Digital preservation of audiovisual files within PrestoPRIME

Matthew J. Addis[1], Walter Allasia[2], Werner Bailer[3], Laurent Boch[4], Francesco Gallo[2], Stephen C. Phillips[1], and Peter Schallauer[1]

[1] IT Innovation, Southampton, UK
[2] EURIX, Torino, Italy
[3] JOANNEUM RESEARCH, Graz, Austria
[4] RAI - Radiotelevisione Italiana, Torino, Italy
{werner.bailer,peter.schallauer}@joanneum.at,
l.boch@rai.it
{allasia,gallo}@eurix.it
{mja,scp}@it-innovation.soton.ac.uk

Abstract. PrestoPRIME is a European project aiming at digital preservation of audiovisual files. In our scenarios we can observe a combination of lifecycles related to content, systems, services, and organisations involved both in the media and technical businesses. Taking as reference the OAIS model, we followed the principle of keeping everything under control, from all perspectives: configuration, resource management, data integrity, content quality and metadata handling. As a result, in the future users will have the opportunity to find and re-use high quality audiovisual material. Project outcomes include tools for simulating and modelling the use of resources and for service governance, software for content quality assessment, services for metadata mapping, techniques for content tracking, model and services supporting the management of exploitation rights. The reference software implementation provides a framework for integrating external tools.

Keywords: PrestoPRIME, Digital Preservation, AV Content Access

1 Introduction

The common thinking about digital files is that they are intrinsically safe because bits are bits. Such an optimistic approach underestimates the real risks which audiovisual (AV) files are going to face over the long term, which might include financial troubles, editorial worthlessness and technical troubles. Additionally, all the costs (investments) for digital AV preservation can be faced more comfortably knowing that a fruitful business model is established.

In this paper we present the achievements of EU FP7 funded project PrestoPRIME [1, 2], focused on long-term digital preservation and access of AV content. PrestoPRIME guidelines, tools and services will be delivered through a networked Competence Centre[5]. An overview of the project is available in [17].

[5] http://www.prestocentre.eu

We identified three main areas of contribution related to digital preservation: *preservation lifecycles)* for content and services (Section 2, *content re-use* (Section 3 and technologies for *actual content access* (Section 4).

The PrestoPRIME precept, "keeping AV content alive", works on the idea that future access and reuse will bring a return on the investments. However, in a long term scenario financial contingencies cannot be neglected and additionally the opportunity to re-use content could be negligible compared to the number of archival items. The reasons are manifold: (1) content might not be editorially appropriate for any re-use (not interesting) or, although possibly interesting, it could not be found because of bad descriptions not matching future contexts; (2) content quality might not fit actual expectations; (3) possible reuse might be jeopardised if the content exploitation rights cannot be established; (4) technical risks including single file corruption (I/O error), content management system fault (multiple file loss), unrecoverable misalignment between file format and tools used for processing (e.g. plugin not found) could affect content access.

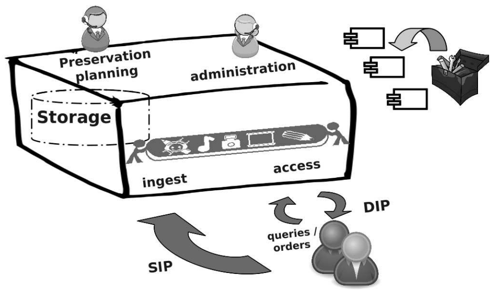

Fig. 1. Representation of OAIS model functional blocks in PrestoPRIME

PrestoPRIME adopted the *OAIS* model [3], developing and providing actual implementations of such a model tailored to AV environment. All the tools and services developed within the project are integrated in the preservation system, as shown in Figure 1. We evaluated the results of other projects in the digital preservation area, such as CASPAR[6], Planets[7] and DCC[8]. One of the PrestoPRIME partners, ExLibris [4], has an OAIS compliant product for sale, named *Rosetta*, a version of which will be tailored to project scenarios. In our specific context the essential part of the *information packages* is made of digital AV materials, with different and more complex preservation issues than the textual documents and still pictures material considered in other works.

[6] CASPAR Project - `http://www.casparpreserves.eu`
[7] Planets Project - `http://www.planets-project.eu`
[8] Digital Curation Centre - `http://www.dcc.ac.uk`

2 Aspects of the preservation lifecycles

When considering the long-term storage of audio-visual data in a preservation service it is important to understand the variety of lifecycles that must be dealt with. How and where is content produced? What is archived? When is it discarded? What about the storage and preservation services, contracts and organisations? The content generally outlives them all and remains even after various economic exploitation rights on the content have expired. There are many other entities that come and go during the scenario of producing and storing content. These include: content rights (e.g. copyright), organisations (rights-holders and service providers), contracts, service level agreements (SLAs) and services for storing, preserving and manipulating the content.

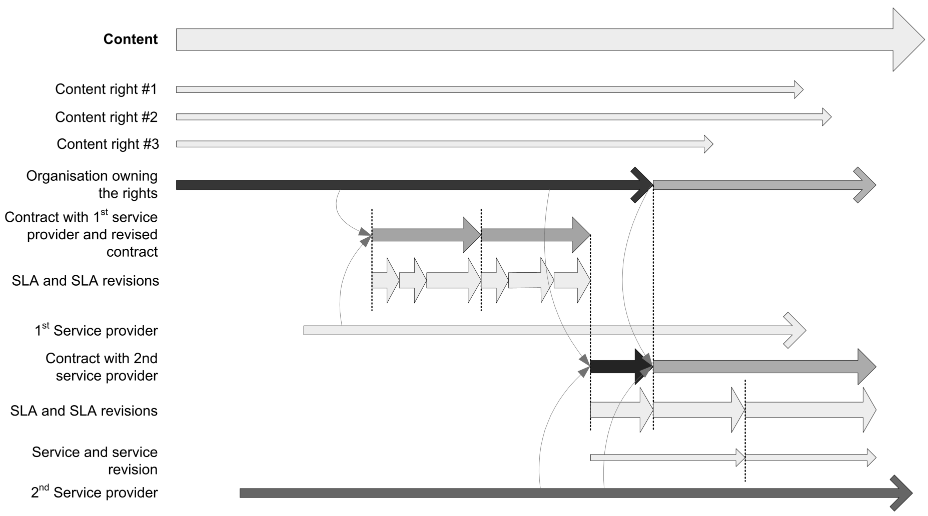

Fig. 2. Illustration of how different entities come and go during the life of content.

This complex set of relationships is worthy of further explanation. Figure 2 illustrates just how decoupled the various transition points between service providers, services and SLAs can be. These *services* and *service providers* refer to both in-house and external installations, managed both internally and out-sourced.

2.1 Content lifecycle

Content is created. Some is kept for a long time, some for a short time and some discarded. For a broadcaster moving to file-based working covering the whole production chain, the majority of files in use at any one time would be rushes. The 'shoot-to-show' ratio of a production can be as high as 20:1 for high-end factual productions or even more. As broadcasters move to using media asset management (MAM) systems, preservation becomes part of the production process with assets and meta-data entering the preservation system throughout the cycle. For new content it will not just be the case of creating a programme and then place it into the archive. For a single programme, different parts of essence and metadata are ingested and accessed throughout the production cycle.

Once data is in the archive it must be actively preserved. This involves not just taking care of the 'bits' using checksums and regular integrity checking in tape and disc storage systems but also migration to new physical carriers and new data formats when necessary. The simulation tool discussed below includes these integrity checking and migration events as they are both costly in terms of resources.

2.2 Service lifecycle

Of particular importance is the lifecycle of a service, which goes through the phases represented in Figure 3: specification, negotiation, provisioning, use and decommissioning

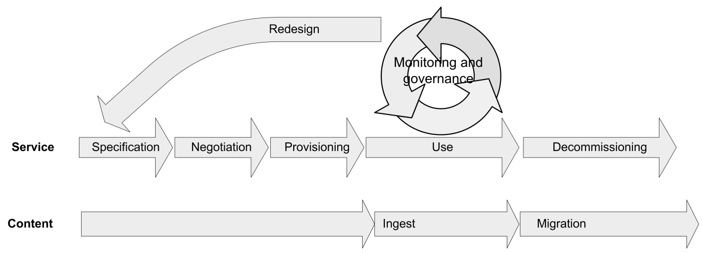

Fig. 3. Service lifecycle.

PrestoPRIME has developed cost and risk estimation tools for storage systems and for media transfer chains. These tools fit into the specification stage of the service lifecycle and, as we will see, into the monitoring and governance of the operational systems.

The CCSDS Producer-Archive Interface Methodology Abstract Standard [3] discusses in detail the stages represented simply as *specification* and *negotiation* in the diagram above. It is a companion document to the OAIS document and defines the actions required from the initial time of contact between the producer and archive (service provider) until the objects are received and validated by the archive. This covers the OAIS "Negotiate Submission Agreement", "Receive Submission" and "Quality Assurance" aspects. More detail on these processes can also be found in deliverable D2.2.1 [9], available at [2].

Once a service is being used and content is ingested, a continual monitoring process should be in operation. Monitoring data should be available to a service governance system (either automatic or manual). Experience of the service behaviour and user requirements will generally feed into an improvement or redesign process and eventually the service will be decommissioned and the data held in the service will be migrated to a new service (either at the same provider or elsewhere). The final point, that data may need to be migrated elsewhere should be emphasised. Moving huge quantities of data from one physical location to another is very expensive (in time and money) and any contract with a service provider must include an exit strategy to cover this eventuality.

2.3 Simulation and modelling

The PrestoPRIME project has developed simulation modelling tools both for hierarchical storage systems and for media transfer chains. We focus here on the trade-offs that must be made in the design of a storage system and the cost of risk of loss [12].

When designing (and redesigning) a system for storing and preserving digital content there are many choices that can be made, for example, number of copies to make, what technology to store them on, how often to check their integrity, whether to use automated or manual processes, and how to balance user needs for ingest and access with internal functions such as media migrations, file scrubbing and replication. These activities take time to execute dependent on resources available (e.g. people, servers, bandwidth), which in turn cost money and become contended with different uses having different priorities. Our approach to simulating this problem starts with a simple but flexible storage model (Figure 4) that has the function of accepting files for storage (writes), returning files from storage (reads) and storing the data inside the files using some form of physical media (hard drive, data tape, optical disc etc.).

The model, which includes a manual/automatic 'controller' that mediates these processes, can be applied to automated hardware/software such as a HDD server, or it can be applied to a more manual process such as data tapes on shelves with people to move tapes to and from the drives. When writing or reading files, various operations such as integrity checks, encoding or applying error correction may be performed by hardware, the filesystem, manually by an operator, or a combination of all of these. Likewise, various failures or errors may occur when reading or writing data or even when it is, in effect, 'doing nothing'. These errors, both latent and extant, range from 'bit rot' in a HDD system through to accidental damage from manual handling of discrete media. The various operations and errors are represented through costs (in time and resource) and error rates for read/write/store actions.

One or more storage systems may be combined into an archive configuration including ingest and access queues for new file arrival and retrieval of stored files. The configuration determines how files are allocated to storage, how they are replicated, and how they are repaired if there are failures. Resources can be allocated to serving ingest, access and copy operations as well as for activities such as integrity checking and repair within each storage system.

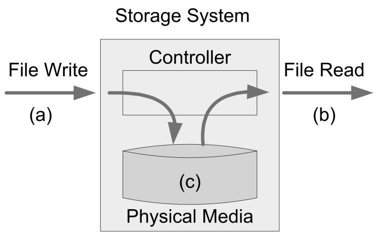

Fig. 4. Generic storage model. The controller could represent anything from a human operator moving data tapes on or off shelves to software with queues, integrity checks and fast disc storage.

The interactive simulation tool[9] takes a discrete event simulation approach. During the simulation, time ticks away (e.g. 1 second of the simulation might correspond to 1 week in the real world) and events are generated (e.g. random corruption of files in a storage system, requests to access a file, new files to be added to the archive). These events then trigger actions, such as a copy/repair process, which are added to the queues of the storage systems involved. A storage system processes items in its queues according to how much resource it has available The available capacity of the resources used by each service determines how many items are processed for each tick of the clock, and at what cost.

The users can interact with the simulation as it progresses, changing the quantity of resources available or changing the policy for data safety. In this way the users are in effect playing a game that helps them understand how to react to and manage events that they might see in practice when operating a real system. For example, there is also an option to simulate 'disaster scenarios': rare but catastrophic events where large fractions of the storage become temporarily or permanently unavailable.

2.4 Service governance

Recognising that automation is the key to lowering costs and reducing the risks caused by fallible human operators, PrestoPRIME is developing digital preservation infrastructure components that go beyond the OAIS functional model and provide automation and execution of preservation policies and processes. For instance, providing managed storage as a service conforming to defined SLAs and Quality of Service (QoS).

The service management software monitors services (both local and remote) using a light-weight asynchronous protocol. Monitoring reports describe usage using metrics where each metric is a reference to an OWL[10] concept that relates it to a well-defined model expressed as an ontology. Metrics can be combined to produce complex QoS terms which are compared against deployed policies and

[9] `http://prestoprime.it-innovation.soton.ac.uk`
[10] `http://www.w3.org/TR/owl2-overview`

may trigger alerts or configuration updates on the managed services to attempt to address deviations in the expected behaviour.

This management layer is generic. The metrics reported by a PrestoPRIME's prototype service in the context of storage include: the amount of data ingested (bytes), the number of frames ingested, the amount of data accessed (bytes), the rate of data ingest and access (bytes per second), the number of file corruptions and frame corruptions observed and the time from an access request being received to the start of the data being accessed (seconds).

The service management software also aggregates information across all services and SLAs and monitors the whether each data service is 'up' or 'down'. Terms in the customer SLAs can be defined by various mathematical operations on the measured metrics. For instance, it is possible to set a fast access bandwidth and then reduce that bandwidth if the customer exceeds a data transfer threshold during a 24-hour period. The software is also able to define and track QoS guarantees such as the mean response time during a monthly period.

Part of the power of the service management system is that it does not have any of the metrics, QoS terms or management terms pre-defined. The metrics are URIs with characteristics defined by an ontology and the QoS and management terms result from configurable mathematical and logical terms [16].

2.5 Continuous refinement

The previous two sub-sections have discussed both the modelling and the operational monitoring and management of services. We are now working on bringing these two aspects of service governance together, feeding monitoring data from the services into the models so that the models can reflect reality more precisely and better inform the operators who define and update the automatic service management policies.

Once the storage model can be synchronised with the state 'now' and be parameterised with probability distribution functions generated from historical monitoring records then its utility in answering important 'what if' questions will be hugely increased. It would then be possible to ask what the maximum ingest rate that could be sustained by the current system was or what additional resources were required to support a new SLA with an associated usage pattern. In addition, optimisations could be performed to discover the optimum scrubbing interval, trading off cost and data safety. By linking the model to reality, the answers generated will be directly applicable to operational policy decisions.

3 Enabling reuse

In order to make archive content accessible beyond a small group of experts, and – even more importantly for many organisations – to provide the economic basis for preservation, archive content needs to be available for re-use. On a technical level, content must be stored on media and in formats that still can be read and decoded, and the quality of content must be known in order to decide the appropriateness for certain distribution channels and/or apply the necessary restoration actions. On the level of cataloguing, sufficient metadata must be available in appropriate formats, in order find relevant content items. The metadata must be interoperable with documentation and search systems currently in use. For solving issues on the legal level, the metadata needs to contain sufficient information about the rights situation of the content and its constituent parts in order to decide whether it can be used for certain purposes.

3.1 Efficient video quality control

Automatic quality control for AV media is important in several steps of their lifecycle. Broadcasters are checking quality after production, after editing and before play-out to various delivery services. Archives are checking for content quality during archive ingest, delivery and migration. Content providers are checking their content for correct encoding and conformance to the required format standard before dispatching to customers. These use cases have in common that mainly technical properties of the material are checked, e.g., stream compliance, wrapper compliance and file header correctness. Today only some content properties can be checked automatically, e.g., blocking or luma/chroma violation. Other relevant content properties and impairments like noise level, sharpness, large dropouts, flickering, unsteadiness and many more can only be checked by humans manually exploring the AV content.

Current solutions for manual video quality control result in extremely high labor costs or, when these costs cannot be afforded, in not assuring the content quality and therefore in limited possibility of content usage and re-usage. The transition to file based production and preservation environments enables automation of quality assurance tasks with the clear goals of reducing costs and of increasing quality of content produced and re-used out of an archive. Our video quality control research in PrestoPRIME focuses on the development of automatic reference-free visual quality detection algorithms and on its practical application in file based production and preservation processes.

Our proposed detection algorithm for video breakup models "normal visual flow" and detects video breakup occurrences by violations of this normal visual flow. This generic approach allows the detection of a very large group of severe visual distortions, e.g. many kinds of dropouts, timebase correction hits, assemble edits, and head clogging. The evaluation on an expert annotated compilation of 51 standard definition videos (452 min) with various challenging defects is possible in real-time and results in recall rates above 90% with a false positive rate of below 1.0 per minute. For reference-free noise level estimation we use an efficient GPU based optical flow technology for motion-compensation of consecutive frames. A noise level measure is estimated by spatiotemporal feature analysis. The algorithm supports various types of noise, from fine electronic noise to coarse film grain noise, and is also able to estimate the noise signal dependency. Preliminary mean opinion score evaluation results show a good correlation between the estimated noise level and the subjective human ratings. In order to allow the application of these developed detectors they have been integrated in two tools.

The *AV-Inspector Analyser* supports fully automatic batch analysis of video files with respect to a set of video impairments like video breakup, noise level, monochrome frames and others. The *AV-Inspector Summary* (see Figure 5) allows for visualisation and interactive exploration of impairment analysis results to support the user in quickly getting an overview of the condition of the material. Temporally condensed views as well as ranked lists of detected impairments support the user in efficiently spotting most relevant impairments in order to initiate quality control actions.

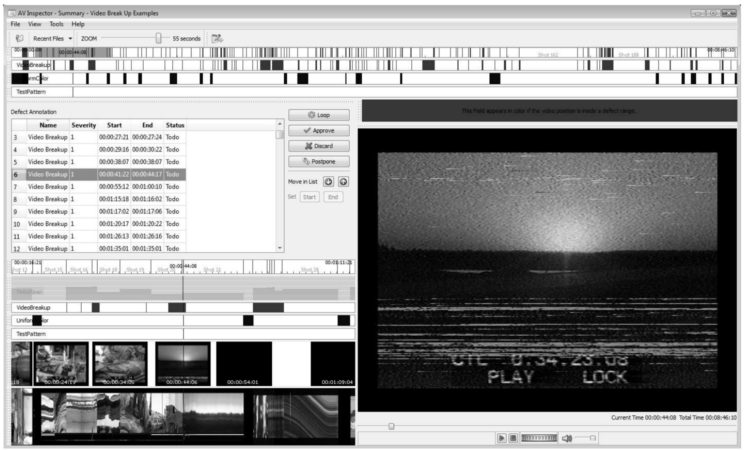

Fig. 5. Screenshot of the AV-Inspector Summary viewer.

To facilitate metadata interoperability and exchange between these and other tools, a comprehensive and MPEG-7 standard compliant framework for the description of visual impairments, including generic and specific impairment descriptors as well as a comprehensive impairments classification scheme, has been developed and is available by Joanneum Research[11]

3.2 Keeping descriptions up-to-date and interoperable

Exchange of content descriptions is the key to ensuring access to AV collections, as well as keeping them up-to-date for future uses and integration with current workflows. Metadata exchange is often hindered by the diversity of used metadata formats. There are several points in the preservation workflow of AV content where interoperability of metadata is important: (1) conversion of legacy technical metadata in preservation scenarios; (2) access to legacy descriptions for performing mapping between material and editorial entities; (3) migration to new documentation models and tools; (4) extracting metadata embedded in

[11] http://mpeg7.joanneum.at

file headers and convert it to the data structures needed in a source/or archive information package of an OAIS [3] compliant system; (5) outsourcing of annotation and access services, with potentially different data models involved; (6) content provision to Europeana[12] or similar portals. The last point is a prerequisite for external access to collections, while the other points are in the internal preservation workflow of an organisation.

Metadata interoperability needs a solution on syntactic and on semantic levels. While the multimedia metadata formats overlap in a core set of covered metadata properties, they are at the same time dissimilar in many ways (e.g., domains, comprehensiveness and complexity). Due to these differences, mappings can only be partial in many cases, such as properties only supported in one of the involved formats, inconsistent properties, etc.

Our approach (for details see [15]) uses a high-level intermediate representation of metadata elements. Metadata elements from specific metadata formats are described in relation to these generic elements. Further mapping templates are used on data type level. The mapping between a pair of formats is derived from these sources. We extended an existing ontology for our purpose, to represent the formal semantics of the high-level intermediate representation. Deriving mapping instructions from the ontology eases maintenance of mapping instructions against hard coded ones. The high-level intermediate representation further serves as a hub for mapping between formats and hand-crafted one-to-one mappings between pairs of formats are avoided. Mappings can be created automatically. Adding new formats does not have side effects.

The core of this approach is the meon [14] ontology which describes generic metadata elements and the relations between them. Meon was originally developed to model metadata elements used throughout the AV media production workflow in a format independent way in order to support content exchange and automation. The meon ontology has been extended to express mapping relations between metadata formats. In addition to the ontology of generic metadata concepts, specific ontologies are created for each format. They follow the same pattern as the meon ontology and include relations between format specific and generic concepts. Mapping instructions for low-level issues (e.g. conversion between data types) are defined as well.

A prototype[13] was implemented following this approach (ontology expressed in OWL-DL and mapping instructions encoded as XSL transformations[14]). Metadata concepts and mapping relations in the ontology are annotated with additional information needed for mapping (i.e., format specific metadata, corresponding XPath and information on data type representation). Using reasoning techniques over the ontology data type templates are inferred for the mapping between different format specific concepts. A SPARQL[15] query retrieves all nec-

[12] `http://www.europeana.eu/portal/`
[13] available at `http://prestoprime.joanneum.at`
[14] `http://www.w3.org/TR/xslt20/`
[15] `http://www.w3.org/TR/rdf-sparql-query/`

essary mapping information from the ontology allowing the generation of XSL template documents with mapping instructions.

3.3 Content tracking and rights

Two issues related to content re-use are how to determine provenance and use of content and how to establish if the content can be exploited.

Systems making use of fingerprint technology for keeping track of AV content are under development by INA. The fingerprint tools, in addition to provide some indication about the structure of the AV contents, support the indentification of repeated clips. Fingerprints are deterministically computed, as "visual hash", from the analysis of video signal fragments form a "digital summary" and permit to track quasi-copy clips, resulting from excerpt editing. In our context the provenance information of each fragment of material resulting from re-use of content is required. Such genealogy can be very complex because it entails the identification of all the sources, including the correct timeline pointers.

PrestoPRIME defined an ontology model for supporting the management of AV rights information within a business domain (see D4.0.5 [6, 2]. This ontology extends ISO/IEC 21000-19 (MPEG Media Value Chain Ontology) and is included in MPEG-21 part 20 (Contracts Expression Language). The rights clearance scenario identified in D5.1.1 [7, 2], showed that a large part of the current effort is put in establishing the rights situation for programmes produced with re-use of archival content. The contracts held by the broadcasters are actually referencing the older sources and thus it is required to follow the provenance tree in order to re-build the complete status of exploitable rights.

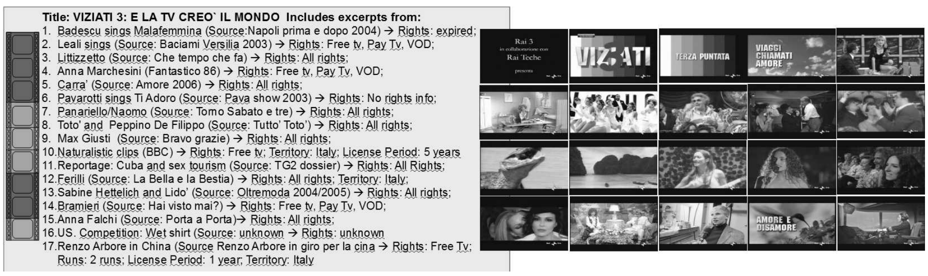

Fig. 6. Example of TV segmentation according to provenance and rights criteria

An example is the RAI broadcasted programme *"Viziati: e la TV creò il mondo"*, which is only made of archival excerpts (see Figure 6). As each fragment comes with it own constraints, and also because of the contracts agreed with the performers, the chance of exploiting the whole programme is almost null, although a partial exploitation would still be possible. In order to address this need the PrestoPRIME rights ontology adopted a URN media fragment notation[16] as the suffix of the Intellectual Property Entity `RelatedIdentifier`. Thus the rights instance document describes faithfully even such a complex rights situation, should provenance information be available and correct.

4 Actual access to high quality material

At the present time many archives are still facing the task of digitisation of legacy analogue materials. This means that for a certain time the preeminent activity of a newly deployed AV digital preservation system is going to be *Ingestion*. However, we can assume that the ingestion phase will be limited to freshly produced items, while content *Access* will gain greater and greater relevance.

OAIS identifies three types of *information packages*, for submission (SIP), archival (AIP) and dissemination (DIP) respectively (see Figure 1). The PrestoPRIME Preservation Platform (P4) is going to provide to its users both *ingestion* and *access* interfaces. The former is implemented by a service accepting in input the SIP, running a complete ingestion workflow. The latter provides a service for posting access requests and queries, and in response makes available the DIP, containing all the necessary links to perform actual access to the materials.

This section gives a brief description of the P4 architecture, which provides mechanisms for integrating internal and external preservation tools. The technology and the approach adopted are presented. The core of P4 software will be available as open source, giving to the archives with technical capabilities a kind of "do-it-yourself" option. Additionally, the PrestoPRIME data model and the information package format are also briefly presented.

4.1 Architecture design

The architecture of the PrestoPRIME preservation platform (P4) was designed according the OAIS model mentioned before, following a Model Driven Architecture (MDA) approach. The OAIS functional entities (Ingest, Access, Administration, Preservation Planning, Data Management and Storage) are mapped onto software components. The design of P4 is already described elsewhere, see for example [17] and references therein, where the authors provide the overall component diagram representing, according to MDA, a candidate Platform Independent Model. In Figure 7 we represent the Platform Specific Model of our information system design.

In order to achieve the highest flexibility and to enable an easy plug-in of new software components into the platform, an Enterprise Service Bus approach has been chosen, where most of the services are published by REST [18] interfaces. Ingest and access of information packages into the preservation system have been implemented as a CRUD web service exposing REST APIs. Further details can be found in PrestoPRIME deliverables D5.2.1 [8] and D5.2.2 [11], available at [2].

The Ingest component is based on a Workflow engine, which allows the implementation of complex workflows involving asynchronous operations, including

[16] according to MPEG21-part17 or upcoming W3C MediaFragment standard

SIP validation, metadata enrichment, quality analysis and indexing. The Ingest of SIP and the Access of DIP makes use of a data model specification (reported in project deliverable D2.1.3 [10] and available at [2]), which takes into account the preservation scenarios identified in the project and makes use of several standards, in order to define the structure of SIP and DIP. The editorial entity is based on METS [13], which is a metadata encoding and transmission standard format capable of wrapping other metadata representations for specific purposes. We identified appropriate metadata formats for technical and descriptive metadata, including MPEG-7 and Dublin Core. The SIP contains the metadata information related to the content, as well as a reference to AV files and other external entities. The ingest interface performs SIP validation and parsing, while the data management component is in charge with storing the information for persistence. On the access side, the platform can retrieve the stored AIP and provide a DIP with different levels of information, taking into account the user request and profile. For instance, a user could be interested only in a limited set of information related to a given AIP (e.g. the information about rights) or she could only be allowed to access part of the AIP information.

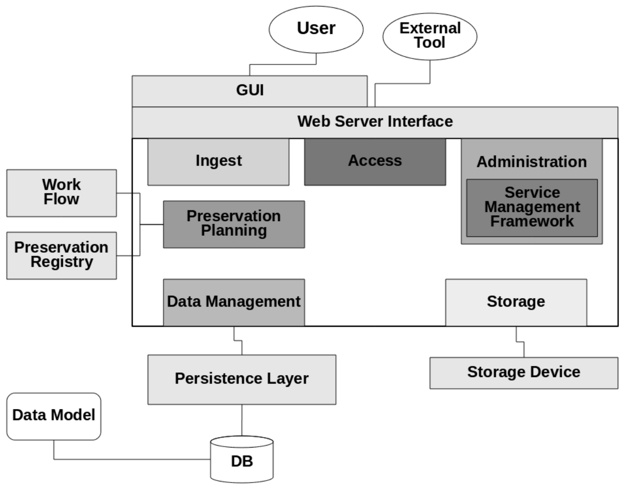

Fig. 7. Architecture Diagram of PrestoPRIME Preservation Platform

It is worthwhile noticing that the Administration component will integrate a Service Management Framework which is charge for monitoring the services and operations within the archive, ensuring that the SLA for each service are fulfilled. Since the OAIS model does not suggest a specific implementation of the persistence module, we have chosen to implement a Persistence Layer, which is an additional component aware of the Data Model and which is responsible for the persistence of AIP. Further details are available in PrestoPRIME deliverables [2].

4.2 Technologies for the implementation

The choice of technologies used for the implementation of the preservation platform was driven by the following requirements: (1) develop an open platform which can be used as reference software of PrestoPRIME results (2) allow the integration of external tools and software for specific preservation activities (3) whenever possible, adopt open source and common technologies. References to the technologies mentioned in the following can easily be found on the Web.

The source code of P4 is written in Java, using Oracle JDK[17] reference implementation, Eclipse IDE[18] and Apache Ant[19] as building tool. Whenever possible we chose dependencies from libraries available as open source or with no restrictive licenses for research purposes. The P4 web server API implements the REST paradigm, using Jersey[20], the Java REST API reference implementation. The web application runs within Apache Tomcat[21], the reference implementation of Java web container. Data persistence makes use of Java persistence API, JBoss Hibernate[22] was used to implement the persistence layer, enabling the use of any JDBC-compliant DBMS (P4 has been tested on MySQL[23] and Oracle XE[24]). Exchange of information through the REST API uses XML for Java objects serialization. We implemented the workflow engine using OMG BPMN[25] for modelling processes and Java open source libraries for the conversion of the BPMN graphical notation to a machine-processable XML representation.

5 Conclusions

We have presented how the outcomes of the EU project PrestoPRIME address the needs of the digital preservation of AV files all along the preservation lifecycle, from the time of planning and setting the strategies to that of operating and providing a successful fruition of archival contents. Using as reference the criteria of the OAIS model used, we identified the key to success for our goal in having the tools for keeping everything under control. The management of the use of resources, including estimating the investment and operating costs, permits a continuous assessment both of the selected strategy and of the capability of our preservation system to provide the agreed service levels to the archive users. The attention to AV content ranges from keeping care of the file bits, which pose the issue of data rate for storage and processing, to looking after audio-video file formats, and the technical quality of the content. Eventually the knowledge about the content, in terms of descriptions, provenance, and

[17] http://www.oracle.com/technetwork/java/index.html
[18] http://www.eclipse.org
[19] http://ant.apache.org
[20] http://jersey.java.net
[21] tomcat.apache.org
[22] http://www.hibernate.org
[23] http://www.mysql.com
[24] http://www.oracle.com/technetwork/database/express-edition
[25] http://www.bpmn.org

rights information is taken into account by the project with the work done on specific models, tools, and services. The second project test-beds held in RAI in Autumn 2011 will assess the implemented preservation platform, integrated with all developed tools and services, based on feedbacks during first test-bed session in BBC in Autumn 2010.

Acknowledgements.

This work was partially supported by the PrestoPRIME project [1], funded by the European Commission under ICT FP7 (Seventh Framework Programme, Contract No. 231161). The project is coordinated by INA (Institut National de l'Audiovisuel) and integrates 15 partners of all domains from France, UK, Italy, Austria, The Netherlands and Israel.

References

1. PrestoPRIME - Keeping Audiovisual Content Alive `http://www.prestoprime.eu`
2. PrestoPRIME Results `http://www.prestoprime.eu/project/public.en.html`
3. OAIS Blue Book `http://public.ccsds.org/publications/archive/650x0b1.pdf`
4. ExLibris `http://www.exlibrisgroup.com`
5. MPEG-7 Extension for AV Defect and Quality Description `http://mpeg7.joanneum.at/`
6. PrestoPRIME Deliverable D4.0.5, "Content Tracking System Proposal and Common Rights Ontology" (Jan 2010)
7. PrestoPRIME Deliverable D5.1.1, "Definition of Scenarios" (Aug 2010)
8. PrestoPRIME Deliverable D5.2.1, "Reference Architecture" (Jun 2010)
9. PrestoPRIME Deliverable D2.2.1, "Preservation Process Modelling" (Feb 2010)
10. PrestoPRIME Deliverable D2.1.3, "Data Model Specification" (Feb 2011)
11. PrestoPRIME Deliverable D5.2.2, "Prototype Implementation" (Dec 2011)
12. Addis, M. et al (2011). "Digital Preservation Strategies: the cost of risk of loss for AV Content". Jan/Feb 2011 edition of the Motion Imaging Journal of the Society of Motion Picture and Television Engineers (SMPTE).
13. METS `http://www.loc.gov/standards/mets/`
14. Martin Höffernig and Werner Bailer. Formal metadata semantics for interoperability in the audiovisual media production process. In *Workshop on Semantic Multimedia Database Technologies (SeMuDaTe)*, Graz, AT, Dec. 2009.
15. Martin Höffernig, Werner Bailer, Günter Nagler, and Helmut Mülner. Mapping audiovisual metadata formats using formal semantics. In *Proceedings of 5th International Conference on Semantic and Digital Media Technology*, Saarbrücken, DE, Dec. 2010.
16. Mike Surridge, Ajay Chakravarthy, Maxim Bashevoy, and Martin Hall-May. Serscis-Ont: A Formal Metrics Model for Adaptive Service Oriented Frameworks. In *Second International Conference on Adaptive and Self-adaptive Systems and Applications*, Lisbon, Portugal, Nov. 2010.
17. Matthew Addis, Walter Allasia, Werner Bailer, Laurent Boch, Francesco Gallo and Richard Wright. 100 Million Hours of Audiovisual Content: Digital Preservation and Access in the PrestoPRIME Project. In *Proceedings of the 1st International Digital Preservation Interoperability Framework Symposium*, Dresden, DE, Mar. 2010.
18. Roy T. Fielding Architectural Styles and the Design of Network-based Software Architectures Ph.D. Thesis, University of California, Irvine, CA, 2000.