

100 Million Hours of Audiovisual Content: Digital Preservation and Access in the *PrestoPRIME* Project

Matthew Addis
IT Innovation
2 Venture Road
Southampton, UK
mja@it-innovation.soton.ac.uk

Laurent Boch
RAI CRIT
Corso Giambone, 68
Torino, Italy
l.boch@rai.it

Walter Allasia
EURIX R&D
Via Carcano, 26
Torino, Italy
allasia@eurix.it

Francesco Gallo
EURIX R&D
Via Carcano, 26
Torino, Italy
gallo@eurix.it

Werner Bailer
JOANNEUM RESEARCH
Steyrergasse 17
Graz, Austria
werner.bailer@joanneum.at

Richard Wright
BBC R&D
56 Wood Lane
London W12 7SB, UK
richard.wright@bbc.co.uk

ABSTRACT

We report the preliminary results of PrestoPRIME, an EU FP7 integrated project, including audiovisual (AV) archives, academics and industrial partners, focused on long-term digital preservation of AV media objects and on ways to increase access by integrating media archives with European on-line digital libraries, specifically Europeana. Project outcomes will result in tools and services to ensure the permanence of digital AV content in archives, libraries, museums and other collections, enabling long-term future access in dynamically changing contexts. PrestoPRIME has a special focus on digital preservation in broadcast environments, where very large files of digital video must be preserved at high quality (suitable for future re-use in an AV production environment) in affordable distributed and federated archives. The adoption of standard solutions for digital preservation processes (metadata representation, content storage, digital rights government, search and access) enables the interoperability of the proposed preservation framework and guidelines. OAIS model was chosen for the reference architecture, METS is adopted as wrapper for metadata representation, while relevant standards (e.g. W3C, ISO/IEC and others) are used for content and rights description. Project outcomes will be delivered through a European networked Competence Centre, to gather knowledge and deliver advanced digital preservation advice and services in conjunction with Europeana and other initiatives.

Categories and Subject Descriptors

H.3.7 [Information Storage And Retrieval]: Digital Libraries; K.4.1 [Computers And Society]: Public Policy Issues—*Intellectual Property Rights*; H.5.1 [Information Interfaces And Presentation]: Multimedia Information Systems—*video*

General Terms

Algorithms, Design, Standardization

Keywords

Digital Preservation, Long-term Preservation Framework, Broadcast, Audiovisual archives

1. INTRODUCTION

There are millions of hours of content in collections of dedicated AV archives, other archives, libraries and museums. The responsible staff has decades of experience in dealing with physical carriers: visible, held in the hand, stored on shelves, played by dedicated devices. Now handling AV content is undergoing a dramatic change, being migrated to files: invisible, stored “in the cloud” and played by unknown technology at a remote machine. How can these collections be efficiently managed, so that files don’t get lost (metadata, file and storage management), do keep their correct relationships with other files (provenance), do maintain their technical (quality), legal (rights) and archival (quality, provenance, rights, metadata) integrity? How can AV content be preserved forever (without going broken almost immediately)? And finally, this all has to be done at lowest possible cost. The purpose of PrestoPRIME is to supply answers, and tools that implement those answers.

The paper is organized as in the following. The *context* is fully presented in Section 2, followed by a description of the *preservation framework* in Section 3. More details on the *architecture and technologies* are given in Section 4. Section 5 is specifically about the issue of *AV rights management*, while Section 6 focus on the consideration on *standards and best practices* and how this is related with the PrestoPRIME work.

2. BACKGROUND

In this Section we first present an overview of the objectives and tasks of PrestoPRIME and a brief description of the covered scenarios about institutions and organisations as holders of AV collections; afterwards a description of relevant features of AV content to be preserved and the impact in case of failure or omission is provided; finally the further complications for archive management coming from the dynamic AV content lifecycle are recalled.

2.1 Overview of the PrestoPRIME project

The PrestoPRIME project addresses the objective of keeping AV contents alive within the digital domain. The prominent requirement is to provide means for implementing a

long-term digital preservation of AV files. Once the preservation is ensured, it is necessary to guarantee to future users the access to archived items for simple fruition or for re-use as part of new multimedia products. The future access is the final goal of today preservation activities.

The digital preservation process for AV material is expensive (compared to the preservation of other content types) and each action performed on the content can affect irrevocably the future access to the content. In order to optimise digital preservation process, a crucial task for PrestoPRIME is the identification of the possible models and strategies to be adopted, according to various contexts and conditions. For each envisageable strategy, the appropriate set of tools are considered, developed, and tested. In addition to the traditional methods based on file format migration, PrestoPRIME is also investigating a multivalent approach [17]. Other important topics for PrestoPRIME are also quality assessment and storage management.

Regarding the access and the metadata, the descriptive information must be kept up-to-date with respects to changing contexts, both by making interoperable the various annotation schemes adopted along time and by involving the users in a continuous process of enrichment. The ability to track the re-use of material, which specific tools, will support various preservation activities, such as identification of best-quality copy and solution of some exploitation rights dilemmas. The ability of handling rights is required for optimising the use of archival items, through which the preservation process is partially funded. This area, often neglected, is also addressed by the project.

PrestoPRIME will not deliver only a set of tools and technologies, but also a complete architecture and framework for integrating all the components, after a process of assessment and validation. The same architecture will be adoptable for both open-source implementations and commercial systems. The results achieved by the project will also exploit the outcomes of other projects and initiatives relevant to the digital preservation, as described in the following. It is worthwhile mentioning that the OAIS model [6] has been chosen as the reference for the architecture design and in the following we assume that the reader is familiar with OAIS terminology and main concepts.

2.2 Scenarios regarding content holders

The content holders organisations addressed by PrestoPRIME are generally those holding AV collections, but the most relevant scenarios which have been identified are related to broadcast archives, Higher Education Institutions and small archives, as described in the following.

Broadcast archives, such as BBC, ORF, RAI, B&G and INA (which are all partners of PrestoPRIME), own large collections of AV material in SDTV and HDTV formats. Their business model includes the roles of content creators, distributors (to the users) and archives. The archive is a relevant possible source of materials for new productions and publications, where archival items are expected to be used in the short, medium and long term period. *Higher Education Institutions*, such as Universities or other research centers, are not strictly focused on AV contents, but the AV content

is more and more related with their activities. The AV collections are often very large and the need to deal with AV preservation requires the acquisition of new technical skills. *Small archives* experience many problems in reaching the critical size for setting up an affordable digital archive process for their assets, therefore they have to rely on services provided by other organisations.

The content holders mentioned above require different approaches to the digital preservation. PrestoPRIME devoted much effort to the dissemination of its results, organizing public events and collecting information also with online surveys, in order to identify which are the most relevant issues and requirements that such content holders have to face with. The results produced by the project will be publicly available as soon as they are finalized, in the form of documents, guidelines and software tools. The aim of PrestoPRIME is also to create a European competence center which will continue beyond the project lifetime and will provide the outcomes of the project, creating a network for all actors involved in the digital preservation.

2.3 AV content in PrestoPRIME context

PrestoPRIME aims to deal with all kinds of AV content, including content originating from specific AV archives (such as broadcast or film archives) as well as AV holdings of libraries and museums. The preservation platform provided by the project (see Section 4) will be accessible by these users and also by other software systems such as other OAIS compliant archives and service providers. An important issue is also the interoperability with existing initiatives from in the cultural heritage community, most prominently Europeana.

Concerning content access, PrestoPRIME adopted a standard approach based on the OAI PMH protocol and MPEG Query Format (MPQF) [7], in order to enable the *query-by-content paradigm*, which has been experimented by PrestoPRIME partners in previous EU projects, such as SAPIR¹.

Broadcast archives have millions hours of AV material (almost not yet digitised) and the amount of content is growing very fastly. Content size depends on compression schemes and related bit rate. This must be considered in relation with the expected quality level. *Master quality* level is good for production purposes and any lower quality can be derived from a master level copy. Reference values for the content size can be given in tens of GB per hour (for example SDTV production is typically at 30GB/hour, HDTV premium production requires quite more). *Exploitation quality* level is good for current commercial purposes and simple production processes, such as news programmes. Archives are expected to work for dissemination at the exploitation level, where size is much lower than for master (ratio can be even 1:10). An other level is the *Broadcast/Publication quality*, which depends on the publication medium and whose size is lower than Exploitation level. Finally the *Browsing quality* level is enough to search, identify and select the content, since the size is negligible compared to the Master level, but it is not useful for user consumption.

The production quality (Master quality level), when recorded

¹<http://www.sapir.eu>

on digital AV-media files, requires a large amount of storage capacity which can be limited only by the adoption of complex compression techniques, which might impact on the quality of the material (lossy compression) or not (lossless compression). Typical values for uncompressed video, for instance, require about 80 GB per hour at standard definition, compared to 30 GB per hour required for a very good compression format. The migration of AV content from one format to another, when dealing with uncompressed videos, is a critical issue. In the long term, even guaranteeing access to AV archive material in a compressed format will only be possible if the required tools, decoders and players, will still be available together with the technical environment which they require for running correctly. PrestoPRIME is deeply investigating these problems, evaluating several approaches, such as the Multivalent approach, which has been experimented in other EU projects, such as SHAMAN [17].

The metadata models used in AV archives focus on both technical and descriptive metadata. The choice of the metadata model often is not determined by the type of content being described, but by the collection holding it, e.g. a broadcast archive will typically use a format designed for AV content, such as EBU Tech3295 [12], while AV content held by a library or museum is often treated like other types of items in the collection. While the models coming from the domain of AV archives are in most cases ready to hold also detailed descriptive metadata, they lack in most cases support for preservation metadata (as can be found in standards from the library community, such as PREMIS [13]) and provenance information.

Preserving digital AV content means keeping it accessible and usable for a long time, including the metadata describing it. This requires migrating metadata to models currently in use and enabling metadata mapping to formats used in the systems which need to retrieve, deliver, display and process archived content. Many current approaches for publishing and mapping metadata are highly by mapping to models that represent least common denominators of the models used in the related collections. New ways of dealing with metadata, using approaches based on emerging Semantic Web technologies and Linked Data could overcome these limitations. Technical and descriptive metadata must include also information related to the preservation plan within the archive. In addition users can add annotations to the content available in the system and user generated metadata are fed back and enrich the existing documentation of the content. This field is also under study in PrestoPRIME, an example for such a user annotation tool is the *Waisda?*² tagging game² developed by project partners.

Finally, AV content is characterised by the time dimension. Playing AV is a presentation of content along its timeline and also editing, seeking a specific point, describing and indexing segments and in general any action or process on AV material is affected by the need to manage the timeline properly. PrestoPRIME will provide a solution for the representation of temporal information based on relevant standards, such as W3C Media Fragments [22] or MPEG-21 DIDL [5]. In particular PrestoPRIME investigated the relevance of temporal

decomposition related to digital rights, where each segments of a given content can be associated to a different license.

2.4 The impact of content loss

The value of lost content is often difficult to estimate. The AV content records the memory of time and places, the voice and the images of people, but content created in the past cannot be re-produced again, unless in fiction terms. The peculiarity of the AV document is also relevant, because there may be lots of copies or versions or other contents which may replace the lost material or maybe not. Additionally the production of new content is generally more expensive than re-using pre-existing material. If appropriate content is not found in the archive, there is a clear impact on the production costs. The proliferation of means of publication of multimedia is economically possible only because of reuse of archive material.

Technical quality is also a value. If the highest quality material is lost and only a lower quality copy is available, it can sometimes be enough, but the future fruition of poor quality material will have great impact, especially for the professional uses. The presentation technologies, i.e. screens and loudspeakers, are going to change and improve ever and ever, therefore preserved low quality material will be at serious risk of appearing old and technically not adequate. This would be a waste for archive and preservation efforts.

2.5 Content lifecycle

The mass digitisation of analogue archive holdings plus the transition to tapeless production for new content means AV archives inevitably face the prospect of adopting file-based solutions using IT storage technology. Audiovisual archives are typically very active places as a result of continual content acquisition, access and reuse. AV archives are increasingly becoming directly integrated into wider content production and consumption processes, yet somehow need to achieve content permanence in world where almost everything seems to be very much transient and concerned with the 'here and now' and not the future. Long-term safe storage of assets for ten, fifty or even a one hundred years is very difficult to achieve in an environment where little stays the same for more than a few years. This dynamic world brings with it the need for archives to accommodate AV content in a wide range of contemporary formats, both for submission and access, and to integrate a wide range of supporting technologies (storage, coding, wrapping, distribution, quality control etc.). All of these tend to have relatively short lifespans and hence migration to deal with technical obsolescence becomes a way of life. Likewise, the access and use of the AV content often varies during its lifecycle, often unpredictably and with unforeseen 'step changes', e.g. as a result of a move to public access. This adds further complications to archive management.

3. PRESERVATION FRAMEWORK

In PrestoPRIME the digital preservation will be extended in order to take into account all the digital contents plus those provided by Broadcasters, which are bringing something new in the field of digital preservation: high quality videos, which requires innovative approaches for monitoring the status, storing and managing the obsolescence. We are facing the

²<http://www.waisda.nl/>

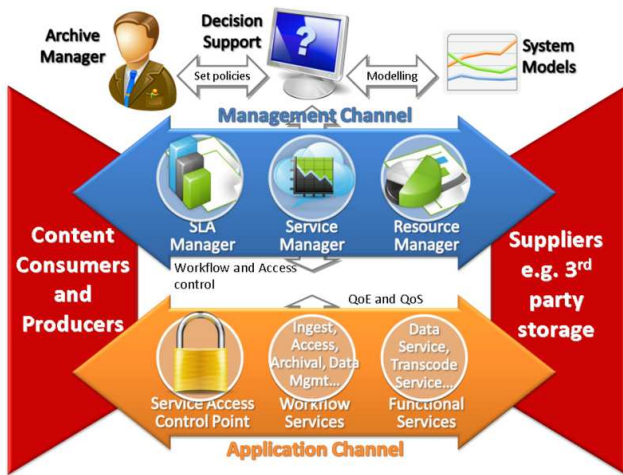


Figure 1: PrestoPRIME architecture layers

challenge of dealing with compressed/uncompressed files, audio/video format changes and the big size files of the high quality / master copy version.

PrestoPRIME is developing tools and techniques to support the transition AV archives are making to file-based content and IT systems. Our technology developments include tools for calculating the long-term Total Cost of Ownership (TCO) of these systems, planning and optimising file format migration, assessing and comparing what storage technologies or services to use, managing the risks involved, selecting suitable preservation strategies (e.g. file format migration, emulation, multivalent), and automating content quality control, e.g. defect detection and analysis.

Recognising that automation is the key to lowering costs (and risks), PrestoPRIME is developing supporting digital preservation infrastructure components that go beyond the OAIS functional model and provide automation and execution of preservation policies and processes, e.g. to make use of managed storage as a service that conforms to defined Service Level Agreements and QoS.

The architecture of the preservation framework is made up of three layers, which are described in the following with reference to Figure 1: application, automated management and manual management layers.

The *application* layer (application channel) at the bottom contains the services that deliver preservation and access, e.g. the tools and services that would be found within the main functional areas of OAIS. By considering the application channel as a set of services, each of which has an SLA and defined QoS, then allows them all to be governed in a consistent way. The *automated management* layer in the middle (management channel) automates the management of the services and also of the customer and supplier relationships. The SLA manager deals with customer SLAs (those of the consumer and producer) which set out the constraints and service level objectives (SLOs) on ingest and access. The Resource manager deals with supplier SLAs (such as out-sourced storage or compute facilities) and with in-house

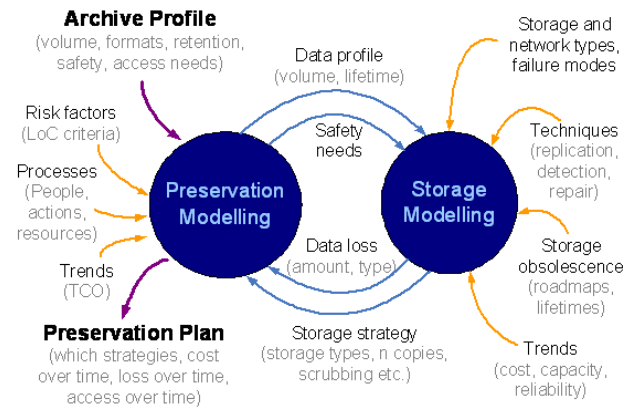


Figure 2: PrestoPRIME preservation modelling

resources. The Service manager balances commitments to customers with resources available internally and from external suppliers. This is an event-decision-action loop, where the decision is made according to a policy. The *manual management* layer at the top (decision support) is where people design, test and set the policies to be executed by automatic management layer. The decision support tools use models of the services which may be updated by real-world experience. This layer is where "intelligence" can be provided by a combination of automatic and manual analysis supported by system modelling tools. The output of this analysis is the management policies. The service manager in the management channel uses these to decide on which action to take to meet its commitments, to plan to continue meeting them, and to mitigate the effects of events (e.g. failures) that cause it to stop meeting them.

The key feature of the PrestoPRIME governance architecture is the ability to decouple the management and the services as much as possible, e.g. through local autonomy where services understand the parts of the policies (e.g. security) relevant to them, therefore being able to make their own immediate decisions. The monitoring and management loop (between the application and management channels) can then be asynchronous with the service manager requesting usage reports (either queued or instantly generated) from the services periodically (as a pull) and then updating the services' policies as necessary in a slower time-frame.

The decision support tools are essentially about preservation planning with an emphasis on risk management and automation of preservation actions, i.e. policy definition. As shown below, the objective is to convert archive's needs (how much content, how long to keep it, how safe it needs to be, and who needs to access it) into a preservation plan (what to do, when to do it, what the consequences will be). The tools combine preservation modelling (e.g. file level preservation approaches) with storage modelling (bit level preservation approaches) and considers the interplay between these two (e.g. choice of file format impacts on storage needs as well as sensitivity to corruption in the storage layer).

The result is a preservation plan (set of automatable policies plus projections of cost, access and loss over time). The policies include preservation actions at all levels, which in-

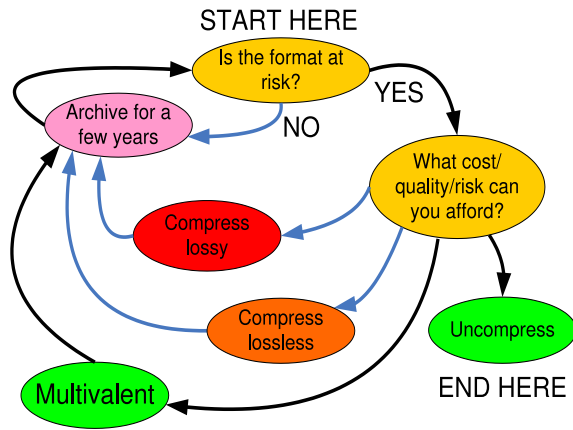


Figure 3: PrestoPRIME file format migration

clude periodic fixity checks and repair at the bit level (e.g. disc scrubbing), plans for large scale file format or storage migrations, how to deal with conflict (e.g. preservation actions such as migration or fixity checking can consume resources that might otherwise be needed to deliver access to content), and security (e.g. authentication, access control, rights management).

In each of these areas, the best course of action is a function of time and hence the policies require ongoing review. For example, the choice of file format to use for a particular AV asset depends on available and projected tool support, storage capacity, budget along with requirements for quality, safety and access. A simplified diagram for video file format migration is shown in Figure 3, further details can be found in project deliverables [14].

4. ARCHITECTURE & TECHNOLOGIES

PrestoPRIME aims to provide an open, flexible and standard system for the digital preservation. Beside the mentioned MDA approach, all the basic functionalities will be exposed as *services* leading towards an ESB architecture.

Following the OAIS specifications and the Object Oriented Programming guidelines, and considering previous experiences such as the CASPAR project, the *PrestoPRIME Preservation Platform* takes into account the best practices of commercial systems such as Rosetta developed by ExLibris [3], one of the current leader in the market place and partner of PrestoPRIME.

The overall approach for PrestoPRIME software development makes use of the well established best practices and design patterns already used for solving software problems. In order to achieve the highest flexibility and to enable an easy plug-in of new software components into the platform, an ESB (Enterprise Service Bus) architecture is adopted. Services are provided by SOAP (Simple Access Object Protocol) each time a specific schema is required leaving REST (REpresentational State Transfer) to manage the protocol necessary to exchange simple XML documents.

The architecture design contains the main OAIS functional blocks mapped into software components: *ingest*, *access*, *administration*, *preservation planning*, *data management* and *archival storage*.

If a pre-existing digital preservation system follows the OAIS guidelines, the PrestoPRIME solution will be easily integrated and can help to manage some specific issues not already covered. Otherwise the PrestoPRIME platform could be used in order to migrate an old archive to a new one, based upon standards and best practices. Finally some isolated tool can be extracted and plugged into an existing system. Recognising that automation is the key to lowering costs (and risks), PrestoPRIME is supporting digital preservation infrastructure components that have the ability to execute preservation policies and processes and make use of managed storage as a service that conforms to defined Service Level Agreements and QoS.

One way to avoid the act of design is to reuse existing designs: according to the MDA [9] approach, the overall analysis and design of the PrestoPRIME preservation platform reference architecture maps the six OAIS logical blocks onto software components making use of available design patterns. *Generative patterns* are used for addressing the specific requirements of each functional block. The design patterns solving the problems are recognized and brought together in order to build up the software component diagram for which Figure 4 shows a simplified overview, pointing out the conceptualised component of each block, where the OAIS activities and functions in each block are omitted, since the diagram is focused on the software implementation (actually the Platform Independent Model [9] design).

A software component not displayed for leaving the diagram easy to read, is a processing layer named *workflow module*, a cross software component used by the others, responsible for managing the workflow that the several processes need during the runtime phases, such as the ingestion and the preservation. A further component not planned in the OAIS model has been added to the figure (top left): the *Preservation Registry*, that the platform has to contact during most of the processes, such as the ingestion (SIPProcessing) and the RiskAnalysis (Preservation Planning). It is made up of three main components: *Format*, *Risk* and *Application*.

The Format component is responsible for the identification of the format and type of the digital content. Associated to the Format we have the *MetadataExtractors*, that are available for a specific Format and type (for example for pdf Format and Type version 2). The Risk component represents the risk associated to a specific Format (and type). The Application component is split into *Editing* and *Rendering* applications, for example respectively for coding video contents or play a movie. These kind of information are stored into the *Preservation Registry* that should be distributed. In our context, there is a Registry for each Archive and it is planned to have a centralized one gathering the information from all the archives, in order to provide a federated service for automatic preservation.

The detailed description of the six OAIS blocks depicted in Figure 4 is beyond the scope of the document. Further in-

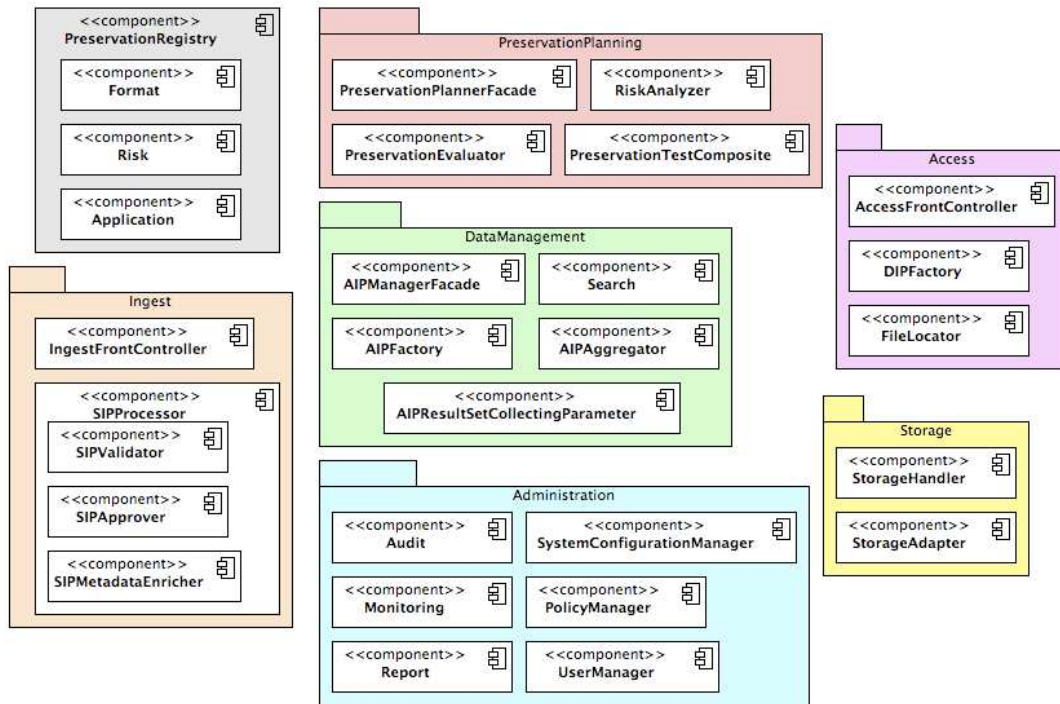


Figure 4: *PrestoPRIME Preservation Platform Component Diagram*

formation can be found in PrestoPRIME deliverables [14], the full details of the architecture design and software implementation will appear publicly in Summer 2010 [15].

5. AUDIOVISUAL RIGHTS MANAGEMENT

Rights clearance and management operations are becoming the major bottleneck in the exploitation of archived AV content, whether analogue or digital. The original contracts often need to be analysed and interpreted by specialists, which considerably increases costs and causes delays: vast portions of the archive remain unused due to uncertainty over rights. The sheer number of rights at the EU level illustrates how difficult it will be to create an automatic rights system.

Rights statements are included in contracts written in free form, without an agreed and unambiguous terminology. Statements can apply to different parts of the content, but as the contract is often drafted before the content is produced, there is no explicit association between shots and rights. Furthermore, rights can have a temporal span and are often reacquired or renegotiated subsequently. Different regulations apply across Europe, expressed with different terminologies, with different approaches based on different legal traditions. The existence of a common rights model would greatly simplify the rights management as it would be straight forward to extract and interpret rights clauses from contracts and map them to the content time line, as any other metadata.

PrestoPRIME aims to develop an infrastructure for media content provenance and tracking, and a European rights ontology, as the basis for an integrated system of rights management throughout the lifecycle from acquisition, to edit-

ing, to archive and distribution. This will be achieved by: developing a *European rights ontology* based and associated data model, based on the analysis of current rights typologies associated with broadcast content in Europe, which will be included in a common rights glossary; delivering a *rights management system* to extract and interpret rights clauses from contracts and map them to the content time line, as metadata and that will index the digital rights metadata within the archive, enabling the search for content based on digital rights information; guaranteeing *interoperability* between archives, other heritage institutions, the Semantic Web and UGC adopting a solution which is compliant with most relevant standards such as MPEG-21 and OMA DRM.

PrestoPRIME is focusing on open standards also by defining profiles, specifying subsets or combinations of two or several standards when necessary. The broadcasters partners will foster the contribution to such standards with the requirements and best practices for contracts and licenses relevant in their context.

6. STANDARDS AND BEST PRACTICES

This section discusses various standards initiatives around preservation of AV content. We put the main focus on formats for metadata and containers, as those are specific to AV content, while standards for system design or retrieval tend to be more generic (being only adapted to AV content in the context of a certain implementation).

6.1 METS

Metadata Encoding and Transmission Standard (METS) [10] is a standard representation for expressing the hierarchical structure of digital library objects, including the names and

locations of the files that comprise those objects and the associated metadata. METS allows the use of externally developed metadata schemes, which can be fit into its two defined metadata sections. METS itself does not care about the descriptive or administrative metadata schemes that are incorporated by implementers.

Some community based standards are recognized by the METS board and one of them is PREMIS [13], used to describe preservation metadata. To use PREMIS together with METS some decisions have to be made in advance as the PREMIS schema was developed in an implementation neutral way and also METS has quite some flexibility within it. Several issues exist, e.g. there exist redundancies among PREMIS and METS, PREMIS schemas can be used in a number of METS sections, whether to use the PREMIS container schema or not, how to deal with format specific metadata within PREMIS. As a consequence on these issues guidelines have been developed between the METS and PREMIS communities [13].

METS is one of the most commonly used containers (“transfer syntax”) for representing different types of packages in an OAIS compliant system (SIP, AIP, DIP). As METS still provides flexibility in terms of the descriptive, preservation and rights metadata used (or referenced), it has been decided to adopt METS also as a container for the packages in the PrestoPRIME system.

6.2 Container formats

MXF (SMPTE 377M) [19] seems to become the first choice as archive container format for AV content. Variants supporting the most common codecs in broadcast production (e.g. MPEG-2, D-10 [18], [20]) are defined and for digital cinema content it also used as the master and distribution container (according to DCI specification, with JPEG2000 encoding). Other appealing properties are the option to include uncompressed essence (if a collection can afford the storage costs) and the option to embed a wide range of metadata. Among the various defined operational patterns of MXF, the Operational Pattern 1a (SMPTE 378M) [21], which defines a file with a single playable essence comprising a single essence element or interleaved essence elements, is most relevant in a preservation context.

6.3 Metadata

We discuss here some relevant standardisation initiatives in the area of AV metadata models, rights models and fragment identification. It is impossible to cover in this papers the wide range of metadata standards for multimedia content, a good overview can be found in [4].

6.3.1 MPEG-7

MPEG-7 [8] (ISO/IEC 15938) is an excellent choice for the description of AV content due to its flexibility and comprehensiveness, but this comes at the price of complexity and potential interoperability problems. In order to partly solve these problems, profiles and levels have been proposed, but the adopted profiles do not support detailed description of AV content and lack semantic constraints. The Detailed Audiovisual Profile (DAVP) [1] aims at describing single multimedia content entities, allowing a comprehensive structural description of the content, including also audio and

visual feature descriptions. The NHK Metadata Production Framework [11] aims at similar goals, also defining an MPEG-7 profile with semantic constraints. These efforts are currently being harmonized by the EBU ECM SCAIE group [2].

MPQF (MPEG Query Format) [7] (ISO/IEC 15938-12) is a query format providing a standardized interface for multimedia content information retrieval systems in three aspects which are input query format, output query format, and query managements. The MPQF is a good candidate for the search and access functionalities in the PrestoPRIME preservation platform, since it specifies the interface through which the users can describe their search criteria with a set of precise input parameters in addition to a set of preferred output parameters to depict the return result sets, specified by the output query format.

6.3.2 MPEG-21

MPEG-21 (ISO/IEC 21000) is a suite of standards which aims at defining a normative open framework for multimedia delivery and consumption for use by all the players in the delivery and consumption chain.

MPEG-21 REL (Rights Expression Language, ISO/IEC 21000-5). REL provides a machine-readable XML-based standard format for licenses and rights representation, and the possibility to define new profiles and extensions to target specific requirements. Profiles defined so far, which address B2C (business-to-consumer) scenarios, are: MPEG-21 OAC (Open Access Content), for mapping Creative Commons licenses; MPEG-21 DAC (Dissemination and Capture) used in the broadcasting domain and mapping TVAnytime RMPI; MPEG-21 MAM (Mobile And optical Media) used in the mobile environment and the Media Streaming Profile, currently used in MXM development.

MVCO (Media Value Chain Ontology, ISO/IEC 21000-19). MVCO is an ontology for formalizing the representation of the Media Value Chain. It couples naturally with the MPEG-21 multimedia framework and been coded as an OWL Ontology. The MVCO is a reference for the definition of a rights ontology in PrestoPRIME.

6.3.3 MXM (MPEG eXtensible Middleware)

MXM (MPEG eXtensible Middleware, ISO/IEC 23006) is a suite of standards developed for the purpose of enabling the easy design and implementation of media-handling value chains whose devices interoperate because they are all based on the same set of technologies, especially technologies standardised by MPEG, accessible from the MXM middleware. MXM reference software enables the integration of standard modules and technologies within the integrated framework.

6.3.4 W3C Initiatives

The *Media Fragment working group* is developing URI based syntax to address fragments of media resources [22]. The specification supports temporal, spatial, track and named fragments and aims to be more general w.r.t. file formats supported than MPEG-21 part 17. The *Provenance incuba-*

tor group [16] deals with describing and tracking provenance of information (esp. considering developments in the Semantic Web and linked data). This is a very important issue in the preservation context and an important basis for rights modeling.

7. CONCLUSIONS

Considering that the amount of AV content only in Europe is already in the range of millions of hours, and it's increasing at a quite fast rate, for a large part of that content the destiny in a long term scale is far from being certain. PrestoPRIME tries to address this issue by tackling on one hand the point of awareness (about content value and risks of loss) and on the other hand the capability to provide preservation services and keeping their costs under control.

AV content is particularly challenging because of its characteristics. The answers from PrestoPRIME cover the areas of preservation strategies and technologies, of defining a flexible integration framework, and maximising the opportunities of future exploitation.

The core functionalities of the PrestoPRIME preservation platform will be released as software libraries under an open license. At the same time a commercial implementation will be available, provided by the ExLibris Company [3]. All the documentation of the platform design and source code will be available publicly at the project website.

In order to provide support to all the users that need information and assistance in setting up a digital repository and a preservation system, the project partners plan to set up a *Competence Centre* which will stay active after the completion of the project.

8. ACKNOWLEDGMENTS

This work was partially supported by the PrestoPRIME project [14], funded by the European Commission under ICT FP7 (Seventh Framework Programme, Contract No. 231161). The project is coordinated by Institut National de l'Audiovisuel (INA) and integrates 15 partners of all domains from France, UK, Italy, Austria, Netherlands and Israel, representing the variety of competencies needed for the AV preservation. The project therefore brings together participants including archive owners, research centres from archive institutions, general research centre and universities, and SMEs industrials for development and integration, including among the project partners the European Digital Library Foundation (EDLF).

Additional information about PrestoPRIME, including public deliverables and software, as well as other related news are available on the project Web site [14].

9. ADDITIONAL AUTHORS

Additional authors: Guy Ben-Porat (ExLibris Group, email: guy.benporat@exlibrisgroup.com), Annarita Di Carlo (RAI CRIT, email: a.di_carlo@rai.it), Stephen Phillips (IT Innovation, email: scp@it-innovation.soton.ac.uk), Daniel Teruggi (INA, email: dteruggi@ina.fr), Elisa Todarello (EURIX R&D, email: todarello@eurix.it)

10. REFERENCES

- [1] W. Bailer and P. Schallauer. The detailed audiovisual profile: Enabling interoperability between MPEG-7 based systems. In *Proceedings of 12th International Multi-Media Modeling Conference*, pages 217–224, Beijing, CN, Jan. 2006.
- [2] EBU ECM, SCAIE group: Automatic Information extraction. <http://tech.ebu.ch/groups/pscaie>.
- [3] ExLibris Group. <http://www.exlibrisgroup.com>.
- [4] M. Hausenblas (ed.). Multimedia vocabularies on the semantic web. W3C Incubator Group Report, <http://www.w3.org/2005/Incubator/mmsem/XGR-vocabularies/>, Jul. 2007.
- [5] ISO/IEC 21000-2. MPEG-21 Standard Part 2: Digital Item Declaration Language (DIDL).
- [6] ISO 14721:2003. CCSDS, Reference Model for an Open Archival Information System (OAIS).
- [7] ISO/IEC 15938-12. MPEG-7 Standard Part 12: MPEG Query Format (MPQF).
- [8] ISO/IEC 15938. MPEG-7 Standard: Multimedia Content Description Interface.
- [9] MDA. Model Driven Architecture. <http://www.omg.org/mda/>.
- [10] METS. Metadata Encoding and Transmission Standard. <http://www.loc.gov/standards/mets/>.
- [11] NHK Metadata Production Framework. <http://www.nhk.or.jp/str1/mpf/english/about.htm>.
- [12] EBU Tech 3295. P_META Metadata Library v2.1, Jul. 2009.
- [13] PREMIS. Preservation Metadata Maintenance Activity. <http://www.loc.gov/standards/premis>.
- [14] PrestoPRIME - Keeping Audiovisual Contents Alive. <http://www.prestoprime.eu>.
- [15] PrestoPRIME Deliverable D5.2.1: Definition and Design of a PrestoPRIME Reference Architecture for the Integration Framework, to be published in summer 2010. <http://www.prestoprime.eu>.
- [16] W3C provenance incubator group. <http://www.w3.org/2005/Incubator/prov/>.
- [17] Shaman: Sustaining heritage access through multivalent archiving. <http://shaman-ip.eu/shaman/>.
- [18] SMPTE 356M. Type D-10 Stream Specifications, MPEG-2 4:2:2P @ ML for 525/60 and 625/50.
- [19] SMPTE 377M. Material Exchange Format (MXF), File Format Specification, 2004.
- [20] SMPTE 386M. Material Exchange Format (MXF), Mapping Type D-10 Essence Data to the MXF Generic Container.
- [21] SMPTE 378M. Material Exchange Format (MXF) operational pattern 1a (single item, single package). <http://www.smpte.org>, 2004.
- [22] R. Troncy and E. Mannens (eds.). Media Fragments URI 1.0. W3C Working Draft, <http://www.w3.org/TR/media-frags>, Dec. 2009.