documents for the French, Italian and Dutch languages, provided by the Echo partners.

The archives consist primarily of radio and television broadcasts and are challenging to transcribe as they contain signal segments of various acoustic and linguistic nature. Two main problems needed to be addressed to deal with the continuous flow of inhomogeneous data. These concern the varied acoustic nature of the signal (signal quality, environmental and transmission noise, music) and different linguistic styles (prepared and spontaneous speech on a wide range of topics, spoken by a large variety of speakers).

The archive documents are different from contemporary broadcast news, which raises several issues for processing them: they are often noisier, the speaking style has evolved, there are epoch-specific proper names that need to be added to the lexicon, and the range of topics is very broad compared to contemporary news.

Training the statistical models of the transcription system requires a large amount of acoustic and text data, representative of the documents to be processed and in electronic format. Finding a sufficient amount of data for this adaptation, especially electronic texts about historical periods, is one of the biggest challenges.

The LIMSI broadcast news transcription system was used to transcribe the French part of the corpus, spanning the latter half of the 20th century. The word error rate with the system designed to transcribe recent data was about 42%. By adapting the models to the acoustic conditions of the Echo data and using additional texts from a French video archive web site (documentary summaries covering the period from 1945 to 1979) reduces the word error rate by about 20% relative.

The ITC-irst broadcast news transcription system was initially applied to transcribe the Italian part of the corpus, consisting of films produced between '30s and '60s of the 20th century. Due to an evident acoustic and linguistic mismatch between the broadcast news and the new application domain, the word error rate was about 46%. By exploiting a limited amount of manually annotated training data from the ECHO domain, adaptation of all components of the transcription system was performed, namely the audio partitioner, the acoustic model, and the language model. The improved system achieved a word error rate of about 30%.

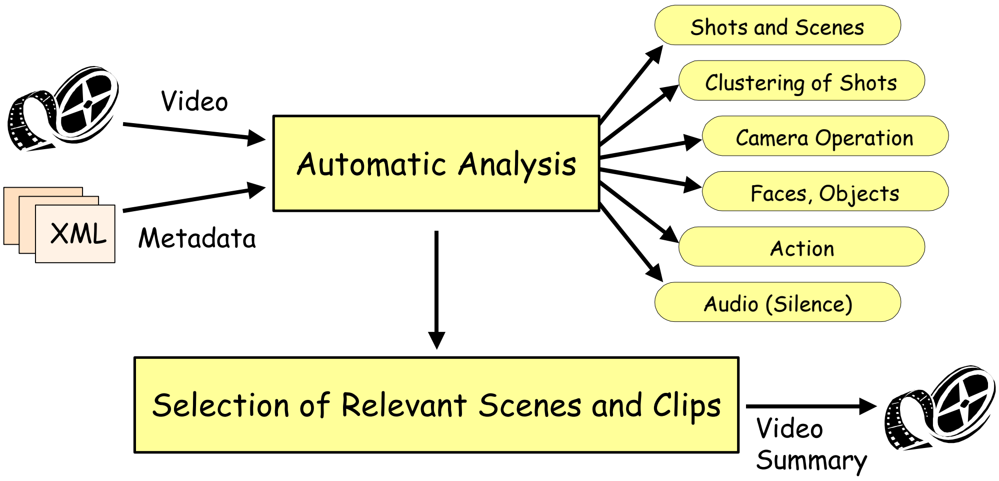
Twente's broadcast news recognition system was applied to he Dutch part of the ECHO corpus, which contains films from the period 1930-1980. Due to the extremely low audio quality, in addition to the linguistic mismatch, initial recognition figures were poor, especially for the older material. Efforts to obtain text data for training language models for outmoded speech by scanning available paper transcripts of some of the material did not succeed due to the low quality of the paper texts (carbon copies), and not enough data was available to apply clustering techniques to train domain specific acoustic models. The focus was therefore put on improvement of the contemporary language models. This resulted in word error rates between between 30% and 40% for more recent ECHO material.

The automatic transcriptions have word error rates in the range of 30 to 40% relative to detailed manual transcriptions in the three languages. However, many of the errors are on function words or are due to lack of agreement (number and gender) which are not important for information retrieval purposes. As demonstrated by document retrieval experiments, the transcriptions are of sufficient accuracy to locate relevant audio segments.

A13. Video summarization

The abstract of an article is a short document summary often used to preselect material relevant to the user. In the age of multimedia it would be desirable to use as video abstracts short clips that contain the essence of a longer video, without any break in the presentation medium. A video abstract or summary is a sequence of moving images that have been extracted from a longer video. Much shorter than the original, the abstract nonetheless preserves its message.

The video summarization application is a component of the ECHO system developed at the University of Mannheim. A summary is created in two steps. The first encompasses the automatic analysis of the video and the calculation of relevant features. If the data of a specific feature are available in the ECHO database, the application uses the information available. The following image depicts the summarization process.



Automatic analysis includes the decomposition of the input video into semantic units (shots and scenes). Semantically rich pieces of information such as frontal faces, moving objects or action scenes are calculated as well.

In the second step, the most important clips are selected. These are assembled into their final form, then the corresponding audio parts are selected. This summary is stored as an MPEG video. Users of the ECHO system may specify the length of the abstract

A14. Indexing of video key frames

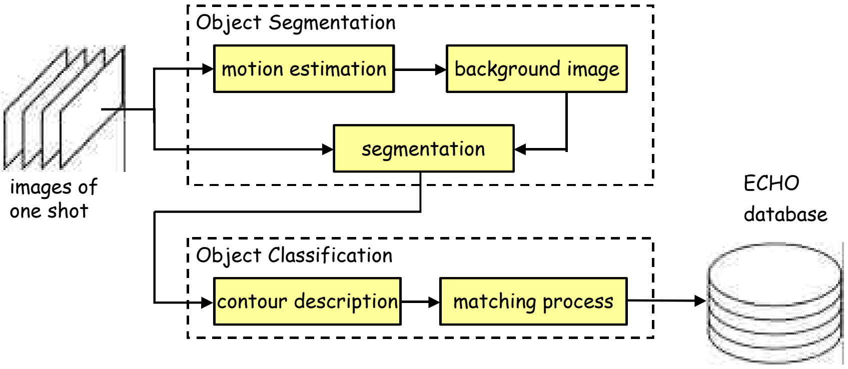
Content-based image retrieval requires an image search engine to find the set of images from a given image collection that is similar to the given query image, where similarity may be very subjective. Most research in image retrieval emphasizes features and invariants of single images. Video matching usually involves comparing two sets of extracted features emphasizing motion and color. In ECHO, a single representative frame from each shot is chosen for story board views and similarity retrieval.

The classification algorithm, which was previously developed at CMU, when applied to the static frames computes the color and texture of the target image. The image is broken into 9 equally sized regions. A 15-bin histogram is taken for the Red, Green, and Blue. Then, six texture histograms of 15 bins each are taken, representing the appearance and "feel" of surfaces. All of these vectors are appended to create a 1215 dimensional vector. This vector is reduced to 50 dimensions by multiplying with a previously computed singular value decomposition, in this case based on prior work with comparable quality video. Each vector is than placed in a tree data structure that allows K-nearest-neighbor searches to produce a ranked set of similar images. The similarity search process is based on the ability of measuring the similarity between two items. This is obtained by defining a *similarity function*. In ECHO we use an approach based on so-called *metric trees* which directly consider only pair-wise distances between objects to recursively partition the search space without considering positions of objects in a multidimensional space - dimensionality of space need not even be known. Furthermore, we have also integrated a search method that supports "approximate similarity search", which provides significant performance improvements at the price of some imprecision in search results.

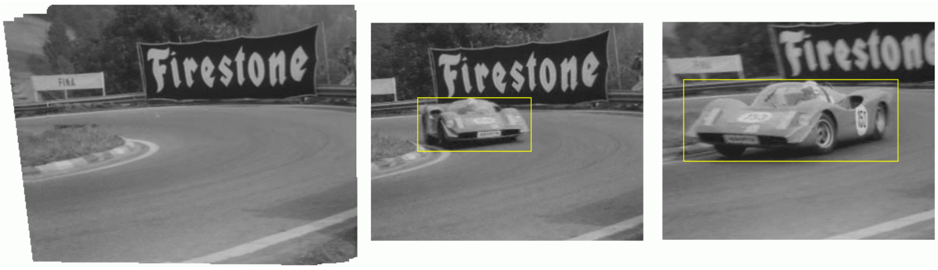
Shortcomings result from the fact that the histograms did not include any spatial information and hence were prone to false positives. They were not well suited for retrieving images in finer granularities, e.g., particular colors or regions. Extended feature extractions that may be included in subsequent versions to improve similarity matching include edge detection, camera motion and motion vectors.

A15. Moving-Object recognition

Moving objects are segmented on the basis of motion cues. Each shot is analysed and the global camera motion (pan, tilt, zoom) calculated.



For each shot, a background panorama image is constructed. The foreground objects in this background image are removed by means of temporal filtering. The object is segmented by comparing each frame of the video to the background image (segmentation).



For each detected object, an efficient shape-based representation is calculated (contour description). This description is then matched to pre-calculated object descriptions. The background image and the recognized objects (cars) are depicted in the following image.

Moving objects are segmented on the basis of motion cues. Each shot is analysed and the global camera motion (pan, tilt, zoom) calculated.

For each shot, a background panorama image is constructed. The foreground objects in this background image are removed by means of temporal filtering. The object is segmented by comparing each frame of the video to the background image (segmentation).

For each detected object, an efficient shape-based representation is calculated (contour description). This description is then matched to pre-calculated object descriptions. The background image and the recognized objects (cars) are depicted in the following image.

A16. Face detection and recognition

Object detection is the problem of automatically locating objects that belong to a category, such as human faces or automobiles, in unconstrained photographic and video imagery. The main challenge is the amount of variation in an objects appearance, e.g., human faces vary widely across persons and images. The pose (geometrical relationship with camera), lighting, shadows, and the properties of the camera and lens all affect appearance, and so the detector must implicitly or explicitly model such variation too. Moreover, the face must be distinguished from any possible visual scenery.

ECHO face detection, previously developed and implemented at CMU, uses a classifier based on a statistical model of appearance to find the face over a limited amount of variation in size, orientation, and alignment. The classifier examines a fixed size image window and makes a binary decision about whether the object is present within specified tolerances in size, orientation, and position. To then be able to detect instances of the face over larger variation in size and position, a detection algorithm scans the classifier over the image exhaustively in position and scale. To detect faces over a wide range of orientations, separate classifiers are trained and scanned for each such orientation.

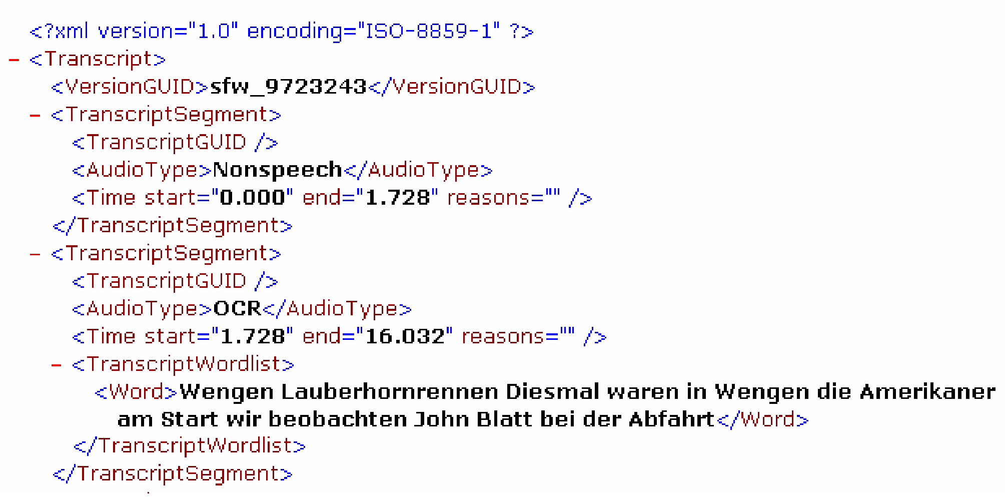
A17. Cross-language information retrieval on audio transcripts

The final ECHO prototype incorporates a text search, which supports search not only on metadata but basically on transcripts. The searchable item will be the entire transcript of a newsreel, the transcript of a single scene, respectively of a shot.
The transcripts will be derived out of the speech part of the newsreel and aligned to the scene-structure of the video. The transcripts will be in Italian, French or Dutch. The transcripts for Swiss newsreels in German language will be derived from written transcripts. As these transcripts are subjected to an OCR process, the quality of text and alignment to scenes and shots will be quite different to the other parts.

A17.1. Document Corpus: «Audio Transcripts»

In the ECHO prototype the a full text search will be performed on newsreels *transcripts*. The most adequate document entity in this collection will be the transcript of a single shot. The transcripts text is aligned to single shots by automatic shot detection. The audio is aligned to the detected shots and saved as text in the ECHO data repository, after automatic audio-text conversion. In other word the searchable item and also the aligned document for our multilingual information retrieval method will be the textual representation of a *single newsreel shot.*
As to the selection of the newsreel according to specific *themes* for ECHO, we will find certain alignment on shot level, although the chosen newsreel do not cover exact the same theme.

Table 1: Transcript in XML



```
  <?xml version="1.0" encoding="ISO-8859-1" ?>
- <Transcript>
    <VersionGUID>sfw_9723243</VersionGUID>
  - <TranscriptSegment>
      <TranscriptGUID />
      <AudioType>Nonspeech</AudioType>
      <Time start="0.000" end="1.728" reasons="" />
    </TranscriptSegment>
  - <TranscriptSegment>
      <TranscriptGUID />
      <AudioType>OCR</AudioType>
      <Time start="1.728" end="16.032" reasons="" />
    - <TranscriptWordlist>
        <Word>Wengen Lauberhornrennen Diesmal waren in Wengen die Amerikaner
          am Start wir beobachten John Blatt bei der Abfahrt</Word>
      </TranscriptWordlist>
    </TranscriptSegment>
```

A17.2. Multilingual Information Retrieval Method

A short introduction to the methods used by Eurospider to achieve a cross-language information retrieval:

In the ECHO Project multilingual information retrieval is applied to audio transcript generated either by speech recognition processing or by OCR processing written transcript material. The segmentation of the searchable item is reduced to single scene, respectively to a single shot. On scene level, the ECHO system can correlate the ranked list of the transcript search with the video material, thus allowing natural language search simultaneously in several predefined languages to retrieve archived newsreel material immediate.

The cross-language retrieval method used can be described in short terms as pseudo relevance feedback on machine translated transcripts. First, the query is run against the machine-translated transcripts in the query language thus obtaining a first ranked list. Machine translating scene/shot level transcripts (rather short documents) produces rather a thesaurus expansion into the target language than a real translation. Second step uses the native language representation of the top documents in the first ranked list for relevance feedback, this time restricting the search only to native language documents; thus obtaining a merge list of all involved language sources.

Additionally to the use of Eurospider’s CLIR search engine another major problem had to be solved. Adjusting speech recognition processed transcript and OCR processed transcript to approximately same level of noise using a lexicon-based approach. This way OCR- and audio transcript can be merge in the ranked list, using the same weighting schema.

A17.3. Document Translation

The OCR/audio generated transcript is machine-translated into all available languages (German/French/Italian/English). We will use for the third prototype a large aligned dictionary for German-French-Italian (TDS/Thesaurus du droit suisse) created in another project. Unfortunately the realm of the TDS dictionary is different, which will lead to a rather “noise” translation. For all pairs with English, we use a simple “off the shelf” machine translation program.

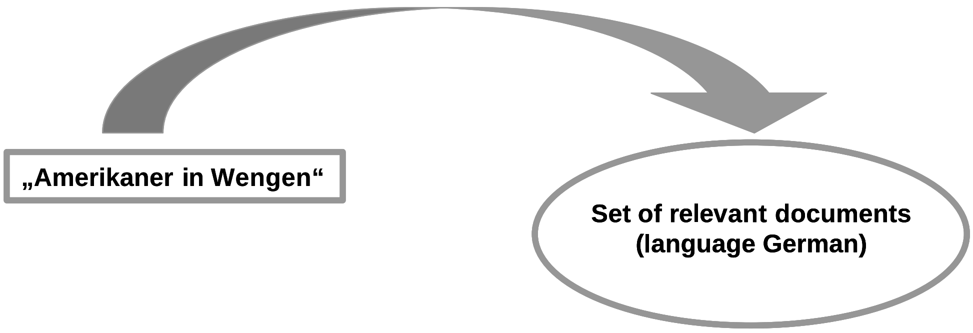
The method using aligned dictionary will also enhance the translated text by synonyms. Rather than machine translation it is an substitution of words by one or more synonyms and then replacing it by the aligned entry in the dictionary. Nevertheless we hope to achieve fair results due to the method of pseudo relevance feedback.

A17.4. Pseudo Relevance Feedback

In the first step is full text search on a subset of transcripts, according to language of the query. The language detection applies the query to the appropriated subset of documents.

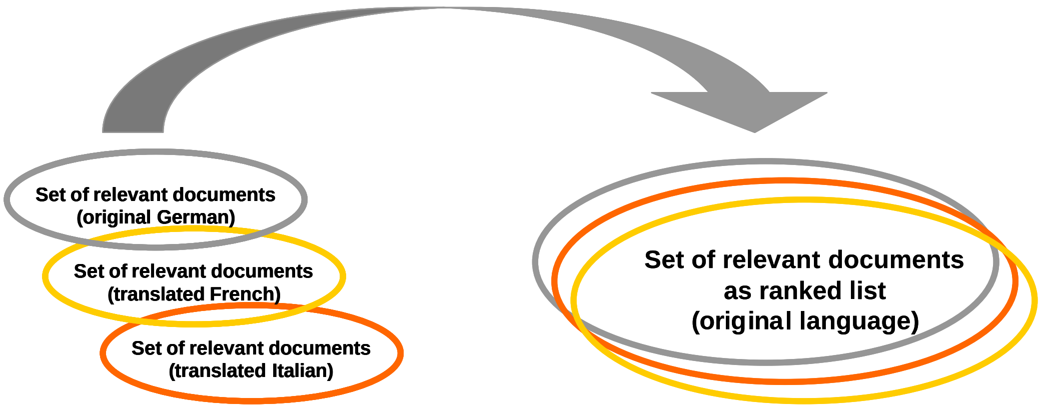
Thus a ranked list in the query language is obtained (mono-language search).

Table 2: Mono-Language Search



In the second step, a set of most relevant documents from the first step is used for the pseudo relevance feedback. Each chosen document is combined with the aligned document in the four other languages, e.g. with there machine translated counterparts. This new set is then used as query on all documents in native language.

Table 3: Pseudo Relevance Feedback


Set of relevant documents (original German)
Set of relevant documents (translated French)
Set of relevant documents (translated Italian)
Set of relevant documents as ranked list (original language)

No (time consuming) machine translating of queries is necessary. The translating is part of the indexing batch-process. The non-trivial problem of merging several mono-language searches is solved with pseudo relevance feedback.

A17.5. Merging the Results of Cross-Language Searches

Merging the results of a mono-language search and of one or several cross-language search(es) is a non-trivial problem for two reasons. First, the Retrieval Status values $RSV_i(q,dj)$ of the different search methods i are on different scales. Second, the numbers of relevant documents in the different languages are unknown. It is possible that the same number of relevant documents exist in each language; but is also possible that all relevant document are in a single language.
To cope with the merging problem we suggest linear transformations of the Retrieval Status Values. For instance, assume that the document d_j has been retrieved by method i because it is in language i, then its retrieval status value $RSV_i(q,d_j)$ is mapped to a common scale $RSV(q,d_j)$ in the following way:

$$RSV(q,d_j) := \alpha_i + \beta_i * RSV_i(q,d_j) \,. \qquad \textbf{(1)}$$

The parameters αi and βi are determined by means of aligned documents and a least square fit which minimizes the sum of the squares of the error of aligned pairs. For instance, assume that dj and dk were aligned because dj covers a story in language h and dk covers the same or a similar story in language i. These two documents obtained the scores RSVh(q,dj) and RSVi(q,dk). Because they were aligned, they should be mapped to similar scores,

$$\alpha_h + \beta_h * RSV_h(q,d_j) \approx \alpha_i + \beta_i * RSV_i(q,d_k) \,, \qquad \textbf{(2)}$$

or in other words: the square of the difference

$$\Delta_{jk}^2 := (\alpha_h + \beta_h * RSV_h(q,d_j) - \alpha_i - \beta_i * RSV_i(q,d_h))^2 \qquad \textbf{(3)}$$

should be minimized, which is achieved by a least square fit. The advantage of this approach is that not only relevant but also irrelevant pairs of aligned documents are used for merging. Of course, non-aligned documents can also be mapped to the common scale using the mappings that were determined by means of the aligned pairs. Results of future experiments will show which percentage of documents have to be aligned in order to accomplish merging the search results in this way.

A18. System evaluation

The purpose of a software product evaluation is to provide any interested party with quantitative results concerning a software product that are comprehensible, acceptable and trustworthy. Existing standards define the features for ensuring software quality. For ECHO, the reference what ISO/IEC 9126, and this defines six features to be taken into account: functionality, reliability, usability, efficiency, maintainability and portability.

Some features concern the software development, others deal with the actual usage of the system. Below, we will go deeper into how the usability of the system was defined.

A18.1. Measuring usability

We need 'usability' in order to achieve system acceptability. Acceptability is most usually thought of as consisting of a number of different factors, in particular the usefulness and ease of use of the system. A consequence of this is satisfactory use of the system and thus acceptance. However, it is important to note that satisfaction is not always a consequence of usability, but ought to be considered a factor in its own right.

These different elements can be defined as follows:

- **Usability:** The extent to which a product can achieve specified goals with effectiveness, efficiency and satisfaction in a specified context of use;
- **Effectiveness:** The accuracy and completeness with which users archive specified goals;
- **Efficiency:** The resources expended in relation to the accuracy and completeness with which users achieve goals;
- **Satisfaction:** the comfort and acceptability of use.

In the ECHO project, measuring points regarding these elements are embedded in different phases of the project. Three prototypes were delivered as part of the development approach. Each prototype extends the functionality of the previous one(s). From an end-user point of view, the most interesting and advanced possibilities of the system were integrated exclusively in the third prototype.

Both the second and third prototype where subject to extensive user testing in two application domains: that of education and entertainment. The tests were conducted on an extensive corpus of audiovisual material, taken from the collections of INA, Sound and Vision, Memoriav and IL. The efforts provided useful feedback for adjustments and further developments of the system at one hand, and more general on the use of advanced searching possibilities on a wider scale.

A18.2. Setting up the test

From an application point of view, the ECHO project developed a digital library service for historical films belonging to large, national, audiovisual archives. The system will allows different user groups to see an event that is recorded and documented in the country of origin and compare it with the way the same event has been documented in other countries.

Apart from the differentiating between different users that use the system, the different purposes for which the system is used also had to be taken into account. The user requirement report that was compiled in the early stages of the project distinguished between the following types of usage for which someone might want to use ECHO:

- Data entry for archiving;
- Retrieval for reference searching, without interest in viewing the content;
- Retrieval and viewing the content by the general public;
- Professional retrieval and viewing the content for scientific study –history or film science-;
- Professional; retrieval, viewing and -non--profit reuse of the content;
- Educational usage.

With this in mind, a test methodology was set up. In test methodology, there is a clear division between the methods that can be deployed and on which features of usability they are able to collect data from. There’s a widely accepted and practised set of usability methods available, each with their specific advantages and disadvantages. After studying the possible testing methodologies, it was decided to use a mixture of conducing interviews, questionnaires and observing the users.

The next stage was to translate these ideas into a more practical programme of work. User-testing follows a more or less standard sequence, leading eventually to a final usability report. These are, consecutively: select users, design test, set up test, run test, analyse results and write the report.

After first three preparatory stages the actual test was conducted following a strict scenario. After a short introduction on the prototype for the whole group of users, users were asked to provide some statistic details, captured by means of a pre test questionnaire. The subject sits at a desk with a computer. Additional seating was provided in the room for the observer(s). The observers are concerned with observing interactions between the subject and the browser. They are responsible for capturing data throughout the test such as task time, error rate, verbal responses and heuristics. They record their findings on specially designed report forms.

The subject is asked to carry out a number of set tasks given to them by the test monitor. They are encouraged to talk through what they are doing during the test to help the observers gather information.

For these tests a number of tasks have been designed to bring into use functions associated with using the application. The aim is to probe the user interface to uncover potential usability problems. The set of tasks has been ordered in a way to expose the subject to the functions and features of the application and assess their responses.

- General task to guide through the system -simple and advanced-;
- Task on national collection -transcript-;
- Task on national collection -boolean search-;
- Task on Theme 4 -also advanced only-;
- Additional time to ‘play’ with the system.

Following the test the subject attends a post-test interview to provide a qualitative assessment on the use of the application. Questionnaires are very often used in usability testing. The questionnaire that needed to be to be filled out is a standard questionnaire for usability assessment in accordance with the ISONORM 9241/10 (developed by Prümper & Anft). Additionally, the subjects were asked some additional questions, in order to complete the measurement of utility and satisfaction of the system. This is a useful source of information and many of the subjects provided information that was not apparent through the task led testing or observation.

The interviews provide a lot of information on qualitative assessment. Finding out what was appealing and easy to use as not normally evident from task testing. Subjects commented more often on what they found difficult during tasks, rather than what was useful or enjoyable.

A18.3. Analysing the results

All the information gathered was analysed and a usability report was compiled.

The objective –quantitative- data gathered from the questionnaires was entered into spreadsheets for tabulating and summarising the data. In the end report, these results are displayed as a whole and broken down for more detail into specific profile user groups. The more subjective data gathered from post-test interviews have been assessed in reference to the list of heuristics appearing earlier in this report. Satisfaction has been measured through the post-test interview and questionnaire.

The results and findings are arranged into their respective sections of Utility for performance measures, Usability for heuristic (human interaction) results and satisfaction for subjective reporting. A number of charts are provided with task achievement data and results from the subjective rating scale in the post-test interview.

B1. Appendix B

This is the task list used during the training sessions

Searching A/V documents

1. Simple guided search - text search-:
   - Select AVDocument, press “Submit”, and select the “Title” field. Type the word „costumi“ in the text search field, then submit.
   - Klick on the document about the visit of Mussolini at “Costumi *del mondo. La romantica cornice della giornata nuziale a Giethorn*”; When has been done? Read the description. How is the maker?

   - On the right side of the screen click “*Version”*, and then click “*Video*”. You see now the so called storyboard, where you find every shot. Chose a keyframe, klick on it.
   - Now the film is started at the position where the keyframe is located.
   - How long is the whole document (in seconds)?

   - Try to retrieve the same document by specifying a more precise query.

   - Try to find some video document with Color attribute=”Color”. Restrict the query selecting IndicationVideoUnit=”Whole”. In this way it will find only the main videos.

   - Try to find some video “mute” video ie, with Silent attribute=”true”.

   - Try to find some AVDocument made on a specified year.

2. Full text search

3. Simple guided search in transcript:
   - Look for documents about medicine in Italy. Go to the Full text interface. Select “TranscriptWorkList”, and type „medicina“
   - Chose a document, then go to the Version
   - Go to AVDocument
   - Go to Version and select “Video” see the storyboard of the Video.
   - Search for the name in the text of the transcript and control if the text corresponds with the query.

4. Simple guided search in the transcript:
   - Type „ferdi kübler“ -famous swiss cycling champion-, then submit.
   - Chose a document, then go to the AVdocument and look at the description in English.
   - Go to the storybord and chose „transcript“ in the box you find on the right side on top.
   - Search for the name in the text of the transcript and control if the text is present.

5. Multilingual search:
   - Imagine you want to retrieve documents on “Daily life” (vita quotidiana), “Work and leisure” (lavoro e tempo libero), “Exhibitions” (esposizioni).
   - Go to the text search and type the word **Exhibitions** on the ThematicKeyword field of the AVDocument.
   - Issue the query, and look at the metadata of the results.
   - Copy the title of one of the documents and issue the query with this title at the video level.
   - Find the document and the video.

6. Free search for a given subject:

- Imagine you have to write a paper on „schools in Europe“ and your source is the ECHO collection. Look for documents that could be helpful for your work. Note the storage-IDs. (school is spelled “scuola” in Italian, “ecole” in French and “schule” in German); try also the English word “School” in the English abstract
- To illustrate your paper you would like to show an anniversary event at a European University. Search for the 100th Anniversary of the ETH (using the english abstract) and look at it.
- Select a picture that could be used in your document and write down the identification of the picture.

7. Free search for a free subject:
    - Search for documents of your choice.
    - And open the editor.

Manual classification of A/V documents

IBM annex tool test

1. Open the (in the same directory where the tool is) palermo.mpg file and wait for the shot detection. Click the tab “shots in the video”.
2. Try the Video Playback window
3. Play a shot then click stop and try to annotate them with:
    - Events
    - Static Scene: background static
    - Objects
    - Keywords
4. Regions
    - check the menu item “Region Annotation”,
    - play and stop the second shot
    - check the word “car”
    - edit the region
5. Lexicon editing
    - Un-check “Region Annotation”
    - Go to the “transportation” and add the child “Bus”
6. Categories editing
7. Saving annotation
8. Saving lexicon
9. Load complete example
10. Generated xml files:
    - palermo.mpeg
    - palermo.mp7.xml
    - palermo.lex.xml

Regia test

Exercise Test1: Update an existent AVDocument

Open the local document test1.xml, perform the following operations and save as new document the result:

1) Add new metadata the root AVDocument. Adding the following fields (the first three fields need to be activated by the menu item “Edit assigned/unassigned flags”):
    - Location
    - Event date
    - Makers
    - Thematic keywords
2) Edit the metadata of the MainVideo. Adding the following fields:
    - IndicationUnit
    - Person
    - Organization

- Silent

3) Close the local document
4) Using the retrieval interface, find and open the document about the “Assistenza all’infanzia…”
5) Try to modify the shots:
    - Merge the first three scenes
    - Select and the delete a group of central scenes
    - Create a new scene inside the new space
    - Modify the shots boundaries of the scene
        i. Using expand key
        ii. Using the timecode
    - Edit the keyframe of the new shot
6) Try the create some shot of the new scene
7) Close the editor and re-open the document. Edit some of the shots and give them a reasonable title.
8) Search some of the word of the title introduced.

Exercise Test2: Creation

Create a new AVDocument from scretch using the video file in Test2 directory. Find and check the document in the Database.

C1. Appendix C

This is the list of organisations that have been contacted in order to promote the course.

Library associations

AAB (Asociación andaluza de bibliotecarios) ANDALUSIA
http://www.aab.es

ABF (Association des bibliothécaires français) FRANCE
http://www.abf.asso.fr

BAD (Associação Portuguesa de Bibliotecários,Arquivistas e Documentalistas) PORTUGAL
http://www.sdum.uminho.pt/bad/

BAILER (British Association for Information and LibraryEducation and Research) UK
http://www.staff.livjm.ac.uk/busjofar/bailer/

BBS (Verband der Bibliotheken und derBibliothekarinnen/Bibliothekare der Schweiz = Associationdes bibliothèques et bibliothécaires suisses) SVIZERLAND
http://www.bbs.ch

BF (Bibliotekarforbundet = Union of Danish Librarians) DENMARK
http://www.bf.dk
Col·legi Oficial de Bibliotecaris-Documentalistesde Catalunya SPAIN-CATALOGNA
http://www.cobdc.org

VDB (Verein Deutscher Bibliothekare) GERMANY
http://www.vdb-online.org
IFLA (International Federation of Library Associations andInstitutions)
http://www.ifla.org

SBS (Svenska bibliotekariesamfundet) SWEDEN
http://www.bibliotekariesamfundet.se

VÖB (Vereinigung österreichischer Bibliothekarinnenund Bibliothekare) AUSTRIA
http://voeb.uibk.ac.at

VVBAD (Vlaamse Verening voor Bibliotheek-, Archief- enDocumentatiewezen) BELGIUM
http://www.vvbad.be

Archives and archive associations

ANAI (Associazione nazionale archivistica italiana)
http://www.anai.org

FIAF (Fédération Internationale des Archivesdu Film = La Federación Internacional de ArchivosFilmicos = The International Federation of Film Archives)
http://www.fiafnet.org/

International Television Association
http://www.itva.org/

On-line Digital libraries magazines

Diario Digital (http://www.diariodigital.pt)

Wissen Digital (http://v.hbi-stuttgart.de/Bibliothek/wd/)

Journal of Digital Information (http://jodi.ecs.soton.ac.uk/)

Transfroming Traditional Libraries (http://www.lib.usf.edu/mdibble/ttl/)

D-Lib Magazine (http://www.dlib.org/)

School Library Journal Online (http://www.slj.com/)

IFLA Journal (http://www.ifla.org/V/iflaj/index.htm)

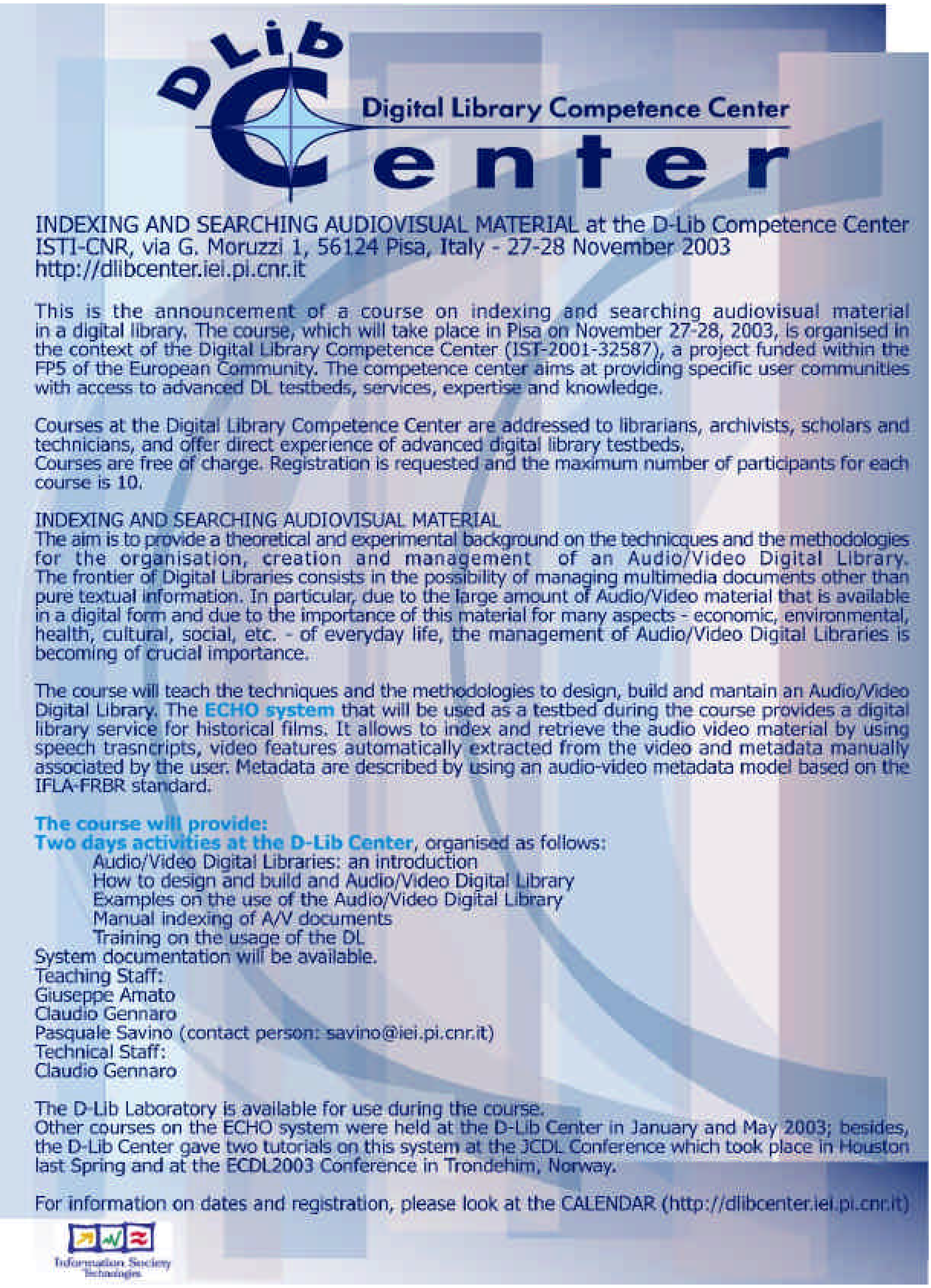
Biblio Tech Review (http://www.biblio-tech.com)

Library Journal (http://libraryjournal.reviewsnews.com/)

Primary Source (http://www.imls.gov/whatsnew/new_imls.htm)

Annex A

This annex contains the flyer used to promote the course



DLib Center

Digital Library Competence Center

INDEXING AND SEARCHING AUDIOVISUAL MATERIAL at the D-Lib Competence Center
ISTI-CNR, via G. Moruzzi 1, 56124 Pisa, Italy - 27-28 November 2003
http://dlibcenter.iei.pi.cnr.it

This is the announcement of a course on indexing and searching audiovisual material in a digital library. The course, which will take place in Pisa on November 27-28, 2003, is organised in the context of the Digital Library Competence Center (IST-2001-32587), a project funded within the FP5 of the European Community. The competence center aims at providing specific user communities with access to advanced DL testbeds, services, expertise and knowledge.

Courses at the Digital Library Competence Center are addressed to librarians, archivists, scholars and technicians, and offer direct experience of advanced digital library testbeds.
Courses are free of charge. Registration is requested and the maximum number of participants for each course is 10.

INDEXING AND SEARCHING AUDIOVISUAL MATERIAL
The aim is to provide a theoretical and experimental background on the technicques and the methodologies for the organisation, creation and management of an Audio/Video Digital Library. The frontier of Digital Libraries consists in the possibility of managing multimedia documents other than pure textual information. In particular, due to the large amount of Audio/Video material that is available in a digital form and due to the importance of this material for many aspects - economic, environmental, health, cultural, social, etc. - of everyday life, the management of Audio/Video Digital Libraries is becoming of crucial importance.

The course will teach the techniques and the methodologies to design, build and mantain an Audio/Video Digital Library. The **ECHO system** that will be used as a testbed during the course provides a digital library service for historical films. It allows to index and retrieve the audio video material by using speech trasncripts, video features automatically extracted from the video and metadata manually associated by the user. Metadata are described by using an audio-video metadata model based on the IFLA-FRBR standard.

The course will provide:
Two days activities at the D-Lib Center, organised as follows:

Audio/Video Digital Libraries: an introduction
How to design and build and Audio/Video Digital Library
Examples on the use of the Audio/Video Digital Library
Manual indexing of A/V documents
Training on the usage of the DL

System documentation will be available.
Teaching Staff:
Giuseppe Amato
Claudio Gennaro
Pasquale Savino (contact person: savino@iei.pi.cnr.it)
Technical Staff:
Claudio Gennaro

The D-Lib Laboratory is available for use during the course.
Other courses on the ECHO system were held at the D-Lib Center in January and May 2003; besides, the D-Lib Center gave two tutorials on this system at the JCDL Conference which took place in Houston last Spring and at the ECDL2003 Conference in Trondehim, Norway.

For information on dates and registration, please look at the CALENDAR (http://dlibcenter.iei.pi.cnr.it)

Information Society Technologies