Deliverable 16.4

Delivery models

DOCUMENT IDENTIFIER	PS_WP16_RAI_D16-4_Delivery Models
DATE	2006-03-15
ABSTRACT	The document describes and analyses the models representing how the PrestoSpace Factory may provide its services to the Archives. A specific component, the PrestoSpace Orchestrator, responsible for managing the relationships with the Archives, has been defined together with the involved entities. The outlines of the relevant software components are given.
AUTHOR, COMPANY	Laurent Boch (Rai – Radiotelevisione Italiana)
CONTRIBUTORS	Michele Vigilante, Walter Allasia (euriX)
KEYWORDS	Archive management, archive inventory, archive exploitation, editorial objects, audio visual material sources, essence and metadata storage, transaction orchestration, preservation restoration and documentation services, orders, batches, delivery options, working options, essence and metadata formats.
RELATED ITEMS	System architecture, storage architecture, system integration, MAD area Business Process model.
NOTE	At the date of delivering this document it has been conceived the intention to merge into a single and more extensive document the present one and current Deliverable 3.1.

DOCUMENT HISTORY

Release	Date	Reason of change	Status	Distribution
0.0	2006-01-05	Draft	Living	Author + gdimino
0.1	2006-01-11	Draft	Living	Project
0.2	2006-01-17	Draft	Living	Project
0.3	2006-01-23	Draft	Living	Project
0.4	2006-01-31	Finalisation	Delivery draft	Author+euriX
0.5	2006-02-01	Added Software tools	Delivery draft	Project
0.6	2006-03-02	Revising on comments	Delivery draft	Project
1.0	2006-03-14	Fixing status of software component section And final Delivery	Delivery	Project
1.0.1	2006-03-15	Added missing header fields (abstract, keywords, related items)	Delivery	Restricted
2.0	2008-02-26	Dissemination level changed to Public	Closed	Public

Table of contents

List of figures

List of Tables

1 Scope of the document

The scope of the present PrestoSpace deliverable, “Delivery Models”, is the scenario of transactions which have to take place between the PrestoSpace Factory and its institutional customers, namely the audio-visual Archives. The analysis on the background turns around the Archive management process.

The PrestoSpace Factory is defined to be a complex set of facilities which are able to provide massive analog-to-digital migration of audio-visual contents, digital restoration, and the documentation and metadata enhancement, obtaining also by means of automatic features extraction, which are going to permit the best exploitation of the archive contents.

The assets of the Archive are therefore the object of the PrestoSpace Factory processes and the accomplishment of the work is achieved with the final delivery to the Archives of new digitised Master Quality Materials – and/or any other requested quality level such as broadcast or browsing - together with the set of identification, description, and indexing information which make up the structure for information and material retrieval within the process of publication of archive contents.

It happens that the delivery models are actually depending on many factors, such as the model of the process which is founded on the Archive-Factory relationship, the specific needs of the Archive organisations, their situations, and their expectations.

Beyond the modelling, the examination of options, the description of processes, the actual possibilities of implementations depend on the availability of suitable software tools.

Some pertinent tools had already been developed for the Documentation process within the PrestoSpace MAD Working Area. The subsequent definition within the PrestoSpace Factory of a specific component responsible for the whole interface towards the Archive has carried to the development of the so-called PrestoSpace Orchestrator component (PSO), the description of which is given in chapter 6.

2 References

[D3.1] Models and Protocols for PrestoSpace Factory process - SYS_D3.1_SAS1_Models_Protocols – version 2.0.9 - *Jean Hugues Chenot (INA), Andrea Bernicchia (Reply), Peter Schallauer (JRS), Walter Allasia (Eurix), Laurent Boch (Rai) – 2006-03-01*

[D12.4] A Strategy and Architecture for Storage - D12.4 Storage Tech Architecture– *Richard Wright (BBC) – 2006-03-02*

[D15.1] Analysis of current audiovisual documentation models - D15.1 MDS1 – version 1.6 – *Christoph Bauer, Florian Rosensprung, Stefan Lajtos (ORF), Laurent Boch (RAI), Philippe Poncin (INA), Conny Herben-Leffring (B&G) – 2005-03-16*

[PSWP] PrestoSpace project web home page - http://www.prestospace.org/ - *2006-01-20*

[WKPD] wikipedia - http://www.wikipedia.org - *2006-01-20*

[PMETA] Tech 3295 - P_META Metadata Exchange Scheme - EBU Tech 3295 (notes) versions 1.x – *EBU - 2002-2005*

[BPM] MAD Area business process model - MAD Area Analysis of Business Process Model MAD Area Analysis of Business Process Model v0.4 – *L. Boch, A. Messina - 2005-07-04*

[MDM] MAD Area data model - MAD Area data model version 0.3 - *L. Boch, A. Messina - 2005-06-10*

[MDF] MAD Area data formats -MAD Area Data Format Version 0.96 - *A. Messina, W. Bailer, P.Schallauer - 2006-02-14*

3 High level transaction models

3.1 Criteria of work

"*The aim of the PrestoSpace Factory is to provide affordable services to audiovisual Archives in order to manage and to allow access to their assets".* [PSWP]

This analysis about the high level transaction models of the PrestoSpace Factory has been carried out so as to identify the distinguishing elements of the varied possible operating scenarios and to understand the implications towards the possible delivery models.

It has been considered particularly important the reference to the basic business process model of an audiovisual Archive, because the role of the Factory is, beyond the technical solutions for digital preservation, to enhance the effectiveness of the Archive mission itself.

Also it has been given importance to the identification of the business actors, that are those entities playing a specific role within a relationships.

To this purpose, it is useful to observe that some services are going to be used in order to fill the gap which prevents the Archives to make the most of their collection, while other services aim to assist the Archive exploitation in a continuous way. In the former case it is easy to identify the Factory and the Archive as the two relevant parts in the transaction relationships, while for the latter the offered services work on behalf of the Archive towards the Archive Users.

A preliminary classification of the Archives, and of their situations, is the pivotal activity for the analysis of the transaction models and several dimensions are going to be interested:

- Archive dimension – Roughly the amount of items, physical media or hours of content. Although the evaluation of this component can vary considerably according to the expectations of future exploitation.
- Status of assets, physical – This gives the weight of the preservation & restoration components of the Factory required services.
- Status of assets, information – Generally the quality of the legacy information.
- Archive organisation – Is the archive a standalone organisation or rather a department providing archive services to a wider organisation? Is the current structure planned to be changed? Which are the skills of the work force?
- Infrastructure & logistics – Which are the current and possible resources in terms of locations and equipments.
- Archive Users – Who are the current users and who are going to be in the future?
- Budget & Founds – Which kind of investments are contemplated? In which timescale?

It is easy to grasp that all the mentioned components are inter-related, as it hasn't been identified a unique solution for digital audiovisual archives.

3.2 Basic archive business process model

The scenario where the audiovisual archives operate includes basically three other kinds of actors, as found in [PMETA] and reported in Figure 1

> *Producers (Content creators)* - These are involved in the Production of programmes and other content and make new material available for publication.
>
> *Archives* - These are involved in the preservation of existing material, that can be re-used as originally produced, or to provide input material for new programmes.
>
> *Distributors* - These are involved in the publication and delivery of material to the end users. They are the ones who have interchanges in the business-to-consumer scenario. They are also known as Broadcasters or Content Aggregators.
>
> *Consumers* - These are the beneficiaries at the end of the media supply chain.

Quote 1 - P_META model actors

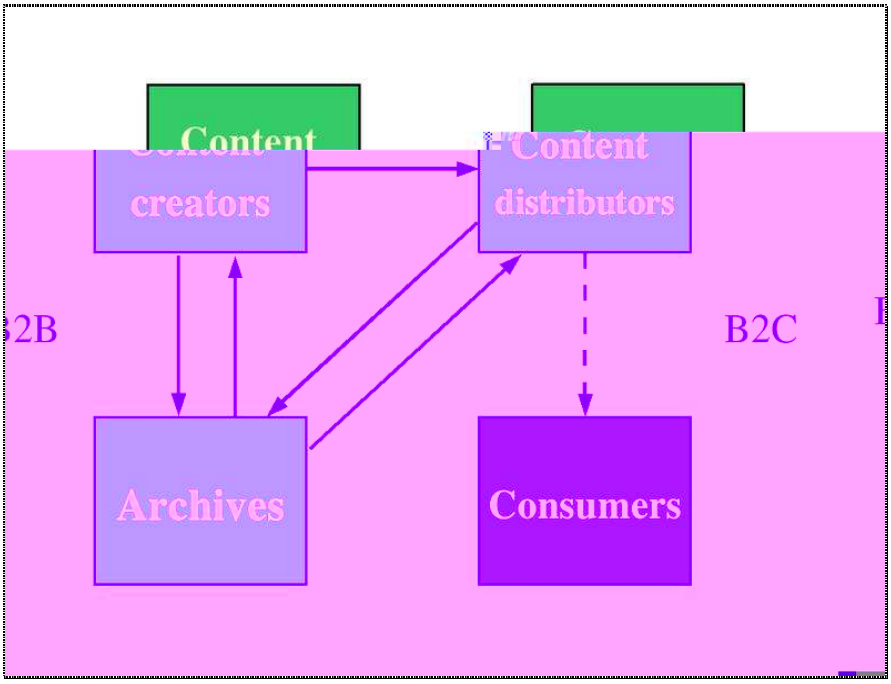


Figure 1 - P_META Process Model

The P_META model was depicted in order to define the operating context of business-to-business information exchange about audiovisual assets. In that model the ***Archive Users*** are implicitly defined to be either Content Creators or Content Distributors, hence excluding the end users. However the model identifies roles rather than organisations and actually most main broadcaster companies cover, totally or partially, all of the three business domains.

Nowadays the Archives organisations are considering qualifying more precisely their users, in order to find the resources necessary to feed the processing loop consisting of preservation and exploitation.

3.2.1 The archive management processes

Within the work area MAD (Metadata Access and Delivery) of the PrestoSpace Project an analysis of the business process model was carried out for that domain, by producing and commenting a set of use cases diagrams, and is available in [BPM].

According to the PrestoSpace objectives the audiovisual archive management is described from the point of view of the digital archive and therefore it includes processes, such as Digitisation and Restoration, which are going to be provided by the PS Factory services. The relevant diagram is given in Figure 2

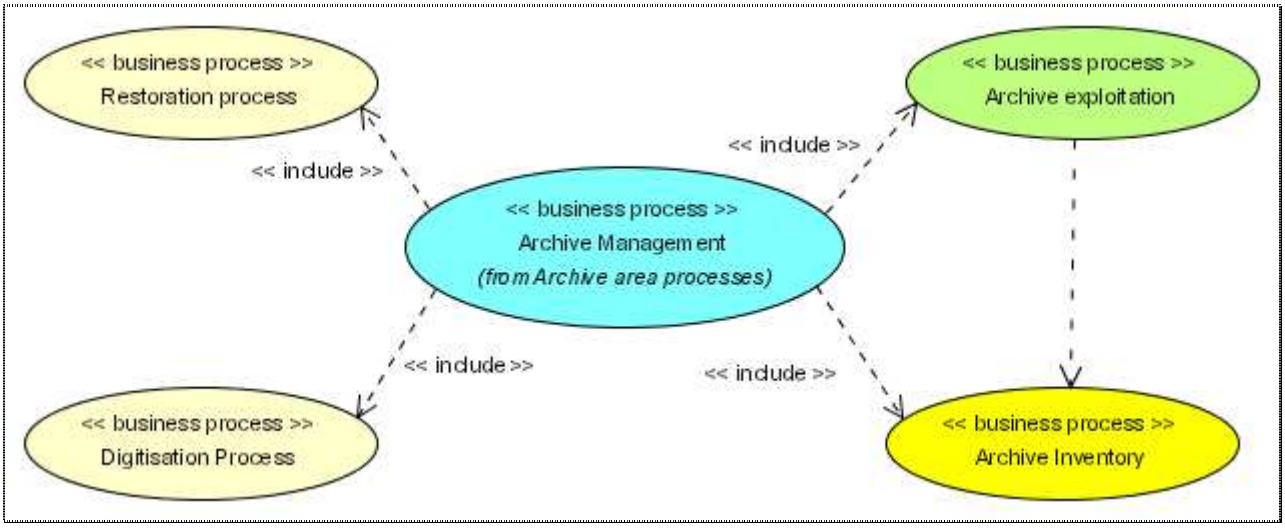


Figure 2 - Archive management processes

The Archive Exploitation process is the objective to which the other sub-processes are finalised. However, in order to actually use the archive content, an Archive Inventory process is necessary, as pointed out by the explicit dependency relation in the diagram. The inventory process is here intended as the way to obtain and maintain the information about which are the owned editorial assets and how they are physically realised by the audiovisual items stored in the physical archive.

3.2.1.1 Archive inventory process

The inventory process, showed in Figure 3, includes as sub-cases the inventory of Editorial Objects, Editorial Collections, and Material Sources, and it is accomplished through the association between Editorial Objects and their realising materials.

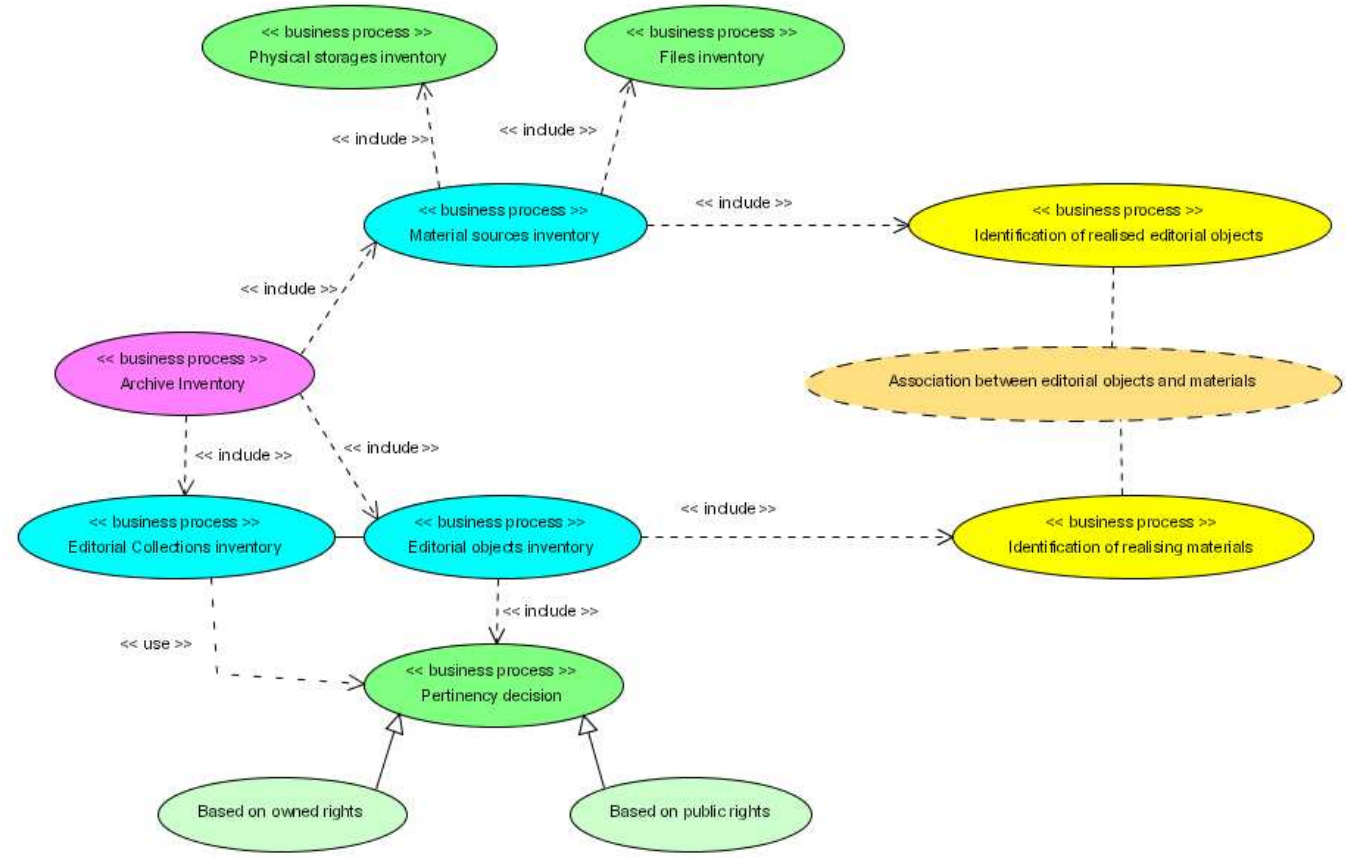
As pointed out in the referenced document, that association requires mutual identification and timelines mapping, a crucial task, the difficulty of which may vary from quite simple to quite complex, and that in practice has been accomplished for the legacy archives with various degrees of accuracy.

3.2.1.2 Archive exploitation

A further detailed diagram for archive exploitation is given in Figure 4. The main objective of the archive user is retrieving either the audiovisual items stored in the archive or the information thereby located in the same place.

The latter gives account of the cases in which archive users are interested in obtaining pieces of knowledge without being directly interested in the audiovisual material somehow connected with it. For example knowing whether certain events happened or not, knowing about the content of a public speech of a famous person, and more in general making reports of certain aspects of social, arts, costume. From this point of view, the information that is living in the archive becomes itself an asset for the organisation owning the archive.

Both the retrieval functions can be distinguished in use and fruition. The use consists in the actual employment and entails exportation from the archive, while the fruition is limited to accessing information and material from the user interface by previewing, browsing, and reading.


<< business process >>
Physical storages inventory
<< business process >>
Files inventory
<< include >>
<< include >>
<< business process >>
Material sources inventory
<< include >>
<< business process >>
Identification of realised editorial objects
<< include >>
<< business process >>
Archive Inventory
Association between editorial objects and materials
<< include >>
<< include >>
<< business process >>
Editorial Collections inventory
<< business process >>
Editorial objects inventory
<< include >>
<< business process >>
Identification of realising materials
<< include >>
<< use >>
<< business process >>
Pertinency decision
Based on owned rights
Based on public rights

3.3 Relevant transaction models

The PrestoSpace Deliverable D3.1 provides a system architecture overview pointing out, from that perspective, the main transactions between Archive and PS Factory and among the PS Factory Units, with specific interest in preservation, restoration, and documentation.

Although the issues are logically not the same, the system architecture, the storage architecture [[D12.4], the workflow, and the transaction model are related. Indeed, also with consideration about the assortment of archives, within the PrestoSpace project it was made reference to a so-called *Negotiator* for understanding the customer requirements and needs and for setting up the transaction model and workflow most suitable for each case.

Figure 5 describes shortly the system architecture defined in [D3.1].

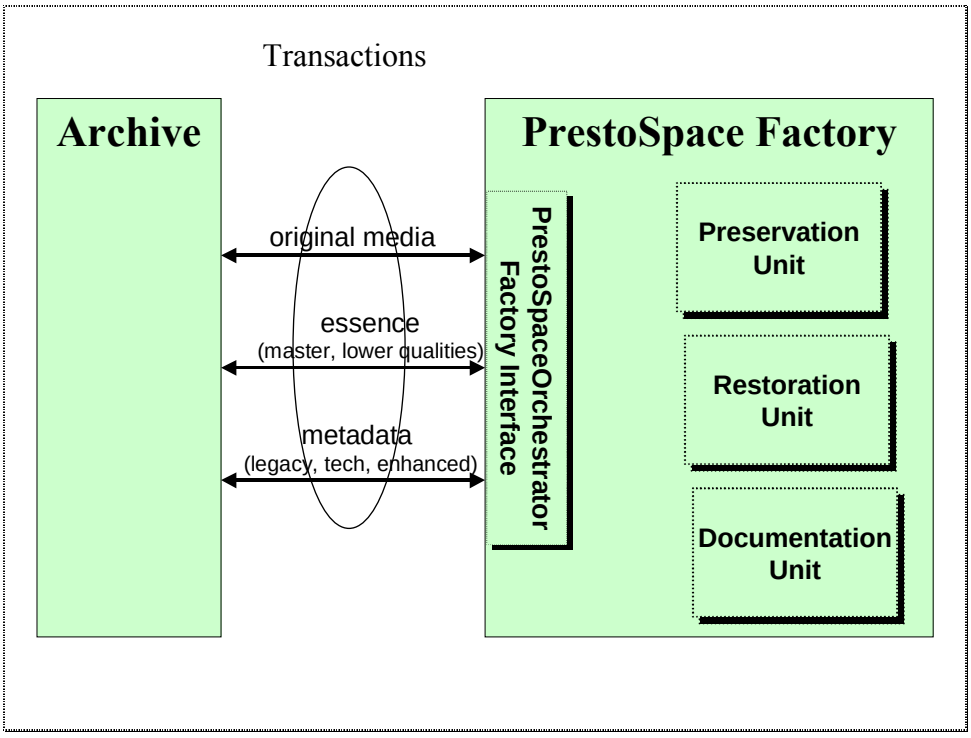


Figure 5 - System architecture overview

The arrows in Figure 5 tell which kind of entities is involved in the transactions. Actual transactions, however, need to be supported by services which will have to follow a defined protocol, use defined formats, and require a wide set of accompanying information which will specify all the required details and working directives (options).

The relevant main transaction models have been analysed according to the following dimensions:

- *Space* – Where the PrestoSpace Factory is located regard to the Archive location. The answer impacts especially on the transfer of physical storage media, but also affects the availability and the bandwidth of any required data connection.
- *Time Cycle* – How transactions and workflow are organised along time. This is important when the amount of archive items requires having many working batches. At a given time one item may be completed, within the Factory, scheduled by the archive, or discarded from the process. Moreover if the relationship between Factory and Archive goes on for a while, it is likely that some modifications of the relationship occur.
- *Termination* – Whether the relationship between Factory and Archive has a natural conclusion or may continue indefinitely. The former answer typically applies to the digitisation of an archive, where analogue tapes or film go through preservation and restoration processes in order to obtain new masters. In this case once all legacy formats are transformed to digital the Factory has completed its task, because all newly created archive items should be originated in digital form. However the Archive may intend to apply for the Documentation services even later, without mentioning services for the archive exploitation (publication).
- *Workflow management* – Whether the Factory has a main workflow management in charge also for interfacing towards the Archive, and thus responsible for the transactions, or not. In the former case the

Factory shall include a component named *Orchestrator*, which will manage the exchanges between the Archive and the PrestoSpace Factory Units. In the latter case each Unit is in point of fact independent and the Archive shall implicitly do the Orchestrator task.

However it has been observed that the practical consequences of any option of transaction model differ considerably according to the kind of entity involved in the exchange, namely those given in Figure 5:

- *Original media* – The physical removable storages where audio and video were originally recorded. Generally they are in an analogue format, although most considerations may apply also to those digital formats that are not "*File Based*", e.g. Digital Betacam.
- *Essence* – The digital audio and video material sources. Without mentioning coding (compression) and wrapping formats, we can say that essence is stored and exchange as binary data. Typically a large amount of data, please make reference to [[D12.4] for further details.
- *Metadata* – The information about the audiovisual material and its editorial content. The scope of the information is quite wide, including entity identifications, descriptions, and technical details. They can be stored and exchanged as binary data; the data formats adopted require text-based representation in XML, which can be lossless, compressed for exchange/storage efficiency. The size of data is quite lower than for essence.

The work carried out within the MAD Area of the PrestoSpace project as what concerns the business process model [BPM], the data model [MDM], and the data formats [MDF], provides a model defining the relationships between those entities and means to keep consistently all link information.

To summarise the resulting concept, the entity which is the fulcrum of interest is the *Editorial Object*, basically *what comprises the expressive, artistic, and communicable aspects of an audiovisual work* (or an editorial constituent part of it), which is realised by at least one instance of audiovisual *Material*, which can be generated by *Sources*, consisting of suitable data recorded on files or media[1]. Figure 6, taken from [BPM], provides an example showing the relationships among those entities. It must be also remarked that the complete scenario of mapping between Editorial Objects and Materials include several particular cases, some of which are described in Table 15.

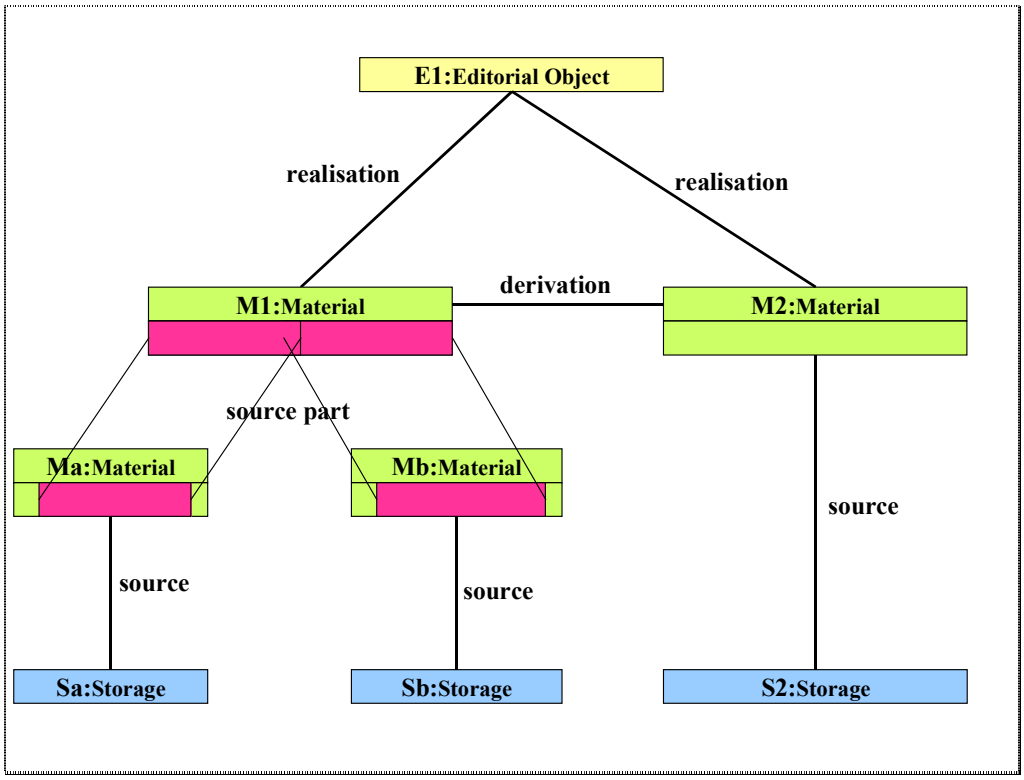


Figure 6 - Objects diagram with Editorial Object, Materials and Storages

[1] At an higher level the model doesn't make much difference between *Files* and *Tapes*, as both are containers for the Material Sources. At a lower level the issue of providing actual *Accesses* to the Sources need clearly to have distinct solutions

3.3.1 Transaction models for original media

Transactions involving the original media are particularly critical when the analogue master storages are moved to the Factory. Indeed inconveniences and failure at this stage may cause the irrecoverable loss of audiovisual material and subsequent lacks for the realised Editorial Objects.

Once a new digital master is available in a file format from the Preservation Unit, the old analogue media become redundant and to give them back to the owner archive is a due action and a further caution.

Factory **on site**

If the Factory, or at least the Preservation Unit, works **on site**, at the Archive locations, the transfer of original media is quite simpler. A hand trolley is sufficient to displace a group of storages from their shelves to the preservation equipments. Usually there is no need to wrap them into a closed box and the transfer time is not that significant.

Factory **off site**

When the two locations are different, the Factory is **off site** and the transfer has a greater level of complexity. A shipment has to be set up. There are a *Sender*, a *Receiver*, and at least a *Carrier*, actors with their defined responsibility. Storages are grouped in batches for the displacing, wrapped in containers (boxes) and the transfer time may not be negligible.

3.3.2 Transaction models for essence

The discussion of transaction models for essence is strictly related to the issues described in [D12.4].

As remarked above, essence data can be duplicated as many times as needed, but they are bulky (to store them is expensive). Transaction for essence means also to transfer the responsibility about preserving those data. In most cases the process can be summarised by the following steps:

- A copy of the essence data occurs from the Sender to the Receiver
- The success of the copy is verified at the Receiver side
- The Sender is authorised to remove the source copy
- Update of the Accesses information

Factory on site with common storage service

In this case not only the Factory works at the Archive location, but also both share some kind of advanced storage service. Essence data transfers may apparently happen such as described above, but what actually takes place in the back end depends entirely on the implementation of the common storage architecture and it is mainly opaque.

Basically for this model the Essence transfer is rather the modification of the status and of the permissions.

Higher level software components must be set up in order to take the greatest advantages from the storage services, avoiding operations, which may carry to waste of capacity or other inefficiencies.

Factory on site with LAN

To have some high bandwidth local network is most easily affordable when the Factory works on site. Nowadays Ethernet 100Mbit/s infrastructures are the entry level and GigaEthernet is almost commonplace.

Most of the security issues can be solved at the boundaries and therefore Factory and Archive systems can access a variety of exchange protocols (ftp, smb, http, file).

The basic assumption is that the two actors have a separate jurisdiction on their respective systems.

Although the default scenario implies data duplication for the Essence transfer, it is possible to avoid this step if both the Factory and the Archive systems can share a File server. In this case the transfer of data ownership can be achieved also by moving file links and changing the permissions.

Factory off site with data network connection (WAN)

When the Factory works off site, it should be possible anyway to set up a suitable data connection.

The differences with respect to the previous scenario are:

- There is the need of a provider for the connection; there are greater security issues and limitations of the available protocols.
- The bandwidth will be typically lower with greater transfer times, possible connection interruptions, and more attention required to data integrity (verification of checksums).

Factory off site without data network connection

In this case we have the Factory off site and absent or poor data connection (at least for the Essence data).

This model can be similar to that already described for the original media, because the exchange will be based on the transfer of some kind of removable media (data tapes, DVD, removable discs, etc). The main difference is that Essence data can be recorded on those media as files and thus the Receiver can extract them at any time.

Common aspects

The essence data are actually binary material sources and are contained in files. To say it with the terminology of the MAD Area data model, Files are a type of Storage for Material Sources (see Figure 6), and so they are also involved in the Archive Inventory Process, as shown in Figure 3.

By default the identifier for a given file is its name (*FileName*), but that is clearly a weak identifier because of the easiness for changing it and the difficulties to set up and run a suitable file name policy.

The model developed within the PrestoSpace MAD Area specifies that a defined Material Source be also identified by the Material instance that it is able to generate ("canonical play"). And the Material instance is identified by the UMID[2]. It is useful to remind that the UMID has to be generated and provided by the entity emitting the Material instance.

When a Material Source is required to be transferred (*getMaterial(UMID), insertMaterial(UMID)*) the owner of the Source has to provide the information about the set of available *Accesses*.

The verification of the File identity is permitted through the control of the *checksum md5*, and the file size information allows the tracking of the transfer progress.

Table 1 contains a couple of examples of the information set provided to enable to access media files.

```
<Materials xmlns:wu="http://www.prestospace.org/MAD/CorePlatform/xml-instance">
    <Material UMID="0x060A2B340101010501010F3013000000DA06F937CE2A4861AA6CE901B6E8E47C">
        <Source>
            <File format="video/mxf" name="shortclip.mxf" num_bytes_qty="987654321"
                  md5_checksum="a2a0cf152dabee9839ab6de659cda899">
                <Accesses>
                    <Service issue="2004-10-13T14:36:47" validity="2004-11-13T14:36:47">
                        <Protocol>file</Protocol>
                        <Port>-1</Port>
                        <Host>localhost</Host>
                        <Context>/shared</Context>
                        <Name>shortclip.mxf</Name>
                    </Service>
                </Accesses>
            </File>
        </Source>
    </Material>

    <Material UMID="0x060A2B340101010501010F3013000000CEDBB351D2384E168C6760123F2518AE">
        <Source>
            <File format="image/jpeg" name="00_00_aa_bb.jpg" num_bytes_qty="7654"
                 md5_checksum="92a899b178f74c3c4ba637dc9acde4a5">
                <Accesses>
                    <Service issue="2004-10-13T14:36:47" validity="2004-11-13T14:36:47">
                        <Protocol>file</Protocol>
                        <Port>-1</Port>
                        <Host>localhost</Host>
                        <Context>/shared</Context>
                        <Name>00_00_aa_bb.jpg</Name>
                    </Service>
                </Accesses>
            </File>
        </Source>
    </Material>
</Materials>
```

Table 1 - Example of information exchanged for a get/insert Material

[2] The UMID is the Material Identifier as defined by SMPTE330M. In textual documents, as XML, it must be recorded as a string starting by "0x" and giving then the UMID value in hexadecimal (each byte is coded with a couple of characters in the range [0-9,A-F]). The UMID has 32 bytes with the a possible extension to total 64 bytes.

3.3.3 Transaction models for metadata

A couple of remarks are a premise for discussing the transactions of metadata:

- We don't mean here metadata as process management information, but on the contrary we care about the information about the audiovisual works (Editorial Objects) and materials. Likely that information is going to be limited when going from Archive to the Factory and enriched on its return path.
- The size of metadata is generally considered negligible compared to that of essence, and that's is almost true. However some attention should be paid because it is quite easy to produce large xml files, when using extensive generic schema (where some "verbosity" is required to avoid ambiguities) together with a great level of information detail. Data lossless compression is the possible solution.

Anyway it comes out that the issues about Factory location and network bandwidth are less important in the case of metadata, provided that some kind of connection is available. One could even argue that in the worst case metadata transfer could be arranged through the exchange of cheap removable data storages, such as CDs and memory cards.

So the transaction models on metadata have historically been designed around the relationship between metadata and essence.

Metadata embedded with Essence

This model is supported by several interesting arguments:

- Size of metadata will not impact on size of Essence+Metadata.
- Embedding constitutes a strong link between Essence and Metadata. You cannot lose your metadata (unless you lose the Essence too).
- You don't need to think about metadata transfer, because that's already solved together with the Essence problem.

Such option is also viable at the present time because of the availability of formats, such as MXF, supporting that feature. Moreover some formats, including MXF, also support the ascription of pieces of information to Material timeline intervals or single points.

Disadvantages on this model do exist and will be discussed below. However it is fair to remind that we are describing only the exchange activity, and that this choice should not constitute a constraint for the persistency models at both source and destination points.

Metadata separate from Essence

The advantages of this model partially oppose the drawbacks of the one described above:

- You don't need to process the Essence wrapper in order to access metadata (writing/reading).
- You don't need to move Essence and metadata at the same time; otherwise you need to wait for the last one to be ready.
- Metadata formats are independent from Essence formats (and thus from the tools specific for the latter)
- Metadata are not linked to any specific copy of Essence (master, rather than lower quality copies).

This option is affordable if metadata provide reliable links to the related Essence.

One adverse argument sometimes complains about the fact that handling the Essence you might not have links to the metadata. That's somehow misleading as an inventoried system must always be able to manage the association between Editorial Objects and Materials (see 3.2.1.1).

PrestoSpace MAD Area approach

The approach put in place by the MAD Area is oriented to keep separate transactions of Essence and Metadata, although both are clearly related.

Taking as a reference the diagram provided by Figure 6, the whole metadata set is contained within *the Editorial Object Document* (*EOD*) either directly or by reference to Material instances. In particular the EOD must contain the information about the Materials that realise the Editorial Object, primarily the UMID, which is going to be the key to be used in order to get the information for accessing the Essence data (material sources).

The transaction of an Editorial Object Document is going to be a validation point, which will determine the status of the Editorial Object respect to its Materials:

- *Complete* – All Materials are known and available.
- *Waiting* – All Materials are known but not all of them are available. An Essence transaction may be in progress.
- *Lacking* – At least a Material instance is not known.
- *Error* – In case of failure of actions directed to change the status form Lacking/Waiting to Complete.

Information stored within knowledge base repositories

The Documentation Units working within the PrestoSpace Factories will include tools for semantic analysis that make use of knowledge and information management services. Those services are originally made up of an “ontology” [WKPD search for ontology (computer science)] and an initial knowledge base (KB) [WKPD search for knowledge base], which is going to increase as a side effect of the semantic analysis process. The expanding may reach then saturation in the number of entities.

The information piled up within those knowledge base repositories is going to be very useful for future information retrieval in the Archive exploitation process. However it must be remarked that it is not possible to specifically assign any information elements of the knowledge base to Editorial Object instances.

The Archive exploitation can be considerably improved by the use of information retrieval tools relying on knowledge base repository. Therefore there is an issue about how to transfer the information between a documentation KB instance and a publication/exploitation KB instance.

The currently identified possible scenarios are subject to change in the future because of the evolution of this technology. The PrestoSpace MAD Area has examined the following cases:

- Services integration – the KBs of Documentation and Publication are strongly integrated and a continuous update mechanism is configured between the two KBs. This option depends on many constraints and would produce a communication independent from the PSO interface, which is not generally recommended.
- Traditional archive exploitation services – In this case there is no KB at the Archive publication end. Therefore the Archive publication has to rely uniquely on the information contained within the EOD.
- KB export as RDF – The documentation KB is exported, whenever requested, as RDF XML file [WKPD search for Resource Description Framework]. This option may rise to constraints in the update process of the Publication KB, however it is recommended because it carries to a greater independence between the PrestoSpace Factory and the Archive exploitation.

3.4 Factory customers profiles

As outlined in the “Criteria of work”, there are several components that generate the assortment of the Archives domain from which the factory customers come. A PrestoSpace Negotiator will be certainly interested in Archive dimension, budgets and timescales, however the final question is *“how far the current Archive status is far form the possible target?”*

3.4.1 Worst case

Archive is more a media repository than a veritable Archive. This occurs when the Archive is not able to run a sufficient management process (Figure 2), in particular concerning the inventory (3.2.1.1). Likely the status of assets is either bad or unknown and the legacy information is very hardly usable. Current archive exploitation is near to nil.

If the size of the “archive” is large and the available budget too, that could be quite a best case, from the factory service provider perspective.

The problem is that, without a starting point good enough, it is difficult to guess which could be the future exploitation possibilities.

A suggestion is to try to define a strategy in order to fall into a better profile, attempting to understand whether it is better to work at preservation without having a minimal inventory or to ask to the archive to do some housekeeping before.

3.4.2 Modernisation case

The profile described here is about Archives aiming to undertake a significant modernisation.

These Archives are currently working, i.e. they have some Inventory and Exploitation processes, but they rely on legacy tools, especially concerning the management of information, which prevent significant enhancement. Often their organisations have already progressed in the information technology, but the Archive itself hasn’t be touched too much because of the complexity in dealing with the legacy systems.

Usually what triggers the modernisation initiative is the need to digitise the analogue media, however there is also the will to obtain substantial improvements with new features, such as multimedia cataloguing, and to fix malfunctions due to old inconsistencies and defective patches.

Of course a wide range of different sub cases fall under this profile. Anyhow the available information permits to estimate the benefits of the increase of exploitation, constituting thus a source of funding for the process.

Being the legacy system old enough, these Archives are prepared to change them almost completely.

3.4.3 Enlargement case

Eventually we have the case of those Archives that achieved a good level of modernisation within the last ten years. Their Archive Management Processes might not rely on state of the art tools, however their systems permit very good performances and their exploitation figures give the evidence of the achieved improvements.

These Archives are in a good situation, but they understand they can get even better. The Archive users, after having taken advantage from the previous modernisation, are full of new expectations. New kinds of users are envisaged.

Likely, the digitisation is not complete, or even late, what an opportunity to also consider new process models and profit of the features provided by the PS Factory services.

These Archives are prepared to change something in their organisation and systems, where they identified bottlenecks, but they particularly require a greater level of integration with their existing resources.

4 Transaction orchestration and delivery

4.1 Criteria of work

As shown by Figure 7, where we can notice that a single factory may have several archives as customers at the same time.

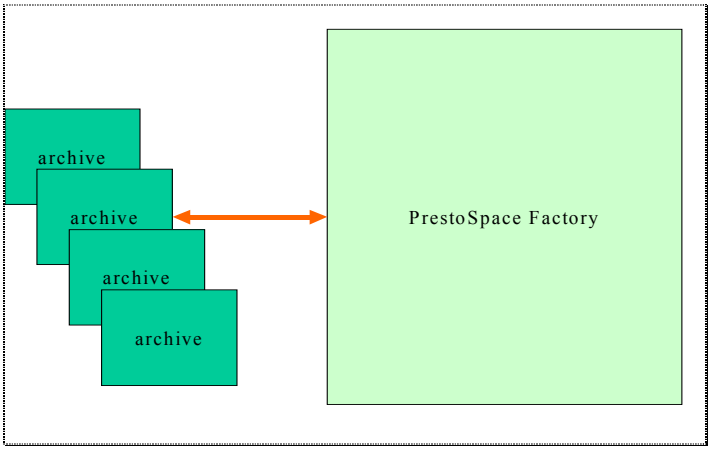


Figure 7 - The Presto Space Factory and the archives

The analysis carried out in chapter 3 explain how wide can be the assortment of combinations of Archives, Archives needs and requirements, and offered services and options.

This complexity has suggested the adoption of a gradual approach, where the transaction orchestration and the delivery are defined firstly by setting the frame of the relationship between Archive and Factory, with a sketch negotiation and the agreement on basic matching between Archive needs and offered services, and then by selecting, among the possible options, the most suitable set of formats and details.

The first stage aims to the definition and agreement on terms of business. It is possible to assume that the PS Factory delegates to a "*Negotiator*" the task of inquiring, together with the Archive, about both the current and target Archive situations and of defining a rough proposal of deployment of Factory services.

The Archive must be made aware that the results of the Factory processing will partially depend on the availability and reliability of information provided in input.

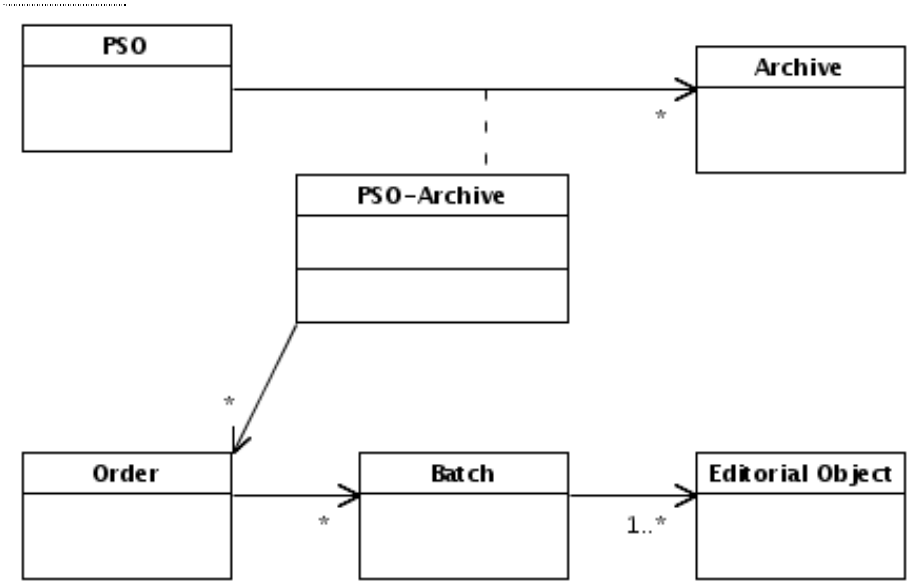
Besides, as the result will heavily depend on the selection of tools, and of the respective working options, the most appropriate scenario need to be identified. Various workflows may be defined for a single Archive depending on the typology of content or material, and also flexible workflows, where one task can be determined on the basis of results from a previous one, are conceivable.

The basic modelling assumption is that the PS Factory has got a specific component, *named PrestoSpace Orchestrator (PSO)*, for interfacing with its customers, the Archives, as shown in Figure 5.

The PSO is the actor who only can be aware of the whole PrestoSpace workflow, and able to collect and provide the suitable inputs (Edobs, Material, and Job-directions) to the various Units.

Besides, the PSO is also appointed for the knowledge about the archive's desiderata and the capability to translate it into the most suitable sequence of activity directives.

The diagram given in Figure 8 shows which are the identified entities, concerned in the relationships between PSO and Archive, which will be mentioned in the following sections.


PSO
Archive
*
PSO-Archive
*
Order
Batch
Editorial Object
*
1..*

4.2 Definition and agreement on terms of business

4.2.1 Interest in services

The parties should preliminary try to verify which are the typologies of service that the Archive is interested among Preservation, Restoration, Documentation, and also services for assisting in the Archive Exploitation.

Concerning the Archive Exploitation, it is important to assess the interest of the Archive in the information retrieval by use of a knowledge base resulting from the documentation process.

4.2.2 Location, termination, and TurnkeySystem option

The issues of location and termination, already discussed in 3.1 and 3.3, are clearly related and there ..4422(e)1(u)1.4422(s

4.2.3 Archive-PS Factory relationship account

The PS Factory has to account its relationships with the Archives, and the component responsible for this task is the PSO.

The following information set is identified and defined:

- *ArchiveDetails* **-** The information set enabling the identification of the Archive for any relevant process into the PS Factory, and permitting to establish any required communication. It includes:
  - *ArchiveName* – which shall identify the archive within the PS Factory - required
  - P_META S14(ORGANISATION_DETAILS) or P_META S12(PERSON_DETAILS) - optional
- *OpeningInfo*
  - *OpeningDate* – Date of creation of the archive information record. It states the beginning of the relationship - required
  - *ValidityDate* – Date after which the Archive relationship is scheduled to expire - optional
- *Status* - This provides the information about the current status of the relationship
  - *Date* – Date when the status has been updated. Possible values for the Status itself are defined in Table 2.

Value of Status	Definition	Caused by
Open	Initial state, it is possible to open Orders	Creation of record
Active	At least one Order is “Active”	Depends on Status of related orders
Finalising	All Orders are either “Closed” or “Finalising”	Depends on Status of related orders
Closed	Final state, no new Orders expected	Either by reaching the *ValidityDate* or by administrative action

Table 2- Value set for Status of the relationship PS-Factory vs Archive

- *ClosingInfo* **-** Information about closing the relationship:
  - *ClosingDate* – Date when relationship ends
  - *ClosingReason* – Reason for closing, the value set is given in Table 3.

Value of ClosingReason	Definition
Completed	All Orders have been completed
Expired	The ValidityDate has been reached
Cancelled	Relationship closed with an administrative intervention

Table 3 - Value set for ClosingReason in the Archive relationship

Table 4 – ArchiveInfo Set for PSO

The status information has to be kept up-to-date by the PS-Factory (PSO).

The Archive is enabled to request the update of the ArchiveInfo set only for modifying the *ArchiveDetails* (not the *ArchiveName)* and the *ValidityDate*.

4.2.4 Quantities and typologies.

The model of the PS-Factory-Archive relationship defined in 4.1, and showed in Figure 8, contemplates that one Archive may have a multiplicity of Orders towards the PS-Factory. Each Order is therefore related to a set of working units sharing their processing requirements, which presumably originate from the typology of both Editorial Objects and Material Sources.

The planning of time and resources allocated by the Factory to an Archive Order is based on the information about the approximate quantities provided at the instantiation of the Order itself.

From those quantities it should be possible to foresee the Order termination date, and details about the organisation of Batches, as discussed in 4.2.5.

Among the Order information set, which the PSO has to account in order to achieve management, we find defined the following:

- **ApproximateQuantities**

  The whole set is optional
  - **Batches amount** – approximate amount of Batches which will be related to this Order
  - **Editorial Objects –** may occur from 0 to unlimited
    - **amount –** approximate total amount of Editorial Objects for this typology
    - **hours –** approximate total duration of Editorial Objects for this typology
    - **typology –** classification of Editorial Objects according to negotiated categories
  - **Storages –** may occur from 0 to unlimited
    - **amount –** approximate total amount of Storages for this record
    - **hours –** approximate total duration of recorded Materials for this record
    - **storageType –** type of Storage (e.g. the Betacam Metal Small cassettes)
    - **condition –** known approximate condition of Storages (e.g. good, bad, very bad)

Table 5 – OrderInfoSet for PSO, section: ApproximateQuantities

The Order information doesn't specify the precise amount of Batches, Editorial Objects, physical storages etc. because it is deemed that those quantities may change during the life of the Order.

So the information defined here should provide quantities in terms of "expected amount" at the time of Order instantiation.

The information may be used in order to perform planning and allocation of resources.

The quantity information is not strictly required, even if useful, almost because it is difficult to guess which information the Archive is easily able to provide depending on its Inventory process.

In the Editorial domain, interesting quantities are the number of items and the total amount of hours, which together with the typology, determine the efforts required especially to the Documentation Units.

The value set for the typology might be negotiated with the Archives. For TV Archives, as given by [D15.1], the reference values are: *news (daily/magazine), sport, fiction, and entertainment*.

In the Material domain, the total amount of storages is important, together with the average storage duration, however this information should be associated to the storage type and to an indication about average storage conditions.

4.2.5 Timing and Batches

The issue of timing and batches has to be defined and the PSO has to manage it at the Order level.

Firstly we have very basic information set for managing the opening/closing and status of the Order.

Then there is the need to specify delivery timing from the Archive to the Factory and the criteria for delivering back from the Factory to the Archive.

Definitions are given in the following frames.

- **OrderName** – Identifies the Order within the PS-Factory – required
- **ArchiveName** - link to the Archive "owning" the Order instance – required
- **OpeningInfo**
  - **OpeningDate** – Date of creation of the Order instance - required
  - **ValidityDate** – Date after which the Order is scheduled to expire - optional
- **Status -** This provides the information about the current status of the Order
  - **Date** – Date when the status has been updated. Possible values for the Status itself are defined in Table 6

Value of Status	Definition	Caused by
Open	Initial state, no batch in progress	Creation of record
Active	Batches in progress, other batches expected	Depends on Status of related Batches
Finalising	Last batch received and work in progress	Depends on Status of related Batches and triggering of last batch
Closed	Final state, no work in progress, no batches expected	Either by reaching the *ValidityDate* or by Completion of all Batches or by an administrative action

Table 6 - Value set for Status of the Archive Order

- **ClosingInfo -** Information about closing the Order:
  - **ClosingDate** – Date when Order gets Closed
  - **ClosingReason** – Reason for closing, the value set is given in Table 7

Value of ClosingReason	Definition
Completed	All Batches have been completed
Expired	The ValidityDate has been reached
Cancelled	Order closed with an administrative intervention

Table 7 - Value set for ClosingReason in the Archive Order

Table 8 - OrderInfoSet for PSO, basic management sections

Also for the Order the *Status* information has to be kept up-to-date by the PSO.

The overall *OrderInfoSet*, cannot be modified by the Archive subsequently its acceptance by the PS-Factory, excepted for the *ValidityDate* and for the *NotificationDetails* described in Table 27. Any changes about the required process have to be requested with the definition of a new Order.

- **BatchDeliveryCycle –** required

  Information about the periodicity of Batches. When a new Batch is typically created by the Archive for the PS-Factory.

  The suggested value set is provided by Table 9

Value	Optional parameter	Description
Daily	time	One batch per day, typically created at time of day
Weekly	day	One batch per week, typically created at day of week
Monthly	day	One batch per month, typically created at day of month
OnArchiveReady	-	At any time the Archive is ready
OnFactoryIdle	-	When previous batch is completed

Table 9 - Suggested value set of BatchDeliveryCycle

- **LastBatchDeliveryDate –** required

  The date after which no other Batch will be inserted into the factory by the Archive for this Order.

- **ReturnDeliveryPolicy** - Required

  Information about the policy of deliverying the output of the PS-Factory back to the Archive

  A suggested value set for this field is given in Table 10

Value	Optional parameter	Description
AsSoonAsAvailable		The PS-Factory may deliver (or make available) to the Archive every produced result as soon as delivered by the relevant service (e.g. the new digitised master files delivered at the end of preservation process)
OnProcessComplete		The PS-Factory may deliver (or make available) to the Archive the results when the whole requested process is completed
OnBatchComplete		The PS-Factory may deliver (or make available) to the Archive the result for a whole batch when it is completed
OnRequest		Archive pulls after inquire
AtDate	date	

Table 10 - Suggested value set for ReturnDeliveryPolicy

Table 11 - OrderInfoSet for PSO, Bathes and Delivery sections

It is possible to remark that the concept of Batch is especially helpful for the Archives, because it makes simpler to manage and tracking a process that may involve many items for a quite long time (e.g. months, even years for very large Archives).

From the PS Factory perspective Batches are less important and they are primarily related to the reception from the Archives.

The Delivery Policy options defined in Table 10 permit to select either a criterion related to Batch or other criteria related to the completion of Editorial Objects processing.

- **BatchName** - Identifies the Batch within the PS-Factory – required
- **OrderName** - link to the Order "owning" the Batch instance – required
- **ArchiveName** – link to the Archive "owning" the Order instance – required
- **OpeningInfo**
  - **AcceptanceDate** – Date of Batch submission - required
  - **ExpectedCompletionDate** – Date by which the Batch is expected to be completed - optional
- **Status** - This provides the information about the current status for each Editorial Object of the Batch
  - **Date** – Date when the status has been updated.
  - **EditorialObjectsInfo** – list Editorial Objects composing the Batch (unbounded) – required
    - **EditorialObjetID –** identifies the Editorial Object instance – required
    - **Date** – Date of last modification of the EOD - required
    - **Status** - Status of the Editorial Object instance, the value set is given in Table 12 - required
    - **Note** – Textual description of the reason for the status – opional

Value of Status of Editorial Object instance	Definition
Acceptance pending (initial)	Instance acceptance is pending. This is the initial state at the registration of the Batch
Accepted	Instance compliant to acceptance criteria. Work not started yet.
Refused (final)	Instance rejected at the acceptance.
Working	Work in progress for the instance.
Stalled	Work started, but not in progress. Some process has called a fallout which is waiting for being managed.
Unsolvable failure (final)	Work started, but some important process couldn't run and the problem couldn't be solved.
Partial failure (final)	Work was not completed because of a failure of a secondary requested service. Most services run successfully
Completed (final)	All requested services run successfully

Table 12 - Value set for Status of Editorial Object instances

- **ClosingInfo -** Information about closing the Batch:
  - **ClosingDate** – Date when Batch gets Closed
  - **ClosingReason** – Reason for closing, the value set is given in Table 13

Value of ClosingReason	Definition
Completed	Batch completed (all Editorial Objects have reached a final state)
Cancelled	Batch closed with an administrative intervention

Table 13 - Value set for ClosingReason of Batch

Table 14 – BatchInfoSet for PSO

We can observe from the examination of Table 14, that the BatchInfoSet allows to account the Batch history from registration, to progress of works of each Editorial Object, till the achievement of a final state.

It is useful to remind the reference model depicted in Figure 8.

4.3 Input options & acceptance requirements

The model depicted in Figure 8 identifies the Editorial Object as the working unit entity for the PS-Factory.

For management and logistic issues Editorial Objects are grouped into Batches that make reference to a defined Order, as what concerns the definition of the required process.

If the instantiation of each entity is correct, the PSO is able to associate each Editorial Object instance to the pertaining Batch-Order-Archive instances.

An Editorial Object instance is validated against the defined acceptance requirements at the moment of registration within the Factory.

The acceptance is based on the following elements:

- basic identification for management
- realisations by Material instances
- metadata information

The basic identification is provided by a dumb string/number assigned to the Editorial Object. Formally that identifier is issued by the Archive, which should be able to recognise it (unique with the Archive domain, the PS-Factory may always add the Archive ID as prefix in order to guarantee uniqueness within the Factory), however the registration service may generate a suitable value in the default case and return it to the Archive.

The information about the "Realisations", showed in Figure 6, is mandatory. The required information relies on Archive Inventory [Figure 3] and thus there may be some difficulties for the Archives to provide it in a satisfactory way, especially, but not exclusively, for Realisations based on analogue media, as explained in Table 15.

When a single Tape gives the Editorial Object realisation, that's the common case, the mapping is completed by giving the start point of the Editorial Object within the Tape Material (a 30 minutes programme can be recorded on a 60 minutes tape). It is not unusual, however, that for recording a single Editorial Object two (or more) Tapes have been used. In this case for each contribution to the Editorial Object realisation a complex mapping record (duration, tape, start on tape, start on editorial object, tracks) is required. Moreover gaps and/or overlaps (more often) are possible. Sometimes two (or more) Editorial Objects are recorded on the same Tape. Also in this case the mapping records are required.

Table 15 – Mapping of Editorial Objects vs. Material

The main issue for the Realisations input is to define the behaviour of the Factory in those cases in which the Realisations information is either incomplete or not accurate.

Concerning the metadata information, it is good to make reference to Table 22, where legacy information is generally "welcome" and recommended as what concerns "identification". Further acceptance requirements might depend on the working options discussed below, which are defined at the Order level.

4.4 Working options & transcoding functionalities

At the Order level there must be the indication of the required services and their main relevant options. The general information set for describing the requested process is given in Table 16.

The services are here grouped per "WorkingArea", namely those pertaining to the PS Factory Units:

- **Preservation**
- **Restoration**
- **Documentation**

However it is possible that similar services (border area) might be provided in more than one Unit.

The common criteria are that the Archive has to decide, when defining the Order, which Services should be activated for that Order, and for each Service it must be possible to indicate some working options.

Generally working options indicated in the Order must be kept as simple and intuitive as possible, while the Service Administrators may only configure technical settings and options.

As an example, a "shot detection" service may have a ***HigherRecall*** option and a ***HigherPrecision*** option (not to be used together), which are easy to be understood, while the actual algorithm thresholds are hidden within the processor configuration.

Service dependencies must be solved within the Factory, which is also responsible for the order by which services must/can perform and for deciding if some services can work concurrently. The specific component in charge for this task is the PSO.

The Service definition must indicate dependencies from other Services and the PSO may also activate unrequested Services in order to fulfil a dependency.

Dependencies may also depend on Service options.

- *ServiceName* – which identifies the Service within the Factory – required
- *WorkingArea* – indicating within which working area the required service should run. This is applicable only for services that are not strictly related to a single working area – optional
- *Options* – a list (Option element may occurs from 0 to unbounded) of boolean flags enabling the options offered by the service and thus overiding the default behaviour
    - *OptionName* – which identifies the Option for the given service
- *Parameters* – a list (Parameter element may occurs from 0 to unbounded) of input parameters providing values which overide the default ones
    - *ParameterName* – which identifies the Parameter for the given service
    - *ParameterValue* – the value for the given Parameter

Table 16 - OrderInfoSet for PSO, section: "ProcessingInfo"

4.4.1 Preservation

The preservation area should provide the digitisation service for the options of which a few suggestions are given in Table 17, although they may be changed in the future after revision of definitions.

- *Digitisation* – Produces digital master from original analog media
  - *HigherQuality* – requires an quality higher than default, accepting a greater resulting size
  - *LowerSize* – requires a target bit-rate lower than default, accepting a lower quality
  - *EmbeddedAudio* – asks for audio data to be embedded with video data into a single wrapper
  - *NotEmbeddedAudio* – asks for audio data to not be embedded with video data into a single wrapper
  - *Inspection* – asks for inspection of original media before digitisation
  - *Preparation* – asks for preparation of original media before digitisation
  - *Crop* – ask for the new master to not include parts exceeding the Editorial Object timeline

Table 17 - Digitisation service

4.4.2 Restoration

The restoration area should provide the homonymous service, for the options of which a few suggestions are given in Table 18, although they may be changed in the future after revision of definitions.

- *Restoration* – provides restoration of digital essence with respect to specified defects. Requires Digitisation.
  - *OnlyAudio* – ask to not perform video restoration
  - *OnlyVideo* – ask to not perform audio restoration

Table 18 - Restoration service

4.4.3 Documentation

The documentation area should provide a wide set of services in order to fulfil requirements of Archive documentation models [D15.1] and outcomes of [BPM]. The list of services given in Table 19 and the respective suggested options are subject to change according to future revision of definition of the documentation services.

- *Speeches* – provide text from audio
  - *ForceLanguage <languageCode>* - ask to override target speech language to the given value
- *EditorialParts* - Find Editorial Parts within Editorial Object
  - *NewsItem* – ask to find specifically news items as for within Newscast
  - *UseLegacy* – ask to try matching with legacy information
  - *HigherRecall* –
  - *HigherPrecision* –
- *CompleteIdentification* - Add identification information which was missing or lacking from legacy
  - *Titles*
  - *Publications*
  - *Credits*
  - *Awards*
  - *Collections*
  - *OriginalClassification*
- *Categories* - Provides subject classification. Requires Speeches and EditorialParts (default)
  - *NoEditorialParts* – enable classification on the Editorial Object as a whole
  - *CategorySet <categorySet>* - ask to use the given defined category value set
- *Topics* – Provides the subject of content. Requires EditorialParts (default)
  - *NoEditorialParts* – enable subject to be found independently from Editorial Parts basis.
  - *HigherRecall*
  - *HigherPrecision*
- *Descriptions* - provide synopsis or itemised description of Video and/or Audio content
  - *OnlyVideo*
  - *Synopsis*
  - *Itemised*
  - *OnEditorialParts*
- *Enrichment* - Provides reference to related published documents
  - *ExternalResource <ResourceURL>* - ask to use the given publication source – can be repeated

Table 19 - Documentation services

4.4.4 Transcoding

Transcoding can be requested by the Archive either as a Factory Service, in which case the PSO is formally responsible for honouring the request, or as a Service offered by a specific Factory Unit.

In 4.5.4 three possible material quality levels are defined:

- Master quality material
- Broadcast quality material
- Browsing quality material

As the Broadcast quality is typically an input to the Documentation Unit, it is reasonable for that Unit to also provide a transcoding from Broadcast to Browsing.

Both the Preservation and Restoration Units should be able, if involved in the process, to provide transcoding from Master Quality to any lower level.

Eventually a specific Transcoding Unit could be added to the Factory architecture.

In addition to the Archive requests, transcoding processes can be needed as a consequence of Service dependencies. For such eventuality the behaviour of the Factory must be defined, by selecting between:

- The produced material will be considered temporary and thus deleted after use within the Factory
- The produced material will be delivered to the Archive as if it was explicitly requested.

Table 20 provides the suggested options and parameters for the transcoding service. Subject to changes in future service definition.

- *Transcoding* – Derives a new Material Realisation by transcoding a Material Source
  - *BroadastQuality* – Requires a transcoding to Broadcast quality material
  - *BrowsingQuality* – Requires a transcoding to Browsing quality material
  - *HigherQuality / LowerSize* –
  - *TargetFormat <format>*

Table 20 - Transcoding service

4.4.5 Content Analysis

The Content analysis services listed in Table 21, which provide several information on audiovisual features and content indexing, may be offered by various Factory Units. It is a task of PSO to dispatch the request to most appropriate processor.

- *VideoAnalysis* –Requires Digitisation
  - *ShotBoundaries* –
  - *ShotSimilarities* –
  - *KeyFrames* – extracted from shots
  - *StripeImages* –
  - *CameraMotion* –
  - *MotionActivity* –
  - *VisualFeatures* – from keyframes
  - *SmallerPictures* – applies to keyframes, stripe images
  - *HigherRecall / HigherPrecision* -

- *DQAnalisys* - Provides defects and quality analysis of digital essence. – Require Master Quality

- *AudioAnalysis* –
  - *HigherRecall / HigherPrecision* -

Table 21 - Other services

4.5 Delivery options

4.5.1 What delivery is about

It is useful to remind, as an outcome of the criteria given in 4.1, that the fulcrum of delivery is the Editorial Object. Clearly the Editorial Object is related to Materials, however the delivery of Material Sources has to be considered in strict relation with that of Editorial Object and cannot be independent. Further details on this concept are given in 4.5.5.

The EditorialObjectDocument provides the links needed for the delivery of any kind of related entity, such as Material Sources and ancillary information.

It must be reminded that semantic analysis tools used within the Documentation Unit may also collect cumulative knowledge, which can be requested for delivery, which is not depending on single Editorial Object instances [see pag. 16]. The currently supported delivery mechanism for such entity is the export of the knowledge base, on request, as RDF file. Other mechanisms are possible, but they can be provided, if possible, on specific request, as they imply some particular configuration constraints

4.5.2 Wrapping formats

The wrapping formats define the packaging of exchanged entities, either essence or metadata or both, and are interesting because of the practicability for use of possible available tools, providing a certain degree of independence from the nature of the contained stuff.

Generally speaking the wrapper is a type of file which is able to host several types of related content. This is specifically used for exchanging all together a multiplicity of components, simplifying the transaction, and sometimes also for providing concurrent access to the various related part.

The examination given here is also depending from the discussion already provided in 3.3.3.

Concerning the Essence, the wrapping formats provide basically the same functionalities of a multiplexed format (a container for both video and audio data, which is used for audiovisual presentation), such as an MPEG Program Stream, with in addition the capability to not require a single specific coding/compression format for the essence.

It should be remarked that wrapping is not generally required. Essence wrappers can be necessary in order to allow synchronous presentation (play) of Material with separate Sources (video from one file and audio from other files, however this is an issue for the player application and a multimedia playback platform not requiring this is conceivable (e.g. Direct Show).

MXF

The reference wrapping format for high quality essence within the PrestoSpace project is certainly the ***MXF***.

That format is based on set of SMPTE standards which define the wrapper itself and the modalities for encapsulating each kind of content according to its specific nature and format.

For example it allows video data to be uncompressed, MPEG coded, DV coded, etc.

The MXF file is also able to contain metadata. However, in addition to the issues already discussed in 3.3.3, it currently puts several constraints on the metadata formats and coding, which will be hopely solved in the future.

It is also helpful to consider the question of MXF *Operational Patterns*, which define a set of suitable limitations to the ways in which the essence data (sources) produce the output audiovisual Material. The mechanism of Operational Patterns permits the definition of several layers of complexity, allowing simpler implementations to be deployed. The drawback is that the selection of an Operational Pattern, or at least to indicate the most complex desired one, is also required.

TAR and similar bundles

A mechanism deserving consideration and often successfully adopted to gather together a set of strictly related files, typically when exchanging metadata but also sets of essence data files - such as key frames -, is the TAR bundle, also named tarball.

Tar (correctly tar, with low case t) is a type of archive, as created by the homonymous command, without compression option applied. A list of files, including directory structures and file system information, are gathered together in one file. Optionally a lossless data compression algorithm is applied thence (e.g. gzip, compress, bzip2).

Further details are available at [WKPD search for tar (file format)] including limitations.

Clearly the *tar* archive doesn't define the nature of its content that needs to be agreed together with the transaction details.

The PrestoSpace MAD approach includes the *tar bundle* as a mean for exchange Materials. This is particularly directed to keyframes images, because it permits to avoid multiple transactions for small data entities. However the same approach can be extended to larger media files.

Moreover nothing prevents to use one wrapper within another one, so MXF files may be gathered together with other files within a tarball.

4.5.3 Metadata standards and options

As outlined in 3.3, the PS Factory has to exchange metadata with the Archives, and it is a bi-directional deal. On one hand the Factory expects minimum identification information, and it generally welcomes any legacy information. On the other hand the Archives are going to receive both technical information, including reporting from preservation and restoration processes, and the results from the documentation process (enhanced metadata).

PrestoSpace Deliverable D15.1 provides a description of some current audiovisual documentation models and some relevant metadata standards.

Now it is difficult to find audiovisual Archives whose legacy information is currently compliant to a fresh metadata standard and requiring the Factory documentation services. In most cases the organisations have their own documentation model, some aiming to keep it or slightly enhance it, other ones aiming to change it more radically.

Some years ago the lack of standards about metadata within the audiovisual scope raised many independent initiatives and efforts. Today the selection of a single standard among the differentiated offers is controversial.

Nowadays the ROI from the complexity of metadata standard development is decreasing and it is believed that the best benefits come from clear definition of meaning and formats and from clever setting of mapping.

In particular it is deemed that some mapping is required at the Archive side:

- When exporting their legacy information

  This is required at least for a minimum identification set. Unmapped information can be transferred anyway as opaque ancillary material information, but without any warranty of correct decoding/understanding.

- When importing the enhanced metadata from the Factory.

  This is required for those Archives intending to keep most of their pre-existing data structures

The Editorial Object Document (EOD) defined within the PrestoSpace MAD Area is able to cover the needed requirements.

The EOD format partially defines how to code some information elements, specifically concerning the relationships between Editorial Object and Materials, and makes reference to EBU Tech.3295 (P_META) and to MPEG7 for specific elements, as described in Figure 9

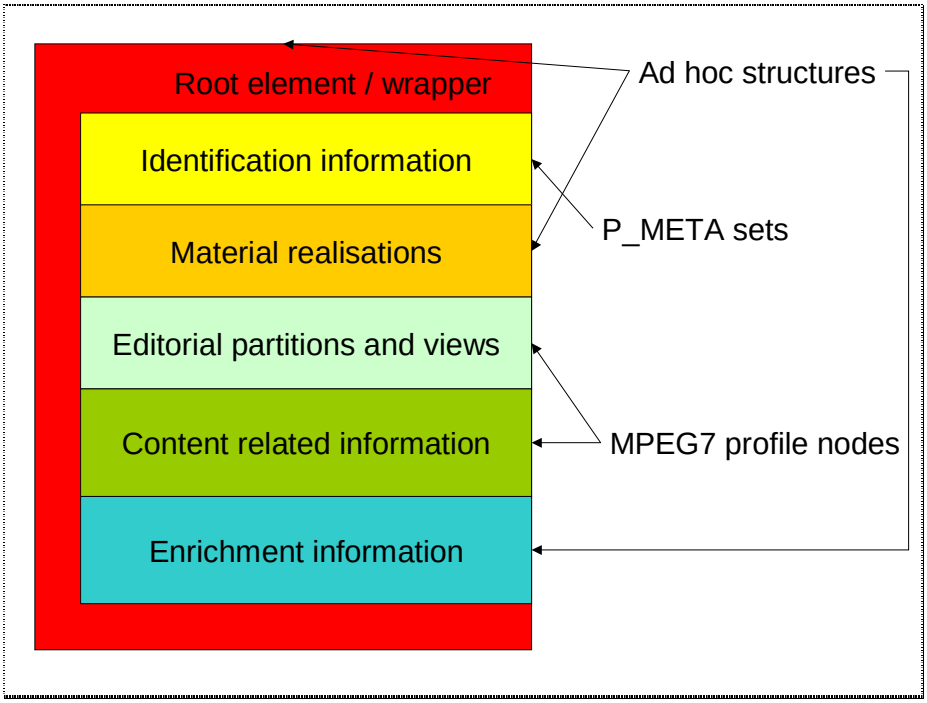


Figure 9 - Structure of Editorial Object Document [to be updated]

It happens that the P_META standard is particularly relevant, although not exclusively, for coding the legacy information, while the MPEG7 standard is used for information more strictly related to temporal decompositions.

Concerning the information coming from the Archives the statements given in Table 22 apply.

The **"Realisations" are mandatory**. The **"Identification information"** (such as titles, identifiers, publication details, and credits), which are legacy, are **"Recommended"**, that is: • Every element is optional by itself • It is recommended to the archive to provide at least some of the possible information • If no other information is provided, the only identification datum for the Editorial Object will the attribute "id", which is an arbitrary dumb string/number. The other legacy information is optional and welcome.

Table 22 – Metadata from the Archives

Delivery options

The EOD is the default delivery format option. Any other alternative option requires a further mapping effort with concrete possibility of loss of meaning and detail, without providing at the other hand so much advantage.

However the existence of performing tools for XSL transformation allows to conceive the implementation of "*export as*" facilities, as long as we consider XML formats. These options may be suitable to those Archives feeling more comfortable to deal with a particular metadata standard.

- *Dublin Core* – It covers sufficiently well a set of information for identified entities. It suffers for loss of detail and it gets in troubles concerning relationships based on timelines and temporal decomposition.
- *SMPTE/DMS1* – The original descriptive schema conceived for MXF embedding. It may be required in case of completely MXF based transactions, because although any descriptive schema could be used in theory, in practice the standardisation procedures and the issue of KLV coding have discouraged other proposals. There are proposals for embedding any kind of XML coded metadata within MXF, which would be rather opaque in the MXF environment, but this effort is dissuaded by the unnecessity of metadata embedding in most cases. Issue: KLV coding.
- *P_META* – Although a complete mapping to P_META is possible, the exact details of such mapping need to be defined. A smoother result could be obtained by a change request for a specific new set.
- *Mpeg7* - For some information elements there would be some loss of meaning and detail, in a very few cases a request of extension to the standard are needed. However the result would be more than acceptable in most situations.
- *other* – other future options may be added afterwards, if considered interesting

4.5.4 Essence formats

The issue of essence formats deals with the compression, coding, and file formats for material sources.

The most important features of an Essence format are:

- Quality – we won't try to define this here.
- Number of bits – This depends on sampling and on compression/coding. The sampling parameters (possible/suitable) depend on the format of the analogue source (please notice that analogue video is already sampled vertically and temporally). The compression/coding can be generally classified as:
    - Uncompressed – Samples are simply coded without trying further to decrease the number of bits
    - Lossless compressed – Data are compressed in order to reduce the number of bit, taking advantage of *redundancy*, but the process is entirely reversible.
    - Lossy compressed – Data are much more compressed taking into account also *subjective redundancy*, this process is not reversible because some information is lost for ever, and thus it may affect both subjective and objective quality.
- Access modalities – This depends on compression/coding. Complex compression schemes may give further constraints to the timeline access points, impacting on arbitrary access for both seeking and editing.
- Format support – This a format management issue. Is the format open? How versions will be managed in the future? From this it depends the actual possibility to access material in the long term.

Master quality material

By "master quality" we mean the material instances with the highest quality, used as reference, and from which any equivalent or lower quality may be derived.

Generally a derivation process doesn't generate a new master, because the quality cannot typically be increased, however a few exceptions exist:

- A lower bit-rate copy obtained by lossless compression can generate another master quality.
- A restoration process, if successful, may generate even a better quality. However one could argue that doing the restoration again, in a different way, could carry to better results. So it is controversial to say that at the output of a restoration process we have a "true master".

The master quality is the post-production application scenario and is the object of Archive long-term storage.

Logically, in the digital file domain, any quality lower that the master could be re-created from the master.

Concerning the Audio, the number of bits generated, once defined the sampling parameters, is so lower than for video Essence that the Audio Master is usually an Uncompressed coding format (PCM).

Conversely, the number of bits for the video is still quite bulky as Uncompressed. Therefore a Video Master is usually either Compressed Lossless or "High-end Compressed", where by "High-end" it is meant a form of lossy compression the parameters of which provide assurance of high quality, on the base of official tests made on critical material.

The identified options for video coding are given in Table 23, where the reported values refer to standard definition even if the given options also support higher (some even any) resolution.

Options supported by PrestoSpace for Master Quality Materials:

- *MPEG-2* - ISO/IEC 13818 – provides high end compression at 15-50 Mbps for standard definition
- *JPEG2000* - ISO/IEC 15444-1 / ISO 15444-3 – provides both lossless and high end compression, at 80-100 Mbps for standard definition (lossless) [3]

Options not currently supported by PrestoSpace (could be provided, if possible on specific request):

- *DV* - IEC 61834 – provides high end compression Intra coding at 50 Mbps for standard definition
  - also 25Mbps
- *MPEG-4* - ISO/IEC 14496-10 / H264 – provides high end compression (better than MPEG2 but still less supported) – **Considered interesting for High Definition Materials**

Table 23 - Video essence compression/coding formats for "master quality materials"

Broadcast quality material

We mean by "broadcast quality" those material instances the quality of which is appropriate for publication through current audiovisual publication media (terrestrial and satellite television, DVDs). Notice that the same instances might not provide that expected quality in the future, when published though more demanding media, such as over higher definition.

The customer Archive may ask to the PS Factory to deliver, in addition to a master quality material, also a broadcast quality material. Actually although it should always be possible to obtain the latter from the former subsequently, it is acknowledged that a so profitable opportunity to produce such a copy is quite attractive, as later it could be more troublesome.

Clearly the selection of an option here is quite less critical than for the master quality, even though the cost of processing would suggest making the wisest possible choice.

In this case typically also audio is compressed and the options given in Table 24 refer to coding schemes that include both video and audio coding. Combination of schemes belonging to different standard families is discouraged.

Option supported by PrestoSpace for Broadcast Quality Materials:

- *MPEG-2* - ISO/IEC 13818 – provides appropriate quality at 4-8 Mbps for standard definition
  - *Audio may also be MPEG-1* ISO/IEC 11172

Options not currently supported by PrestoSpace (could be provided, if possible on specific request):

- *MPEG-4* - ISO/IEC 14496-10 / H264 –better than MPEG2 but still less supported
- *Windows Media 9* – Former proprietary Microsoft codec, currently submitted to SMPTE for standardisation
- *Dirac* – Open Source (GPL) wavelet codec developed by BBC/RD, published on SourceForge
- *Theora* – Open Source (GPL) codec developed by Xiph.Org Foundation (part of OGG project)

Table 24 – Audio/video essence compression/coding formats for "broadcast quality materials"

It should be remarked that a broadcast quality material could be produced anyway internally to the PS Factory, as it is appropriate for Content Analysis within the Documentation Unit.

3

ISO/IEC 15444-1:2004 | ITU-T Rec. T.800 defines a set of lossless (bit-preserving) and lossy compression methods for [..] digital still images.

ISO/IEC 15444-2:2004 defines extensions to part 1

ISO/IEC 15444-3:2002 specifies the use of the wavelet-based JPEG2000 codec for the coding and display of timed sequences of images (motion sequences), possibly combined with audio, and composed into an overall presentation. In this specification, a file format is defined, and guidelines for the use of the JPEG2000 codec for motion sequences are supplied.

Browsing quality material

By “browsing quality” we mean those material instances the quality of which is sufficient to appreciate and recognise the content, albeit some impairments might be quite perceptible. Those materials are often used as “proxy” for their higher quality corresponding ones when there are lesser resources, in terms of bandwidth and storages, or when there is no intention to make available a better quality copy, as for free sample preview. In the PrestoSpace domain this level is interesting for the Publication Platform

As for the Broadcast Quality, also the Browsing Quality Material could be obtained in any time from higher quality copies, but the Archive might find suitable to ask to the PS Factory the production and delivery of this kind of Material, especially for saving on the cost of processing in another time.

Also in this case the options given in Table 25 refer to coding schemes that include both video and audio coding. There is not a specific support by PrestoSpace, as the format choice doesn’t imply particular constraints within the Factory.

Note that typically the listed picture coding schemes allow image downsampling (horizontally, vertically, and even temporally), which is typically required for browsing quality material.

Here the target bit-rate currently falls into the range 0.1-2 Mbps for all schemes, and therefore there is a quite broad range of possible quality results.

- *MPEG-2* - ISO/IEC 13818 – not state of the art scheme for this target application
- *MPEG*-1 - ISO/IEC 11172 -
- *MPEG-4* - ISO/IEC 14496-10 / H264 –
- *Windows Media 9* – Former proprietary Microsoft codec, currently submitted to SMPTE for standardisation
- *Real* - proprietary video codec developed by RealNetworks
- *Dirac* – Open Source (GPL) wavelet codec developed by BBC/RD, published on SourceForge
- *Theora* – Open Source (GPL) codec developed by Xiph.Org Foundation (part of OGG project)

Table 25 - Audio/video essence compression/coding formats for “browsing quality materials”

Again it must be remarked that such browsing quality material might be produced anyway internally to the PS Factory as appropriate for Annotation within the Documentation Unit.

4.5.5 Exchange details

This section develops in further detail the issue of transactions between PS-Factory, according to the System Architecture showed in Figure 5, with the focus on the delivery task.

The transaction activity needs to be supported by an orchestration between the parties implementing a defined protocol and allowing the management of the relationship by the PSO. The orchestration itself is realised by means of exchanges according to defined formats, while the specific transactions must be supported by the availability of suitable protocols.

The PS-Factory makes available, through the PSO interface, a set of Exchange Services, listed in Table 28, which permit the implementation of the model showed in Figure 8. A full exchange implementation is completed by a couple of mechanisms, for which it is necessary to identify the protocols that are appropriate and can be proposed to the Archives:

- *A notification mechanism* – which allows the communication from the Factory to the Archives about status modifications, process completion or failure, and other events
    - e-mail – the archive may indicate an e-mail address set up for receiving notifications
    - sms – a mobile telephone number may be given for small message service
    - rss – rich site summary
    - web service – the archive may provide the coordinates of a web service set up for receiving notification
    - other rpc mechanism – any other mechanism set up by the archive and based on remote procedure calls
    - any other – anything else suitable, e.g. telephone, fax, standard mail,
- *A Material Source exchange mechanism* – for actually delivering material sources. Considered protocols are:
    - file – this is useful only to provide access to resources once they are stored on the same host.
    - smb – Server message block. Protocol for sharing directories, allows both mounting a file system and getting files as for ftp
    - nfs – network file system. Protocol for sharing directories allowing mounting a file system
    - ssh – Secure shell. Protocol for secure remote host access, allows file transfers
    - http (or https)- hypertext transfer protocol (or http secure). Protocol used for web based exchanges, allows file transfer
    - ftp – File transfer protocol
    - carrier – used only for removable media that are physically exchanged by means of a carrier.
    - bundle – indicates that the requested resource is contained into a bundle (e.g. a tar), the access of which must be granted via another protocol.
    - storage - indicates that the requested resource is contained into a removable media, the access of which must be granted via another protocol.

- **NotificationDetails** – indicates information on requested notification services – optional
  - **Events** – may occur from 0 to unlimited
    - **EventType** – the type of event for which notification is required, the value set is given in Table 26 - required
    - **NotificationMechanism** – the type of mechanism for the notification (e.g. sms, web service) - required
    - **NotificationDestination** – where the notification must be sent (e.g. email address, phone number, web service url) – required

Value of EventType	Definition
OrderStatus	Any modification of status at the Order Level
BatchStatus	Any modification of status at the Batch Level Which includes any modification of Status of the Editorial Objects belonging to the Batch
BatchCompletion	When all Editorial Objects have reached a final state
BatchProblems	When an Editorial Object comes into a non completely successful final state
NewMaterialAvailable	When there are new material sources available at the PSO interface
UnspecifiedProblems	When a serious problem preventing completion and success occurs

Table 26 - Value set for EventType in subscribing Notification

- **DeliveryOptions** – optional
  - *ExportMetadataAs <format>* - asks for exporting metadata according to a format different from EOD. Implies partial loss – optional
  - *DoPublication <PublicationService>* - asks to perform direct delivery to the given publication system – optional

Table 27 - OrderInfoSet for PSO. Notification details and other delivery options

The Order information has to include also the notification details and other possible delivery options that are given in Table 27.

The Notification details are optional because they are given only if there is a subscription for notification services, which can be modified along the Order lifetime.

Actually it may be argued that notification is even redundant because the Archive can always inquire for the status of each entity. However it could happen to be useful for either avoiding the inquire polling mechanism or to be informed about the occurrence of troublesome events.

Currently the Delivery Options of Table 27 include:

- the possibility to ask for a different metadata format, which is anyway discouraged [4.5.3].
- the possibility to directly perform the publication of results on the given Publication Service, which is normally a task of the Archive. This option could be used either in the case of the "*Turnkey System*" [4.2.2] or whenever the Archive-PS Factory relationship is set up suitably to automatically feed the publication system. The details necessary to configure the latter case are not defined here.

It is useful to remark that the options about essence and wrappers discussed in 4.5.2 and 4.5.4 are actually pertinent to the processes which create the requested material sources - namely preservation, restoration, and transcoding – and thus have to be indicated within the request of the relevant services.

Eventually it has been considered the suitability of delivering the overall set of information and essence related to an Editorial Object wrapped up into a single bundle, and it has not been found particularly appropriate because:

- Original media would be necessarily kept out
- The size of the bundle is mainly determined by the master quality materials sources and becomes considerable.
- The relationship Editorial Object – Material is not forcedly one-to-one, as explained in Table 15, and the same Material Source could be requested to be delivered more than once.
- The Knowledge Base information, discussed at page 16, would also need to be considered apart.

Level / Entity	Service Name	Parameters	Description
Archive	Register	ArchiveInfoSet	A new Archive is registered at the PS Factory
Archive	Close	ArchiveName	Asks to the PS Factory to STOP all processes concerning the Archive
Archive	Update	ArchiveInfoSet	Modification of *ArchiveDetails* (except for *ArchiveName*) and/or *ValidityDate*
Archive	Delete	ArchiveName	Asks to the PS Factory to remove all records about Archive.
Archive	Inquire	ArchiveName	Returned: ArchiveInfoSet
Archive	GetKnowledge	ArchiveName, [Date,ListOfURIs]	Returned: RDF file (or file access information) containing the whole KB gathered for the given Archive. Optionally the provided RDF may contain only the update of the KB after a given date and/or only the information related to the entities identified by the given list of URIs
Order	Register	OrderInfoSet	A New Order is created for the given Archive
Order	Close	OrderName	Asks to the PS Factory to STOP all processes concerning the Order
Order	Update	OrderInfoSet	Modification of *ValidityDate* and/or *NotificationDetails*
Order	Inquire	OrderName	Returned: OrderInfoSet
Batch	Register	BatchInfoSet (it must include the list of EditorialObjectIDs)	A New Batch is created for the given Archive and Order.
Batch	Get	BatchName	Returned: List of EditorialObjectDocuments
Batch	Close	BatchName	Asks to the PS Factory to STOP all processes concerning the Batch
Batch	Inquire	BatchName	Returned: BatchInfoSet
Batch	Clear	BatchName	Enables the PS-Factory to delete all records about the Batch instance, including all Editorial Objects and associated Material Sources
EditorialObject	Register	BatchName, EditorialObjectDocument	Asks to register an Editorial Object instance. Success will be pending until check of acceptance requirements
EditorialObject	Get	BatchName, EditorialObjectID	Returned: EditorialObjectDocument
EditorialObject	Inquire	BatchName, EditorialObjectID	Returned: Status of process on the EditorialObject instance
EditorialObject	Clear	BatchName, EditorialObjectID	Enables the PS-Factory to delete all records about the EditorialObject instance, including associated Material Sources
Materials	Insert	ArchiveName, MaterialAccessDocument	Provides to the PS Factory the Access information for importing Materials
Materials	Get	ArchiveName, ListOfUMIDs| EditorialObjectID	Asks to the PS Factory to provide Access information for delivery of Materials. If an EditorialObjectID is given in input, it asks for all Material related to an Editorial Object. Returned: MaterialAccessDocument

Table 28 – Exchange services provided by the PSO to the Archive

The definition of the services implies a sequence of service calls that the Archive has typically to follow in order to interact properly with the PSO:

- RegisterArchive
- RegisterOrder
- RegisterBatch
- InsertMaterials – actually this depends only from RegisterArchive. It must be remarked that a single Material instance may contribute to the Realisations of multiple Editorial Objects.
- RegisterEditorialObject
- Inquire – on any entity, use depends also on selected ReturnDeliveryPolicy
- Get EditorialObject – any time the significant process status is reached according to the selected ReturnDeliveryPolicy
- GetMaterials – the Material Identifiers (UMIDS) required in input can be found within the Editorial Object Document. Other versions of the same service allow getting access to all Materials related to an Editorial Object instance identified in input (or to group of Material Sources sharing the same criteria of relationship with the Editorial Object).
- ClearEditorialObject
- GetArchiveKnowledge – any time it is felt useful to update the KB of the Archive exploitation facilities

Internally to the PS-Factory, the PSO component is responsible for dispatching tasks to the Units, with the modalities defined in [ref D3.1], and in accordance with the Order definition.

5 Discussion issues

5.1 Original media identification

In section 3.3.1, when discussing about “Common aspects”, the need of the identification of the Original Media through a TapeNumber, readable also as barcode, was stated.

From the perspective of the Factory, specifically of the Preservation Unit, there is the concern about the reliability of such identification because:

- The identifier can be missing. Especially for “Worst case” Archives.
- The identifier may not be unique. Risky when there are several Archives with their organisational tape numbers.
- The identifier may be according to a non suitable format. Especially for bar code reading.

There is the temptation of a new identification, made within the Factory. However this option can be seen unfavorably by the Archives, which require the capability to always map to the legacy identifiers.

5.1.1 Lacking archive inventory

One further issue is raised for the “Worst case” Archive, which may not have original media identification at all.

In such situation the Archive is not able to perform and require a complete tracking of its assets.

From the perspective of the Factory it would be interesting to be able to provide some kind of “CompleteInventory” service.

6 Software components outlines

6.1 Presentation

The delivery functionalities of the PrestoSpace Factory are provided by the PSO component. From the perspective of the description of the software tools, the Architecture Overview, already outlined by Figure 5, is presented in Figure 10 with a more detailed description of the PSO architecture.

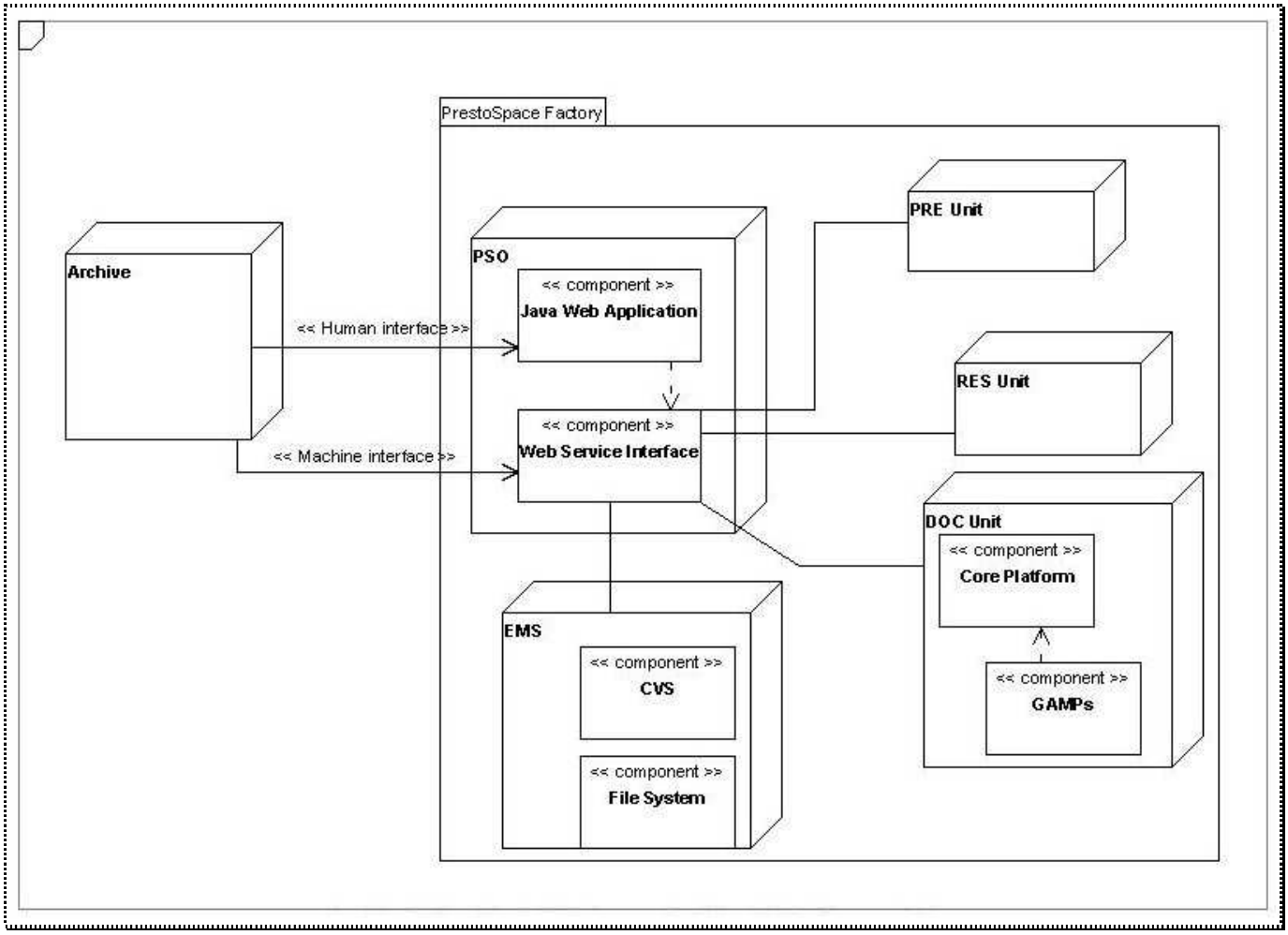


Figure 10 – PrestoSpace Factory & Orchestrator Architecture

The PSO object is mainly composed of two components: a Java Web Application, that's used to give access to the PrestoSpace Factory to human beings, and a Web Service interface. This interface is the core of the PrestoSpace Orchestrator; in fact it is used by the Java Web Application in order to perform all the operations. The Web Application provides only a human interface, but the entire logic is implemented by the Web Service interface.

A further representation is given in Figure 11, which provides a component diagram specific of the PrestoSpace Orchestrator component, including the main provided and required interface ports.

The details on implementation and interfaces are subject to change according to the evolution of the definition of interfaces and formats internal to the PS Factory and depending on the consideration of possible alternative tools.

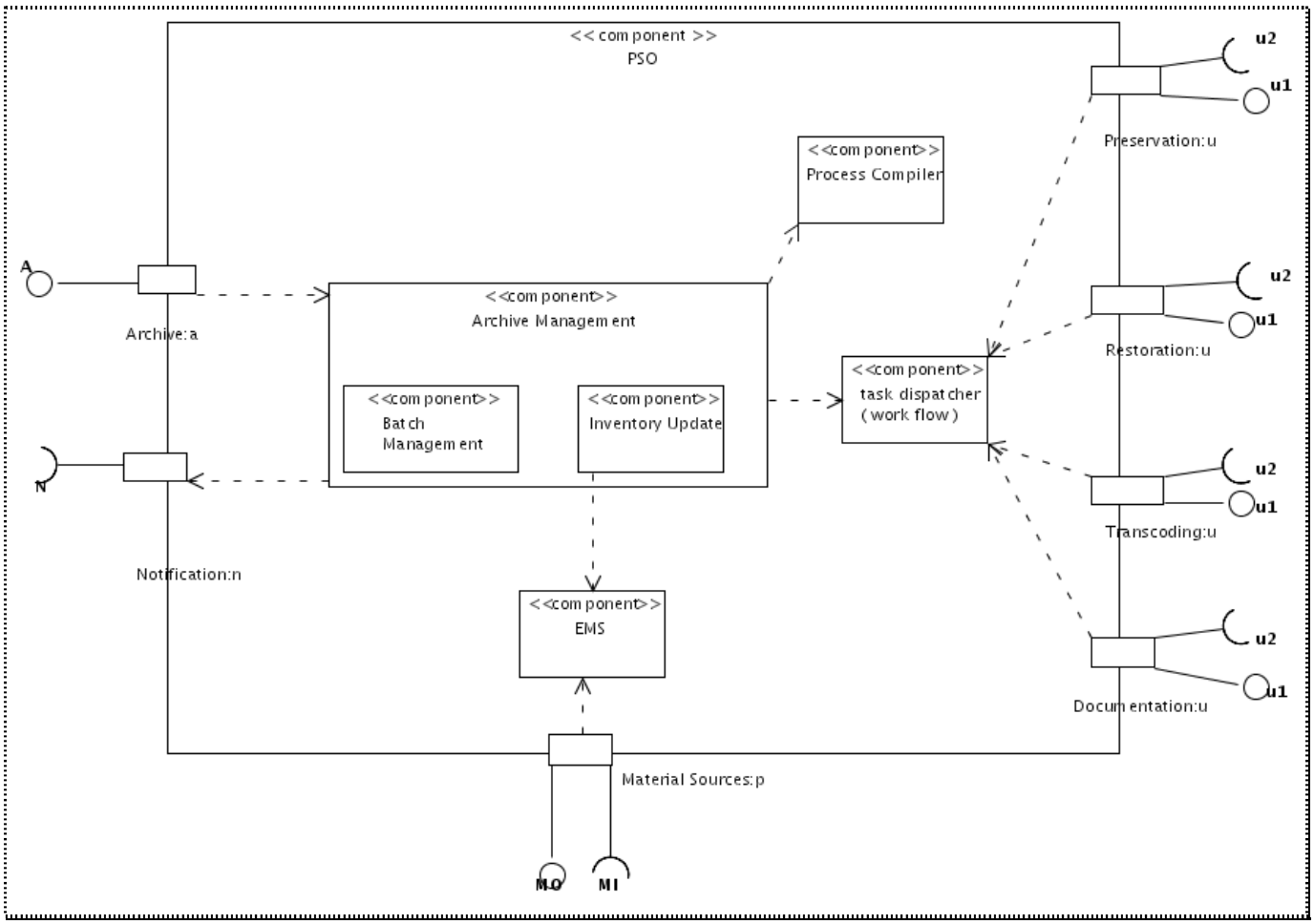


Figure 11 – Diagram of PSO component

Interface *A* is provided to the Archive in order to implement the methods listed in Table 28. It is a Web Service interface, which is derived from the analogous interface already developed within the PrestoSpace MAD Area for the Documentation Core Platform that was dealing with Editorial Objects and Materials only.

In all cases where Notification services are requested by the Archive, the PSO will require an interface *N* to which communicate about subscribed events, as defined in Table 27.

Interfaces indicated as *MI* and *MO* are used for exchanging Material Sources media files, and are offered to both the Archives and to the other PS Factory components (Units). The mechanism currently supported is the same adopted within the Documentation Unit, where standard file transfer protocols, such as those listed in 4.5.5, are used after an exchange of material access information in *GET* mode. Therefore a file transfer service is provided for getting the materials out of the factory (delivery) and a service is required for inserting the material in the factory.

Other mechanisms for the management of media file transfer are under consideration, because they may become attractive for the Archives organisations. An example is the Media Dispatch Protocol (MDP) [WKPD search for MDP], which is still in a work in progress status, although it seems very interesting.

The interfaces between PSO and the various Factory Units are presented in both “provided” and “required” mode in order to keep some generality, because the exact constraints given by the Unit interfaces are not completely defined. In the case of the Documentation Unit, however, the “required” case (*u2*) is currently supported.

6.2 PSO components

6.2.1 Archive Management

The Archive Management component is that on which the main Archive interface is based and is responsible of managing the relationship with the Archives and the information about Orders, Batches, Editorial Objects, and Materials.

It includes a Batch Management component specifically appointed for keeping up-to-date the status of Batches and of each of their Editorial Object components.

The Inventory Update component is devoted to the inventory process which is continuously executed in order to always maintain the consistency of the relationship Editorial Objects vs. Materials, and thus of Materials and Material Sources.

As Original Media are not stored on the EMS, because they are not computer files, it is the Inventory component that has to trace the location of this kind of Material Source.

6.2.2 Process Compiler

The task of the Process Compiler is to elaborate the processing directives requested by the Archive within the Order registration in order to produce the most appropriate list of tasks to be executed by the Factory Units.

The mission includes the understanding of service dependencies, the addition of implied services, and the definition of the order of processes, deciding also if some services can work concurrently.

6.2.3 Task Dispatcher

The Task Dispatcher component acts as a workflow manager by assigning tasks to the Units and being able to account about the status of work in progress, failure or success of each task instance.

It also provides an administration interface for the management of fallout situations.

6.2.4 Essence and metadata storage

The EMS component is responsible for storing the essence and metadata, and for providing access to these resources according to defined rules and protocols.

The metadata information is stored relying on a CVS system, which gives the control of read & write permissions and also provides the management of versions, permitting to recover the documents from each stage of process and the roll back from erroneous updates.

The essence media files are not within the CVS management, because of efficiency issues and because the defined process rules don’t permit to update the Material Sources, but only to create new Material instances.

7 Description of outcomes and conclusions

From the discussion carried out along this document it is possible to earn that the scenario of the relationship that has to be set up between Archives and PrestoSpace Factory can be characterised by many and various components, which can impact on the adoption of delivery models.

The assortment of Archive necessities and expectations is naturally reflected in the PS Factory positioning, affecting also the degree of integration between the two parties, especially in those cases where the PrestoSpace services are required also for the Archive exploitation.

The outline of the system architecture has permitted to identify the relevant transactions (original media, essence, and metadata), which need a framework of logical infrastructure, in order to be actually implemented.

The pragmatic PrestoSpace approach is summarised by the propensity to identify a default recommended solution, by remaining however available to consider, if possible, the Archive specific requests, not neglecting to remark any undesired drawbacks given by unusual choices.

For the PrestoSpace Factory a specific component responsible for managing the relationships with the Archive has been identified, the PrestoSpace Orchestrator.

Moreover a model describing the entities involved in that relationship has been defined and the information elements permitting the management of both processes and transactions have been identified and defined.

That has carried to the definition of services that have to be provided by the Factory (PSO) to the Archive, and the description of a protocol that should be followed by the parties.

8 Glossary

Term	Abbreviation	Definition
Archive		Organisation responsible for the preservation of audio-visual material
Archive Order		A collection of processing and transaction requirements which define the behaviour of the PrestoSpace Factory on related working units. An Order is issued on the request of an Archive and one or more Batches of working unit can be related to it.
Batch		A defined set of working units the process of which is collectively accounted.
Broadcast Quality Material		the material instances the quality of which is appropriate for publication through current audiovisual publication media
Browsing Quality Material		the material instances the quality of which is sufficient to appreciate and recognise the content, albeit some impairments might be quite perceptible.
Carrier		Organisation or Person responsible for transfering removable storages, especially original media, between two locations
CVS	CVS	Concurrent Version System, a system able to manage different version of the same artefact, keeping track on changes. It provide the methods for checking-out and checking-in documents.
Editorial Object	EDOB	An audiovisual work, or an editorial constituent part of it, from the perspective of its artistic, communicable, and expressive aspects. It is identified as the PrestoSpace Factory Working Unit.

		For each Material instance, identified with the UMID, a list of access mechanism options (at least one) is given. The access information shall include protocols, ports, filename, file size, file checksum, and other details. Account information such as user login and passwords are NOT included.
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. Properties of Material Source (esp. digital) are the encoding scheme and the number of bits.
PrestoSpace Orchestrator	PSO	The component of the PrestoSpace Factory responsible for interfacing towards the Archives, for accounting all information about the processes on working units, and for dispatching activity tasks to the various Units
Realisation		The relationship accounting how an Editorial Object is realised by one or more Material instances.
Shipment		The delivery of a bunch of removable storages, especially original media, either from Archive to the PrestoSpace Factory or viceversa
Storage		A container for a Material Source. Either a computer file or a physical removable device, such as tapes.
Transcoding		Process of deriving a Material instance from another one by decoding the Material Source and producing a new one with a different encoder. The process typically affects the properties of the both the Source and Material
Transwrapping		Process aiming to change the file format of a Material Source container, without modifying the Material Source itself. The Essence is the same, while how it is organised within the container may change. Therefore the quality of Material is the same.
Turnkey System	TKS	The TKS is a lightweight System specifically tailored for small size archives. An “all inclusive” solution, where the PrestoSpace provides the main Factory services together with tools for Archive Exploitation, within a single system.
UMID	UMID	The UMID is the Material Identifier as defined by SMPTE330M. In textual documents, as XML, it must be recorded as a string starting by “0x” and giving then the UMID value in hexadecimal (each byte is coded with a couple of characters in the range [0-9,A-F]). The UMID has 32 bytes with the a possible extension to total 64 bytes. The UMID has to be generated and provided by the entity emitting the Material instance.
Working Unit		The item on which a given process is focused. From the PrestoSpace Factory perspective the Working Unit is the Editorial Object. Other components of the PrestoSpace Factory may have different working units.

Table 29 - Definitions of terms used in this document