Audiovisual cultural heritage
From TV and radio archiving to hypermedia publishing

Gwendal Auffret and Bruno Bachimont

Institut National de l'Audiovisuel (INA)**, Direction de la Recherche Prospective
4 Avenue de l'Europe, 94366 Bry-Sur-Marne Cedex
email{gauffret, bbachimont}@ina.fr

Abstract. In this article, we present a model of digital audiovisual (AV) library. We describe how AV library users need to be provided not only with accurate and efficient ways to retrieve images and sounds, but also with new environments allowing to read and interpret these images and sounds as *AV documents*. We show how library users perform an active reading of documents by contextualizing them using corpora of structured meta-information. This documentation consists of documents elaborated from previous readings of this AV content, such as producers' files, critics, etc. It provides a good alternate representation as defined in [34]. We propose a model allowing library users to read AV documents not only along their documentation but *from* their documentation. This model is based on concepts from the electronic publishing world: it defines different levels of editorial control over the semantics, the structure and the layout of documentation and, in the end, allows the automatic generation of hypermedia applications, which we can be used as a new and efficient AV reading environment by library users. We also describe a prototype implementing parts of this model.

1 Introduction

For a long time, books have been considered as the only "real" cultural artefacts, whereas mass media in general, and TV and Radio in particular, were regarded as popular means of leisure, without any real cultural value. But recently, audiovissual (AV) publications have been more and more recognized as part of our cultural heritage. New patrimonial AV libraries are being built, which appear to be quite different from traditional broadcast archives. Users of such libraries are scholars, journalists, school teachers and pupils who perform what we call an "active reading" of AV documents: they read and interpret these publications in order to write and publish their own essays, assignments or articles. This type of

** INA, Institut National de l'Audiovisuel, is the French TV and Radio Archives. It has been archiving French TV since 1949 and French Radio since 1929. Its repository contains more than 3 millions documents, which comes up to 400 000 hours of video and 500 000 hours of audio programs.

reading requires specific means of access to AV documents and to their context of production, publication and reception.

In this article, we propose a model for the creation of such digital AV libraries. First, we provide a definition of AV documents, which distinguishes them from AV streams and AV storage units. Secondly, we analyze the tasks performed by library users on such documents and we show that a digital AV library cannot be limited to a Video On Demand system or an image bank as many projects tend to do. In particular, we show that digital AV libraries can be considered as *large scale publisher of structured documentation*. As a result, we describe how the whole metadata generation process can be organized following four major steps: the inferential consistency step (semantic definition of descriptors), the descriptive consistency step (definition of the description scheme and content indexing), the editorial consistency step (definition of the documentation structure) and the layout consistency step, which takes advantage of document management technology such as style-sheets in order to publish automatically hypermedia presentations from the documentation structure. Finally, we provide an example of this electronic publishing chain and we show how it can be used in AV digital libraries to provide new means of browsing of TV and radio documents to their users.

2 AV document: a definition

We propose to extend the notion of document provided by R. Furuta in [18] in the following definition:

Document : self-contained unit representing an identified intellectual contribution and published on a media for some specific purpose. A document exhibits, to a certain extent, an intentional structure which defines how the elements of its content are organized along its axes[1]. This structure is interpreted by a reader as a testimony of the original publishing purpose.

The above definition raises a question: how can the document unit be identified? This apparently simple question is, indeed, in itself, a crucial problem for TV and radio libraries. If we consider the case of textual libraries, it appears that books, which are editorial units, often delimit also document units: the textual storage unit (*i.e.* the physical exemplary of the book) matches the extent of the intellectual contribution of the author, as well as the extent of the object used by readers to appropriate and interpret the content of the document. Unfortunately, as long as we are dealing with AV content, this matching disappears: a document can be stored on multiple tapes or film reels and these are not immediately readable by human beings, they have to be manipulated mechanically to re-produce images and/or sounds which are shown on a screen, and/or played using an amplifier. From the origin, the temporal nature of AV content imposes specific constraints on reading by separating the following three elements:

[1] Namely time and space for AV documents.

AV stream: an AV stream is any linear temporal succession of sounds and/or images following a specific rhythm that makes it understandable for a reader. This definition covers for instance media such as cinema, TV or radio.

AV storage unit: as AV streams are temporal and continuous, they cannot be stored as such, they require spatial storing devices that will allow the re-creation of a certain piece of the AV stream by the mechanical manipulation of spatially represented information, we call these storing devices "AV storage units". Example of such storage units are film reels, beta SP or VHS video tapes, MIDI or MPEG-1 files, etc.

AV document: an AV document corresponds to a segment of an AV stream testifying of a specific editorial practice, which is stored on one or more AV storage units. For instance, the 8 O'Clock news program of France 2, stored on a half Beta SP tape, can be considered as a document testifying of the editorial practice of this specific broadcaster.

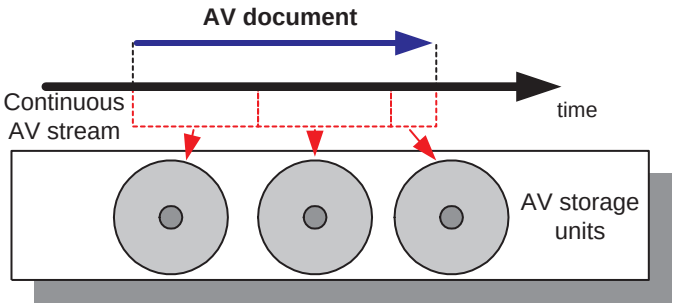

Fig. 1. AV streams, AV storage units and AV documents

These distinctions, which are illustrated on Figure 1, show that AV documents cannot be restricted either to an AV stream (since streams are continuous and we have defined documents as discrete units), or to AV storage units[2]. In the end, it is necessary to take in account the editorial practice in order to abstract

[2] Even if this approach constitutes the underlying model of many research projects today [22],[35], it cannot be accurate in the context of AV libraries. Indeed, AV storage units are technical artefacts, which are changing through time (a damaged film reel can be transcoded on a modern video tape) whereas the documents, themselves, remain unchanged (the 6 O'Clock news program of 1st of July 1957 remains the same document may it be stored on Beta SP or on VHS tapes). Moreover, quite often, the storage unit boundaries are different from the document boundaries (a film is most of the time on multiple film reels) and different storage units can be used to store the same document for different purposes (the archive might store a high bitrate version for long term conservation and a low bitrate version for netcasting). In the end, building a digital AV library on a model that identifies AV documents

the document. Let us consider, for instance, the two repositories represented at INA, the public service broadcasters' archives and the legal deposit library: when describing the same stream (namely French public TV and radio), it appears that they do not consider the same document units and structures. Indeed, the first archive is concerned with what has been published by the producer and it stores tapes on which are the original TV and radio programs created for the broadcasters, whereas the other is concerned by what has been published by the broadcaster [16] and therefore, it stores recordings of the AV stream that has been actually broadcasted, including the advertisement and the potential unexpected interruptions of the stream.

This example shows that there is no such thing as an "objective" (or even commonly accepted) AV document unit and that AV documents are, in fact, constructed by the library from an *interpretation of the original editorial practice*.

3 Requirements on a digital AV library

Once segmented and stored, documents must be described by librarians before being provided to users. Indeed there is no such thing as a "full AV stream search" mechanism to directly retrieve and manipulate elements of the AV content since images and sounds are specific semiotic forms which, contrary to text, do not provide direct access to any discrete and semantically relevant unit. There is no equivalent to symbols, letters and words in AV streams, nothing that would be regarded by a large majority of users as semantic units and which could be used as a basis for search and computation processes.

Many research projects try to address the complex issue of searching and retrieving AV content. In particular, as the current flourishing literature in the field of multimedia shows, providing access to AV content repositories implies not only to be able to count on stable compression and streaming standards such as the ones developed by MPEG [28], but it also implies thorough research work in the field of descriptor extraction [2], database technology [19, 29], server delivery [11], information retrieval methods [13, 35, 17] and user interfaces [14]. These useful and relevant technical answers most of the time have one common goal, namely : provide new and efficient ways to search, select and retrieve moving images and sounds in digital repositories.

Of course, this type of usage is crucial for many people and, in a sense, it corresponds to some of the traditional goals of TV and Radio archives, which were originally intended as a resource repository for producers and broadcasters. However, we claim that restricting the requirements on a digital AV library model to this usage would be merely considering AV libraries a huge Video On Demand (VOD) systems or image banks, which they are not. Indeed, library users are not only looking for images and sounds, they are involved into a certain reading task and they look for AV *documents* which they use as primary source

to AV storage units might seem an easy technical solution but raises many problems on the long term.

material for their own work. This type of AV reading is not passive as the one anyone experiences in front of a TV set, it is an "active reading", *i.e.* a reading activity which leads to the writing of a new document (may it be audio-visual or not). Such an active reading implies the thorough analysis,the contextualization, the interpretation and the rewriting of the document through annotation [33], which, if traditional a relatively easy for text (even if, in the digital domain, new tools are needed, as shown in [12]), remains scarcely developed for images and sounds. There are few tools targeting scholarly research on AV documents as, for instance, the FRANCK system [34].

In order to formalize these user-driven constraints, we provide here a set of requirements which apply to any digital AV library model. In our opinion, in order to provide an efficient service, a digital AV library should allow users to:

1. *search for AV documents*: use efficient library tools (such as catalogs, thesauri or ontologies) or information retrieval methods to look for documents in the repository. This requires a coherent indexing of the content of AV documents;
2. *browse AV documents*: access the content of AV documents and perform a non-linear reading (or viewing/listening) using traditional VCR functions (play, pause, stop, back, forward, etc.) or any mean of direct access allowed by the digital nature of the document;
3. *navigate in AV documents using metadata structures*: access directly the content of AV document using efficient navigational tools such as temporal and spatial summaries and abstracts, tables of content, glossaries, etc. which guide the interpretation and help grasping the overall content of an AV document.
4. *interpret AV documents in context*: access the documentation of the AV document. Indeed, an AV document, as any semiotic production, is never a standalone object, it is always inserted into a communication process implying a production and a reception context. Along this chain, a lot of documents are created as a result of previous readings, which concern the AV document: from the author's project to the critics' articles and the production file, the rights management file, the script, the pictures taken during the shooting, the report of the sound recording session, the original shootings which have not been edited, etc. All these documents constitute a contextual corpus which can be used to guide the interpretation of the content of the document and to contextualize it from diverse points of view;
5. *annotate AV documents* : write down their own interpretation of the AV document using annotations and use these annotations as a mean of browsing and documenting the AV content;

4 Our model: the digital AV library as an hypermedia publisher

To fulfill the above requirements, we propose to apply the model illustrated on Figure 2. This model is based on the hypothesis that the digital AV library

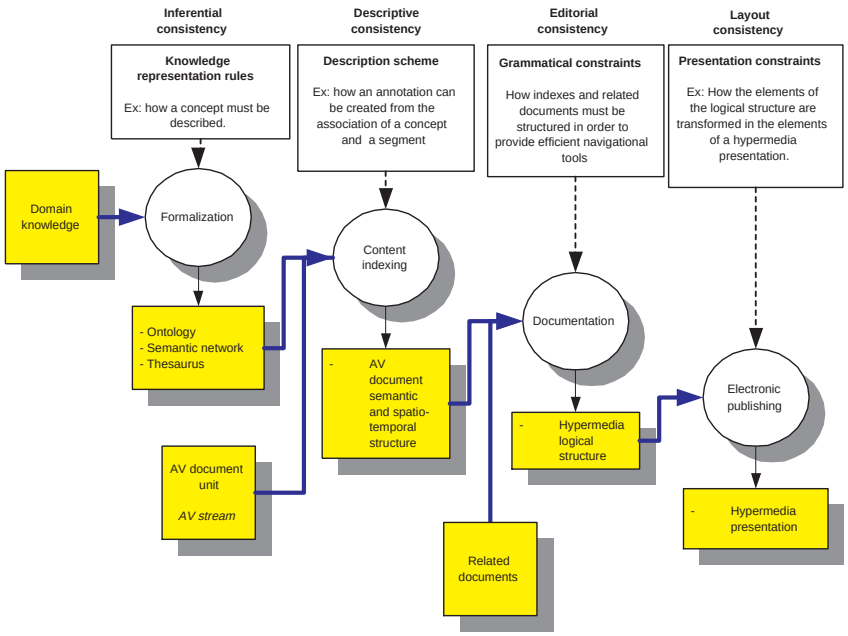

Fig. 2. Life cycle of the documentation

is, in fact, a large scale publishing house generating structured documentation and that this documentation can be used as an efficient mean of hypermedia browsing and contextual interpretation of the AV document content. Our model is composed of a cascading set of constraints and validation steps which allow a multi-level consistency check of the metadata published[3].

First, at the *inferential consistency step*, the library defines the semantics of its descriptors as well as the semantic relations among these descriptors. This step corresponds to the formalization of the domain knowledge. Secondly, at the *descriptive consistency step*, the library defines rules concerning the application of descriptors to spatial and temporal segments of AV documents and applies these rules during the indexing process. Third, at the *editorial consistency step*, the library defines how the indexes should be combined in order to create navigation tools such temporal, spatial or conceptual table of contents of the AV document and how related documents produced alongside the AV document production chain can be gathered and linked to the description as a piece of a hypermedia logical structure. This step corresponds to what we call the *docu-*

[3] Please note that this figure should not be regarded as a chronological representation of the library design process. Indeed, there are many feedback cycles among these different logical steps. For instance, it is impossible to define descriptors at the inferential consistency step without having studied before the user requirements for browsing and search AV documents at the layout consistency step, etc.

mentation of the AV document content. Finally, at the *layout consistency step*, the library defines how the documentation and the AV document must appear on screen as a hypermedia application. This step corresponds to the publishing of the documentation as an AV reading device.

4.1 Inferential consistency step: defining the semantics of descriptors

The first concern of any digital AV archive is to control the semantics of its descriptors. Descriptors are symbols, linguistic terms or even icons [15], representing a coherent set of semantically relevant elements of the content of AV documents. The interpretation of their semantics must be controlled as much as possible, if we want metadata to be interpretable and computable in an consistent and accurate way.

Exactly as the document unit, the semantics of descriptors is an institutional choice: one specific library uses a term in its own specific sense, which might be different from other libraries. This is particularly true of broadcast archives, since the tradition of AV document description as well as the practice of bibliographic exchange in that field is far from being as developed as in the field of text. AV librarians and library users have not yet agreed on a common background such as the tag sets provided by the *Text Encoding Initiative* [10], for example. Each AV library has its own view of AV document content. Therefore, their is only one way to ensure consistency among descriptions and consistent interpretation by users: each AV library must provide clear (and as unambiguous as possible) definitions of the specific semantics of its descriptors. We propose to use an ontology for this purpose. An ontology is traditionally defined as the "specification of a conceptualization" [20], and is used to represent the concepts associated with a domain. In this article, we will follow the definitions provided by [7], i.e. descriptors are linguistic terms. The semantics of these descriptors is defined by the location of the terms in an "is-a" tree. Once the terms used in the domain are defined by experts using this tree structure, they can be used as primitives for any formal representation of the domain knowledge.

We are currently working on the creation of an ontology, which would allow the explicit representation of indexing methodologies used by the community of AV archivers. We plan to express such ontologies using knowledge representation formalisms such as Resource Description Framework (RDF) (See http://www.w3.org/Metadata/RDF), which would then allow us to use inference engines to apply intelligent querying methods on the AV archive repository.

4.2 Descriptive consistency step: expressing description schemes and validating descriptions

Once descriptors are defined, they have to be associated with AV documents. When AV documents were stored in analog format, descriptors were associated with the whole document unit. In a totally digital world, however, it is possible to perform a much more precise indexing and to associate descriptors with segments

of the AV document (a part of a document characterized by spatio-temporal coordinates). This type of indexing is often referred to as *content indexing*. In a first approach, we define an *index* as the association of a specific descriptor with a specific segment, which is quite similar to the notion of Primitive Annotation proposed by Prié & Al [30]. Indexes make relevant substitutes for actual segments of AV documents: users can search, retrieve, manipulate segments by simply manipulating indexes, which are symbolic, and thus easily computable.

In a library, the indexing is mostly created manually by documentalists who interpret the content of documents. As we want the descriptions to be as systematic as possible in order to allow efficient automatic processing, we need to constrain the creation of indexes: we use *description schemes*, which define limitations on the way a specific type of descriptor can be attached to a specific type of segment, limitations on which types of descriptors must or might be instantiated in order to validate the description, and limitations on the spatio-temporal borders of the segments indexed and their spatio-temporal relations. In this sense, a *description scheme* formalizes the indexing policy of the digital AV library with regards to a certain type of AV documents.

Defining structural constraints on indexes The description scheme contains not only the list of the index classes to be used in a specific type of description, but it also defines some constraints on the instantiation of these index classes. These can be divided into two categories:

- Cardinality constraints : constraints on the instantiation and cardinality of specific types of index. For instance, it can be specified that the description of news programs must contain one or more segments indexed by the "Anchor Man" descriptor.
- Axial constraints: AV segments are organized along axis which, for audio-visual material, are temporal and spatial. Description schemes define constraints on the instantiation of indexes along these axis, *i.e.* spatio-temporal constraints. For instance, it can be specified that any segment indexed by the "shot" descriptor should be temporally included into a larger segment indexed by the "scene" descriptor, or that, in the context of a document of type "news", the anchor man "follows" or "overlaps" the titles, etc.

The composition of cardinality and axial constraints defines a spatio-temporal grammar of the AV description which can be used for validation as it is done with SGML/XML Document Type Definitions (DTDs) for textual documents.

Representing formally description schemes and descriptions using XML In [4], we proposed to encode description schemes using an model developed at INA, *Audiovisual Event Description Interface* (AEDI).

Using DTDs to encode descriptions schemes In the context of the ACTS project DICEMAN[4], we encoded AEDI in an XML-based syntax. We extended the XML

[4] http://www.teltec.dcu.ie/diceman/

DTD mechanism to express index types and constraints. This XML expression of our model was simple to implement and it has been a satisfying proof of concept which it has been presented to the MPEG-7 standardization body [6, 1].

However, using DTDs to represent description schemes on AV documents appeared to be restrictive. Indeed, DTDs have been created for the encoding of a limited set of well identified grammatical constraints applying on a unit axis: the linear axis of characters composing a text. The transposition of these contraints into a spatio-temporal grammar proved to be complicated. It forced us to fix the semantics of DTDs in a new domain which was not bijective with the original one. For instance, the validation on temporal axis requires the expression of specific constraints such as Allen relations for instance [3]. On the other hand, the XML DTD did not provide any "&" connector. As a consequence, we had to use the "," connector for separating element types in the content model. But what does this connector mean in a spatio-temporal context? For spatial segment, we can easily imagine that it is a succession constraints, but what does succession mean when it comes to spatial objects? In the end, some DTD rules could not be translated in a spatio-temporal environment and, as a consequence, remained over-constraining, whereas some others could not be expressed and controlled only by DTDs and required a second level of parsing.

Using an XML schema to encode description schemes and description In the end, we decided not to use DTDs and to create our own format for the expression of description schemes. This format is based on an XML serialization of the AEDI model. XML is now just the syntax we use to declare the constraints on descriptions as well as the descriptions themselves.

In description schemes, users define the axis of the document in a coordinate system quite similar to HyTime FCS [25]. Then, they specify the classes of elements which are usable in descriptions such as:

- descriptor classes: descriptors are description elements with a name and attributes;
- axial descriptor classes: descriptor classes with a content model on the axis of the descriptions. The instances of axial descriptors will be attached to segments of the document and constitute the core of the description tree structure. It is possible to specify if an axial descriptor has implied or explicit bounds on its axis, if some of the bounds are inherited from its parent or must be computed from its children's bounds, if it is possible to define an order relation among its children on such or such axis, etc.;
- value containers: attribute-value pairs, where the walue can be a standalone object (ex: title:string), a list or structure of objects (ex: Filmography:film+);

Moreover, description scheme designers can express constraints on the instantiation of axial descriptors. For instance, it is possible to define that children of a specific axial descriptor class should not overlap on such or such axis, or that axial descriptors of class A and axial descriptors of class B should have such or such Allen relation (ex: they should always 'meet') on a certain axis, etc.

Our model for description scheme allows the easy specification of simple structures such as traditional shot, scene and sequence trees (in this case, it is quite similar to an SGML DTD) as well as the specification of very complex n-dimensionnal structures for other uses.

Once the description scheme is defined, descriptions conform to this description scheme can be expressed using our XML schema as a list of empty XML elements related one to the other using links. As a result, elements can be written in the XML file in any order, which was not the case when we used DTDs since we had to identify the linear succession of characters to one axis of the document, namely time, to be able to validate our constraints. In our new format, we are much more independent from the textual constraints of XML.

Once generated, descriptions can be validated against their description scheme using a validating java parser. This parser has already been implemented and should be used for the second phase of the ACTS-DICEMAN project.

4.3 Editorial consistency step

Most of today's projects in the field of digital AV libraries stop their design at the previous step. Once indexes are anchored and metadata structures created using knowledge representation technology, they can be stored in a database and queried by information retrieval and automatic reasoning engines. The result of such a query is a piece of AV stream which can be watched by users. Sometimes, some elements of the meta-information related to the document (such as the title and the author) are provided to users in order to help them contextualizing the images and sounds they are looking at or listening to. However, we claim that AV libraries are more than such image or sound banks inasmuch as they do not only create but also gather and *organize* metadata in order to help the users' reading and interpretation tasks. We call this phase, during which the library creates a logical structure contextualizing the AV document, the *documentation* phase.

Gathering related documents from the production to the reception First an AV document, as any document or any semiotic production, is never a standalone object, it is always inserted into a communication process implying a production context and a reception context [34]. Along this chain, there is a lot of documents created about the AV document, from the author's project to the critics' articles and the production file, the rights management file, the script, the pictures taken during the shooting, the report of the sound recording session, etc. All these documents, which are collected by the AV library, constitute in fact a *contextual corpus* describing the content of the document from diverse points of view. From an interpretation point of view, providing access to such documents is crucial since, thanks to this corpus, twenty years after its creation, scholars can analyze the content of a TV or radio document accurately by referring to its actual context of production and reception [5].

Organizing indexes into navigation tools Indexes created by the library cannot be integrated as such in this contextual corpus. Indeed, as stated above, indexes are just independent pieces of information attached to the content of AV documents, they are used for finding content, but not to read it. When used as reading devices, indexes have to be organized in structures. Our hypothesis is that, similarly tables of contents, glossaries and indexes for texts, it is possible to create specific index structures for AV documents which can be used as efficient navigation tools.

These navigation tools remain mostly to be invented and will certainly evolve with the stabilization of new AV reading usages in digital AV libraries. However, we can already point out some of them which have already been tested and proved to be useful, such as:

- *navigation along one or more coordinate axes*: indexes are related to segments of the document content. These segments are, themselves, locators related to one or more axes following the dimensions of the document (namely mathematical time and space for AV documents). As a consequence, it is useful to provide views on indexes that would be organized along these axes. The traditional example is the temporal view, where the temporal or spatio-temporal indexes are grouped and organized by order of begin time in the AV document.
- *navigation by class of descriptor*: indexes related segments to descriptors. We can therefore create a view of the AV document content that is organized by descriptor class. It is possible to group in the same structure element all the segments indexed by descriptor instances of the same class (*e.g.* collect all the segments indexed by the concept "Anchor Man"). This defines a sort of glossary of the AV document. Moreover, the organization of the descriptors in the ontology can be reused to organize such a glossary as a thesaurus.
- *navigation by projection upon a specific set of indexes*: it is possible to decide that a specific set of indexes is the best segmentation or the best navigation clue in the document and, therefore, to create a structure that would "flatten" the stratification by projecting the different indexes from the different strata on one specific set of segments corresponding to one or more chosen descriptor. With this type of approach, it is possible, for instance, to build a shot-based annotation of the AV document by projecting every descriptor on the segments indexed as shots all the other descriptors.
- *navigation following the structure of another document*: many documents from the contextual corpus have a structured content which, once related to indexes, offers an efficient mean of navigation through the AV document. This is the case for transcripts or commentaries, for instance. It is therefore possible to create specific index structures that would be projected upon the structure of a specific related document from the documentation.
- *creation of video summaries using templates*: the idea here is to gather relevant segments of a corpus and to create a summary based the elements which are supposed to be the more relevant for a specific need. This approach can be based on assumptions on the montage strategy, which are transposed in

automatic tools for the indexing of images as in [26], on a speech recognition process and a recomposition from a keyword search [21], or be totally controlled from a fixed template created by the library as in [27].

We are currently developing an integrated system that would allow us to experiment and assess there efficiency in different reading task as well as to help specific structures emerge, especially for pedagogical uses of AV documents

Documentation or hypermedia logical structure? As more and more AV content is produced directly in digital form and as multimedia metadata standards such as ISO MPEG-7[5], the joint EBU-SMPTE task force[6], W3C Metadata[7] or the CEN/ISSS[8] or DAVIC[9] are emerging, the creation of such metadata structures will become necessary in order to help users find their way in all the AV data available and the different indexes generated automatically or annotated by hand.

Moreover, the different pieces of the contextual corpus might be soon available in digital form and transmitted directly with the AV stream in a standard way, along the production and distribution chain. Once received by the library, it will have to be transformed and adapted to fit into the documentation structure of the institution. As a consequence, in the perspective of a totally digital archive, metadata and data can be linked and stored alongside on the same digital medium in order to be manipulated by computer programs similarly. This means that the traditional separation between metadata and data is disappearing and that, in digital AV archives, *metadata becomes data*.

This might sound trivial if we consider that, in textual libraries, metadata and data have always been expressed using the same semiotic form, text [8], but for AV archives and libraries, this digital convergence is a major move. Indeed, AV documentalists have to consider as one single object things that have traditionally been regarded as radically different. They were previously storing pieces of AV streams in AV storage units and documenting them with text: the AV document was an heterogeneous object which could not be manipulated as a single entity. Its unity remained virtual and, so to say, unreachable. Now that technology allows to store all the elements of the AV document as a single structure on a single digital media, AV librarians discover that they are storing and providing access to composite interrelated networks of images, sounds and texts, namely *hypermedia logical structures*.

[5] MPEG-7 ("Multimedia Content Description"). See http://drogo.cselt.it/mpeg/

[6] The European Broadcasters' Union/ Society of Motion Pictures & Television Engineers task force has created a metadata dictionary which has been published as an international standard in 1998 and should be adopted by the EBU and NATO from 1999 on. See http://www.ebu.ch/

[7] World Wide Web Consortium, see http://www.w3c.org/Metadata/

[8] CEN Information Society Standardization System is currently involved in the Metadata for Multimedia Information initiative (MMI). See. http://www2.echo.lu/oii/en/metadata.html

[9] Digital Audio Visual Council, see http://www.davic.org/

We are using SGML [23] to structure and validate the documentation. Moreover, we integrated some of the concepts and mechanisms from HyTime [25], such as coordinate systems, in our architecture in order to be able to represent the spatio-temporal anchors of the descriptors. In the end, we obtain one single hypermedia logical structure which can be processed for publishing purposes as shown in the next section.

4.4 Layout consistency step

Once all the documentation has been gathered and organized as a SGML encoded hypermedia structure, what can we do to provide our library users with an efficient access to this information?

We propose to generate a hypermedia presentation from this logical structure, which would provide interactive and dynamic means of reading AV documents *from their documentation*. However, we do not think it is reasonable to imagine that a large scale hypermedia application such as the document reading interface of an AV digital library can be generated using traditional hypermedia technology which tend to focus on "one shot" productions such as traditional cultural heritage CR-ROM production [32]. Indeed, these productions are expensive to built. They are often closed and it is very difficult to insert new elements without having to change the whole structure. In a context where new AV documents are being described and documented every day, we have to find another approach.

Our approach is inspired by editorial techniques applied in the electronic publishing industry. Indeed, as described above, the digital AV library applies many constraints and control to ensure that its metadata is consistent and coherent. As a result, we are provided with a highly structured and organized material that can be processed in a systematic way and, in particular, transformed automatically in hypermedia presentations following sets of rules. This is the point of view of Lloyd Rutledge & al [31], who show how SMIL[10] hypermedia presentations can be automatically generated from HyTime structures using DSSSL style-sheets [24]. We are currently processing our SGML structures using tree-transformation scripts and a commercial software. In the future, however, we would like to use norms such as DSSSL to built a complete standard electronic publishing chain which would be independant from the software market.

5 Implementation issues

As a prototype implementation of our model, we have explored the automatic generation of JavaScript applications using an SGML transformation tool. An example of such a presentation running in a client-server environment is provided on Figure 3. This interactive version of a documentary program produced by INA and France 3 has been created in collaboration with the director of the

[10] SMIL (Synchronized Hypermedia Presentation Language) is a W3C recommendation for the specification and the transmission of hypermedia presentations on the web. See http://www.w3.org/TR/REC-smil/

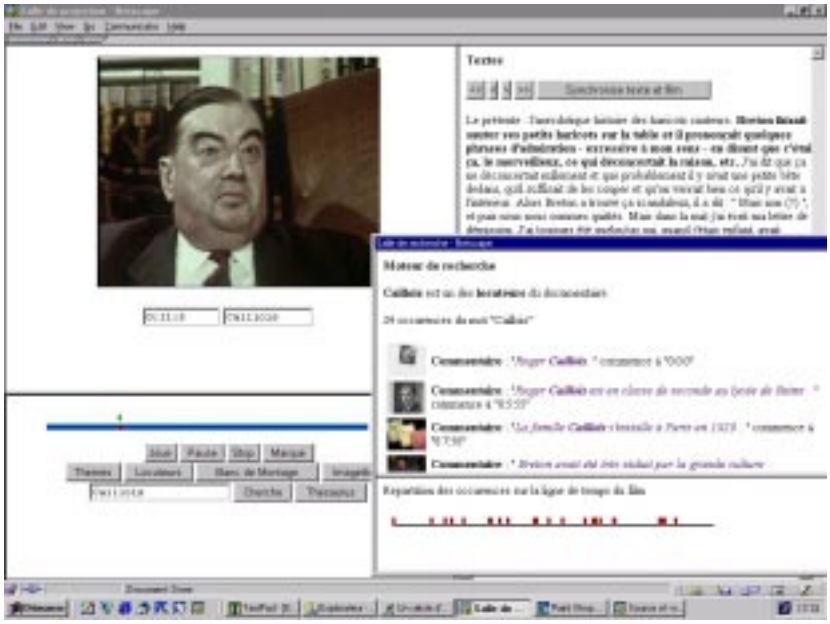

Fig. 3. An automatically generated hypermedia presentation

film. We defined an ontology of descriptors and a strategy of description which was used during the production process. Moreover, we gathered documentation created during the shooting and the editing and we encoded it in SGML along the production chain. In the end, we obtain a hypermedia presentation which, in a limited version (copyrights problems for that kind of production remain huge!) has been web-casted on http://www.ina.fr/Production/Studio/caillois.en.html exactly at the time the TV program was broadcasted on the French public channel France 3.

Users accessing this application can, of course, simply watch the AV stream using traditional VCR functions, but also read the transcript, which is aligned to the timeline. This transcript is linked, sentence by sentence, to the AV stream and can be used as a basis for full text searches or hyperlinks, as it is done in [21].

Moreover, users can access a thesaurus of keywords provided by the archive and by the director of the documentary and look for segments of AV stream indexed by these keywords. They can also look for the interventions of specific locutors and combine keyword searches to select more accurate segments.

The original footage of the interviews have also been made available on-line. When users switch from the final AV document to a footage, the elements which have been selected for the final editing appear in red font in the text. This type of interface provides access to the origins of the document, which is extremely

relevant for scholars working on TV production methods and strategies, for instance.

Finally, it is possible for users to bookmark temporal references by adding an annotation on the timeline of the document. These annotations, as well as all the indexes provided by the archive, can also be used for creating a new user-centered editing: the user selects segments and decides to look at the document only through these particular segments. In such case, the application creates a SMIL document on the fly and provides it to the server which plays only the selected segments.

6 Future Work

A lot of research and implementation work remains to be done before our digital AV library model is complete. In particular, we plan to explore the following issues:

- *Assess the scalability of the model*: our approach has been tested for the moment on a prototype scale and broadcasters were interested by the result and the concept. We still have to assess it on a larger scale. First by developing it from the beginning till the end (a lot remains to be done in the field of the ontology creation for instance) and secondly by testing it thoroughly with the production department on large scale projects;
- *Links between knowledge representation and document structure*: in our model, the inferential consistency step is obviously based on assumptions and models from the knowledge engineering world, whereas the other steps are directly inspired by the electronic publishing community. With the development of the web, these two paradigms become closer and closer: people need to manipulate more easily the semantics of their documents and document management systems become a major target for artificial intelligence technology. However, the combination of semantics and grammar is not as easy as it may seem and there is still a long way to go towards the "Semantic Web" announced by Tim Berners-Lee [9]. As a consequence, we are currently working on a framework to express and manipulate the constraints of the electronic publishing world into the formalisms and representation paradigm of the knowledge engineering world;
- *Multiple points of view presentations*: a hypermedia presentation of a AV document such as the one illustrated on Figure 3 can be described as "video-centric". Most of the presentation is driven by the AV stream, the other types of available data being considered as satellites. In the future, we would like to enlarge this approach by allowing the exploration of the same documentation structure through diverse entry points. In particular, each element linked in the original base can be considered as the focus, or the centre of the browsing. However, browsing a video from a text and browsing a text from a video is not quite the same thing. Therefore, we are working on dynamic style-sheets, which could compute, from the characteristic of the element in focus, the appearing on screen of description elements and of links;

- *Multiple delivery:* we are currently generating hypermedia presentations automatically, but, as we are working with a SGML encoded corpus, nothing forbids to imaging other style sheets that would allow the automatic generation of a paper book or an audio tape at the same time, which could be considered as multiple deliveries on the same AV document structure. We are analyzing this opportunity with INA's production department.
- *User annotations:* even if we integrated some level of annotation functionality in our prototype we would like to create more generic tools allowing users to personalize their browsing of the AV document by adding their own annotations (i.e. their own indexes) on the top of the editorial structure provided by the library. In such a case, they would be able to perform in fact their own editing of the document and can read it from this new point of view.

We are currently working on a project for the use of documented filmed theater for literature courses in high schools and we think this experience will allow us to experiment some of these points and to assess our main hypothesis.

7 Conclusion

In this article, we introduced our model of digital AV libraries. We distinguished four major steps in the design of the AV libraries: the inferential consistency step, the descriptive consistency step, the editorial consistency step and the layout consistency step. At each of these design steps, we showed that the library applies specific control mechanisms on the semantic of its descriptors, the validity of its indexing, the documentation structure and, finally, the way this structure can be processed to generate automatically generation of hypermedia presentations. We provided an example of implementation and we describe how the publishing of such user interfaces by digital AV libraries would allow non linear reading of AV documents *from* their documentation.

In our opinion, the underlying concepts of this model, though developed specifically for the purpose of patrimonial digital AV libraries, can be applied to much larger contexts. Indeed, as long as users of new a repository need precise contextual information for the interpretation of the documents (as in archives dedicated to industrial projects for instance), the type of reading task we are targeting will be necessary and, therefore, the same requirements will apply. Users of such systems cannot be provided only with a video or an audio player and a search engine, they need more structured and interactive presentations which can be created at low cost only by apply thorough control on the indexing and documentation process in order to allow efficient hypermedia publishing.

Acknowledgements

The authors would like to thank Vincent Brunie for his suggestions for improvements to this paper.

References

1. ACTS-DICEMAN. Diceman ddl. Technical report, ISO MPEG contribution, Lancaster Meeting, feb. 1999. Official submission of the ACTS DICEMAN project to the MPEG-7 standardization.
2. P. Aigrain, D. Petkovic, and H.J. Zhang. Content-based representation and retrieval of visual media : a state-of-the-art review. In *Multimedia Tools and Applications special issue on representation and retrieval of visual media*, 1996.
3. J.F. Allen. Maintaining knowledge about temporal intervals. *Communications of the ACM*, 26/11:832–843, 1983.
4. Gwendal Auffret, Jean Carrive, Olivier Chevet, Thomas Dechilly, Rémi Ronfard, and Bruno Bachimont. Audiovisual-based hypermedia authoring: using structured representations for efficient access to av documents. In Klaus Tochtermann, Jörg Westbomke, Uffe Kock Wiil, and John J. Leggett, editors, *ACM Hypertext '99*, pages 169–178, Darmstadt, 1999. ACM Press.
5. Gwendal Auffret and Yannick Prié. Managing full-indexed audiovisual documents: a new perspective for the humanities. *(To appear in) Computer and the Humanities, Special Issue on Digital Images*, 1999.
6. Gwendal Auffret, Rémi Ronfard, and Bruno Bachimont. Proposal for a minimal mpeg-7 ddl for temporal media. Technical report, ISO MPEG contribution 3782, Dublin Meeting, jul., 1998 1998.
7. Bruno Bachimont. Mpeg-7 and ontologies: an editorial perspective. In *Virtual Systems and Multimedia 98*, Gifu, Japan, 1998.
8. Tim Berners-Lee. Axioms of web architecture: 3) metadata architecture. Technical report, W3C, January 6 1997.
9. Tim Berners-Lee. Challenges of the second decade. In *WWW Conference*, Toronto, 1999.
10. Lou Burnard and C.M. Sperberg-McQueen. *TEI P3 : Guidelines for Electronic Text Encoding for Interchange*. Lou Burnard and C.M. Sperberg-McQueen, 1994.
11. K. Selçuk Candan, E. Hwang, and V. S. Subrahmanian. An event-based model for continuous media data on heterogeneous disk server. *Multimedia systems*, 6:251–270, 1998.
12. François Chahuneau, Christophe Lécluse, Bernard Stiegler, and Jacques Virbel. Prototyping the ultimate tool for scholarly qualitative research on texts. In *8th Annual conference of the UW Centre for the New Oxford English Dictionary and Text Research*, Waterloo, 1992.
13. Shih-Fu Chang, John R. Smith, Mandis Beigi, and Ana Benitez. Visual information retrieval from large distributed online repositories. *Communications of the ACM*, 40/12:63–71, 1997.
14. M. G. Christel, David B. Winckler, and Roy C. Taylor. Improving access to a digital video library. In *INTERACT '97, 6th IFIP Conference on Human-Computer Interaction*, Sydney, Australia, 1997.
15. M. Davis. Media streams : an iconic visual language for video annotation. In *IEEE Symposium on visual languages*, pages 196–203, 1993.
16. Francis Denel, Genevieve Piéjut, and Jean-Michel Rodes. *Le dépôt légal de la radio et de la télévision*. INA-publications, Paris, 1994.
17. Jonathan Foote. An overview of audio information retrieval. *Multimedia Systems*, 7:2–10, 1999.
18. Rick Furuta. What can digital libraries teach us about hypertext? *SIGLINK newsletter*, 6/3:7–9, 1997.

19. W. I. Grosky. Managing multimedia information in database system. *Communications of the ACM*, 40(12):73–80, 1997.
20. Nicola Guarino. Formal ontology, conceptual analysis and knowledge representation. *International Journal of Human-Computer Studies*, 43(5/6):625–640, 1995.
21. A.G. Hauptmann and M.A.Smith. Text, speech, and vision for video segmentation : the informedia project. Technical report, Carnegie Mellon University, 1995.
22. Jane Hunter and Renato Iannella. The application of metadata standards to video indexing. In *Second European Conference on Research and Advanced Technology for Digital Libraries*, Crete, Greece, 1998.
23. ISO. *ISO 8879:1986, Information processing - Text and office systems - Standard Generalized Markup Language (SGML)*. ISO, Geneva, 1986.
24. ISO. *ISO/IEC IS 10179:1996 Document Style Semantics and Specification Language (DSSSL)*. ISO, Geneva, 1996.
25. ISO. *ISO/IEC IS 10744:1997, Hypermedia/Time based Structuring Language (HyTime) AFDR Meta-DTD*. ISO, Geneva, 1997.
26. Rainer Leinhart, Silvia Pfeiffer, and Wolfgang Effelsberg. Video abstracting. *Communications of the ACM*, 40/12:55–62, 1997.
27. Craig A. Lindley and Anne-Marie Vercroustre. Intelligent video synthesis using virtual video prescriptions. In *International Conference on Computational Intelligence and Multimedia Applications*, Churchill, Victoria, 1997.
28. J. Mitchell, W.Pennebaker, C. Foog, and D. Le Gall. *MPEG Video Compression Standard*. Chapman and Hall, New York, 1996.
29. E. Oomoto and K. Tanaka. Ovid: design and implementation of a video-object database system. *IEEE Transactions on Knowledge and Data Engineering*, 5(4):629–643, 1993.
30. Y. Prié, A. Mille, and J. Pinon. Ai-strata: A user-centered model for content-based description and retrieval of audiovisual sequences. In Springer-Verlag To appear in LNCS, editor, *First Int. Advanced Multimedia Content Processing Conf.*, Osaka, 1998.
31. Lloyd Rutledge, Lynda Hardman, Jacco van Ossenbruggen, and Dick C. A. Bulterman. Practical applications of existing hypermedia standards. In *ACM Digital Libraries '98*, pages 191–198. ACM Press, 1998.
32. Peter S. Samis. The evolving state of the art cd-rom: the national museum of american art and les impressionnistes. *Archives and Museum Informatics*, 12(1):3–16, 1998.
33. Jacques Virbel. La performative textuelle. In *Le texte en mouvement*. Presses Universitaires de Vincennes, 1987. In French.
34. Ken Yap, Bill Simpson-Young, and Uma Srinivasan. Enhancing video navigation with existing alternate representations. In *International Workshop on Image Databases and Multi Media Search*, Amsterdam, 1996.
35. B.L. Yeo and M.M. Yeung. Retrieving and visualizing video. *Communications of the ACM*, 40(12):43–52, 1997.