

FP7-ICT-2007-3-231161



## **Deliverable D5.1.1**

### **Definition of Scenarios**



L. Boch (Rai)

2010-01-25

## Document administrative table

Document Identifier	PP_WP5_D5.1.1_Scenarios_R0	Release	0
Filename	PP_WP5_D5.1.1_Scenarios_v1.02.doc		
Workpackage and Task(s)	WP5 Integration and validation Task 1 – User-driven scenarios		
Authors (company)	Laurent Boch (RAI)		
Contributors (company)	Jean-Hugues Chenot, Daniel Teruggi (INA) Richard Wright (BBC) Johan Oomen, Marius Snyders (B&G) Christoph Bauer (ORF) Walter Allasia, Francesco Gallo (Eurix) Matthew Addis (ITI) Lora Aroyo (VUA) Günter Muehlberger, Hubert Feichter (UIBK) Georgia Angelaki, Jill Cousins (EDLF)		
Internal Reviewers (company)	Christoph Bauer (ORF), Jean-Hugues Chenot (INA)		
Date	2010-01-25		
Status	Final Draft		
Type	Deliverable		
Deliverable Nature	Report		
Dissemination Level	Public		
Actual Deliv. Date	2010-01-25		
Abstract	<p>This document provides the description of contexts and cases, to be used as the base reference in the subsequent stages of integration and evaluation of the project's outcomes. Preservation for broadcast archives, rights clearance, the perspectives of higher education institutions and of small &amp; medium AV archives, adding AV content to Europeana, information enrichment by users, and the perspective of the service providers are the main topics of this set of scenarios.</p>		

### DOCUMENT HISTORY

Version	Date	Reason of change	Status	Distribution
0.xx	2009-04-27	First Outline	Outline	Confidential
0.xx	2009-06-18	First working draft <sup>1</sup>	Working draft	Confidential
1.0	2009-12-18	Submitted for internal review	Final draft	Confidential
1.01	2010-01-14	Revised after internal review	Final draft	Confidential
1.02	2010-01-25	Finalised	Release	Public

<sup>1</sup> subsequent updates in "working draft" status are not reported

## Table of contents

<a href="#">Scope.....</a>	<a href="#">6</a>
<a href="#">Executive summary.....</a>	<a href="#">7</a>
<a href="#">References.....</a>	<a href="#">10</a>
<a href="#">1.Introduction.....</a>	<a href="#">13</a>
<a href="#">1.1 Adopted approach.....</a>	<a href="#">13</a>
<a href="#">1.2 Progress of work.....</a>	<a href="#">13</a>
<a href="#">1.2.1 Feedbacks from workpackages and first draft.....</a>	<a href="#">13</a>
<a href="#">1.2.2 Criteria.....</a>	<a href="#">13</a>
<a href="#">1.2.3 Workshop with the user group.....</a>	<a href="#">14</a>
<a href="#">1.3 Finalisation.....</a>	<a href="#">14</a>
<a href="#">2 Preservation for Broadcast Archives.....</a>	<a href="#">15</a>
<a href="#">2.1 Setting the scene.....</a>	<a href="#">15</a>
<a href="#">2.1.1 Reference quality levels.....</a>	<a href="#">15</a>
<a href="#">2.1.2 Archive and production.....</a>	<a href="#">16</a>
<a href="#">2.2 Questions addressed.....</a>	<a href="#">17</a>
<a href="#">2.2.1 Digitisation and Ingest.....</a>	<a href="#">17</a>
<a href="#">2.2.2 Preservation of AV properties.....</a>	<a href="#">18</a>
<a href="#">2.2.3 Preservation models and strategies.....</a>	<a href="#">18</a>
<a href="#">2.2.4 Storage technologies and strategies.....</a>	<a href="#">20</a>
<a href="#">2.2.5 Content based quality assessment.....</a>	<a href="#">20</a>
<a href="#">2.2.6 Risk assessment.....</a>	<a href="#">21</a>
<a href="#">2.2.7 Relationships with providers.....</a>	<a href="#">22</a>
<a href="#">2.2.8 Access and delivery.....</a>	<a href="#">22</a>
<a href="#">2.3 The role of broadcast archive.....</a>	<a href="#">22</a>
<a href="#">2.3.1 Mission in the organisation.....</a>	<a href="#">22</a>
<a href="#">2.3.2 The archive manager.....</a>	<a href="#">22</a>
<a href="#">2.3.3 Other roles within broadcast archives.....</a>	<a href="#">23</a>
<a href="#">2.4 Relationships with other roles.....</a>	<a href="#">23</a>
<a href="#">2.4.1 Content creators and content distributors.....</a>	<a href="#">23</a>
<a href="#">2.4.2 Providers of services, equipment and software.....</a>	<a href="#">23</a>
<a href="#">2.5 Description of cases.....</a>	<a href="#">24</a>
<a href="#">2.5.1 Preservation of material digitised as D10.....</a>	<a href="#">24</a>
<a href="#">2.5.2 Preservation of material as uncompressed.....</a>	<a href="#">25</a>
<a href="#">2.5.3 Preservation of material derived from 90's HDTV.....</a>	<a href="#">26</a>
<a href="#">2.5.4 Preservation of material derived from Film.....</a>	<a href="#">26</a>
<a href="#">2.5.5 Preservation of material SDTV with DV coding.....</a>	<a href="#">27</a>
<a href="#">2.5.6 Preservation of material currently produced in HDTV.....</a>	<a href="#">27</a>
<a href="#">2.6 Discussion topics.....</a>	<a href="#">28</a>
<a href="#">2.6.1 Compression.....</a>	<a href="#">28</a>
<a href="#">2.6.2 Definition of obsolete format.....</a>	<a href="#">29</a>
<a href="#">2.6.3 Compression options for analogue sources.....</a>	<a href="#">29</a>
<a href="#">2.6.4 Metadata in AV wrappers.....</a>	<a href="#">30</a>
<a href="#">3 Rights Clearance for Access and Fruition .....</a>	<a href="#">31</a>

3.1 Questions addressed.....	31
3.2 Users and roles.....	31
3.2.1 Content creators.....	31
3.2.2 Archives.....	31
3.2.3 Content deliverers.....	32
3.2.4 Rights holders and agents.....	32
3.2.5 Consumers/general users.....	32
3.3 Descriptions of cases.....	32
3.3.1 From contracts to rights information set.....	32
3.3.2 Rights clearance and content provenance.....	34
3.3.3 Archive user accessing rights information.....	34
3.3.4 Rights clearance for reuse by content creators.....	36
3.4 Implications for test materials and evaluation.....	37
4 The perspective from other organisations with AV archives.....	38
4.1 Higher education institutions (HEI).....	38
4.1.1 Characteristics of the HEI scenario.....	38
4.1.3 Main questions addressed.....	39
4.1.4 Users and roles.....	39
4.1.5 Quantities.....	40
4.1.6 Storage and access strategies.....	41
4.1.7 Considerations on formats.....	41
4.2 Exploitation models for medium and small archives.....	42
4.2.1 Questions addressed.....	42
4.2.2 Users and roles.....	43
4.2.3 The “broker” approach.....	43
4.2.4 The “distribution agreement” approach.....	44
4.2.5 Success factors.....	45
4.2.6 Technical implications.....	45
5 Adding AV content to Europeana and improving its usability.....	46
5.1 Questions addressed.....	46
5.2 Main scenario.....	46
5.2.1 Use case 1a.....	46
5.2.2 Use case 1b.....	47
5.2.3 Background information: Europeana technical requirements.....	47
5.2.4 Actors.....	48
5.2.5 Open questions.....	49
5.3 Improving searching and browsing functionalities for AV.....	50
5.3.1 Use case 2.....	50
5.3.2 Background information.....	50
5.3.3 Open questions.....	50
6 Information enrichment for Access and Fruition.....	51
6.1 Context.....	51
6.1.1 Tagging game.....	51
6.1.1 The starting point.....	51
6.1.2 Question addressed.....	52

6.1.3 Users and roles.....	53
6.1.4 Description of cases.....	53
6.2 Creation of contextual data.....	53
6.2.1 The starting point.....	53
6.2.2 Questions addressed.....	54
6.2.3 Users and roles.....	54
6.2.4 Description of cases.....	54
7 The perspective of Service Providers.....	55
7.1 Service providers for preservation.....	55
7.1.1 Users and roles.....	55
7.1.2 Description of cases.....	55
7.1.3 Possible offered services.....	56
7.1.4 Metrics for service definition: service level agreements.....	57
7.2 Service providers for Search and Access.....	60
7.2.1 Users and roles.....	60
7.2.2 Services offered for search and access of digital contents.....	60
7.2.3 Reference protocols and standards for content search.....	61
7.2.4 Description of scenarios and use cases.....	61
7.2.5 Service level agreements (SLA).....	61
7.2.6 SLA Metrics.....	62
7.2.7 Integration with other services.....	63
7.2.8 Relationships with other roles.....	63
Glossary.....	64

## Scope

The PrestoPRIME main objective is to provide means allowing to keep audiovisual contents alive within an unlimited time scale, which means from now to the far future.

In order to achieve this ambitious goal, it is necessary to consider a wide range of tools, technologies, and environments, which can only be effective, even cost-effective, if they are integrated and deployed in an appropriated and focused manner. This is also because each integration solution has to be evaluated against its theoretical outcomes by means of tests and validation activities.

Generally speaking, the context in which the PrestoPRIME technology will operate is made on one hand of **preservation activities**, which provide infrastructures, strategies and processes for content protection, and on the other hand of **fruition activities**, in which the audio visual contents become useful and the whole system appears eventually as a beneficial asset. Furthermore the processes have to be mentioned by which the collections are enriched with incoming materials and also with new information (metadata) which might improve the fruition of already archived items.

However, a concrete implementation is only possible if a clear (and sufficiently detailed) view of how the needs, reflected from the actual situation of AV archives and libraries and their uses, can be rendered into business processes, with interrelated activities and decisions.

A way to build this view is through the development of **scenarios**, i.e. storylines which tell about contexts of use of a “system”, which can be used as a meeting-place between developers/technology-providers/system-integrators, and “consumers”, in order to permit actual satisfaction of the context needs.

Scenarios, inescapably, depend on the perspectives from which the context is perceived by a given “actor”; there are various types and subtypes of actors as organisations as well as individuals having specific roles in their organisations. So “broadcast archives”, “content creators”, “content distributors”, “institutions”, “service providers”, “business managers”, “technical managers”, “rights holders”, are possible stakeholders and all their views contribute to compose several “pictures” which will be used on the integration and evaluation workbenches.

## Executive summary

This document provides a set of scenarios with description of contexts and cases, to be used as the base reference in the subsequent stages of PrestoPRIME, like the integration and evaluation of the project's outcomes.

The adopted criteria have been that of taking into account the views of various actors, according to their role — broadcast archives, service providers, content creators, institutions, small archives, business and technical managers, rights holders, end users — and trying to consider the context evolution henceforward, starting from the current situation and drawing a path from now to the future.

The **preservation scenarios for broadcast archives** are complex because current preservation processes are affected by erstwhile decisions and activities, such as migrations of master material performed in the tape-based domain. Within PrestoPRIME, we consider the preservation of AV-media files, but many broadcast archives have still to deal with the issue of digitisation.

In the long term there is a divergence between archive choices and production choices on the AV formats. The archive is mainly responsible for preserving the master copy, but also for how to make material affordably and quickly available from archive to production. To have “access copies” ready in the most common exploitation format in advance is a possible strategy, although it entails the need to update these accordingly to the changing production requirements.

A set of questions are addressed on the topic of digitisation and ingest. Ingest is recognised to be a critical time for decisions which will affect the archival item subsequently. The selection of an initial archive format and the possibly required adaptation from the submitted format are the key factors, together with the initial validation and the setting of the future preservation policy. What is to be done with original AV properties (aspect ration, resolution, colour space) in case of format conversion?

Strategies are the subject to discussions about preservation models and storage. While the decision points, such as tool, format and option assessment, have been identified, there is a lack of criteria and related parameters required for the evaluation of trade-off and decision. Basically the comparison of cost is the suggested approach, although the cost estimation for each option is an actual difficulty.

Questions have been raised also on content based quality assessment, risk assessment and modalities for access and delivery.

The cases specifically identified consist of: the preservation of (1) materials digitised and kept as D10, (2) uncompressed materials, possibly originated in various ways, such as through playback and capture from SDI, (3) materials derived from Films, (4) SDTV materials with DV coding, and (5) materials currently produced in HDTV, such as XDCAM-HD422.

The access and fruition of archived AV items have always to face the issue of **rights clearance and handling**. This scenario addresses rights handling for the purpose of the general exploitation of the content of an AV-Archive, when conceived for long term preservation, but with the aim of being the object of continuous fruition.

The scenario addresses at first how rights information on an AV work should be prepared for submission to a preservation system, or for updating the same rights information set, as a consequence of some new rights negotiation.

The representation of rights information is taken into account with all the required relationships between rights and the timeline of an AV entity.

Other questions address how to support rights clearance in case of AV entities created with the re-use of pre-existing materials and how to support from the rights perspective the re-use of archival material for delivery to content creators.

It is also considered how information on rights can be supported in search & retrieval activities and how rights information originated by various organisations can be shared or compared.

There are many organisations with AV archives, such as **higher education institutions** (HEI) and **medium and small archives**, which operate AV preservation in contexts quite different from that of the typical broadcast archive.

Universities and generally Higher Education Institutions are systematically collecting, purchasing, and producing AV material. Researchers use AV as teaching objects for illustrative purpose, students for courses or even entertainment. HEI have generally an up-to-date IT infrastructure and dedicated staff, but there is a lack of best practise models for the digital preservation of multimedia objects.

In the case of HEIs, digital preservation implies dealing with a rather large number of formats and metadata models. The main questions are organisational ones, but they are also connected to a number of technical issues and address how to smoothly integrate an AV digital preservation process.

Many medium and small size archives typically lack one or several of the needed demands for preservation and exploitation. Not reaching the critical size to launch preservation and digitisation plans, they fail to make their contents accessible for commercial or cultural purposes.

This scenario aims on the approaches to exploitation models involving both large and small archives and the complementary actions that can be undertaken with mutual benefits for all participants. It is based on the visibility of an established large archive and the potential richness of small collections.

A “broker” approach, where a large archive provides a single point of access for a large amount of collections, and a “distribution agreement” approach with service extended to some or all preservation aspects, are proposed; main challenge is identifying conditions for the described approaches to be successful.

The vision for **Europeana** is to provide a unique point of access to Europe’s distributed cultural heritage. Audiovisual content is an important part of Europe’s collective memory depicting cultural, historical, artistic and socio-economic events that span the 20<sup>th</sup> and the beginning of the 21<sup>st</sup> century. Although in Europeana AV contents currently make up only 3% of the total of searchable digital objects, in the content strategy launched in summer 2009, a high priority has been announced to attract and make more AV content searchable and accessible.

For museums, libraries and archives to contribute content to Europeana, they need to comply with the technical requirements of Europeana, i.e. make available their metadata in a way compatible to the Europeana data specifications.

As it is practically impossible for Europeana to work individually with every European content provider, so Europeana works with an intermediate layer of content providers, the aggregators. Those aggregators collect metadata from their group of content providers; they have the possibility to apply multilingual tools and services to improve the usability of content. This gives the opportunity to rank content higher in search engines, attracting more users to the individual content provider’s sites.



This scenario addresses various questions related to the particularities of AV, such as searching and accessing a particular segment within an AV entity, in order to consider their accommodation to the new Europeana data model.

**Information enrichment** makes the basis of future long term exploitation of the AV archives. Information useful for permitting future access was already provided to the archive together with the original submission. The idea here is that such information can be continuously complemented and kept up-to-date with the contributions of the archive users by the use of new specific tools.

The related scenarios approach this objective addressing on one hand the user involvement with a “video annotation game”, and on the other hand the creation of contextual information, by speeding up manual annotation process with natural language processing techniques and resorting to professional or semi-professional users with knowledgeable lay (wiki-oriented Web 2.0 users).

In both cases the main questions deal with the issues of incentives, standards, estimation of added value, validation of results, instructions to be given to the users, modalities for crediting the contributors, modalities for using automatic tools, modalities for providing feedback within the archive process.

**The perspective of service providers** is split up between preservation, and search and access.

**Preservation Service Provider (PSP)**, in the context of this document, means an entity that provides an online service (access is over a network) for the long-term storage, preservation and access to AV assets owned by someone else. The concept doesn't need to involve a third party, because a department might provide a preservation service to various business units within the same company.

This scenario distinguishes between preservation services – such as ingest, access, integrity validation, format migration - and lower-level infrastructure services – such as storage, computing, and networking resources. It covers some, but not all, aspects of Media Asset Management.

In addition to listing a set of services, which might be used in combination to address the various use cases, this scenario specifically focuses on the issue of trustworthiness regarding the preservation performances, and points to the fact that, when the problems get evident, it might be too late to change decisions on the choice of a service provider. How can you entrust digital content to a provider for long-term safe-keeping and have little or no visibility of how that content is being stored? The negotiation of a service level agreement (SLA) between customer and provider of the service is crucial.

SLA terms, which contain obligations on both sides, have to cover various areas – such as performed functions, modalities, bounds of performance, metrics for assessment, penalties.

Complex **search** functionalities for accessing distributed AV materials can be provided to users through specific providers. On the users side such services should offer the possibility to submit queries in various modes – such as by full text, by descriptive or rights metadata, by content – while the search capabilities have to be supported by appropriate indexing services. Search services too are defined by SLA terms, which can be measured according to relevant metrics – such as service availability (up and running time), search response time, quality of results (recall and precision measurements).

## References

Bookmark	Reference information
D10Specs	SMPTE – SMPTE 356M - Type D-10 Stream Specifications – MPEG-2 4:2:2P @ ML for 525/60 and 625/50 – 2001 – <a href="http://www.smpte.org/standards">www.smpte.org/standards</a>
DCI	DCI - Digital Cinema System Specification Version 1.2 – March 07 2008 – <a href="http://www.dcinovies.com/DCIDigitalCinemaSystemSpecv1_2.pdf">http://www.dcinovies.com/DCIDigitalCinemaSystemSpecv1_2.pdf</a>
DCMT	DCMI – DCMI metadata terms – 2008-01-14 - <a href="http://dublincore.org/documents/dcmi-terms/">http://dublincore.org/documents/dcmi-terms/</a>
DHT	W. Allasia, F. Gallo, M. Milanese and R. Schifanella "Indexing and Retrieval of Multimedia Metadata on a Secure DHT", Informatica - Special Issue on Multimedia Information System Security, Vol. 33, No. 1, March 2009, pp. 85-100
DVSpecs	IEC – IEC-61834 Helical-Scan Digital Video Cassette Recording System Using 6,35 mm Magnetic Tape for Consumer Use - Part1: General specification - Part2: SD Format for 525-60 and 625-50 Systems – <a href="http://www.iec.ch">www.iec.ch</a>
	SMPTE – SMPTE 306M 6.35-mm Type D-7 Component Format —Video Compression at 25 Mb/s and 50 Mb/s — 525/60 and 625/50 – 2002 – <a href="http://www.smpte.org/standards">www.smpte.org/standards</a>
	SMPTE – SMPTE 314M Data Structure for DV-Based Audio, Data and Compressed Video 25 and 50 Mb/s – 2005 – <a href="http://www.smpte.org/standards">www.smpte.org/standards</a>
EBUPNP	EBU - P/NP Networked Production (web page last visited 2009-12-03) – <a href="http://tech.ebu.ch/groups/pnp">http://tech.ebu.ch/groups/pnp</a>
EDLF	EDLF – European Digital Library Foundation Purpose (web page last visited 2009-12-15) - <a href="http://dev.europeana.eu/edlnet/edl_foundation/purpose.php">http://dev.europeana.eu/edlnet/edl_foundation/purpose.php</a>
EFG	ACE and EDLF – European Film Gateway (web page last visited 2009-12-15) - <a href="http://www.europeanfilmgateway.eu">http://www.europeanfilmgateway.eu</a>
EGUID	Europeana 1.0 - Metadata Mapping & Normalisation Guidelines for the Europeana Prototype version 1.2 – 2009-08-07 - <a href="http://www.version1.europeana.eu/c/document_library/get_file?uuid=58e2b828-b5f3-4fe0-aa46-3dcbc0a2a1f0&amp;groupId=10602">http://www.version1.europeana.eu/c/document_library/get_file?uuid=58e2b828-b5f3-4fe0-aa46-3dcbc0a2a1f0&amp;groupId=10602</a>
ESE	Europeana 1.0 – Europeana Semantic Elements specifications version 3.2.1 – 2009-11-06 - <a href="http://version1.europeana.eu/c/document_library/get_file?uuid=c56f82a4-8191-42fa-9379-4d5ff8c4ff75&amp;groupId=10602">http://version1.europeana.eu/c/document_library/get_file?uuid=c56f82a4-8191-42fa-9379-4d5ff8c4ff75&amp;groupId=10602</a>
EUR	European Union - Council Conclusions on the Digitisation and Online Accessibility of Cultural Material, and Digital Preservation – 2006-12-07 - <a href="http://eur-lex.europa.eu/LexUriServ/LexUriServ.do?uri=CELEX:52006XG1207%2801%29:EN:NOT">http://eur-lex.europa.eu/LexUriServ/LexUriServ.do?uri=CELEX:52006XG1207%2801%29:EN:NOT</a>
HDTVspecs	EBU – Tech 3299 - High Definition (HD) Image Formats for Television Production – 2004 - <a href="http://tech.ebu.ch/docs/tech/tech3299.pdf">http://tech.ebu.ch/docs/tech/tech3299.pdf</a>

Bookmark	Reference information
	<p>ITU – BT 709 - Parameter values for the HDTV standards for production and international programme exchange - 2002 – <a href="http://www.itu.int/rec/R-REC-BT.709">http://www.itu.int/rec/R-REC-BT.709</a></p>
	<p>SMPTE – SMPTE 274M - 1920 x 1080 Image Sample Structure, Digital Representation and Digital Timing Reference Sequences for Multiple Picture Rates – 2008 - <a href="http://www.smpte.org/standards">www.smpte.org/standards</a></p>
	<p>SMPTE – SMPTE 296M - 1280 x 720 Progressive Image Sample Structure – Analog and Digital Representation and Analog Interface – 2001 - <a href="http://www.smpte.org/standards">www.smpte.org/standards</a></p>
JPSEARCH	<p>ISO/IEC JTC1/SC29/WG1 WG1N4108, "JPSearch Executive White Paper", December 2006, See also "Overview of JPSearch: a Standard for Image Search and Retrieval" Dufaux, Frederic ; Ansorge, Michael ; Ebrahimi, Touradj Presented at: 5th International Workshop on Content-based Multimedia Indexing (CBMI-2007), Bordeaux, France, June 2007.</p>
MPGSpecs	<p>ISO (and ITU) - ISO/IEC 13818-2 (aka ITU-T H.262) - Generic coding of moving pictures and associated audio information: Video – 1996 - <a href="http://www.iso.org/iso/iso_catalogue/catalogue_ics/catalogue_detail_ics.htm?csnumber=22990">http://www.iso.org/iso/iso_catalogue/catalogue_ics/catalogue_detail_ics.htm?csnumber=22990</a> <a href="http://www.itu.int/rec/T-REC-H.262">http://www.itu.int/rec/T-REC-H.262</a></p>
	<p>ISO (and ITU) - ISO/IEC 14496-10 (aka ITU-T H.264) Coding of audio-visual objects -- Part 10: Advanced Video Coding – 2003 - <a href="http://www.iso.org/iso/iso_catalogue/catalogue_ics/catalogue_detail_ics.htm?csnumber=37729">http://www.iso.org/iso/iso_catalogue/catalogue_ics/catalogue_detail_ics.htm?csnumber=37729</a> <a href="http://www.itu.int/rec/T-REC-H.264">http://www.itu.int/rec/T-REC-H.264</a></p>
MPQF	<p>ISO/IEC 15938-12. See also Matthias Gruhne and Ruben Tous and Mario Döller and Jaime Delgado and Harald Kosch MP7QF: An MPEG-7 Query Format In Proceedings of the 3rd International Conference on Automated Production of Cross Media Content for Multi-channel Distribution (AXMEDIS 2007), November 2007, Barcelona, Spain</p>
MXFSpecs	<p>SMPTE – SMPTE 377M - Material Exchange Format (MXF) - File Format Specification – 2009 – <a href="http://www.smpte.org/standards">www.smpte.org/standards</a></p>
	<p>SMPTE – SMPTE 378M – Material Exchange Format (MXF) – Operational Patter 1a (Single Item, Single Package) – 2004 – <a href="http://www.smpte.org/standards">www.smpte.org/standards</a></p>
	<p>SMPTE – SMPTE 379M – Material Exchange Format (MXF) – MXF Generic Container – 2009 – <a href="http://www.smpte.org/standards">www.smpte.org/standards</a></p>
	<p>SMPTE – SMPTE 381M - Material Exchange Format (MXF)– Mapping MPEG Streams into the MXF Generic Container – 2005 – <a href="http://www.smpte.org/standards">www.smpte.org/standards</a></p>
	<p>SMPTE – SMPTE 382M - Material Exchange Format – Mapping AES3 and Broadcast Wave Audio into the MXF Generic Container – 2007 – <a href="http://www.smpte.org/standards">www.smpte.org/standards</a></p>
	<p>SMPTE – SMPTE 383M - Material Exchange Format (MXF)– Mapping DV-DIF Data to the MXF Generic Container – 2008 – <a href="http://www.smpte.org/standards">www.smpte.org/standards</a></p>

Bookmark	Reference information
	<p>SMPTE – SMPTE 384M - Material Exchange Format (MXF)– Mapping of Uncompressed Pictures into the Generic Container – 2005 – <a href="http://www.smpte.org/standards">www.smpte.org/standards</a></p>
	<p>SMPTE – SMPTE 386M – Material Exchange Format (MXF) – Mapping Type D-10 Essence Data to the MXF Generic Container – 2004 – <a href="http://www.smpte.org/standards">www.smpte.org/standards</a></p>
	<p>SMPTE – SMPTE 387M – Material Exchange Format (MXF) – Mapping Type D-11 Essence Data to the MXF Generic Container – 2004 – <a href="http://www.smpte.org/standards">www.smpte.org/standards</a></p>
	<p>SMPTE – SMPTE 390M - Material Exchange Format (MXF) – Specialized Operational Pattern “Atom” (Simplified Representation of a Single Item) – 2004 - <a href="http://www.smpte.org/standards">www.smpte.org/standards</a></p>
	<p>SMPTE – SMPTE 422M - Material Exchange Format – Mapping JPEG 2000 Codestreams into the MXF Generic Container – 2006- <a href="http://www.smpte.org/standards">www.smpte.org/standards</a></p>
OAIPMH	<p>The Open Archives Initiative Protocol for Metadata Harvesting (OAI-PMH) is a low-barrier mechanism for repository interoperability. Data Providers are repositories that expose structured metadata via OAI-PMH. Service Providers then make OAI-PMH service requests to harvest that metadata. - <a href="http://www.openarchives.org/pmh/">http://www.openarchives.org/pmh/</a></p>
OAIS	<p>CCSDS - Reference Model for an Open Archival Information System (OAIS) – January 2002<sup>2</sup> - <a href="http://public.ccsds.org/publications/archive/651x0b1.pdf">http://public.ccsds.org/publications/archive/651x0b1.pdf</a></p>
PSDM	<p>PrestoSpace - Examples of Editorial Object Document Format (web page last visited 2009-12-15) – 2008-01-31 <a href="http://www.crit.rai.it/attivita/PrestoSpaceFormats/PrestoSpaceEdobDocs.html">http://www.crit.rai.it/attivita/PrestoSpaceFormats/PrestoSpaceEdobDocs.html</a></p>
PSD19.01	<p>PrestoSpace – D19.01 - External and Internal Models and Protocols for the PrestoSpace Factory – 2008-02-25 - <a href="http://prestospace.org/project/deliverables/PS_WP19.pdf">http://prestospace.org/project/deliverables/PS_WP19.pdf</a></p>
QTFF	<p>Apple – QuickTime File Format Specification – 2007-09-04 - <a href="http://developer.apple.com/mac/library/documentation/QuickTime/QTFF/qtff.pdf">http://developer.apple.com/mac/library/documentation/QuickTime/QTFF/qtff.pdf</a></p>
SDTVspecs	<p>ITU – BT 601 - Studio encoding parameters of digital television for standard 4:3 and wide screen 16:9 aspect ratios - 2007 – <a href="http://www.itu.int/rec/R-REC-BT.601">http://www.itu.int/rec/R-REC-BT.601</a></p>
VEL	<p>Velsen, L., M. van ans Melenhorst – “User Motives for Tagging Video Content” - Adaptation for the Social Web workshop, Hannover, Germany – 2008 - <a href="http://www.lexvanvelsen.nl/VanVelsen&amp;Melenhorst2008.pdf">http://www.lexvanvelsen.nl/VanVelsen&amp;Melenhorst2008.pdf</a></p>
Z3950	<p>Z39.50 is a national and international (ISO 23950) standard defining a protocol for computer-to-computer information retrieval</p>

<sup>2</sup> A new version of this document is under approval procedure in CCSDS. The concepts referenced in this deliverable are not affected by the forthcoming release.

## 1. Introduction

### 1.1 *Adopted approach*

A characteristic of this work, within such a complex project as PrestoPRIME, is that of having to provide some reference to the discussions carried out in various tasks, quite before the official release date of the deliverable. This issue of task dependencies has constrained the approach adopted for the development of this definition of scenarios. The main positive consequence has been the involvement of a large number of partners for contributions, feedbacks, opportunities of discussion, presentations in workshops.

The first phase of the work has consisted in the organisation of an internal workshop for setting the scene, a collaborative development of ideas by the use of a wiki, and the development of a first set of scenarios. The goal of the first phase was to provide usable outcomes for the other workpackages with the shortest delay.

The internal workshop was a sort of kick off of the activity and was focused on two main objectives:

**Presentations of tools and technologies** – in order to share the understanding about the possibilities given by the state of the art referenced in the description of work of the project.

**Sketching the first set of scenarios** – this was for identifying the topics and the contexts which could be covered, and for assigning tasks to the possible contributors.

The wiki tool was initially used for collecting the outcomes of the first workshop; afterwards each individual contributor could provide his/her description of a scenario based on the specific perspective of his/her organisation. The advantage of the collaborative tool has been that of offering the opportunity to provide additions, modifications, suggestions and comments without the delay of individual editorial work and distribution. The style of contributions was informal.

### 1.2 *Progress of work*

#### 1.2.1 **Feedbacks from workpackages and first draft**

The wiki tool was also used by the contributors of the technical workpackages for discussing and providing the relevant feedbacks. Their criteria were based on identifying, within the outlined scenarios, indications useful for each area of their work.

The benefit for the development of the scenarios was a first recognition of gaps and a better understanding of how to address questions related to the forthcoming work.

From the other perspective the persons in charge of the WP tasks could take notice of the first main clarifications of contexts and concepts.

This fruitful activity provided the conditions for the development of the first working draft.

#### 1.2.2 **Criteria**

It was necessary to transform a set of individual or organisational contributions into a set of scenarios suitable for wider contexts. The adopted criteria in addition to taking into account the views of the various actors, according to their role (broadcast archives, service providers, content creators, institutions, small archives, business and technical managers, rights holders, end users), was to consider the context evolution henceforward, starting from the current situation and drawing a path from now to the future. This means to regard at both the available and upcoming technical solutions.

It was attempted to find a balance between the more general cases and some interesting specific ones, always strengthening the references to standards and open approaches.

The meaningfulness of the cases was also estimated according to related quantities (e.g. spread of a format), the countries in which they are relevant, the wideness of the scope.

### 1.2.3 Workshop with the user group

In order to improve the quality of the work, to involve much more the partners in the common activities, and to foster the relationships between the Project and our target user group, a specific workshop was organised, during a dissemination event held on 5<sup>th</sup> - 6<sup>th</sup> October 2009 in Vienna, where about 100 persons were attending.

The second day was specifically about scenarios which were presented to the participants grouped according to four areas:

- **AV preservation** – in particular for broadcast archives
- **Service providers** – an attempt to provide a definition and measure of “trust”
- **Access & Fruition** – from various perspectives
- **Other archives** – the problems of 70% of European AV content beyond broadcast archives

The workshop gave the opportunity to collect and discuss interactively several inputs in the form of:

- identification of real current problems
- report of experiences
- perceptions of the future needs
- what may be expected from PrestoPRIME

## 1.3 Finalisation

The harvested contributions, of various kinds, required then to be re-organised within a single document.

Chapter 2 tackles the big area of AV preservation by the broadcast archives and includes contribution from the current experience of BBC, RAI, B&G, and ORF.

Chapter 3, mainly based on RAI contributions, provides a description of a set of rights clearance scenarios, showing how an AV professional organisation has to deal with rights information in order to support access and fruition.

Chapter 4 gives the overall perspective of other organisations holding AV collections. Section 4.1, contributed by University of Innsbruck, gives details on the situation of HEIs (higher education institutions), while section 4.2, contributed by INA, analyses the possible exploitation models of medium and small AV archives.

The scenario depicted in chapter 5 is about the modalities for a general networked access to AV content, through more or less specialised points of access. The major contributor of this part was EDL Foundation with inputs from BBC and RAI.

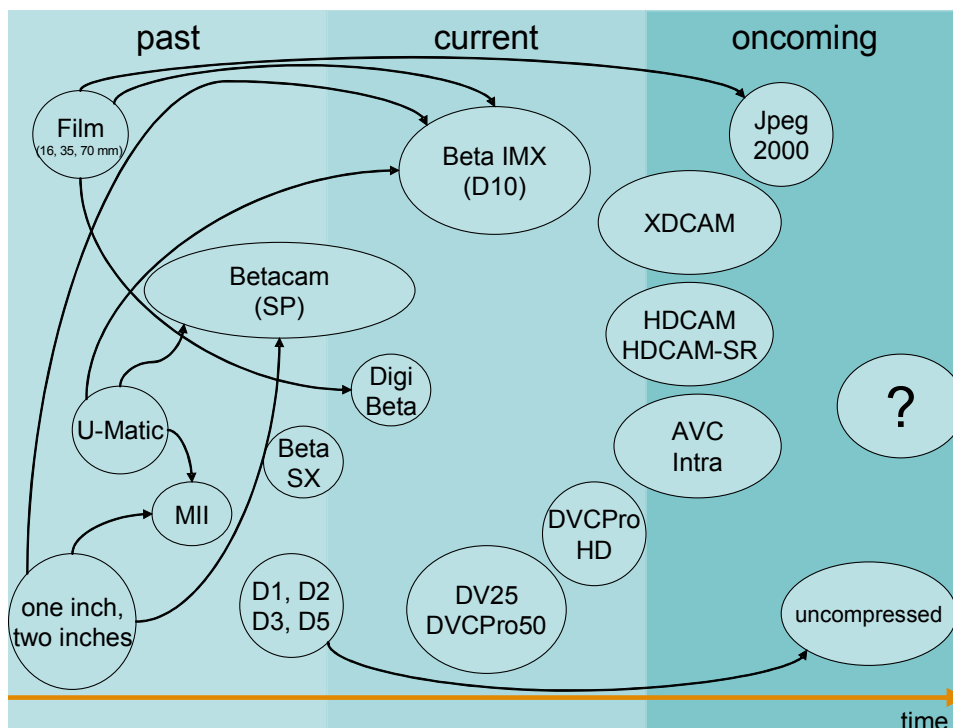
In order to keep access and fruition up-to-date along time, a couple of scenarios on information enrichment, input from B&G and VUA are presented in chapter 6.

Eventually the perspectives of service providers, regarding both preservation (section 7.1, ITI) and search and access (section 7.2, Eurix) are given in chapter 7.

## 2 Preservation for Broadcast Archives

### 2.1 Setting the scene

The introduction of the digital technologies in the areas of production and publication has caused important changes in the broadcast domain. The resulting scenario is quite complex because the initiatives undertaken to tackle preservation by the various broadcaster organisations result from the historical background of each of them.



**Figure 1 - scenario of formats in the broadcast domain along time**

The current preservation process for the broadcast archives is affected by processes occurred in the past, such as migrations of master material performed in the tape-based domain (even in the analogue domain), and on the current solutions and equipments needed for recording, storage, and playback. This scenario is about preservation of AV-media files, but many broadcast archives still have to deal with the issue of digitisation.

#### 2.1.1 Reference quality levels

In PrestoSpace, D19.0.1 [PSD19.01], three levels of quality for AV materials are defined, summarised here:

**Master quality level** – the highest quality<sup>3</sup>, used as reference, and from which any equivalent or lower quality may be derived.

**Broadcast quality level** – the quality appropriate for publication through “current” audiovisual publication media.

**Browsing quality level** - the quality is sufficient to appreciate and recognise the content, albeit some impairments might be quite perceptible. Those materials are often used as “proxy”.

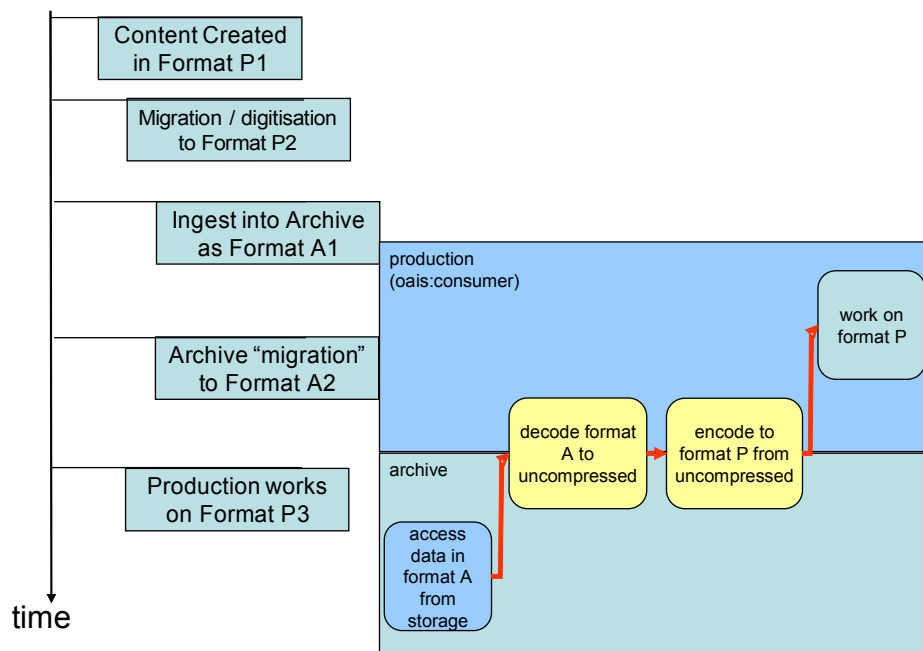
In PrestoPRIME the concept of “long term preservation” refers to the Master quality level.

<sup>3</sup> it shouldn't be taken for granted that the quality is definitely high, because the content might have been created in a low quality format or better quality copies may not be (anymore) available

The **Exploitation quality level** is defined here as the quality appropriate for simple production activities and subsequent transcoding to most publication formats. This is an intermediate between Broadcast and Master.

### 2.1.2 Archive and production

The problem of divergence between archive choices and production choices in the long term is sketched in Figure 2. The question of “delivery burden” – how to make material available from archive to production affordably and quickly – introduces the possibility of preparing in advance **Access copies**, typically at exploitation quality level, which entails further format migration issues (need to re-create them in a different format), together with concerns about the delivered technical quality with respect to the production needs.



**Figure 2 - archive vs production divergence and "delivery burden"**



## 2.2 Questions addressed

The raised questions can be grouped according to which area they address.

### 2.2.1 Digitisation and Ingest

Ingest [Ingest] is the process by which AV-content is taken under the responsibility of an Archive Service and preserved within an AV-media File. When the source of AV-content is not in form of a media file<sup>4</sup> (e.g. it's recorded on video-tape), a digitisation process [Digitisation] is required (immediately) before ingest. Questions to be taken into account are presented in Table 1

topic	questions
selection of ingest AV format <sup>5</sup>	How to take into account the current production format, the original quality level of material, the costs of the required storage capacity, the impacts on the delivery processes?
SIP and AIP	Should the process of transforming a SIP into an AIP include a format conversion for the Master copy?
quality levels	Which other quality levels should be produced within the ingest process?
mismatches	What are the impacts of previous preservation activities in the non-file-based domain ? e.g. when the apparent quality of AV content and the quality of playback during the digitisation process are not consistent. What are the consequences of mismatches between digitisation and ingest criteria? In case, from ingest perspective, the digitisation should have be done differently.
digitisation logs and SIP	Which outcomes of the digitisation process (i.e. parameters, configurations, logs and reports) should/must be provided to ingest (and thus included into SIP)?
ingest flows	For each item to be archived there will be the need to track the ingest flow from which it is originated and to be able to report on the integrity and authenticity of the item. Can there be variants of SIP formats according to the type of ingest flow? And who will be in charge of the burden of adaptation, if required?
exceptions	What interaction with the Provider for exception management? What to do with format validation problems? Correct or Reject? What to do if required information is missing? Default values or Reject? What to if format parameters differ from the agreed ones?
setting of preservation policy	How to associate each item to be archived to a defined (or new) preservation policy? Is that part of SIP? Including: rules and permissions to update rights information and descriptive information, migration options and SLA for access/delivery

**Table 1 - questions addressed on digitisation and ingest**

<sup>4</sup> Content created within a file-based production environment is born digital and doesn't require digitisation

<sup>5</sup> both wrapper and AV coding/compression scheme

## 2.2.2 Preservation of AV properties

Along the history of AV formats some technical properties of AV materials that had been stable during a certain period of time then changed. The issue of their preservation and/or how and with which criteria they should be treated within migration and delivery processes, has to be considered. For the properties mentioned below, the questions raised are: keeping the original properties or not? What to do with these properties in case of format conversion? How to deal with these properties and digitisation and/or ingest time? How to deal with properties which may produce distortions or troubles for presentation?

component	properties
Video	lines per picture and (for digital formats) pixels per line image aspect ratio – for SDTV it's typically 4:3 and for HDTV is typically 16:9, presentation issues depends also what is supported by the screen pixel aspect ratio <sup>6</sup> . HDTV formats have squared pixel, according to current screens and displays. This is not the case for SDTV. frame rate, interlaced/progressive number of bits per sample (8bits, 10bits) Colour space
Sound	monaural, stereo, 5.1 frequency range, sampling rate and bits per sample

**Table 2 - AV properties to be considered for preservation**

## 2.2.3 Preservation models and strategies

The main questions raised within this area deal with the identification of criteria and related parameters useful for the configuration of a preservation system for broadcast archives. Models permitting the evaluations of trade-off points are required. A model describing the process (and the related costs) of the long term maintenance of a decoder tool is also required; this is necessary for a correct cost comparison between the migration processes and the development of the Multivalent solution.

Figure 3 shows a migration flow chart specifically about the **video coding format**, in which the main questions, located just before the decision points (diamonds), are:

1. How to assess a tool or application as working reliably with a given format?
2. In case of concerns: which are the criteria for deciding an approach on the tool/application? The broadcaster archives have some knowledge about the preservation of a working environment for an existing software tool, such as a decoder, through the virtualisation of the operating system, but they have no experience of the Multivalent approach (which has been/is object of study in Project SHAMAN)
  - Models indicating decision thresholds for the Multivalent option, and also taking into account the development costs for the various considered format, are required;
  - Multivalent provides the possibility of including innovative “behaviours” which could enhance the output of the mere decoding process (noise removal, concealment of impairment); how should these potential advantages be taken into account?

<sup>6</sup> this quantity depends on number of pixels and image aspect ratio

3. How to assist decisions on migration?

- How to balance between size and quality?
- How to consider migration costs and tool/application costs?
- Under which conditions is migration to uncompressed formats a profitable option? In which cases could it be recommended now? Which would be the access/exploitation issues?

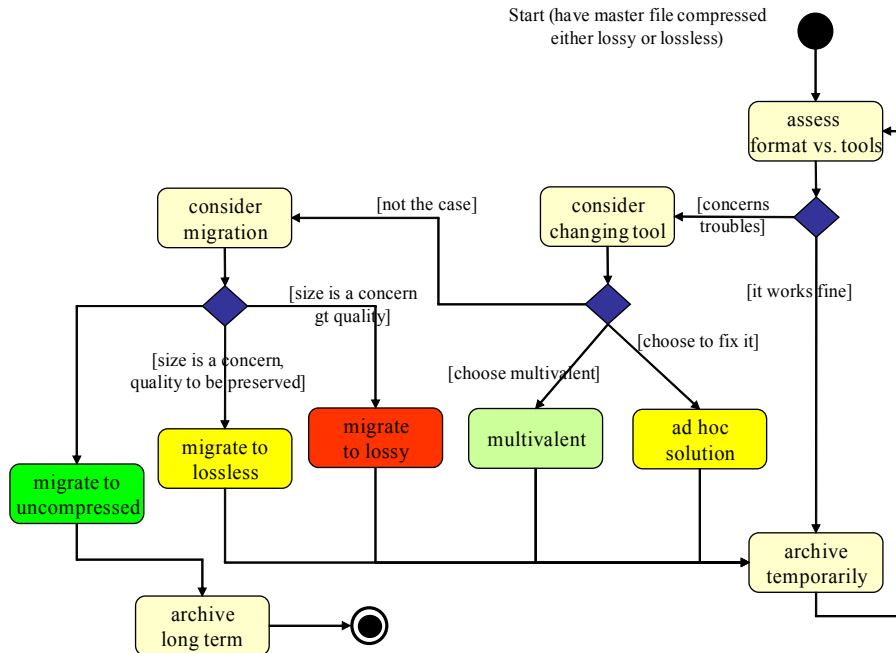


Figure 3- the migration game: a flow chart for keeping digital video content alive

Some broadcast archives are quite conservatives regarding the format of their master material; they are focused on the benefits they can have “here and now” rather than on the possible future threats. How to ensure that the conditions for a “no migration/keep original format” approach are fulfilled? How can the situation be identified in which a migration to a lossless format (such as JPEG2000) would be beneficial?

In the future broadcasters will be tempted to migrate to the new production format, likely lossy/high end, in order to take the (immediate) advantage of a simpler delivery process; how to also introduce in the decision factors related to longer term costs?

**The concerns about the wrapper formats** are of great importance too, with respects to the reliability of tools and applications. Many current compatibility problems are due to wrapper variants and to limited support of tools and applications to the specified format features. Transwrapping [Transwrapping], by definition, doesn’t affect the content quality, nevertheless important information (metadata) might be lost.

Currently, the broadcast archive domain is focused on the MXF wrapper format [MXFSpecs], which actually consists of several sub-cases, according to specified operational pattern and mapping to the specific types of content.

In other domains [Chapter 4] MXF is not widespread and QuickTime [QTFF] will likely be the starting point.

## 2.2.4 Storage technologies and strategies

The current preservation activities within a broadcast archive don't give clear indications about storage strategies except for the concerns about the required investments and that of supporting the delivery to the production environment. Questions to be taken into account are presented in Table 3.

topic	questions
business and logistic domain	Which are the criteria and parameters for deciding about outsourcing storage services? Which costs models should be considered? which are the guidelines for defining SLAs? How accurately can the storage costs be predicted? How to evaluate changing a provider with the costs and effort for relocation?
technical solutions	How to assess the storage service reliability for access and data integrity? How should data redundancy and data distribution be set up? Which storage technologies can be recommended and in which cases?
infrastructure	How to consider the costs of elements such as location, power supply, cooling, and networks? How do the impacts of these parameters change according to variations of contexts?
storage capacity	How to correctly evaluate the possible advantages of optimisation efforts in the use of storage capacity?
AV peculiarities	Which pre-process should be recommended (from the storage service perspective) in order to minimise the risk of data loss and for optimising the delivery services? On which extent should the storage services be AV-aware?

**Table 3 - questions addressed on storage technologies and strategies**

## 2.2.5 Content based quality assessment

The content based quality assessment permits the identification of a large set of problems which can occur along all the life cycle of the products, from capture to (post)production till broadcasting/publication. Some questions are presented in Table 4.

topic	questions
digitisation / ingest	Quality assessment at the ingest step has the objective of ascertain the quality of received material, which is the result of all previous processing. How should the quality assessment affect the ingest process?
master migration	How can content based quality assessment contribute to the validation of a migration process? Is this limited to lossy format?
transcoding for access/delivery	How can content based quality assessment contribute to the validation of a transcoding process?
video	Which are the most important video quality parameters to be recorded? Black frames, frozen video, bit-rate sufficiency? Synchronisation with audio?
audio	Which are the most important audio quality parameters to be recorded? Levels, loudness? Synchronisation with video?

**Table 4 - questions addressed on content based quality assessment**

### 2.2.6 Risk assessment

By understanding how and why data is lost, the persons in charge of archive responsibilities are able to validate existing processes, to plan and prepare for new ones, and to anticipate future problems. There are several dimensions in which the broadcasting archives feel the need to address this topic:

**“Label” of being a “trusted repository”** – broadcast archives want to get this because it is beneficial for requests of funding or for offering services to third parties, such as small archive (see §4.2); how can this rating be obtained? Which are recognised criteria for a good risk assessment? Which innovative criteria should be recommended to be considered as well?

**Risk assessments and SLA** - as the organisations are often in the position of being customers of service providers for crucial parts of their preservation process, how can the risks assessment rating be included in the definition of SLAs and as a component of the criteria for partner selection?

**Time window for risk analysis** - the risk assessment run internally can produce the definition of a risk management process focus on the long term AV preservation objective. Which is the most appropriate time window for the risk analysis?

**Difficulties for effective risk management implementation** - Every event, expected or not, that pushes the Archive managers to decisions having a long term impact, should trigger a comparative risk analysis among the options in order to get a further decision element; how to tackle the difficulties of implementing an effective risk management?

- The identification of threats is at risk to be incomplete; how can the work already done in other projects such as DRAMBORA, TRAC CC, OAIS, be considered?
- The estimation of the probability of occurrence of a given threat is often not trusted as being significant. Every highly probable risk in the long term becomes certitude, while the low probability risks tend to be ignored. How can the reliability of such models be estimated?
- The estimation of impact is also far from being precise in all cases; how can improvement be obtained?

**Once a risk is identified** - the possible options are: (1) to reduce the probability of the threat, (2) to set up countermeasures for repairing or minimising the consequences, (3) quiet acceptance; how are the decisions supported by models for costs comparisons?

### **2.2.7 Relationships with providers**

Broadcast archives dealing with service and equipment providers expect guidelines and criteria for taking decisions and for evaluation of the following topics and questions:

- model and criteria for resorting to external providers
- guidelines for contracting SLA and how to implement SLA monitoring (by broadcast archive)
- models and criteria for estimating the reliability of an offered service
- how to ascertain the conformance of the output of required equipment/software to the relevant format standards? or the same for the tool itself?
- how to ascertain that the output of provided tools will be usable with newer versions of the same tool?
- how to establish/contract SLA for maintenance services?

### **2.2.8 Access and delivery**

Content creators and content distributors are, from OAIS model [OAIS], the professional consumers of the broadcast archive contents. They expect answers for the following issues.

It is required to find easily what is known to be in the archive, the possibility to have fast preview of material for identifying relevant segments and estimating the editorial and technical quality.

The modalities and formats of access have to be defined. The resources required to transwrap/transcode formats from Archive to Production environments have to be identified, i.e. will the task be in charge of the Archive or of the Consumer? How much are the strategies on archive formats (ingestion, migration) depending on the Production formats (which changes over time)? The delay from request to availability for Production is crucial and, if the waiting time for completing the process is considerable, then it is required to have at least partial access to material with the shortest delay. How to get archive material to production affordably and quickly?

## **2.3 The role of broadcast archive**

### **2.3.1 Mission in the organisation**

The broadcast archive is a main “business actor” into the process model of the audiovisual industry, together with the “Content Creator” and the “Content Distributor”. The mission of the broadcast archive is to preserve existing AV material and to make it accessible for re-use either as originally produced or as input material for new AV productions. The broadcast archives are continuously enriched with new items coming from the other two business actors.

The peculiarity of the broadcast archives making reference to the national broadcasting organisations (in particular the public ones) is that they are the repository of content created along with the full history of Television.

### **2.3.2 The archive manager**

The archive manager is responsible for taking decisions on AV archive formats and on preservation strategies in general.

This role is concerned on one hand with budget issues, dealing with the balance between available funding and operating costs, and on the other hand with the mission of providing an up and running archive service in terms of availability and usability of material, including the technical quality.

### 2.3.3 Other roles within broadcast archives

The broadcast archives include various professional roles which may vary from organisation to organisation depending on how the archive process is currently configured. An incomplete general list will contain:

- **archivists** – able to control the processes, organise the work (batches) and manage exceptions
- **annotators** – able to produce, validate, and manage the information about archive content, the archive inventory and indexes
- **AV-technicians** – skilled at dealing with AV material and related equipment. More specific is “maintenance staff”
- **infrastructure staff** – people attached to run and maintain any required infrastructures (IT, networks, power, cooling)

## 2.4 Relationships with other roles

### 2.4.1 Content creators and content distributors

**Content creators** need to get AV material from the Archive as an input to newly created items. Once a production is complete, they have to export from the production environment to the Archive, unless they deliver exclusively to the “Content distributors”.

**Content distributors** have to get Archive items, in order to perform new publication, either on traditional media (broadcasting) or on any other alternative publication mechanisms (DVD, Internet, VOD...).

The broadcasting format is usually different from Archive and Production formats, however the encoding to this format is typically performed by the broadcasting system itself. So the main issue is the capability of On-air server to import archive format minimising the number and/or the cost of transwrapping [Transwrapping] and/or transcoding [Transcoding] processes. The situation w.r.t. this issue might be different for other types of publication.

### 2.4.2 Providers of services, equipment and software

Broadcast archives can be seen as “Preservation Service Providers” for content creators and distributors. However the broadcast archive can decide to rely on services provided by other organisations, either inside or outside the company, which can be classified as follow:

- **storage services** – providing reliable storage capacity.
- **services for data transfer** – here in terms of network and bandwidth for ingest and access processes
- **processing services** – for performing the transwrapping and transcoding processes required for migration and delivery, and for other computational activities related to preservation, such as format validation and check of data integrity
- **management services** – for running and monitoring the services and for handling the exceptions

The relationship between the broadcast archive and the provider department or organisation should be based on Service Level Agreements (SLA). (see §7.1.4)

The broadcast archives which decide to run archive services by themselves have to deal with manufacturers, sellers and generally providers of the necessary equipment and/or software, which may be either specific to the professional AV domain or to the general purpose IT domain.

## **2.5 Description of cases**

### **2.5.1 Preservation of material digitised as D10**

The digitisation interface is **eVTR**<sup>7</sup>, which means that the player device has an Ethernet interface through which it provides an FTP service, so that the AV-content recorded on the tape is virtually available as MXF/D10 file that can be downloaded. If the bandwidth on the network is sufficient and the host storage is fast enough, the whole download can be obtained with a single playback in real-time, otherwise the player automatically stops, empties the transfer buffers, and resumes correctly the reading of the tape.

The source format can be any kind of Betacam tape. However it must be remarked that while for IMX tapes the AV content is delivered AS IS (no transcoding), for analogue tapes the process implicitly includes A/D conversion and coding.

The target format is MXF/D10 file, a variant of MXF file format as defined by SMPTE 386M, which specifies how to store SMPTE D10 essence data in MXF using the Generic Container (SMPTE 379M). The essence data format for D10 is specified by SMPTE 356M and is composed of MPEG-2 Video 4:2:2 I-frame only (at 50Mbps) and one 8 channel AES3 audio stream. The video coding is of type "lossy/high end"

The digitisation process can be preceded by a tape cleaning process, providing a cleaning report, and the playback/download process can provide log information about RF levels and dropouts.

The whole workflow can be automated, including handling of tapes by a robotic, and have a good productiveness (hours digitised per day), but the quality of the process requires a good detection and management of exceptions.

A validation of MXF file is required and concealment can be necessary.

Manual quality check (person viewing result helped by the collected data) can be added.

The storage required for the MXF/D10 media file is about 30GB/hour.

The ingest process subsequent to digitisation includes a mapping between Material and Editorial Entities [PSDM] and it is possible to trim MXF files in order to both get a one-to-one relationship between Editorial Objects and AV-media files while reducing the AV data size by removing outlying useless parts (colour bars, black video, recording of rewinds).

Browsing quality materials, or copies at other quality level, can also be produced during the ingest process.

The set up of a storage strategy requires a balance between these requirements: (1) the preservation of master must be effective, (2) access to master must be granted to current production environment, (3) financial resources must be available.

The very preliminary solution is to store the MXF files on data tape (e.g. LTO4) (with redundancy copy) and let the production to be run on video-tapes as before digitisation, until a complete file-based production environment is set up.

---

<sup>7</sup> all the alternative interfaces for digitisation of Betacam tapes as D10 require a corresponding AV interface on the computer host and a (software) tool for D10 encoding.



The choice of MXF/D10 as archive format is suitable also for the case of delivering a copy of Master by means of XDCAM Professional Disc (no-transcoding), as the XDCAM devices are becoming more and more used in both SD and HD television production environments.

However until an online or near-on-line storage service is eventually in place, the delays for getting access to material from off-line data tape are not generally acceptable.

The strategy is to introduce on-line copies of the Master AV-media files, initially for limited contexts of use, which can be widened gradually once economically viable and supported by network infrastructure.

The long term capability to provide access to MXF/D10 files and their Mpeg2 video streams is based on the fact that the formats are open standards (SMPTE and ISO) with currently available software tools for unwrapping and decoding as open-source, and on the supposed capability to make them running and working also on future systems. This capability has to be verified periodically.

## 2.5.2 Preservation of material as uncompressed

Materials digitised as uncompressed may be originated by one of the following sources:

**acquisition of analogue sources** – the A/D conversion was performed by input hardware equipped with analogue AV interfaces (e.g. components)

**acquisition through SDI** of:

- digital component sources – the uncompressed AV data are transferred through SDI without bit change. Examples of sources are D1 (8bits) and D5 (10 bits) tapes
- other digital sources – where the AV data transferred through SDI are obtained by the player with bit change (decoding and/or signal transformation). The actual source might have been encoded with compression, such as Digital Betacam, or coded in another manner, such as D3 which is an uncompressed digital composite (PAL).

The differences in digitisation also involve the audio, w.r.t. to the number of tracks and the fact of being originally digital or analogue.

The main characteristics of this case are:

**resulting file size** – taking as a reference the video as 4:2:2 digital component, recording only the active part, and assuming four audio channels PCM uncompressed, the resulting required storage capacity is about 76GB/hour (while simply recording the whole SDI data, supposing embedded audio would require about 120GB/hour)

**file format options** – there are various options for saving AV uncompressed, thus for each digitisation activity there could be a specific decision about the format of the wrapper. This carries to issues of format homogeneity with possible need of transwrapping processes. Possible wrappers for uncompressed (1) **MXF / Uncompressed** – for example according to SMPTE 384M GC-UP and SMPTE 382M GC-AESBWF (2) **AVI** (3) **Quicktime/MOV** [QTFF].

Possible mechanism for delivery to production are: (1) print-to-tape or other carried through SDI; (2) to production environment through SDI; (3) transcoding on demand to production file format; (4) delivery of access copies in the production file format (exploitation quality level) instead of uncompressed master.

### 2.5.3 Preservation of material derived from 90's HDTV

This is actually a sub-case of Preservation of material as uncompressed.

In the late eighties-nineties in Europe, mainly for experimental purposes, an HDTV format was defined which is obsolete today (2009). This is ITU-R 709 part 1. A HDTV format with display aspect ratio of 16/9, and 1920x1152, but also 1440x1152. The latter raster was derived by multiplying by four the SD raster defined by ITU-R 601 (720x576).

This originated also the way to record material in this format by using a group of four D1 recorders (the first digital VTR for ITU-R 601, uncompressed coding) under the control of a single system, named Quadriga<sup>8</sup>, used also for the playback.

Thus a little collection of material was created, made of test sequences but also HDTV works shot with HDTV tube cameras, and recorded on D1 tapes in quartets. This material was in danger because of lack of availability of players and for the additional complexity of merging the four parts of the picture properly.

The Ingestion has been performed by recording through the SDI interface connected to a D1 player<sup>9</sup>. The uncompressed audio and video data (about 300GB/hour) are kept as original master copy stored on inexpensive (redundant copies) USB removable hard disc to be kept offline.

For current use (access copy) an XDCAM (lossy/high-end) copy has been obtained, giving a provisional solution to the issues of adapting the resolution, the video levels and the colour space, which are different between old and current HDTV standards.

### 2.5.4 Preservation of material derived from Film

Most broadcasters used to produce news recording on Film in the 1950s and 1960s and have collections of materials mainly in 16mm, 4:3, black & white (with audio recorded apart mostly on SEPMAG). A possible digitisation option for this type of material, taking into account the quality of source, is using an analogue telecine equipment providing components signals to D10 recorders, and then obtain MXF/D10 files as described in 2.5.1.

However the collections of the broadcast archives also include more valuable film materials as 16mm, 35mm and 70mm in colour, for which a digitisation option is through film-scanning or HD telecine

The characteristic issues of this case are related to:

- **decisions about video resolution at digitisation and for archiving** – depends on source type, may be 2k, 4k. The reference here is Digital Cinema [DCI]. Need to consider a relationship with television standards (e.g. HDTV).
- **decision on DPX** – which is the first output of film-scan, uncompressed, according to SMPTE 268M. Should DPX be an archive format option? Audio would be separate.
- **decision on file format and coding** – the MXF wrapper would be suitable for the uncompressed option and for the JPEG2000 encoding (lossy or lossless), need to be able to reference an SMPTE standard (even a future one) or is DCI specification sufficient?

The resulting required storage capacity vary from about 120GB/hour (current DCI spec. lossy JPEG2000) to more than 2TB/hour for an uncompressed format. This requires a bit of clarification.

---

<sup>8</sup> two different ways of splitting the whole HDTV picture in four parts to be recorded as ITU-601 were actually used. The documentation of these schemes is essential in order to be able to create again the composed image correctly.

<sup>9</sup> critical to ensure the alignment of the four files created by the four played tapes

### 2.5.5 Preservation of material SDTV with DV coding

A survey of EBU P/NP [EBUPNP] about the use of file-based AV in production has observed, together with the already mentioned D10 format, a wide use of all the formats of the DV family of coding schemes, such as DVCPRO50 and DV25. Although the broadcast archives represented in PrestoPRIME don't have large collections in those formats, the case deserves to be considered as well. Moreover this format has got important success also for home and semi-professional contexts.

The DV specifications [DVSpecs] originally derive from IEC-61834, have video lossy compression (DCT based, Intra frame) and uncompressed audio. The most high end variant, supporting 4:2:2 format and with lower compression rate is DVCPRO50.

The DV data can be saved as files in various wrapping formats, including the simple raw DV-stream, which is supported by many editing tools. From the long term archive perspective it will be important to consider the MXF reference specification SMPTE 383M GC-DV.

The required storage capacity ranges from 13GB/hour for DV25 to twice for DVCPRO50. Currently there are no signs of forthcoming format obsolescence.

### 2.5.6 Preservation of material currently produced in HDTV

Currently approaches to HDTV production generally imply the use of a compressed format, which is expected to be one among:

- **AVC-Intra 100** – codec compliant to H.264/MPEG-4 AVC – wrapped as MXF OP-atoms (SMPTE 390M) on the recording devices
- **MPEG2 422Profile Long GOP** – specifically in XDCAM
- **MPEG4-SP 422** (HDCAM-SR) 10bit (SMPTE 409M)
- **JPEG2000** (profile to be stabilised)
- **DVCPROHD** – HD version of DV 422 coding (aka DVCPRO100) – coding according to SMPTE 370M and wrapped as MXF OP-atoms (SMPTE 390M) on the recording devices.

The approach described here considers to keep the original production format within MXF wrapper, in order to take advantage in the ingest- and delivery-processes, at least in the short term. However, as these ones are relatively young formats, the capability of handling them must be verified more often because the possible success of an incoming competitor format may speed up their obsolescence.

The compression rate is here higher than those adopted for SDTV and the required storage capacity ranges between 25 and 45 GB/hour.

While in a single archive one of the mentioned formats will tend to be preponderant, in a context of networked archives there is to consider the issue of format homogeneity regarding both the coding scheme and the wrapper properties.

Concerning the MPEG2 Long GOP case, there may be the issue for the delivery process of producing excerpts, with arbitrary in and out points, without transcoding.

In the current HDTV production environments, a survey of EBU P/NP has observed also the use of a couple of proprietary formats, which are based on very high end lossy compression coding and were originally intended only for use during video editing :

- DNxHD – Proprietary Avid
- PRORes422 – Proprietary Apple

From the broadcast archive perspective proprietary formats are not acceptable and the recommendation is for either go to uncompressed HDTV (if affordable) or transcode to an archive/production format. If in the future those coding schemes are proposed for standardisation, with open description of the format, this statement might be re-considered.

## 2.6 Discussion topics

### 2.6.1 Compression

Compression is a process which is performed for reducing the capacity or the bandwidth required for storing and/or transfer the data sources of AV materials. The compression rate is the size of the original data divided by the size of the obtained data. The process is partially reversible so that a decoder tool can obtain back the AV data in the original format. If the data are exactly the same we have **lossless** compression, otherwise it's **lossy**. The compression techniques can be very crude, such as downscaling the sampling properties and keeping a formal uncompressed format (e.g. downscale HD to SD), or conversely complex and refined, aiming at removing insignificant (to some extent) parts of AV content.

The relationship between quality and compression rate is complex, because of the various possible compression algorithms, and only on strictly equal conditions we can say that a lower compression rate correspond to a higher quality.

The properties of the types of video compression schemes are summarised in Table 5

compression type	compression level	compression rates	properties
no compression uncompressed		1	This preserves the full quality obtained from a given source, the quality of which determines the resulting quality.
lossless		about 2	The quality should be the same as uncompressed (otherwise not really lossless).
lossy	high end	from 2 to 16	The obtained quality should be sufficient for most current production processes.
	medium	from 6 to 36	The obtained quality should be sufficient for current publication and for limited production purposes
	low	greater than 36	The obtained quality will be sufficient only for limited publication processes

**Table 5 – types of video compression and properties**

## 2.6.2 Definition of obsolete format

For analogue and digital video tapes the concept of obsolescence is related to the carrier (the storage medium) and the devices used for the playback. Omitting to mention the status of the tapes, once the player devices are out of production then the format can be safely considered obsolete, and a very good premonitory event is when the recorder/player devices are out of production.

For AV-media file the question is different.

If we consider the possibility of accessing the AV content, that is to unwrap and decode properly the file format and get AV in the uncompressed domain for further use, then under some circumstances the format will never become obsolete. The simple constraints of a format known and open (full specification available) and of strict compliance to the specifications are theoretically sufficient to allow the building of a decoder from scratch. This would not be cheap of course, however if we consider that at a given time a working decoder had to exist we can easily reach the conclusion that there are several strategies to maintain the capability of handling a format:

- maintenance of the decoder, having the source
- maintenance of the environment in which the decoder works, or emulation of that environment
- adoption/development of a decoder working into an environment specifically designed to be easily maintained (Multivalent approach)

With such an argument the obsolescence of a format becomes a question of maintenance costs and costs comparison among different solutions and the optimal point will likely change over time.

This is only true for the Master format, however, because the preservation of its original quality has always to be considered.

For non Master copies, i.e. Broadcast/Publication quality, Browsing quality and proxies in general, the outcome is different. A format becomes obsolete when it is no more adequate to its purpose.

An example can be given with Mpeg1-Video. It is decodable and there are no evidences that it won't be decodable in the next years. However it is no more used for the production of new items and it is not considered appropriate for any specific purposes (not any more) such as proxy or web publishing, for which it could have been adopted in the past. So it is obsolete.

But if you have items for which Mpeg1-Video is the Master copy (who knows why), it is better to be more careful, especially if the target format would be a lossy one. It is better to have an approach to find the cost trade-off point.

**What about the wrapping formats?** Simple wrapping migration doesn't affect AV quality. However wrapping formats may provide support to access functionalities, desirable to be preserved. The preservation system has to verify the possible consequences of decision about migrating or not from a given wrapping format.

## 2.6.3 Compression options for analogue sources

We consider three formats of analogue AV tape, VHS, U-MATIC, and Betacam SP, regarded respectively as low, medium, and high quality sources for SDTV. The use of compression for digitisation of these sources is currently a subject of discussion, although a lossy/high-end choice is generally considered a safe one. A greater compression would be certainly possible for VHS and U-MATIC although the result could be affected by the presence of noise.

#### **2.6.4 Metadata in AV wrappers**

The MXF AV wrapper specification defines mechanisms for embedding various types of metadata within the MXF files. Some metadata (structural) are necessary for the effective and efficient access to AV content, while for all the other types of metadata (preservation/rights/description) the suitability of MXF as a carrier for metadata persistency is a matter of discussion. MXF is not expected to be a format for SIP or AIP [SIP, AIP], while for limited use case it may be considered as an option for dissemination [DIP].

The issue of embedded metadata is also applicable to other wrapper formats, such as Quicktime [QTFF].

The general consideration for the archive process is however to have all metadata extracted at ingest time, in order to make the information available for access independently from the access to large AV-media-files.

A question raised is that of consistency, in the long term, between extracted and embedded metadata, because information can be enriched and modified along the archive process. Although this is still a matter of discussion, a possible approach consists in removing all non necessary information from the wrapper and being able to insert it again on delivery, if required.

For the MXF case, an alternative approach, somehow equivalent, consists in separating physically the header from the body at ingest time. This would entail the issue of creating consistent MXF output at delivery time.

### 3 Rights Clearance for Access and Fruition

The access and fruition of AV archived work always has, either implicitly or explicitly, to face the issue of rights clearance and handling. The challenge can be more or less complex, if we consider the broadcast archives or HEIs or other kinds of organisation or even the general user, however everyone should be aware that his/her rights can be limited and that the same protection can be granted to his/her work.

#### 3.1 Questions addressed

This scenario addresses generally the issue of implementing efficient and effective rights handling for the purpose of the general exploitation of the content of an AV-Archive, when planned for long term preservation, but with the aim of being the object of continuous fruition.

More in details we consider here:

- How rights information on an AV work should be prepared for submission to a preservation system, or for updating the same rights information set as a consequence of some new rights negotiation.
- Representation of rights information taking into account all the required relationships between rights and the timeline of the AV entity.
- How to support rights clearance in case of AV entities created with the re-use of pre-existing materials and how to support from the rights perspective the re-use of Archive material for delivery to content creators.
- How the information on rights can be supported in search & retrieval activities.
- How the rights information originated by various organisations can be shared or compared.

In the background there is the issue about unambiguous interpretation of rights information.

#### 3.2 Users and roles

##### 3.2.1 Content creators

In this scenario the role of content creator may cover various aspects:

**Authors, Directors, Actors, Performers, Contributors** – Each person (or organisation) who for any reason contributed to creation of the AV entity is a right owner w.r.t. his/her contribution and, generally speaking, the rights granted to the production company may have different limitations.

**The Production Company** – (person or organisation) – This is the original general owner of the all the rights, except for those limited by the agreements with the other content creators. This role may be covered in partnership in the case of co-production. A broadcaster organisation can cover this role.

##### 3.2.2 Archives

Various types of organisation can act as archives: broadcasters, national institutions, HEI. Specialisations to be considered in the rights perspectives are:

**Rights negotiators** – The person/team in charge for purchasing rights and for trading rights towards customers. (purchase and trade can be separate responsibilities)

**Rights clearers** – The person/team responsible for giving green or red light for a requested use of an Archived AV-entity.

### 3.2.3 Content deliverers

Some content deliverers make publications only, while others such as the typical broadcaster, also act as archives and content creators. In addition to the **Rights negotiator** (same role as for the archive), we have here the **Rights logger** who is responsible for reporting on publication run, when required by the type of right.

Another important role is that of the **Commissioning company**, which (having the business plan on publication activity) commits to a Production company for an AV-production. Broadcasters often act in this role together (being also Production company) in co-production with the provider content creator.

### 3.2.4 Rights holders and agents

Any user considered in this scenario can be a **rights holder** to some extent. However it is important to remark that there can be individuals or organisations acting specifically in this role, because it is a business opportunity or because they are the creators of real life events from which AV entities are originated (e.g. UEFA, Olympic Games and Theatres).

Rights holders (especially authors) often transfer the burden of negotiating rates and terms for use of their works to **rights agents**. In all European countries there are many collecting societies trading licenses and collecting and distributing royalties among their members.

### 3.2.5 Consumers/general users

The general public is thought to be only consumer of AV works; this implies obligations from their part and they should be put in the position of knowing about the licenses and the meaning of the rights terms. Anything further in this direction is anyhow out of the scope of this scenario.

Other aspects however should be remarked.

The general public is a stake holder, to some extent a rights owner and as a member of the community of citizens (national, European) they can claim access to their cultural heritage and to all the AV-works for which some type of rights are/should generally granted to everybody.

Furthermore the users are also part of the big crowd of content creators (case of UGC).

And eventually the general public is protected by the law against any unequable use of images (or other multimedia content) in which they might happen to appear. In case of specific verdicts (or as preventive actions) Archives have to take into account additional constraints to the rights situation.

## 3.3 Descriptions of cases

### 3.3.1 From contracts to rights information set

The rights cases start with programme commissioning at creation, when contracts between production and performers, as well as production and commissioning company are signed.

The receiving organisation (in this case the same one running the archive) will normally get all the rights relevant to their business (or even own all the rights if it is an internal production).

Performers will typically ask for adding constraints, which however will be limited to their contribution. An example is a limitation to "Linear television" for a given number of runs (or within a time interval) and a web publication limited to a specific context (that of the TV programme).

The issues are now :



**Translating the contract terms** in a record in a way, that the information can be handled in the future without ambiguities. This implies:

- reference to contract (or including the contract directly)
- reference to terms as defined by an authoritative glossary
  - type of rights
  - rights constraints (territories, languages, time interval, publication counts, sub-licence, exclusivity)

**Associating rights information to AV.** This is going to be tricky since at the time of setting up the contract the AV item did not exist yet, and even a credible script might be missing. A reference to the AV timeline is mandatory, however a strong link is possible only once the product is real.

- If in contrast to the assumption the AV item was already available (purchase case), the association between rights description and AV timeline may not be much simpler.
- Each association record between rights and AV contents on the timeline (a **short**) must be homogeneous in terms of rights granted, constraints, and contract reference.

**Supporting the case when the information set is incomplete:** some clarification may still be pending, providing a warning and allowing future completion or update.

## SUMMARY OF CASE

### First submission

1. NYNIAU<sup>10</sup> is, as an organisation, one of the parties involved in contracts about rights on AV works. Assuming there is an electronic version of the contracts, for instance as PDF documents (a perfect replica if the original is on paper), it is assumed that contract documents are somehow available through restricted access and are identified by an ID (e.g. a URI). This ID will be used for referencing the contracts in the Rights information set.
2. The Rights information set has to be part of the Submission Information Package (SIP) of an AV work which is going to be archived by NYNIAU
3. The creation of the Rights information set related to the AV work is done by means of:
  - a. Analysis of contracts related to the AV work: This is assumed to be a human activity unless the contract has been produced according to a standard template.
  - b. Recovery of pre-existing rights information related to pre-existing archived AV components reused for producing the AV work: This might be a semi-automatic task.
4. Once validated the Rights information set is ready for successful submission.

### Update

1. The Rights information set about the already archived AV work needs to be updated (purchase of new rights, changes in laws, etc)
2. The archived Rights information set is retrieved
3. A new Rights information set is prepared and submitted for update
  - a. Overwrite – complete substitution of the rights information set
  - b. Append – new rights added

### Export

---

<sup>10</sup> NYNIAU (we in Welsh) is the dummy name of the fictive organisation used in the description of the case

1. The Rights information set of an archived AV work will also be part of the Delivery Information Package (DIP)

### 3.3.2 Rights clearance and content provenance

This case refers for a rights clearance query about an AV-work, for example a TV programme broadcasted at a given date, which is not consisting of only original material but also includes shorts extracted from contributions recorded years before (e.g. it could be another information or entertainment TV programme or another type of feed).

The query indicates a number of target publication events/mechanisms (which may specify the platform, the territory, and the dates).

For the provision of a correct answer the rights clearance system must take into account the rights situation of the original sources:

- either the rights information set of the target work is constantly aligned to the sets of the sources, having all the various shorts perfectly identified
- or it will be necessary to make a complex search of content provenance

The issue here is that most AV archives are among the second case which makes the task harder to achieve. Hence there is a clear reference to how this could benefit from the usage of tools for AV material matching (e.g. fingerprint).

### 3.3.3 Archive user accessing rights information

A user searches in archive some audiovisual materials, aiming to a usage known to him/her. Before using the audiovisual materials found in archive, the user needs to know if the exploitation rights are available. There are several ways to obtain the rights information, depending on the user profile (permissions to access system tools and information) and on the archive policy :

**The user is part of the same organisation that owns the Archive** (in this case the user can claim the same rights of exploitation which are held by the Archive) :

- A RAI producer/director is working at a documentary on “Chinese Emperors” to be produced as a Home Video. He would like to use an excerpt of the TV movie “Marco Polo” (ed. 1988, director: Giuliano Montaldo, actor: Ken Marshall - Marco Polo). He locates this TV movie in the Rai Archive but before using it he has to know if the Archive holds the proper Video Rights. The rights-management-system provides him with the current the rights situation. If the result is positive for his purposes, he asks for the delivery of material; else he may identify the rights holder for the purchase of a license to use or even decide not to use this material and seek for something alike.
- A producer needs some clips of an European city to produce a internet quiz (On the web site of the Archive) “Name This Town!”. In the Archive there are many shots of European cities but only for a few of them the Archive holds the rights for exploitation via the internet. So he sets content information and rights information as parameters of the query in order to get only those clips which refer to the given city, and for which the archive holds the rights that he needs.

**The user is external** to the organisation that owns the Archive and the policy of the Archive doesn't allow the external users to access the rights information:

- The external User (professional or not) searches the archive for material he is interested in. He wants the “Commedia ABC” to schedule it on his Satellite Channel. He finds the production and, before asking for a copy of it, he has to ask to the Archive if he could buy this production and the according Satellite Rights. The clearance rights team of the Archive uses the rights-management-system to retrieve

the situation of rights. The result may be positive if the Archive holds the rights for satellite broadcast. The answer will be negative if the archive does not own these rights. The answer could also be negative if the Archive, although holding the appropriate rights, does not want to grant it (because the Archive is planning some use of that programme, e.g. the publication of a home-video) In the latter case, the choice of the Archive is based not only on rights availability but also on business and opportunities decisions.

## SUMMARY OF CASE

### **Internal user of NYNIAU**

1. A producer or a planner of broadcast scheduling, member of NYNIAU, is making searches in the archive.
2. He/she finds something suitable for the intended purpose
3. The user asks for verification of rights (clearance)
4. The system provides a graphical presentation. As the rights information set is complex, the user will have to browse it in order to find the type of use he/she is looking for.
5. If clearance result is positive, the user may ask for material delivery
6. Otherwise the user might find (if available) information (or link) about how to ask for the missing rights

### **Alternative version of “Internal user of NYNIAU”**

1. A producer or a planner of broadcast scheduling, member of NYNIAU, is making searches in the archive
2. He/she requires to use rights as part of the search criteria, in order to avoid wasting time for subsequent verifications
3. The user specifies rights search criteria in terms of type of rights/conditions, territory, validity windows<sup>11</sup>. The complexity of such criteria might be a simplification of the underlying rights model.
4. The user submits the query and, except for cases due to further details the rights clearance will be positive.

### **User external to NYNIAU**

1. A user external to NYNIAU has granted access to the NYNIAU's archive (or a portion of it) and is performing some search activity. It is not known if he/she is an AV professional or a home user.
2. The user finds something interesting and would like to have access to the AV material (publication quality level assumed)
3. The system performs a rights clearance task and automatically ascertains if NYNIAU holds some rights which might be granted to the external user
4. If the result is negative, the external user gets a kind apologies message
5. Otherwise there is a multiple choice which depends on the policy of NYNIAU with respect to external users
  - a. he/she is invited to fill a request to be submitted to NYNIAU
  - b. he/she gets a list of rights that NYNIAU has decided to offer to the external users, if the AV-item belongs to a pre-defined set.
6. The user follows the instructions provided by NYNIAU.

---

<sup>11</sup> a (date-)time range within which a right is granted.

### 3.3.4 Rights clearance for reuse by content creators

A **professional user** (e.g. video producer) wants to make a documentary on “1968”.

He makes some research on the Europeana website and finds a lot of archive materials: photos, books and audiovisual material.

He has interest in three audiovisual documents about “The Prague Spring”, “The night of the Barricades” (12-13 May, Paris) and “The Battle of Valle Giulia” (1 march, Rome).

These materials are located in two different archives (Archive#1 and Archive#2). Europeana provides him with the name and contact details of both. The user contacts the archives and asks for support and rights to make a documentary. He informs the archives (sales and commercial rights department) that he wants to deliver this documentary via “TV broadcast” and “Internet”.

If the meaning of the rights definitions are shared among the partners (since they make reference to a single authoritative glossary), it will be possible for the User to get materials from various sources with compatible licensing terms; otherwise (current case) the continuation of the scenario will be the following:

USER

Asking for “TV broadcast”-rights, in his understanding he wants the rights for the analogue or digital transmission by terrestrial, cable and satellite including transmission of signals by closed circuit and video on demand. The “Internet” includes for him Internet streaming and the download from web to mobile device (e.g. podcast).

ARCHIVE#1

For the Sales and Commercial Rights Department of Archive#1 the definition of **TV broadcast** is:

- analogue or digital transmission, broadcast or exhibition of visual images intended for reception by conventional domestic or home television excluding video on demand. They include in “TV broadcast” only linear transmission.
- Internet is a particular kind of a Multimedia on-line rights and it includes streaming and download service.

ARCHIVE#2

For the Sales and Commercial Rights Department of Archive#2 the definition of TV broadcast and Multimedia is:

- TV broadcast means the analogue or digital transmission, broadcast or exhibition of visual images intended for reception by conventional domestic or home television. The transmission can be by terrestrial, cable and satellite including transmission of signals by closed circuit and video on demand.
- **Internet** right includes only streaming service but **does not include download** on PC or any other device.

On the basis of those different definitions the two commercial archive departments answer to the user that they will sell him the material together with the rights for “TV broadcast” and “Internet”.

User and archives sign a contract agreement and the user produces the documentary.

When the user offers the documentary on VOD services, the Archive#1 (holder of the rights on a part of contents forbids the use of their archive material on VOD. (The user must require separate licensing arrangements for offering VOD. He needs to acquire the right of making available to the public)

When the user offers the documentary on Internet Download services, the Archive#2 (holder of the rights on a part of contents) forbids the use of their archive material via this way.

The user infringement is due to the fact that the two archives and the user don't have a **shared glossary defining rights and exploitation** on the basis of their negotiation.

### ***3.4 Implications for test materials and evaluation***

In order to have an effective validation of this scenario, the test materials should include:

- Archive television programmes of various editorial formats and genre:
  - entertainment
  - fiction
  - documentary
  - sport
- Archive television programmes which were at least partially made by re-using archived programmes
  - at least approximate EDL information regarding re-used material is expected
- For each AV work there should be a credible rights information set, related to the rights held by the archive organisation. The rights information should address publication and derivation rights granted on conditions about territory, dates, publication runs, access policies, delivery mechanisms, publication medium, etc.

## 4 The perspective from other organisations with AV archives

### 4.1 Higher education institutions (HEI)

Universities and generally Higher Education Institutions (HEI) systematically collect or purchase analogue and digital AV material. All university libraries in Europe have reasonable collections of video and audio tapes of the last 20 or 30 years and are nowadays faced with the problem of preserving this content. But researchers also collect AV material which comes from “official sources” such as broadcasters, but often also comes from “grey” sources, such as other researchers, research institutes, cultural organisations, companies, or private persons.

HEI also systematically produce analogue and digital AV material for research, education and fine arts. Researchers create video and audio files for their studies, as part of their professional work. This material is always unique and represents in some cases extraordinary scientific and cultural value. Researchers also produce AV material as teaching objects for their courses; they use records for illustrative purposes as well.

Eventually HEI provide their AV infrastructure to their students and these students then produce AV material for courses or for entertainment and fine arts, e.g. theatre or arts performances.

The main strategic goal of the scenario should be to set up a best practise model for the digital preservation of multimedia digital objects within HEI.

#### 4.1.1 Characteristics of the HEI scenario

HEI have some specific characteristics which we need to understand if we want to define the scenario in more detail. The following characteristics need to be mentioned:

- HEI generate or collect multimedia content for the sake of research or cultural documentation, for illustration within student courses or directly as learning objects. In other words: HEI are not dedicated to broadcast the material for a wide public. They do not face the situation that a production needs to go “on air” at a given time.
- HEI enjoy some copyright privileges. Again in contrast to other domains, the use of AV material for research and education is eased by international and national legislation. Based on the directive on copyright of the European Commission HEIs are in general allowed to collect, to preserve and make available copyrighted digital material in clearly defined usage situations for their designated community.
- HEI do often have professional or semi-professional e-infrastructure for creating multimedia content, for storage and for making it available. But due to the liberal constitution of HEIs, researchers and students may not be forced to produce multimedia content according to a fixed workflow or according to a compulsory data model. Especially researchers are always free to decide which devices, formats and standards they will use (if any) and in most cases HEI can only encourage their members instead of defining a mandatory in-house workflow and quality assessment model.
- HEI do have in general an up-to-date IT infrastructure, consisting of professional hardware and software and dedicated staff, especially including large server infrastructures, storage devices, and storage routines. But they do very rarely have a digital preservation system in place which allows them to manage digital objects in the way foreseen in the OAIS model. One has to take these characteristics into account if PrestoPRIME wants to develop feasible solutions for this domain.

## 4.1.2

### 4.1.3 Main questions addressed

How can we handle the fact that we are not in the situation to force the content creators to use a given format, to apply normalised metadata standards and to commit to approved quality standards? The digital preservation system must therefore be able to deal with a rather large number of formats and metadata models. Especially if we think on the necessity to edit and render proprietary file formats this requirement will be a challenge. The inclusion of file format registries, or on-the-fly conversion to standardised or widely used formats have to be considered and tested within the scenario.

How can digital preservation systems smoothly be integrated into the IT infrastructure of HEIs? How much effort will researchers and students be willing to invest themselves, how much support has to come from multimedia units? Due to the fact that the production at HEI is often “project based” and not “programme based”, usability issues are key.

Who are the natural addressees of multimedia preservation and which user background will be necessary to run a digital preservation system? E.g. will the university library, the computing service, or an e-learning institution be the natural partner of a digital preservation system within a HEI? Or could there be a model with distributed roles and tasks? These questions are organisational ones but are also connected with a number of technical questions, in terms of background knowledge, training effort needed, interfaces to existing applications within the libraries or the IT infrastructure environment.

### 4.1.4 Users and roles

With respect to AV domain the roles identified for the HEI cases are:

**Content Creators** - Researchers and students are the two main groups of content creators. It has to be emphasised that both user groups do not have a professional education in multimedia production and management, mainly autodidacts are involved. Moreover both user groups are completely independent in their decisions and may not be forced by the HEI management to use certain kinds of production or digital preservation systems.

In some respect also administrative staff such as the IT department responsible for an AV studio act as producer of multimedia content or support researchers and students. In this case some professional background can be expected.

As emphasised above, even if a dedicated unit for digital and multimedia productions exists, it will not be the case that all material is produced within this unit. In contrast a major portion will be produced completely independently, implying the possibility to bypass any defined “official constraint” on the AV production standards.

On the other hand it is to be expected that a digital preservation system would offer the chance to centralise and unify the multimedia production at HEI. In this case the digital preservation system and accordingly the administrative unit which takes the responsibility would need to be able to actually guarantee the long term availability of all multimedia digital objects which are integrated into the archive. This would allow for setting up quality standards and constraints. But it needs to be tested if this approach is viable and will be accepted by the content creators.

**End users** – In this scenario content creators are also end-users : again researchers and students are the target groups for the multimedia productions. Only in rare cases a broader audience is envisaged, e.g. for student entertainment or fine arts productions.

**Production and preservation managers** - Due to the semi-professional approach content creators are in many cases also the production managers and their own

preservation managers. In many HEI dedicated preservation managers are not available yet, which is directly connected with missing digital preservation strategies in general.

Currently digital preservation is within the responsibility of the content creators. Therefore the responsibility lies mainly either at personal level (the single researcher or student) or on the level of research institutes. If a future digital preservation is to be established, it is deemed necessary to set up a central system and a single administrative unit should be in charge of the overall responsibility. Three administrative units can be identified as potential candidates for establishing a digital preservation system at a typical HEI:

- **University libraries** – They have multimedia collections, mainly purchased materials, and the staff is trained to describe the items according to library standards such as MAB<sup>12</sup> or MARC<sup>13</sup>, while structural and technical information is lacking or poor. The AV preservation process mainly consists of a backup copy and a file system storage, because few libraries have enough people trained in digital preservation, which is considered, especially for multimedia collection, a minor duty.
- **IT computing centres** – The IT departments at HEI provide the infrastructure for many applications, such as databases, business systems, content management systems, wikis, e-learning systems, etc. Obviously storage systems play an important role here as well (e.g. at the University Innsbruck some hundreds of TBs can be managed at several quality levels). In contrast to the university library the IT department is strongly technology focused and does not have any experience in metadata management. In fact the IT department would refuse taking over any responsibility on the actual management of content, and sees its role only at an infrastructure level.
- **eLearning and “new media” units** - In the last 10 years eLearning units were established at nearly all HEI. As courses are now accompanied with additional digital material, some hundreds of researchers and thousands of users need to be supported and a strong emphasis is put on user training and user support. These eLearning units are either separate units, e.g. connected directly to the management board of a university, or are in many cases part of the IT departments as is the case of the University Innsbruck. Only in very rare cases they are part of the library. These units are in general quite flexible, but unfortunately the knowledge on digital preservation and metadata management is poor and not part of their daily work. On the other hand these units are rather often involved in multimedia productions from the very beginning since the content creators for multimedia productions are in many cases also active users in eLearning systems and therefore a natural connection is given.

#### 4.1.5 Quantities

At the University of Innsbruck there is a collection of more than 60000 multimedia data carriers summing up to around 88000 hours, with an increase rate of some thousands new carriers per year, and distributed over more than 20 administrative units.

If we look at the European level, with some thousands of HEIs, millions of researchers and more than 10 millions of students, we can extrapolate the figures from Innsbruck so that it is very likely that this scenario comprises millions of hours of content.

The largest part of this material is still in analogue format and would require a digitisation process for which a proper strategy needs to be defined. However newly created material is generally in a digital format therefore a digital workflow including preservation aspects needs to be set up.

---

<sup>12</sup> MAB (Maschinelles Austauschformat für Bibliotheken): an alternative of MARC format used in German speaking countries

<sup>13</sup> MACHine-Readable Cataloguing <http://www.loc.gov/marc/>



#### 4.1.6 Storage and access strategies

The approach taken into account for decisions on storage strategies is related to the expected access needs.

Fundamentally the users of HEI AV repositories would seldom require access to the master copy of a preserved item and thus a secondary storage solution is chosen, although implying greater access delays, because of being cheaper. In the case of the University of Innsbruck the current system is based on tape memory (IBM hardware and software) and can reach a total capacity of hundreds of Terabyte.

The primary storage is then allocated for the lower quality copies, which cover the user access needs for both browsing and fruition. As a reference indication, the current “YouTube” quality<sup>14</sup> is considered sufficient in most cases.

Furthermore, the master quality level is going to be limited to materials considered more important from an “editorial perspective” and for **unique** materials.

#### 4.1.7 Considerations on formats

Typically HEIs don’t have the practice yet to adopt a single production format for the whole institution and it seems that any authoritative decision in this direction might even be considered as “an encroachment on education and research”. As the current practice is to archive the AV-media-files as they were created or purchased, this is going to be a future preservation issue because of the great number of various formats to be supported.

However restrictions on supported formats are considered necessary for efficient preservation. On one hand an approach to this issue is that of a greater involvement of the Archive in the production stage, but on the other hand many questions such as quality, file size, compression, access bandwidth, although considered crucial, are not sufficiently under control of the internal staff and there is a need for authoritative guidelines.

---

<sup>14</sup> video coded in FLV at a few hundreds Kbit/s, CIF resolution .

## **4.2 Exploitation models for medium and small archives**

Large national institutions, such as INA in France, in charge of the preservation and exploitation of the national audiovisual collections in their country, have already an exploitation model in place, which on one hand implies strong initial investments, mainly concentrated in the preservation process necessary to bring all the contents to a digital environment, but on the other hand produces regularly an important income, and therefore supporting the whole archival process.

This is not the case for many medium and small size archives, which typically lack one or several of the needed requisites for preservation and exploitation, mainly in reaching the critical size to launch preservation and digitisation plans needed to make their contents accessible for commercial or cultural purposes.

This scenario targets the approaches to the exploitation models involving both large and small archives and the complementary actions that can be undertaken with mutual benefits for all participants. It is based on the visibility of an established large archive and the potential richness of small collections.

### **4.2.1 Questions addressed**

The main question is that of identifying the conditions for the described approaches to be successful. This implies to consider both technical and commercial aspects of the scenario and the identification of parameters the value of which might be decisive (WP2).

Besides it would be useful to understand which aspects of the relationship between small and large archive have to be considered as for a case of "Service Provision" (which all the related issues such as SLA's) and which ones should be regarded as for a "Commercial Partnership" (both share holders of the same business). The aim here is however to configure models where large archives become Service Providers for preservation, storage and exploitation, while the small archives can develop strategies for access, still conserving the property and benefits of digitisation..

These new collaborations among actors of the audiovisual domain present a certain number of problems, which are typical within any strategy for merging collections:

How to make metadata structures and their contents interoperable (language, documentation procedures)?

- Adoption of a standard integration mechanism?
- Agreement on a fixed information set, and metadata format, based on the large archive procedures?
- Separate structures?

How to harmonise content formats and carriers?

- Small archives comply with large archives procedures and formats?
- How to deal with digital contents and the transcoding that may be necessary?

Other issues may arise, such as:

- How long will the contents be stored and exploited?
- What are the preservation guarantees?
- Where and what is the Original?
- Does the small archive receive a high resolution digital master so to prove that they fulfil their archival missions?
- How to organise and structure and information system that deals with different content provenance, revenue distribution, right management systems?

The large archive involved in this kind of actions would prefer to have a homogeneous set of contents and metadata in order to globally deal with the deposited contents; however this approach may not be the only possible one.

All these issues have implications in the transactions and, whenever commercial archives are concerned, expected revenue.

#### 4.2.2 Users and roles

This scenario is based on the assumption that the size of an AV archive determines, or is at least a crucial factor, for the capability to run autonomously digital preservation and exploitation versus the need to recur to external resources.

Thus we have:

**Large archival institutions** – These are mainly broadcast archives, private or public, however there may be other types of institutions. Their major activity implies the ability to:

- Transfer the original analogue content to digital high quality media
- Code to broadcast production and browsing formats
- Have the content on information systems for immediate access and delivery
- Have all rights information digitised as metadata, which is required for commercial exploitation
- Have in place services (tools and teams) for providing on-line access for professionals and delivering content with commercial and rights management. This may also include delivering content on public websites.
- Develop efficient maintenance capacities as well as migration strategies.
- The role taken by this actor in the scenario can be partially that of the “**Service Provider**” towards the small archives, depending on the selected approach.

**Medium and small sized AV archives** – The holders of AV collections for which running an independent archive process is a critical issue. The possible reasons are:

- Lack of technical knowledge – this may be for migration/preservation or for setting up commercial activity.
- Lack of funding for digitisation or migration – missing the initial investments which enable the archive to reach the level to produce income.
- Lack of rights management capability – which is critical for commercial exploitation.
- Poor knowledge of the collection and poor metadata, with little capacities to improve it.
- The collection is too small.
  - cannot be considered alone a valuable source of contents.
  - the obtainable income would not be sufficient to sustain the activity.

#### 4.2.3 The “broker” approach

The first case consists for the large archive to become a “broker” for other smaller or even large archives, providing a single point of access for a large amount of collections, permitting transversal searches via a common interface. The Europeana portal was conceived to give a unique point of access for metadata and captions coming from cultural digitised collections from European institutions, (<http://europeana.eu/portal/>) excluding any commercial use or downloading of the contents since it is not a repository, just an access point; whenever a user wants to access content, a link to the original repository is established.

This model is applied within other institutions, where the visibility of an existing portal can be shared with other content-holders, which develop a similar commercial activity. A browsing quality copy is held within the large archive's portal so users can vision the contents. Whenever the user wants to make a transaction, a link connects him with the content-holders website.

The small archive keeps the responsibility of holding the contents in their original format and takes in charge the delivery, rights management and billing processes. The benefit for the "broker" is a percentage-fee for each transaction generated.

In this case the large archive acts as a service provider for search and access functionalities.

To some extent this case is similar to the scenario in Chapter 5, broker activities can be mono-language oriented or internationally based; in this context dealing with multilingual access to interfaces and metadata is an essential issue.

#### 4.2.4 The "distribution agreement" approach

Several sub use-cases are considered within this approach, depending on the extent of the service provided by the large archive. It can range from depositing an exploitation copy in a specific format within the large archive, to taking in charge the whole process that goes from the analogue media to digitisation, preservation, storage, metadata description, rights assessment, restoration and exploitation of the contents.

In this case the small/medium archive would physically deposit their assets within the large archive for archival and exploitation. The rights clearance, the delivery and the billing are all in charge of the large archive organisation, which will receive a larger income than that of previous case (as the small archive would be generally credited only of being the original content holder together with being a rights owner if applicable).

The different sub use-cases imply different kinds of contracts to be established, entailing important variations in costs:

**Complete digitisation/archival and management process delegated to a large archive:** which includes preservation strategy, digitisation process, metadata extraction and structuring, identifying right holders, commercial and public exploitation.

**Complete management of digital content delegated to a large archive:** which includes storage and preservation of the Master quality. In this sub-case the large archive is actually running the archive process for the small one.

**Only storage of the exploitation copy is assigned to the large archive:** here the large archive provides all the means for running the commercial exploitation, which includes the storage of the AV materials at the quality level identified to be appropriate (it might be lower quality than master), while the preservation responsibility is left to the small archive.

The first sub use-case is indeed extreme, however existing. Small collections (as individual producers) may totally delegate the archival process of exploitation to a large archive, which will exploit and preserve the contents within a long-term contract. Small collections may not even have the means to afford the cost of such a process, in this case the large archive can take this charge upon themselves, recovering the costs from sales, and repaying royalties to the owner when exploitation costs are covered. Initial investment costs to the small archive are then minimal, the contents are accessible through a trusted repository, and still may generate income with time.

#### 4.2.5 Success factors

The scenarios described above may become very successful, especially if the success factors below are present:

- The partners have to anticipate a mutual benefit. The most obvious potential incentive is a financial one; if the contents have a clear potential value, an agreement is easier to reach... But in some cases the interest for a large archive to host the contents of a smaller one may go beyond the financial value of the contents, for example filling a gap in the coverage of a large collection may be interesting for the large archive, even if the expected commercial revenue is low. In some cases, the expected return may be only in terms of public image, and not financial.
- The collections share some commonalities, justifying a common public access point, and avoiding the overhead in managing a collection of totally different type and contents.
- The contents share a common language.
- The media are in acceptable format and condition.
- The costs in making metadata available and interoperable stay within reasonable limits.
- The rights statuses are reasonably clear.

The absence of one or several of these factors does not prevent a successful collaboration, but may render it more difficult to achieve.

#### 4.2.6 Technical implications

As presented earlier, the process of merging an archive within another one may present several technical issues related to the management of a different set of formats and metadata structures. The issues concerning the merging imply:

- Quality assessment: a HD collection has different implications than a collection from the sixties
- Format transaction: can the small archive impose a format or does it have to follow the large archive standards
- Decision on formats for the digital master
- Transcoding: if the contents are already digital, which processes are to be used to bring them to the exploitation format
- Identifying the collections: from a technical point of view, contents are merged, however they have to be differentiated and probably with different transaction rules

## **5 Adding AV content to Europeana and improving its usability**

The Europeana Beta portal was launched in October 2008 at the Council of the Ministers of Culture in Brussels by the European Commission president Manuel Barroso. The vision for Europeana as introduced by the Commission is to provide a unique point of access to Europe's distributed cultural heritage [EUR].

Audiovisual content is an important part of Europe's collective memory depicting cultural, historical, artistic and socio-economic events that span the 20<sup>th</sup> and the beginning of the 21<sup>st</sup> century. It is estimated that there are around 50 million hours of AV content held in broadcasters and film archives across Europe. Of this only a very small portion is digitised while online accessibility largely remains locked for the citizens. In Europeana, AV content currently makes up only a 3% of the total of searchable digital objects. In the content strategy launched in the summer 2009, a high priority has been set up to attract and make more AV content, searchable and accessible.

### **5.1 Questions addressed**

The current use case scenarios addresses the following questions:

- How can content owners and aggregators make more AV content available to Europeana and to end-users?
- How can AV material be better represented in Europeana?

### **5.2 Main scenario**

#### **5.2.1 Use case 1a**

A regional TV broadcaster in Poland runs a website where he makes searchable footage and clips from his various collections.

His archival material is digitised at a high resolution, well documented and easily manageable through his Digital Object Management System. Via this he can also easily export files and metadata in various formats.

He has built in the online platform B2B and B2C services.

Users can preview the contents at low resolution and make use or buy footage material or complete programs under different conditions.

He has a particularly interesting collection of interviews with World War 2 concentration camps survivors.

The users can search on his website and view the clips in low resolution.

A producer chooses to buy a clip from him for a documentary film he is making on the life of the film director Roman Polański.

A history teacher buys a DVD documentary on the Mauthausen-Gusen (produced by the Polish TV-company) camp to show in his class. The school network in Germany negotiates the use of such clips to produce teaching online educational materials to be used in the closed regional school network.

A researcher from the States uses the clips from the Polish TV-Archive combining them with relevant material he found on the INA website in France.

An individual from Argentina is able to find information about the history of their family by watching a clip from the Polish TV-archive.

Another TV channel from Israel contacts them for a co-production of a documentary using archival content.

An IT company building an interactive application for the Mauthausen-Gusen camp contacts the channel to buy the rights to reproduce the interviews for the application that will run on mobile phones.

The regional polish TV broadcaster is aware that it is very difficult for all these users to locate his content online when searching on Google.

He knows though that opening up access to his collections online through Europeana can bring a lot of international traffic to his site and open up possibilities for him to further exploit his holdings. How can he contribute content to Europeana?

### **5.2.2 Use case 1b**

A cultural institution in Belgium holds DVDs with the performances of the resident modern dance group.

They have a website where they stream 3-minute clips from 5 performances, and users can order the DVD with the complete performance online.

They only have little information about each performance stored on an Excel sheet, they don't have a specialist IT person and know nothing about documentation and different metadata standards. How can they contribute content to Europeana?

### **5.2.3 Background information: Europeana technical requirements**

For museums, libraries and archives to contribute content to Europeana, they need to comply with the technical requirements of Europeana, i.e. make their metadata available in a way compatible to the Europeana data specifications. To ensure interoperability amongst cultural heritage institutions in Europe, Europeana has developed the Europeana Semantic Elements data model [ESE] and workflows for the ingestion of contents.

ESE consists of the Dublin Core (DC) metadata elements, a subset of the DC terms (defined at [DCMT]) and a set of twelve elements which were created to meet Europeana's needs. For integration of AV content in Europeana, compatibility with the ESE and subsequent data models is necessary. Europeana is also working on a new Data Model with the aim to be able to accommodate domain-specific and rich content information.

In addition to the mapping to ESE it is necessary for a normalisation process to be carried out on some values to enable machine readability. In the initial implementation of the Europeana prototype, much of the mapping and normalisation was carried out centrally in the Europeana Office. This work will increasingly be passed on to content providers or aggregators and there are Guidelines [EGUID] provided to support those tasks.

Besides the metadata, content providers provide a thumbnail of the actual digital object. In the case of 2D images and text, this is a low resolution representation of the digital object. However, the requirements have not been specified for moving images and sound objects. Some providers provide the thumbnail of a keyframe.

To actually view the digital object, users are directed to the content providers' site; therefore ownership of the object always lies with the content owner.

## 5.2.4 Actors

**Content Provider** is any organisation that provides digital content for access via Europeana and the metadata that enables the access. As shown in the use cases the amount and type of content, the technical infrastructure, the output formats and the documentation available can vary significantly among content providers.

It is practically impossible for Europeana to work individually with every European content provider due to the enormous amount of work in the harmonisation and normalisation of the metadata needed. Europeana therefore works with an intermediate layer of content providers, the aggregators.

**Aggregator** is an organisation that collects metadata from its group of content providers and transmits them to Europeana, helps content providers with guidance on conformance with Europeana norms and converts metadata if necessary. The aggregator also supports the content providers with administration, operations and training.

There is a multiplier effect in aggregation that cannot easily be reached by individual organisations. Aggregators have greater possibility of applying multilingual tools and services for improved usability of content. They also have the possibility to rank higher in search engines than individual content providers and therefore attract more users to the individual content providers' sites.

Aggregators on the other hand need to receive metadata in a harmonised way or need to harmonise them themselves in order to achieve search across large amounts of heterogeneous data. In this respect, the richness of the source information is often diminished for the sake of simplicity and efficiency in the mappings between different metadata systems.

Aggregators are central to Europeana's business model. Vertical and horizontal aggregators are being set up to target a specific topic or geographical area. Aggregators make their contents and services available to Europeana, and in return receive their content enriched and an increased exposure of their contents to an international audience:

- **Horizontal aggregators** aggregate content across domains (e.g. national aggregators like Culture.fr or regional aggregators such as Erfgoedplus.be)
- **Vertical aggregators** aggregate content from a single domain. Examples of vertical aggregators are TheEuropeanLibrary.org and the European Film Gateway. Vertical aggregators can work at regional, national or international levels.

Thematic aggregators can either be domain specific or cross-domain.

Examples of vertical pan-European audiovisual content aggregators are: EUscreen for broadcast and TV archive contents, European Film Gateway for film archives and cinémathèques, Dismarc for audio and sound archives. Figure 4 depicts how aggregation of AV content in Europeana works.

**European Film Gateway:** initiated by ACE (Association des Cinémathèques Européennes) and the EDL Foundation [EDLF], the EFG project [EFG] aims at developing a portal providing direct access to about 790.000 digital objects including films, photos, posters, drawings, sound material and text documents. Content will be provided mostly by film archives and cinémathèques, which are partners in the project. The content will also be searchable via Europeana.

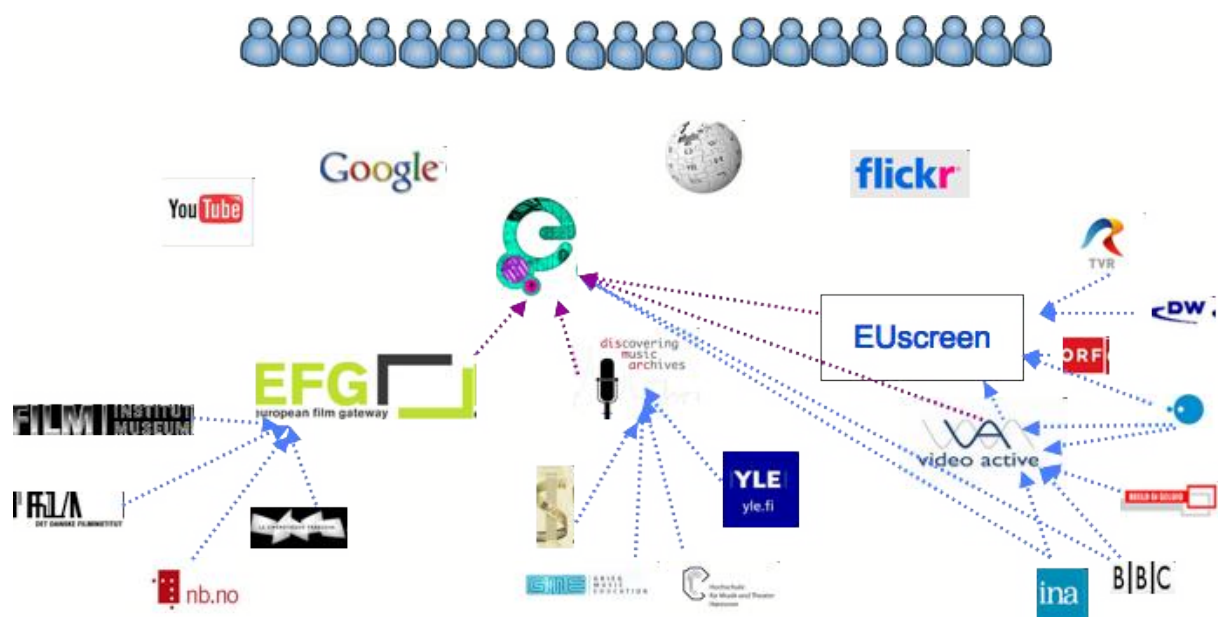
**EUscreen:** Like EFG and DISMARC, EUscreen is a European project that aims at achieving a highly interoperable digitised collection of television material. It is supported by FIAT/IFTA, the EBU and the EDL Foundation.

The project aims at responding to the lack of interoperability among television archives; the non-existence of proven scenarios for the use of audiovisual material; the complexity of rights issues and the lack of contextualisation. EUscreen will tackle these problems by



building a network of content providers, standardisation bodies and specific user groups around the task of providing multilingual and multicultural access to television heritage. The 40.000 objects that will be made searchable over the course of 3 years will also be searchable via Europeana. Statistics showed that 60% of the total traffic to Euscreen's predecessor site, VideoActive, was coming from Europeana.

**Dismarc:** DISMARC provides an access point for under-exposed European cultural, scientific and scholarly music and audio from a variety of content providers such as archives, broadcasters, museums, universities, research institutes and private collectors. More content is being contributed by archives as DISMARC expands. Audio is searchable, discoverable and (where the content owner provides access to audio) listenable.



**Figure 4 - Aggregation of audiovisual content in Europeana**

These aggregators have their own specifications with regards to how content is made available to them. However, they deliver metadata to Europeana harmonised and normalised according to the Europeana rules.

### 5.2.5 Open questions

How can interoperability amongst the different audiovisual content providers and aggregators be better achieved?

How can content providers export content compliant with the Europeana and the AV content aggregators data specifications from their local systems more easily?

Can PrestoPRIME help any type of AV content provider to find information about Europeana, help them locating an appropriate aggregator and provide tools and services that can help them easily exporting their metadata and previews to Europeana?

What tools and services can PrestoPRIME provide to aggregators to help them in the process of harvesting/ aggregation, normalisation and harmonisation from different content providers and exporting?

### **5.3 Improving searching and browsing functionalities for AV**

#### **5.3.1 Use case 2**

A user performs a search in Europeana for “Prague revolution” and retrieves 12 video results.

He clicks on one of the results and further, clicks on “view in original context” to view the full object at the content provider’s native system.

Europeana currently offers exactly the same metadata text-based search and viewing possibilities for all types of objects, i.e. texts, images, videos and sounds.

Which functionalities and services should be provided to improve accessibility and usability of AV content from Europeana? What are the requirements for these?

#### **5.3.2 Background information**

As mentioned before, Europeana only harvests metadata and thumbnails of the digital objects and any actual viewing of the full digital object takes place at the content provider’s site.

Any functionality and service built for better displaying and using AV content should respect these restrictions.

#### **5.3.3 Open questions**

Would it be possible to search via Europeana and directly access a particular segment in an AV work? What are the requirements?

Should Europeana try to incorporate other non-textual basic ways to search and browse AV content, i.e., using similarity, by humming a tune, etc? Which ones? What are the requirements for doing so?

What are the specificities of audiovisual content that need to be better accommodated in the new Europeana Data Model?

How can AV content be further enriched in Europeana?

How can links to audiovisual objects be made available in a persistent way?

How can the notion of “preview” be specified for audiovisual content in a way that takes copyright implications into consideration?

The more metadata are harvested from different sources, the higher the risk of having duplicate content in an aggregator’s database; there are currently no sufficient technologies for identifying and de-duplicating content. Can PrestoPRIME propose solutions for the identification of duplicate content? Would the fingerprinting technology be a solution to this?

## 6 Information enrichment for Access and Fruition

### 6.1 Context

These scenarios deal with the objective of a greater long term exploitation of the AV archives through the improvement and enrichment of the information which provides the basis of future access.

The basic common assumption is that for the fruition of archived documents, it is important to have information elements to be used in the search activity which will give to the user the necessary access information.

Of course some information useful for permitting future access was already provided to the archive together with the original submission. The idea here is that such information can be continuously complemented and kept up-to-date with the contributions of the archive users by the use of new specific tools.



Figure 5- GUI of video annotation game



Figure 6 – GUI of video annotation tool

#### 1.1. Tagging game

##### 6.1.1 The starting point

Researchers of the Telematica Institute<sup>15</sup> have listed incentives of users to tag video content. These include: future retrieval, contribution and sharing, attract attention, play and competition, self-presentation and opinion expression [VEL]. . “Flickr The Commons”<sup>16</sup> already covers some of the incentives. To turn the incentive “play and competition” into the foundation of a new service, B&G has taken the initiative to develop a video tagging game, in close collaboration with software developer Q42, KRO broadcasting and VUA Inspiration for this game came from the ESP Game (<http://www.gwap.com/>), originally conceived by Luis von Ahn of Carnegie Mellon University.

The rules of this game are simple. A user can chose from different channels (newsreels, talk shows, reality TV, documentaries etc.) on which videos a running consecutively. The user can enter the game by selecting a video channel of preference. He joins the players already present in this channel (if the channel is empty, a bot<sup>17</sup> consisting of a previous session will be present).The players cannot communicate, but their presence is listed next

<sup>15</sup> <http://www.telin.nl>

<sup>16</sup> <http://www.flickr.com/commons>

<sup>17</sup> an application running an automated task over the Internet

to the video (similar to a chat-room). Everyone is watching the same video, at the same time. Their task is to agree on a word that would be an appropriate label for what is visible on-screen. All players enter possible tags; once a tag is entered by at least two players (not necessarily at the same time), that tag is agreed upon. Players receive points, based on the temporal agreement with other users and several standard game scoring concepts (time bonus, combos, etc.). User participation is vital to the success of the video labelling game, so a lot of attention is given to provide attractive game play, i.e. the experiences during the interaction with the game. Figure 5 shows the first game interface in Dutch language proposed by B&G and VUA.

### 6.1.2 Question addressed

This scenario basically addresses the problem of verifying the intuition that the user's knowledge can be exploited by the archives providing benefits to the access process in exchange for the offered captivating activity. The main questions addressed are presented in Table 6.

topic	question
incentives	What are incentives for people to tag video content?
standards	What are the candidate standards for video annotation and deep linking?
added value	How can we use the knowledge of outside users to annotate video content? What is the added value of crowd-sourcing the annotation of video content? How to measure the obtained benefits? Which kind of AV documents is most appropriate for this scenario? (from archive perspective and from player perspective)
validation	What are the approaches to validating the quality of contributed data? Should we think about comment rating and ranking, or a reputation system for users?
directions	Which directions should be given to the users for annotation: Tightened to describe video content (e.g. there is a black cat on a table) or Intended to contextualise the AV document (e.g. Elvis Presley singing the Yellow Rose of Texas, dancing and playing the banjo)?
improvements	What is the expected/achievable quality of referencing annotation to the video timeline? Is it possible to include and evaluate elements of sound annotation? Which problems and which advantages can be expected when running the game in a multi-lingual context (players annotating each one in his/her language)?

**Table 6 - question addressed for the video annotation game**

### 6.1.3 Users and roles

The roles relevant to this scenario are:

**Content providers and archives** – They generally would be the beneficiaries of the scenario. As PrestoPRIME partner, B&G is going to be the pivot for the evaluation and tests. Other material is available from the Dutch broadcaster KRO.

**Institutions providing network access** – For instance Europeana. They can be simply the access point redirecting users to content holders offering the game or they might be interested in offering the game themselves.

**Games developers** – The attractiveness of the proposed activity will strongly depend on the quality of the offered game from both technical and artistic perspectives. Better results can be expected from organisations and individuals with specific experience and skill.

**Providers of evaluation and exploitation services** – This is a role which could be taken directly by the archives. However the involvement of specialised actors is suggested, because of the intrinsic experimental nature of the scenario. Within PrestoPRIME this can be done by VUA.

**Players** – This is a key role in the game. The quality of the outcomes will strongly depend on their motivations and on their background.

### 6.1.4 Description of cases

Some steps can be identified:

1. Selection of material, definition of annotation objectives, identification of players
2. Definition of game setting and rules
3. Running the game and collection of reports
4. Analysis of results, identification of possible improvement and further test mechanisms
5. Analysis of exploitability of result, evaluation of benefits.

## 6.2 Creation of contextual data

### 6.2.1 The starting point

This scenario builds upon the results of a national research project CHOICE (<http://ems01.mpi.nl/choice-wiki/>) project where VUA and B&G have been collaborating on for the past three years.

The initial motivations are very close to those of CHOICE: contributing to speeding up the manual annotation process by applying natural language processing (NLP) techniques to generate candidate semantic categories that appear in the selected thesauri from (textual) context information. Context information provides peripheral insights into an object; how it was perceived, how it was created, how it relates to other objects made during the same era and so on. Having access to these sources enables users to expand their explorations into greater depth.

In the audiovisual realm, examples of sources to be somehow linked to objects include: commentary sheets, external reviews, broadcast schedules, viewer ratings and awards, and those sources attached to related objects (e.g. repeats of excerpts or whole of the same object). Within this scenario, possibly relevant statements and setting descriptions from the textual context information will be offered to the human annotator for approval or rejection. Whether a fragment of the context information is (possibly) relevant for semantic annotation is determined by checking whether concepts from relevant thesauri or from the

metadata belong to the related video. Machine learning and statistical methods for natural language processing and information extraction are applied to determine which terms from fragments or sentences will be used in the statements that are offered to the annotator.

## 6.2.2 Questions addressed

The main questions addressed within this scenario are presented in Table 7:

topic	question
incentives	What are incentives for people to tag video content?
credits	How to credit the user's contribution appropriately? Is there a specific rights issue? Consider creative common statements.
use of NLP	How to use NLP and learning techniques to derive relevant semantic categories from texts related to broadcasts? This is also a question about tool selection, tool placement (within architecture/archive model) and timing of their usage
use of results in Archive	How are the results fed back within the Archive process? Integration in the OAIS archive model: results handled within AIP, made available with DIP, used for indexing, search, and browsing
technical evolution of the domain	How to keep the pace of continuous progresses in this domain? (e.g. advances in W3C <a href="http://linkeddata.org/">http://linkeddata.org/</a> )
validation	What are the approaches to validating the quality of contributed data? Should we think about comment rating and ranking, or a reputation system for users?

**Table 7 - questions addressed for the contextualised data**

## 6.2.3 Users and roles

The roles relevant to this scenario are:

- **Content providers and archives** – The potential beneficiaries of the scenario
- **Users as annotators** – Either professional or semi-professional users, with knowledgeable lay (wiki-oriented Web 2.0 users)

## 6.2.4 Description of cases

The video annotation tool might have an interface such as that shown in Figure 6, which includes a form for answering the basic “W questions”: what, where, who, when. The usage is supported by auto-completion in each field (multiple/single vocabularies) and semantic search for videos and fragments. The resulting annotation will consist of mapped concepts from vocabularies and user tags. The working mode is single user (asynchronous).

## 7 The perspective of Service Providers

### 7.1 Service providers for preservation

**Preservation Service Provider (PSP)** in this context means an entity that provides an online service (access is over a network) for the long-term storage, preservation and access to AV assets owned by someone else. So, for example, the service provider might be a facilities house such as Technicolor that provides a service for storing and keeping safe the digital AV contents of a local archive or museum.

Note that the concept of Preservation Service Provider as described here could equally apply to a service oriented model of preservation within one organisation, i.e. it need not involve a third party. For example, within a broadcasting organisation, the IT or archive department might provide a preservation service to the various business units that create or consume content, e.g. production, post, distribution etc.

This scenario distinguishes between preservation services (e.g. ingest, access, integrity validation, format migration etc.) and lower-level infrastructure services that might be used by the preservation service provider (e.g. storage, compute, networking resources) and covers some, but not all, aspects of Media Asset Management.

The activity of the Preservation Service Provider might include activities supporting cataloguing and indexing, services for maintaining the semantics of the archived entities, services for storing/preserving/accessing descriptive or rights metadata. However the focus is pointed here to what is needed for keeping large AV-media-files safe together with the associated technical and preservation metadata, and it specifically addresses the emerging model of the active AV-archive, characterised by the exploitation of networked and file based infrastructure, where the Archive processes is involved from the point of the first creation of the content.

#### 7.1.1 Users and roles

Preservation service providers have to relate with Content holders, Users of the services, other Service Providers.

#### 7.1.2 Description of cases

Some ways in which preservation services might be used include:

- A large archive within a national broadcaster has the internal systems for preserving the content that it creates and the ability to deliver this content to users within the business (e.g. for use in production). It does not however have the capacity to meet an increasing need for external access to the archive (e.g. commercial sales or public access), including making master quality content available over wide area networks. For safety (following the maxim of two copies in two places), the archive uses a preservation service provider to host a copy of the content already in the internal systems. The service provider also makes that content accessible for external customers/users of the archive, taking advantage of the good network connectivity that the service provider has.
- A small archive (e.g. a few thousand hours of video, no dedicated staff) doesn't have the in-house expertise or systems to host its own digital archive or perform appropriate preservation actions, e.g. migrations or integrity checking. The archive uses the preservation service provider to host and preserve the holdings of the archive including providing online access to the contents back to the archive.

- Several small archives collaborate together to create a shared archive of their holdings. They recognise the economies of scale and increased efficiencies of working together on building a shared system and common processes. They host their shared infrastructure using a third party co-location site and operate the system as a preservation service that they all can access.
- A large national archive has a lot of experience with digital preservation and has, through massive government investment, a modern 'state of the art' digital preservation infrastructure. The archive wants to explore new business models that allow the wider community to make use of this infrastructure whilst also allowing the national archive to fulfil its remit of capturing and preserving national heritage. For example, the national archive offers to hold and preserve the content from regional or private collections in return for the rights to use and make accessible this content. In this way, the national archive wants to become a preservation service provider.
- A national broadcaster wants to use an external company to install and operate their new digital archive and preservation infrastructure. Whilst the systems are hosted and operated within the broadcaster's facilities, they are managed and maintained by the third-party. The broadcaster has service level agreements in place with the third party for the preservation services that it provides.

### 7.1.3 Possible offered services

What services are needed? The list below is an attempt to describe a set of services in a modular way that allows one or more of them to be used in combination to address the various use cases:

**Ingest** - the ability to upload content to the service provider for subsequent storage and access. It should be possible for items uploaded during one or more sessions to be correlated with each other as part of collections.

**Access** - the ability to get a copy of content already uploaded to the service provider. This includes access to individual items or access to items aggregated into collections.

**Decoding / Transcoding** - the ability to change the format of the content, e.g. as a preservation action such as moving from DV to lossless JPEG2000, or to create access copies (also called proxies/mezzanine/edit versions) needed for users of the content.

**Dewrap / Transwrapping** - the ability to change the wrapper/container format for the content but leave the essence the same (no decoding and re-encoding of the audio or video). For example, moving from AVI or QuickTime to MXF (or perhaps in the future AXF).

**Integrity checking/validation** - the ability to independently verify the integrity of the content held by the service provider, e.g. by running content corruption detection tests or by using digital signatures or checksums.

**Annotation** - the ability to enrich metadata to the content that is needed by the users of the content or for the service provider to perform preservation actions on the content.

**Access Control** - the ability of the owner of the content to set rules on who can access the content, when and in what form.

**Transferral** - the ability to package up and deliver one or more items of content in a particular format and send it to some destination, which could be another archive or another service provider. Once the content has been unequivocally verified as successfully transferred, the copy in the archive is deleted.

**Removal** - the ability to delete content from the service



**Audit/Report** - the ability to request an audit (technical, financial, process) of the contents held by the service provider on behalf of the content owner. Technical = what is there, what format is it in etc. Financial = how much is the service costing, what are the charges. Process = what actions have been performed on the content whilst at the service provider.

#### **7.1.4 Metrics for service definition: service level agreements**

Trustability on preservation performance might get evidence when it will be too late to change decisions and to change of service provider. Content holders need guidance for selection of services and of their providers. On the other hand the good service providers are interested in making clear the boundaries of what they can supply.

The question addressed here is how can you entrust digital content to a provider for long-term safe-keeping and have little or no visibility of how that content is being stored, yet still be able to access the content whenever you want and most importantly verify its integrity.

A service level agreement (SLA) is a negotiated agreement between two parties: the customer of the Service and the provider of the Service. A SLA may cover all the Services offered by a Provider (it would be long and complex) or there could be a SLA for each individual service.

An assumption here is that only through the interfaces of the services, subject to management, monitoring, and control, it is possible to insert/modify/access the Archive items.

The SLA should cover the following areas:

- The function performed by the service, i.e. what it does
- How to interact with the service. i.e. how to use it
- Obligations on both the provider and consumer of the service
- Agreed bounds of performance (QoS) for the service
- How to measure the delivery of the service, i.e. what metrics apply
- How deviations are handled (exceptions), i.e. what happens when things go wrong
- Penalties or similar clauses if the SLA is breached by either side.

The SLA contains obligations on both sides. The obligations of the service provider are usually obvious, e.g. to provide a service with appropriate functionality and required levels of availability and performance. However, obligations can also exist on the consumer, e.g. not to submit more than an agreed amount of content to the provider in a given month, submitting content in an agreed format, not to attempt to circumvent security measures etc.

The SLA defines what happens when things go wrong and what compensation/penalties may apply. No service provider is ever perfect, so it should not be assumed that the service will always be delivered according to the SLA. Indeed, there may be cases where the service provider deliberately chooses to breach an SLA (e.g. it is economically more viable to meet 95% of commitments and pay a few penalties than it is to meet 100% of commitments and have to invest in more expensive or large scale infrastructure to resource the service).

The ability to define the functionality of the services from the PSP in an SLA is clearly important and is not considered further here. This section looks at some of the non-functional aspects of the SLAs needed, in particular QoS and any constraints (e.g. limits on the number of items or data volumes that can be ingest or accessed each week) that may be needed on either side.

For example, all actions on the Preservation Services (ingest, access, processing) by a customer will cause a workload on service provider. This workload will need to be managed by the service provider not only to ensure the services remain with the SLAs where possible, but also to ensure any internal activities (backup, integrity checking, migration etc.) are not compromised.

From the perspective of the consumer of the service, the benefits of specifying performance and constraints in SLAs include:

**Agreed and controlled ingest of content.** Defining exactly what can be ingested into the Preservation Service Provider and when, can help the consumer plan and manage the submission process. This helps avoiding queues and back-logs at the consumer side, both of which have the potential to put content put at risk. Priority might also be set for different types of content so that the Preservation Service Provider can process submissions from the customer in the right order.

**Agreed and controlled rates of access to content.** This is important as it helps avoiding delays, conflicts, and unpredictable QoS when the consumer wants to access his content (there can be a large number of users with differing and competing needs that go beyond the capacity of the Service and hence contention needs to be managed).

**Defined QoS for preservation actions performed at the Preservation Service Provider.** Many of the Services listed above allow a consumer to perform actions on their content when it is stored at the PSP, e.g. transcoding and transwrapping. With large volumes of content at stake, e.g. 100,000+ hours, the performance of these services is important so that there is certainty that they can be applied and will complete within necessary timescales.

From a Preservation Service Provider perspective, SLAs will typically need to include:

**Limits and priorities for different types of user and/or content access/supply.** This is about limiting the total workload on the service or helping to ensure that the workload is manageable, i.e. it doesn't have huge peaks or periods when the service is idle. This allows the PSP to make planned and managed use of their resources and hence have confidence in meeting user needs. It may also be needed to protect or 'ring fence' any specific resources needed for preservation actions – e.g. to ensure that enough capacity is maintained to do a format migration whilst consumers continue to submit or access content. Only if there is a way for the PSP to control workload can they manage priorities for safety, availability and accessibility of the content they hold on behalf of their customers.

**Security** is also another obvious thing to include in an SLA, i.e. who can use which services and when. Specifying and applying appropriate security measures is a complex area and should be driven by use of an appropriate methodology, e.g. OCTAVE CERT, and is not covered further here.

**The duration of SLAs** also needs consideration. Suppose an archive has a 10 year plan to digitise its analogue holding and store these using a PSP along with any new 'born digital' content it acquires during this time. The amount of content to be stored in year 1 might be small, e.g. 100TB, but by year 5 it could be 10PB and by year 10 it could be 100PB (e.g. because all new born digital content is in SuperHiViz by then!). Should there be one SLA that in year 1 contains the commitments for year 10? No archive can predict the future that well. Should new SLAs be created on an annual basis, but then there is a risk that the PSP decides not to provide the service that the consumer requests at that time. The specific technical implementation of a service might also change during the lifetime of the service, e.g. new data transfer protocols, or new formats supported for

transcoding. These are not necessarily known at the time the service is first provisioned and an SLA set up. Should each change be done through a new SLA, or do we need a mechanism to allow SLAs to evolve/change over time?

## SLA Metrics

All terms in a SLA need to be measurable – otherwise there is no way of knowing whether the service is being delivered as agreed. Each term in the SLA should be quantified, including the way to measure it. In other words, an SLA should include a set of metrics.

Metrics can be used for:

- monitoring (e.g. “how much data was transferred last month”),
- constraining (e.g. “a client should not be allowed to store more than 100TB”),
- QoS (e.g. “the response time should never exceed 50ms”) and
- Charging (e.g. “£3 per GB per year”).

Some metrics likely to be relevant include:

- Number of concurrent users of the access service.
- Number of items ingest per month (e.g. 50 videos per day)
- Average ingest rate of data (e.g. 100 MBit/sec, 5 hours of MPEG2 50i per day)
- Maximum and minimum data rates (e.g. greater than 10Mbit/sec but less than 1Gbit/sec)
- Maximum volume of data transferred each week (e.g. 10TB per week, 100 hours of SD using JPEG2000 lossless).
- Total storage used (e.g. 200 TB per year, 100000 hours of SD in MXF D10).
- Availability (e.g. 99.99% of files will be available 24x7, the ingest service will be online for at least 99.9% of the time over a 1 month period).
- Data integrity (e.g. 100% of bits in a file will be unchanged after 2 years)
- Transcoding volume (e.g. 5000 execution-hours of ffmpeg running on a 1GHz CPU, transfer of 1000 hours of MPEG2 50i to JPEG2000 lossless)
- Transcoding rate (e.g. 100 CPU hours for ffmpeg per week, transfer of 10 hours of DV to H264 each day)
- Frequency of access (e.g. File X accessed 3 times in the last week)

Some of these are instantaneous measurements, e.g. the current data rate being achieved in a transfer, some are averaged over a window of time, e.g. the amount of data ingest per day, and some are cumulative, e.g. total storage used or total amount of transcoding done. Many of the metrics can be expressed in either/both ‘application terms’, e.g. hours of programme material, and ‘resource terms’, e.g. GB of data. Mapping between the two is not always straight forward, e.g. estimating the number of CPU-hours to do a transcode from format X to format Y.

## **7.2 Service providers for Search and Access**

In this Section we focus on the search and access of audiovisual contents based on the services offered by specific providers. We discuss mainly how to provide users with complex search functionalities for distributed audio visual materials.

### **7.2.1 Users and roles**

In this scenario a Service Provider represents an organisation, institution or any other entity offering a network service for searching and accessing digital contents. The service provider is connected for example to broadcast archives, digital libraries or other institutions providing networked access to audiovisual content.

The Provider can offer several different services with different purposes and for different consumers.

In the following we mention different examples to demonstrate which are the most important scenarios and applications relevant for the PrestoPRIME project.

### **7.2.2 Services offered for search and access of digital contents**

This scenario describes how to apply for searching preserved materials provided by digital libraries as well as broadcasters and UGC repositories. Search functionalities are enabled by local or/and distributed indexing of archived materials (Registry).

Results of the query will be the metadata (descriptive metadata and rights) of the searched digital item or a pointer to the actual repository or archive. Furthermore the service could provide also a functionality beyond the simple exact match, making use of similarity techniques in order to propose to the user a list of candidate digital content closely related to the searched one, eventually ordered to a ranking list.

In the following we describe the high level functionalities of possible services provided for the search and access of audiovisual material by end users.

The main functionalities offered to the user by the search service could be the following:

- full text query
- query by descriptive metadata
- query by rights metadata
- query by content
- complex and unified query

The main functionalities that the service provider should implement in order to offer the search capabilities are:

- descriptive metadata indexing
- rights (IPR) metadata indexing
- multimedia content indexing (by features)

Based on PrestoPRIME requirements, the indexing techniques should change according to the size of the metadata as well as digital contents to index and the localisation and topology of the archives.

Hence the service provider may adopt relational or object oriented databases for local and small sized registries, while Structured Peer-to-Peer (P2P) overlays based on Distributed Hash Tables (see for example [DHT] and references therein) are more suitable for large size and topologically distributed registries. The user interfaces exposed by the services providers offering searching features should be designated according to the User experience best practices in the context of distributed search engines (e.g. Google, Yahoo, Quaero, ...).

### 7.2.3 Reference protocols and standards for content search

The service provider offering search services could adopt a standard approach or create a custom proprietary solution for distributed searches. In PrestoPRIME we aim to deploy services based on widely used standards. In the context of Digital Libraries, the most relevant standard protocol for content search is the one proposed by Open Archive Initiative (OAI) named Protocol for Metadata Harvesting (PMH) [PMH] and the older Z39.50 [Z3950] maintained by the Library of Congress. In multimedia contexts the most relevant standards are those proposed by the Working Groups in MPEG and JPEG: the MPEG Query Format [MPQF], and the JPSearch [JPSEARCH].

### 7.2.4 Description of scenarios and use cases

In this Section we imagine a few possible scenarios focused on search and access of audiovisual material exploiting the service providers. Possible ways in which the search and access functionalities could be used are:

- a broadcaster (Content Provider) is searching for a specific audiovisual content to be used for commercial purposes. The content is searched using the Service Provider functionalities (full text and structured metadata search) and the results contain information about the content (Format, metadata, ...) the location (Archive), the related rights (License).
- a user (Consumer) is interested for particular contents for educational purposes, such as scientific documentaries, movies or other audiovisual contents from historical archives, etc. The access to such contents is guaranteed by the search services over several Content Providers. The search service provider is able to forward the query in the distributed network of available registries. The user queries the search service provider by full text feature and retrieves the list of available digital contents ordered by relevance.
- a digital library (Content Provider) has a digital content of which it is not aware if a specific preservation planning has been already taken. By means of the search service provider (text, metadata and content query) the digital library will be notified that the searched digital content is already owned by another digital library which is preserving it.
- a digital content creator (UGC) wants to be informed if someone is making use of its contents or if some archive is hosting them. The user applies for a query on the search service provider (query by content) and gets the results listing the available registry storing or referring his contents.
- a broadcaster wants to verify if their digital contents are going to be distributed over unauthorised channels and stored in unaware digital libraries. The broadcasters apply for a query by content to the search service provider, getting the list (if exists) of digital libraries hosting their contents.

### 7.2.5 Service level agreements (SLA)

The use cases reported above describe an agreement between the Service Provider and the Consumer. Service level agreements (SLA) take into account the Services to be offered by the Provider and also the so called SLA metrics enabling the assessment of the quality of the Services. Different metrics can be considered, depending on the kind of service and agreement.

In PrestoPRIME we will analyse and evaluate in detail the relevant SLAs and Metrics in Work Package 3. The results will be reported in Deliverable 3.1 that can be referred to for further details. Therefore in this Section with focus mainly on SLAs and related metrics

which are more relevant in the context of a search service provider offering access to audiovisual contents from an Archive.

In the following we assume that the Provider offers networked Services via Web servers and therefore the evaluation of the Services according to the SLA is focused on this particular scenario.

The SLA should take into account the following issues:

- search functionality and its availability on the server must be specified in the SLA (e.g. server maintenance and evolution producing service interruption must be in accordance with the SLA)
- access of the content using a Web server
- the Consumer will use the provided material according to the limitations provided in the SLA
- such limitations can be expressed in a License of Use which is signed by the Consumer before accessing the audiovisual content
- the Providers will guarantee an appropriate Quality of Service according to the SLA for what concerns the bandwidth, content availability, search functionalities, ...
- the additional information available to the Consumer during the search and before accessing the audiovisual content are specified in the SLA
- the Provider is in charge for an appropriate information about any law, copyright and use limitation (including also restriction related to the Country) for the audiovisual content
- the Provider will guarantee the correct interaction with the relevant Archives in order to provide up-to-date audiovisual content

### 7.2.6 SLA Metrics

The SLA metrics relevant for a service provider have already been mentioned above. Additionally, the following metrics could be considered as crucial for a search service:

- **service availability** - the service should be guaranteed up and running above a given threshold, e.g. 95% of total time
- **search performances** -taking into account the maximum response time for a query
- **search results guarantee** -the quality of the results retrieved after a query can be evaluated, based on recall and precision measurements

The search service could also be affected by other activities related to the archive. For example, after the ingestion and validation phase of new contents have been completed, the availability in the search results depends on the time required by the archive for storage and indexing operations.

The metrics above should also be tuned according to the real search capabilities offered by the provider. A search provider could implement only the query by content ID or by simple metadata (title, author, ...), while an other one could provide advanced functionalities, such as complex metadata search (based on content descriptive metadata, such as speech transcription, rights information, ...) or maybe even a query-by-content functionality (based on feature extraction from multimedia content). For example, moving from exact match queries to similarity queries, could affect the search performances and the quality of the results and such advanced resource-consuming functionalities should be appropriately regulated in the SLAs.

### **7.2.7 Integration with other services**

Search services should be integrated with the access services (described also in the previous sections) dealing with both content and rights, and should be provided stand-alone as well as integrated in a complete solution offered by preservation service providers.

### **7.2.8 Relationships with other roles**

The service providers described in this Section offer access and fruition to other actors involved in the value chain of the audiovisual material. The following roles have to be taken into account in the scenarios implementation:

- Generic user searching for digital contents for educational or entertainment purposes
- Professional user searching digital contents for creating new items, for business purposes, looking for complex search features and also requiring the information related to the rights clearance
- Broadcasters needing complex search functionalities, also enabling the search for additional and structured descriptive metadata
- Digital Libraries looking for advanced search capabilities

Actors described above, playing the role of content consumer in search scenario, can also play the role of providers, producers and distributors. The search service can be figured out as a cross functionality enabling digital content exploitation with different business models.

## Glossary

Term Name	Definition
4:2:2	The subsampling scheme for video components, in YCbCr colour space, for which the two chrominance components are sampled at half the sample rate of luminance: the horizontal resolution of chrominance is halved, the vertical resolution is the same.
AES	Audio Engineering Society
AIP	<b>Archive Information Package</b> within the OAIS reference model The whole set of data which makes an Item within the archive as the object of the preservation activity.
Access	The possibility for an archive user to actually start the Fruition of archive content. For a single Item this includes: the possibility to find it (a search activity can be successful), the possibility to verify its properties through an examination activity (possibly including a level of preview), and the possibility to obtain all the necessary delivery information (credentials, permissions, resource locator, etc). Moreover access is effective only if the conditions for the fruition are completed satisfied.
DIP	<b>Dissemination Information Package</b> within the OAIS reference model The whole set of data which makes the bundle of the archive Item which is made available by the archive to a consumer user.
DPX	<b>Digital Picture Exchange</b> . See also <a href="http://en.wikipedia.org/wiki/DPX">http://en.wikipedia.org/wiki/DPX</a> This is a file format adopted for the result of film scanning.
Digitisation	The process by which the data sources of AV material are recorded and saved as computer files, according to an appropriate format, from a non file based source, such as AV recording on either analogue or digital video tape. This should not be confused with the migration from a analogue tape format (e.g. Betacam SP) to a digital tape format (e.g. Betacam IMX or Digital Betacam). The digitisation process is particularly critical and important when it produces a new master.
EBU	European Broadcasting Union – A confederation of broadcasting organisations, most of which are public service broadcasters, based in Europe and the Mediterranean
EDL	Edit decision list – An ordered list of references to AV source material, time-point-in, time-point-out (or duration) from which a target material can be obtained in order to conform to the final editing.
Editorial	Related to what the expressive, artistic, and communicable aspects of an audiovisual work are. In the PrestoSpace project we defined the concept of <b>Editorial Object</b> and editorial constituent part. In the AV scope the general concept of



Term Name	Definition
	<p><b>Editorial Entity</b> is mentioned in opposition to Material.</p> <p>An editorial entity is realised by an instance of AV Material, which can be generated by Sources (or Material Sources), consisting of suitable data recorded on files or media.</p>
Exploitation	<p>A quality level of AV material, generally lower than master, which permits a range of professional uses of the materials, including typically publication and simple editing.</p> <p>This level, appropriately defined in details, may be considered the default one for producing a DIP for professional/commercial use, unless master level is expressly required.</p>
FTP	File Transfer Protocol
Format	<p><b>Referred to AV-media files</b>, means the information set for the identification of type of wrapper (file type, standard specification, subtype) and the type of coding schemes for AV content (including significant details).</p> <p><b>Referred to carriers</b>, means the information set for the identification of the type of cassette and or recording medium, which is in turn necessary to identify the appropriate recorder and player devices.</p> <p><b>Referred to AV Content</b>, means the information set for the identification of the main technical properties of AV-Material (e.g. for video width and height, aspect ratio, frame rate, for audio number of tracks, sample rate, sample depth)</p>
Fruition	<p>The process by which a user satisfies needs with some AV archive content.</p> <p>It includes: getting the information related to an AV-work, having a preview of the material getting a presentation playback of the material (with/without control), getting control of a material source instance and making use of that (playback, publish, re-use in new work, etc).</p> <p>For each fruition aspect the user will have got granted access and the relevant rights.</p>
HDTV	High definition television. A (digital) television system with properties of AV-Material higher than those SDTV. E.g. video resolution 1920x1080. See [HDTVspecs]
HEI	<p><b>Higher Educational Institution</b></p> <p>Universities and other institutions providing third level education. They host teachers, researchers, and students. They deal with AV works, which they produce themselves or purchase, relevant to research, cultural, teaching/learning activities. They have some competence for managing their AV collection, but this is not their specific expertise.</p>
ISO	International Organization for Standardization
ITU	International Telecommunication Union
Ingest	The process by which the archive gets an input item in the form of SIP

Term Name	Definition
	and, on success, becomes responsible of its preservation and accessibility.
Item (archive - )	<p>The AV element which is considered by the archive and their users as the main independent object of the preservation for which we can think of atomicity. So either the preservation of the whole item is successful or it has to be considered failed.</p> <p>The relationships between AV archive items and editorial entities can vary according to defined archive policies.</p>
LDTV	<p>Low definition television. A (digital) television system with properties of AV-Material higher than those SDTV. E.g. video resolution 352x288. Format examples are CIF (Common Interchange Format – ITU-T H261) or SIF (Source Input Format – defined in Mpeg1 i.e. ISO/IEC 11172)</p>
Lossless (- compressed)	<p>An AV Material component (Audio or Video) is compressed lossless when the coding scheme permits to record/save data sources with fewer bits than those required for uncompressed but it is possible to reconstruct exactly the original material (no bit change).</p>
Lossy (- compressed)	<p>An AV Material component (Audio or Video) is compressed lossy when the coding scheme permits to record/save data sources with fewer bits than those required for uncompressed but it is not possible anymore to reconstruct exactly the original material (bit change).</p> <p>The decoded (reconstructed) material may however look quite the same as the original (subjective criteria) if the information removed were not perceptible (under certain circumstances) to the viewer (or listener).</p> <p>Limited to video, the lossy compression is considered <b>High End</b> if the obtained quality is sufficient for most current production processes (multi-generations, complex editing, graphics), <b>medium</b> if the quality is sufficient for current publication and for limited production purposes, <b>low end</b> if impairments are easily noticeable (and even annoying) and the quality is sufficient only for limited publication services.</p>
Master	<p>The <b>Master</b> is the available material instance realising an AV-work, with the highest quality.</p> <p>In principle this should be the <b>original material</b>, as it was created, as any further process can modify and affect the resulting quality irreversibly. However this refers also to the format in which an AV-work is <b>consolidated</b> at the end of a production process, independently from the formats used during the productions stages.</p> <p>The master quality level is the highest quality, used as reference, and from which any equivalent or lower quality may be derived.</p> <p>Especially in the AV archives scenarios, the quality of the Master copy is not necessarily high, because the content might have been created in a low quality format or better quality copies may not be (anymore) available.</p>
Material	<p>An audiovisual entity which can be obtained by the playback of some AV recording. Material properties are the duration, the type and number of tracks, the video aspect ratio, and other one. Not to be confused with</p>

Term Name	Definition
	Material Source.
Material Source	The data which are the coded representation of audio-video recording. Material can be obtained by a Material Source by means of a decoding process. Examples of properties of Material Source (esp. digital) are the encoding scheme and the number of bits.
Migration	<p>The process through which a whole set of AV-items is submitted to a relevant and critical modification of state. The outputs of the migration supersede completely the inputs, which generally might be destroyed or assigned to a different role.</p> <p>Typically the migration processes are critical in terms of the cost, the required time, and risks.</p> <p>Referred to format, it means that some properties of the outputs will be different from those of the inputs, and this might include quality issues concerning the AV-content.</p> <p>Referred to storage, it means the access mechanisms might be changed, however all the properties of the AV-content are preserved.</p>
MXF	Material Exchange Format: a container format for professional digital video and audio media defined by a set of SMPTE standards.
NLP	Natural language processing.
OAIS	<p>OAIS is the ISO reference model for <b>Open Archival Information System</b>. The OAIS reference model is defined by a recommendation of the Consultative Committee for Space Data Systems.</p> <p>The OAIS reference model basically defines that a <i>provider</i> user submit to the archive a SIP which is ingested to become a AIP which in turn can be transformed into DIP in order to be made available to a <i>consumer</i> user.</p>
PSP	<p><b>Preservation Service Provider.</b> An entity that provides services for long-term storage, safeguard and access to AV assets owned by someone else.</p> <p>In this context the access to AV material is meant provided over a network.</p>
Point of Access	<p>A facility provided over the network for giving the first access to AV-Content (and other types of digital content) actually hold by multiple organisations. The organisation providing this service.</p> <p>The users experience regarding the search and the first examination of the results is that of having a single large collection of items to explore, offered through a homogeneous and consistent interface.</p>
Preservation	<p>The process by which the digitised AV contents are kept safe and usable in the long-term.</p> <p>This includes the safeguard (integrity) of the material sources and the capability of providing back the materials for the expected use.</p> <p>The preservation objective can be reached by setting up various and different preservation strategies, which must be selected according to the specific cases.</p>

Term Name	Definition
	Note that in PrestoPRIME context the Preservation process applies after digitisation while in PrestoSpace it referred to AV-Content in general, including thus the preservation of legacy carriers.
Quality level	<p>The rough definition of the properties of AV-Content in terms of quality according to the expected use of the AV-Content.</p> <p>It may refer also to the format as the quality level obtainable by its adoption.</p> <p>Mentioned quality levels are:</p> <p><b>Master</b> - highest reference from any equivalent or lower quality can be derived</p> <p><b>Broadcast</b> - the quality appropriate for <b>publication</b> through current mechanisms</p> <p><b>Browsing</b> - the quality sufficient to appreciate and recognise the content, albeit some impairments might be quite perceptible</p> <p><b>Exploitation</b> - a quality lower than Master level but still appropriate for limited production activity and for being the source from which publication formats can be obtained</p>
SDTV	<p>Standard definition television. . A (digital) television system with properties of AV-Material that meet standard values of legacy analogue television. E.g. video resolution 720x576.</p> <p>Format reference standard is ITU-R BT 601 [SDTVspecs]</p>
SIP	<p><b>Submission Information Package</b> within the OAIS reference model</p> <p>The whole set of data which makes the bundle of Item which is ingested by the archive taking the responsibility for its preservation.</p>
SLA	<p><b>Service Level Agreement.</b> The terms agreed between the provider and the customer of a service which define the obligations on both sides.</p> <p>It includes the definition of the service and the conditions within which it is operated.</p>
SMPTE	Society of Motion Picture and Television Engineers
Scenario	<p>A description of the context of use of a system, from a given perspective, intended to identify the questions that must be addressed in order to meet the context needs.</p> <p>It includes the indication of the relevant roles and the descriptions of the typical cases.</p>
Timeline	<p>The temporal dimension of the AV content along which the presentation sequence is defined.</p> <p>An excerpt of AV material is identified by its duration and its position on the timeline of the Material from which it is extracted.</p> <p>Any itemised description of properties related to AV content generally requires the reference to the timeline position.</p>
Transcoding	The process through which a material source is transformed from a coding scheme to another one, implying a modification of the obtainable materials if the target coding scheme is a lossy one.

Term Name	Definition
	<p>The quality obtainable from a transcoding is either equal to or lower than the original quality.</p> <p>Transcoding is often complemented by transwrapping</p> <p>An improvement of quality can only be obtained from a more complex AV processing, including aspects of restoration, and it's liable to subjective evaluation.</p>
Transwrapping	<p>The process through which the format of the container for material sources is transformed into another one without modification of the coding scheme, implying that the obtainable materials are kept unchanged and the quality is the same.</p>
UEFA	Union of European Football Associations
UGC	User generated content
Uncompressed	<p>A coding-scheme format for which the digital AV material sources are saved with exactly the same bits with which the directly obtainable materials are digitally represented.</p> <p>If this coding is adopted since the creation then the highest quality for the given format is achievable.</p> <p><b>Note:</b> in the broadcast contexts the video coding schemes YCbCr/4:2:2 are traditionally considered uncompressed although the chrominance components are sub-sampled with respect to luminance and thus the number of bits of these coding-schemes is lower than that required by recording the same video as RGB.</p>
VOD	<p>Video On Demand – A mode of programme distribution by which an audiovisual content is delivered by means of a service that permits the subscriber to view it at a time chosen by the subscriber (without reference to a schedule pre-established by the service provider) – When the term is referred to Rights, it requires a more exact definition.</p>
VTR	Video Tape Recorder