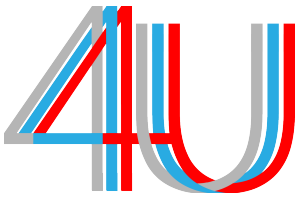

Project Acronym: Presto4U
Grant Agreement no: 600845
Project Title: European Technology for Digital Audiovisual Media Preservation

D6.3: Economic Impact Analysis: Analysis of the economic impact of RTD on the CoPs preservation actions

(Updated Version 1.1)

Project funded by the European Community in the 7th Framework Programme

Table of contents

Scope

The scope of this Deliverable is to analyse the impact in terms of economy of Research and Technology Developments within the preservation actions. It will also analyse the cost for each Community of Practice to undertake preservation actions on their assets and how the impact of Research can reduce costs depending on the characteristics of each domain.

In order to achieve these two objectives the Deliverable has analysed past actions as well as gone through existing information about CoPs and made new rounds of interviews and discussions concerning the cost of preservation and where could there be substantial reductions if RTD could provide specific issues. A general introduction presents the difficult context of the cost of digital preservation, mainly when analysed through the large array of content holders covered by the nine communities of practice in which the project is structured.

The second part, presents the analysis of the economic impact of RTD on preservation actions. This action was undertaken through the analysis of past European research projects having produced tangible results which were either industrialized or introduced within companies to improve and accelerate preservation work. It permits to understand the difficulties of industrial take-up but equally the impact some developments have had on the preservation activity and how it can produce on the medium and long-term substantial economies. There have been important results that have introduced new tools and systems in the market, however the road to industry is not an easy one due to the difficult context in which audiovisual preservation is placed, where priorities are low and funding difficult to obtain.

The third section presents through an analysis of the different operations in the general workflow of Preservation. This analysis is done in terms of the different blocks of the workflow and then Community per community in order to identify where the blocking points in the workflow are and where decisions may be difficult or where financing demands are higher.

The Deliverable provides a final vision in which, whatever advances may be made from a technical, research or implementation point of view, the main obstacle is the lack of financing programs which may permit to many content holders to jump from a day to day more or less efficient action on their contents, to a long-term plan and the capability to afford this action.

UPDATE 11/12/14:
At the beginning of December 2014 a new revised version of D6.3 has been released. The major changes concern an expansion of the contents of Chapter 1 and an extended implementation of specific use-cases in chapter 2, with the insertion of new and more extensive information as examples of impact from research results.
Chapter 3, based on the analysis of each CoP, has instead retained its old format.

Executive summary

During the last 15 years, there has been a strong development of financed researched actions either fostered by the European commission or through National programs in different countries. These actions have permitted to understand, design, conceive and develop tools that have then been introduced within different sectors and markets. Some of these tools have had a direct impact on activity, like the *Samma Tools*, mainly when they permitted a new processing activity within a domain, others have permitted to build systems based on results in order to permit within a company or sector an important improvement in the quality and performance of preservation actions.

The structuring of the Presto4U project in different Communities of Practice, each with different profiles and needs, has permitted to gain a clearer perspective on where they are on the road to preservation and what their needs are, on the short and on the long term.

1 The cost of Digital Preservation

The Presto4U project was conceived to promote and develop tools and actions in order to foster at the largest possible scale an uptake preservation initiatives mostly among audiovisual communities where this action is still pending. Content holders consider preservation as an essential action towards longevity, however even when having a precise technical roadmap to advance they are often stopped by the cost associated to the action and the difficulty in obtaining funding mainly in a highly economical constrained context as the actual one.

The first and most important step is to evaluate the cost of preservation through time; this provides a good roadmap in which costs have to be closely associated to the incurred risks that can be taken (the risk of delaying one year a migration process for example). From an institutional point of view the most difficult issue is to transform preservation and migration in heritage obligations, which needs constant content care and financial input.

Several initiatives exist which provide cost models for preservation, often applied to digital preservation as a whole.

The PrestoPRIME project (www.prestoprime.eu) has provided tools for preservation planning, risk analysis and cost evolution analysis through time: http://prestoprime.it-innovation.soton.ac.uk/

and mainly a digital preservation planning tool which can be very useful to institutions in their long-term approach:

http://prestoprime.it-innovation.soton.ac.uk/planning-tool/accounts/login?next=/planning-tool/

Another example is provided by CAROL (Centre for Archival Resources on Legislatures) driven by the Library of Congress in the US, which provides a series of tools for cost calculation, mainly under the form of excel sheets.

http://www.mnhs.org/preserve/records/legislativerecords/carol/costmodels.php

Other initiatives like JISC (http://www.jisc.ac.uk/) or the Digital Preservation Coalition (http://www.dpconline.org/) provide useful information and guidance to understand digital preservation.

The common issue with cost models and cost tools is the high range of variables that intervene in the calculation. This may make their application to real cases quite difficult to implement due to the number of uncertainties or unknown values regarding future perspectives. Independently of the internal variables within an enterprise or institution, the evolution of technology and market are a deeper source of variation through time; even if there are cost predictions in storage for example, the roadmaps may abruptly change.[1]

[1] The excellent article by Richard Wright, Ant Miller and Matthew Addis on the « The Significance of Storage in the 'Cost of Risk' of Digital Preservation » provides a good view on how storage and storage cost evolve through time: http://www.bl.uk/ipres2008/presentations_day1/21_Wright.pdf

1.1 The cost through time

Active or passive, preservation has always a cost. The difficulty lies in how cost is distributed through time. Preservation is a long-term action, which needs continuous funding with variations through time; it can be seen as a passive or active action.

For passive preservation (where objects are kept in a vault and are rarely accessed), once the functioning and staff costs are established there is a slow regular increase in cost through time depending on the increase rate of collections and sometimes some specific expense related to infrastructure evolution. This form of preservation only makes sense if the value of the contents doesn't decrease through time and is usually done for the preservation of physical objects. For audiovisual contents, the obsolescence cost has to be added to the functioning costs and this slope can be very steep for analogue audiovisual contents.

For active preservation, distribution of costs through time looks more like a chain of mountains, starting with a very steep and long mountain called digitisation, which is the passage from the analogue to the digital world. This process may take years for large collections and can imply doubling the functioning budget for years and hiring temporarily or definitively new staff with different qualifications. Once the big initial mountain period is over, the archive goes somehow back to a regular functioning situation, however with many internal changes:

- Staff competences are no more the same
- The technological environment is different and not permanent
- Regular check and control actions have to be planned in order to avoid decay and loss
- Databases have to be built and interconnected in order to manage the contents, their integrity and to associate any kind of needed information

After a period of time that can last from 5 to 7 years, another mountain of smaller dimensions will appear called migration, meaning the transfer of contents from the first information system (which is becoming obsolete) to a new one: this often implies extra work and staffing, as well as buying the new environment and the carriers needed to make it functions. These small mountains will reappear every 5 or 7 years, with maybe a higher peak, if some transcoding has to be done at some point.

The first advantage of this process is that with time the process starts to run smoothly, risks are minimised and migration becomes a kind of routine, except from a financial point of view. However a long-term funding plan can be established.

The second advantage is that when installing an active preservation process, contents are easily accessible with no potential damage for their integrity or preservation. Access associated to value generation can be a strong incentive for preservation and has often triggered funding.

This implies that archives need to have irregular budgets through time in order to follow the preservation "landscape" and prepare for major events at precise moments. This is possible in many cases, however it can be crucial for many archives mainly due to

difficulties in anticipating the purchase of information systems or staff-increase. These will be needed at a precise moment in the archive's preservation plans, which may be difficult to anticipate in difficult financial periods but risky to postpone in time. As can be seen in the PrestoPRIME preservation planning tool, postponing a major migration can have an important impact in the integrity of preserved contents due to the risk of failures within the information system.

1.2 Managing punctual important budgets

Long-term preservation is based on continuous and regular funding inputs, which permit functioning planning, technology and carrier evolution, and staff managing through time. In other words, long-term preservation is deeply associated to long-term funding.

A situation currently appearing in many countries and organisations is the grant of an important extra budget on one year, which is not carried on the following years. These punctual increases are of course welcome and useful, however progresses may be lower than expected, since they will only cover a part of the needed investment. A difficulty added to this situation is the fact that in some administrations budgets cannot be transferred from a year to the following one, meaning that the structure will have to spend the increase in a very short amount of time. This situation can be very negative when it comes to large budget increases, since it implies that there will be a need for external services, requiring call for tenders, further lengthening the process. In order to prevent these uncomfortable issues, archives have to be prepared beforehand with their preservation and migration plans in order to be ready to apply them as soon as some extra financing appears.

It also depends on the nature of tasks, which need to be undertaken. If financing is dedicated to analogue-to-digital migration, it is always a long and complex action where time estimations and cost evaluation are complex. If it is dedicated to buy a new Information System and the associated carriers, it is feasible to do it in a year's time.

1.3 What can be expected from commercial activity

For commercial institutions working in the B2B domain preservation can be a challenge, due to the cost it implies on the long-term and the complexity in managing it. However commercial actors manage to make business out of their contents. It is then a strong temptation for archival institutions to commercialise their contents when they have the appropriate rights. The question here is: how much revenue can an archival institution expect to make out of their contents?

For B2B activity, this can be an important source of revenue with some variability depending on the nature of the archive, the potential value of its contents, and the facility with which the right clearance can be undertaken (content-holders may have rights but not a structured database to manage them). A rough estimation is that a well-organised, digital archive, with a good Internet selling platform, could expect to make from 25 to 30% of their revenue through selling contents to professionals. For B2C this percentage drops down to a very low level ranging from 3 to 5% of the annual budget.

B2B activities can be a good complement for the activity of an archive, however it is far from covering the costs of digitisation and preservation of assets. For B2C benefits are marginal and this activity is more done for providing an improved image than from a financial standpoint.

B2C activities also can generate revenue under different financing models. A recent presentation at the FIAT/IFTA conference in Amsterdam gave an overview on the perspectives for cultural institutions with this kind of activities.[2] The outstanding point is the evaluation of the potential value of a collection. When a Museum puts its main paintings online, it doesn't have the same value potential as a library digitizing birth records.

1.4 The cost of outsourcing

Outsourcing is a promising solution for Preservation. Services are only starting to exist proposing diverse actions dedicated to the preservation of audiovisual assets. Actually, services are oriented towards archiving in the strict sense of the word, keeping valuable information in a secure location, with sufficient guarantees of long-term accessibility (up to 50 years). This is not yet a preservation service, which keeps and checks contents within a usage perspective as well as providing access. However actors in the domain are moving towards such types of solutions.

Outsourcing Preservation can be an interesting approach for content-holders wishing to provide a long-term perspective to their assets, while not possessing the capacities and knowledge to do it internally. This is a major evolution in archival perspective; archives always considered that their responsibility was to be close to the contents they keep and to constantly keep an eye on them. This has less sense in a digital world, and even if initially archives wanted to keep an eye on their hard disks or data-tapes, this attitude has strongly smoothed down through time and it is thinkable today to externalise this responsibility. For *CoPs* not directly engaged in Digital Preservation, this can be an effective approach, where preservation will look like an effortless action.

Outsourcing has a cost, but more easily manageable since the migration and checking processes are within the Service Providers' responsibility who will charge for an annual cost per Gigabyte associated to a number of files to preserve and thus provide a fixed cost evaluation for an archive through time.

This is why archiving as a service has taken so much time to come out. From an economical point of view it is interesting for the content-holder since he can plan on the long-term his financing needs or pay an initial amount that will cover preservation through a certain number of years. However for the service provider, the business-model may not be easy to establish; the high reliability and integrity level demanded for such actions can make them highly expensive for content-holders or economically uncertain for the service organisers. Often archival services are being associated to storage services, easier to monetize and less risky from and integrity and longevity point of view. Last but not least, the life expectancy of the service company itself is also a source of risk; when outsourcing

[2] http://fr.slideshare.net/DanielTeruggi/fiat-conference-amsterdam-2014-creating-value-from-your-archives-online-revenue-perspectives-and-possibilities-for-cultural-institutions

with a 20 years perspective there is the uncertainty related to the company and the guaranties of recovering the contents if ever the company stops its activity. This issue generates hesitation among content-holders mainly when they have a longevity mission and they find it risky to transfer the mission to an external provider without longevity and security guarantees.

1.5 Living in the Cloud

The Cloud is another way of outsourcing Preservation. The Cloud, as a concept, has many definitions: it can be an internally designed environment where contents are distributed among a number of identified units providing distributed and redundant storage. This situation already exists in many companies and may look totally transparent for users.

Dedicated Clouds as services is an important future development for archiving, since, as well as for archiving companies, the responsibility of the service provider is identified and established through a service level agreement[3]. In France there is a project of a national cloud as a storage and archival device dedicated to administration.

The big Cloud is another issue: the idea that contents are chopped in pieces and distributed in computers all around the world generates a sense of lack of control to content-holders. However there are already many services on the Cloud, many of them dedicated to storage. A storage environment is not a preservation environment: there are high risks of loss since the provider has no responsibility in securing the contents through time. However the Cloud is getting more reliable and its high level of redundancy provide higher guaranties through time.

Cloud services have the advantage of being potentially more economic than dedicated preservation services. They constitute a promising issue for the future, mainly for small and medium collections and for all those who want to externalise this task.

1.6 Conclusions

Economical issues are seen as an obstacle for all collections. Archiving and audiovisual content exploitation generate revenue but at low speed and in a long-tail perspective. Probably the best attitude is to have a clear understanding of the implicit costs of digital preservation and to analyse the costs and benefits of any action.

Digitising and Preserving are not quickly done actions; they take time to organise, to secure the appropriate funding, to train and prepare staff and to decide the functioning and use of the archive or collection once they are digital. The advantage of the digital world is to provide a new environment for publication, distribution and exploitation and thus to bring a totally new perspective for value; however this value is not immediate and has a low revenue potential.

[3] An interesting document analyses the potential impact of the Cloud in the broadcast domain: https://www.prestocentre.org/library/resources/coming-storm-report-impact-cloud-broadcast

2 Analysis of the economic impact of RTD on the CoPs preservation actions

2.1 What has Research in the Preservation domain brought to the audiovisual community

Many National and European projects have addressed Digital Preservation over the last 15 years[4]. These projects started mainly around the problems related to Digital Libraries and progressively were expanded to other domains, in particular the audiovisual domain, which, since the mid-nineties, has attracted increasing interest for research centres, companies and research funding programs.

The subsequent projects developed addressed different aspects of Digital Preservation and provided important methodological and technical advances to all communities concerned with digital contents and permitted to advance on the methods, the needed tools to address it and the associated services. The evolution of Digital Preservation strategies was a slow one, starting with an analysis of the challenges and implications in 1996 by the taskforce on "Preserving Digital Information" which established the key concepts, requirements and challenges as well as proposals on how to address at a global level solutions for it.[5]

Since 1996, European, American and coupled projects have been developed, through the Framework Programs for Research and Technology Developments[6], (from FP4 to FP7) strong American initiatives launched by the Library of Congress and the National Digital Information Infrastructure and Preservation Program[7] and through international cooperation like the InterPARES initiative[8]. The FP programs while working on methodology and services, strongly contributed to research and development of tools and devices for Preservation and Digital Preservation as well as establishing strong transversal initiatives like the Competence Centres, in charge of the information and coordination in different domains.[9]

This programs and projects, while developing a global knowledge and comprehension of the inherent problems of digital data; activated research, industrial uptake and finally

[4] Since the mid-1990s, major European projects have researched different aspects of the problem: digital audiovisual production (*Euromedia*, *Amicitia*); film and video restoration (*Diamant*, *Aurora*, *Brava*); analogue media digitisation (*Presto*, *PrestoSpace*); long-term preservation of information objects (*SHAMAN*) and digital media (*PrestoPRIME*). A range of digital preservation projects have included results with some relevance to audio and image content: *Caspar* and *Planets*, *Aparsen*, *Arcomem*, *LiWA*, *Keep*, *Protage*, *Scape* and *Timbus*.

[5] Donald Waters; John Garrett (1996). *Preserving digital information: Report of the task force on archiving of digital information*. CLIR. ISBN 1-88733450-5. Retrieved November 15, 2012.

[6] http://en.wikipedia.org/wiki/Framework_Programmes_for_Research_and_Technological_Development

[7] http://www.dcc.ac.uk/resources/external/ndiip-national-digital-information-infrastructure-and-preservation-program

[8] International Research on Permanent Authentic Records in Electronic Systems: http://www.interpares.org/

[9] PrestoCentre was initiated through an FP7 project, PrestoPRIME which permitted the conception, construction and running of the initial form of the Centre: www.prestocentre.eu

services for the community in order to propose reliable solutions on the actions to undertake.

The results of these programs are of different nature, and may have integrated the industrial circuit or remained on a shelf for further exploration. This is inherent to the nature of research and the large variety of objects it produces. In many cases European projects have been the "blending" component which permitted initiatives to exist; in the domain of digital preservation there are many small and spread initiatives which need to be linked together in order to generate the necessary impulse which exists, for example, in large industrial companies, where the research department transmits results to the development department, who then works with the marketing and the industrialization sectors to finally arrive to a product or service. A financed project creates the glue among different institutions, which will assume these actions. Whenever one link of the chain partly or completely fails, it can block the whole process and this has often been observed in different initiatives.

A general typology of results brings up:

- Machines or systems ready to use and industrialized or semi-industrialized and accessible on the market

- Machines conceived for an internal use within a company or sector which is operational however not available still on the market as a product

- Open Source results available permitting any user to take it, adapt it and integrate it in their workflows or activities

- The overall impact of a project (like PrestoSpace) on the understanding and methodologies of preservation. The process industrialization has permitted companies to build up services and use the project's models in their activities[10]

As it may be seen, results can be of very different nature and impact depending on the kind of uptake they meet. In some cases results have been achieved; however their industrialization has encountered either lack of industrials willing to take-up the challenge or a very competitive market. It must be clearly stated that the objective of industrials is not to commercialize the results of research (even if they participate to a project), but to generate business on a large scale. As an industrial stated once: "I'm not looking to recuperate my investment in a product, but to multiply it twenty times at least; if I don't have this perspective, I won't advance towards its industrialization"[11].

[10] The Belgian company *Memnon* http://www.memnon.be/Yourmediaarchivingsolutionspartner/tabid/82/language/en-US/Default.aspx has successfully used the results of this project to structure their workflow and organize the digitization process

[11] The president of HI-store Technologies in France, 2007

2.2 Use-cases showing successful results of Research being industrialized

In order to understand the typology of results different examples of project results will be presented in order to follow the process that leads form research outcomes to their integration.

2.2.1 Use-case 3: P4 Platform

The P4 Platform was one of the strong results of the PrestoPRIME project; it contains all the necessary blocks that need to be integrated in order to run a full preservation platform (P4 goes for PrestoPrime Preservation Platform). It combines all single tools developed in the project in one single Digital Preservation System. P4 provides functionalities for ingesting, updating, accessing and managing files and metadata, introducing an adapted SIP-structure based on METS. P4 provides as well a common GUI for all steps.

As an implementation of the OAIS model, it offers the basic services and interfaces for the Ingestion and Access of/to AV items, which will be accepted from the Producers in the forms of Submission Information Packages (SIP) and delivered to Consumers as Dissemination Information Packages. Software tools developed by the partners or third party tools can be integrated according to the defined Reference Architecture. P4 software is developed in Java language. The human users will typically access to the services through their normal browser.

The list of services and tools within the Platform are:

- **iModel** tool for simulating the evolution of large scale storage systems
- **Web-based storage planning tool**, storage optimisation and planning tool
- **MX**
- **FTechMDExtractor**, technical metadata extraction tool for MXF files
- **LTFSArchiver** service for efficient use of LTFS technology for storage and access
- **Video quality assessment**, automatic algorithms for file quality analysis and for the surveillance complexe processes
- **Metadata mapping service** high level representation of metadata systems permitting the mapping of different formats
- **Metadata validation service,** metadata syntax validation through XML schemas
- **Waisda ?** crowd sourcing tool to capture and collect user generated metadata
- **YAZ tag gardening** structuring tool for data generated by users
- **RightsDraw** rights management system
- **Lightweight fingerprint technology** and **fingerprinting services** fingerprint technologies applied to the analysis and tracking of very large volumes of audiovisual contents
- **D10SumChecker** Checksum and fixity calculation tool for very large audiovisual files
- **DRACMA**, i.e. Distributed Random ACcess to Media Archives
- **MServe** flexible framework for ingestion, access and processing of audiovisual files
- **Multivalent media engines** audiovisual formats analysis engines (OGG, MPEG, MXF)

The P4 platform was implemented in Open Source available initially on the PrestoPRIME website and now on the PrestoCentre website.

Fig 1: image of the user interface of the P4 Platform with some of tools related to preservation

The EURIX company was responsible for developing the P4 (PrestoPrimePreservationPlatform) under open-source license, a framework for digital preservation OAIS compliant, able to plug-into the several tools and software provided by the project partners[12].

2.2.1.1 Implementation at the University of Innsbruck

The P4 platform has been successfully implemented within the University of Innsbruck[13], a year after the end of the project. The university was looking for a DMAM system for their collections of audiovisual material, which had been recently digitized. These collections were kept by several departments, which did not have the resources to run a digital asset management system. At the same time, their central computer service was a bit reluctant to take over the responsibility since -apart from setting up the technical infrastructure– it requires a lot of efforts in terms of user service, consultancy with metadata, etc.

The library did not see itself either as the main service unit for audiovisual media. Nevertheless - and this was the strongest reason – users were interested in AV material, both for research as well as for teaching. Also the pressure to digitize analogue material is strong and money is (sometimes) available. Therefore a central system where all digitized material could be preserved and accessed made really sense for the university.

[12] EURIX Italy; www.eurix.it under the impulsion of Walter Allasia and Francesco Gallo
[13] My thanks to Dr. Guenter Muehlberger from the University of Innsbruck who provided a clear description of the process

P4 was their choice since they knew the platform well through the PrestoPrime project of which they had been partners. Looking back, they are very happy they did not use a commercial system (there was a commercial system available in the project, the Exlibris Rossetta system). P4 fulfilled their needs (though they use only fragment of all the Modules, which were finally integrated) and they have it under their own control. It was very important also to know that the company that was in charge of the development of the platform was available if adaptations were really necessary.

Currently the university of Innsbruck is using the system over a collection of several thousands of hours. They would like to include more material, but they would need more resources, which are currently not available at the University. They consider that now, thanks to the project and its results, they have competence on audiovisual material and a system where we can rather easily integrate more audiovisual material. Since their university does not have a preservation plan or strategy, it's hard to predict which role P4 will play in the future. From our point of view we would be happy to do more with it, but from the decision makers view the AV collections are not their first priority (the project did not really change this).

The problem with the university of Innsbruck and many libraries is that decision makers tend to invest in large commercial systems, not really adapted to the audiovisual needs. Open Source was then an efficient and economic way to develop and adapt a DMAM with little resources, which can now support researchers in their needs. The large commercial solutions are not affordable and would also not pay for the cost due to the number of files, people involved, etc. The University is too small and the requirements are too specific to be attended by a commercial system.

2.2.2 Use-case 2: LTFS Archiver

One of the blocks indispensable for the P4 platform was developed by RAI in order to generate an efficient and long-term system for the management of their contents. **LTFS-Archiver** is an open-source service providing means to benefit of LTO storage technology, with LTFS, with and without automated libraries. This supports the scenario of having Content stored on LTO tapes only (master level) and being able to perform all the needed archival processes (integrity, migration, access, partial retrieve).

LTFSArchiver provides the means for the fruition of LTO/LTFS storage technology in digital audiovisual preservation contexts. It can be used independently through its web interface or can be integrated in wider frameworks, such as the PrestoPRIME Preservation Platform (P4) or other client applications, through the defined interfaces.

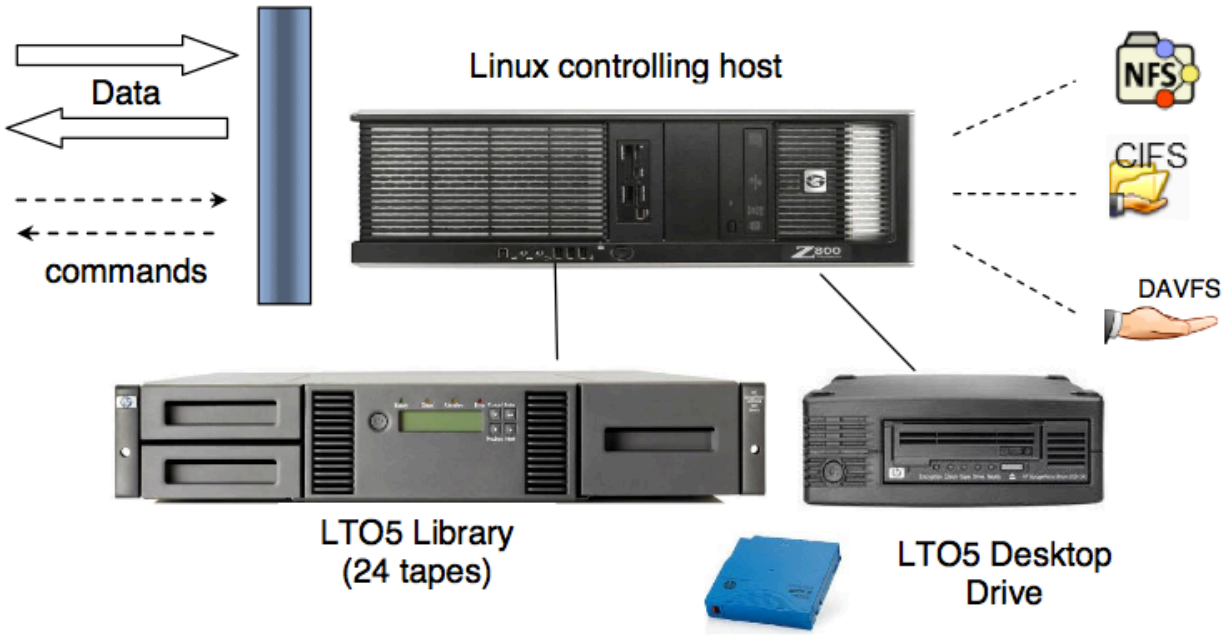

Fig 2: General view of the architecture of the LTFS Archiver

2.2.2.1 Implementation at RAI[14]

1) Which were the identified needs of RAI which led to the development of the LTFS Archiver?

LTO tape as cost effective storage solution is a solution with solid consensus.
The LTFS layer above provided the mean for ensuring open long-term access to the stored data. However LTO tapes, drivers, and libraries are not common hardware objects to deal with and a recurrent dilemma appear. Do we have to rely on external providers for having a complete working solution or is there a level in which we could make our own contribution? The answer is not always the same.
There are contexts in our company in which we develop services that need to be integrated and most of this work relies on open-source tools and/or platforms. In such cases we have the possibility to develop something by ourselves. The main need is to make easy integration of other services, e.g. in digitization processes, with a delivery for storage on LTO/LTFS within a context of preservation.

Other needs are more “itemized”:

- Need to abstract from specificities of LTO technology
- Need to handle information of inventory of archived material
- Need to handle fixity information and support regular fixity checks
- Need to organize pools of LTO tapes according to the user context
- Need to permit access to tape also for Partial Restore operations

2) Was there any product in product that could realize similar tasks? Why was the LTFS different?

[14] http://www.crit.rai.it/EN/attivita/opensource/ special thanks to Laurent Boch

We had the certainty of answering all of our needs. Especially the ones not yet identified, the product permitted us to forecast future uses.

3) How complex was it to develop the Archiver and put it to work?
Not that much ! The model is kind of “extreme programming” So you develop, you deploy, you test, you find bugs, you solve bugs, you deploy… until you don’t find any new bug…

4) What are you currently using it for today?

It is one of the two ends of our Digitization process. The other end is a delivery to the enterprise production system, which has got the HSM storage solution.

5) What advantages has it brought to your activity?

We implemented an archival solution, which mitigates the risk of vendor lock, that doesn’t put any further constraint of future access (you can but you don’t need to use the LTFS Archiver for retrieval). RAI is now in the process of massive digitization of its digital carrier contents into digital files, this kind of tool has all the needed characteristics in terms of implementation since it is robust and can be implemented in future systems.

2.2.3 Use-case 3: Optical record player

One of the oldest media consist of records on which sound was directly engraved. Until the arrival of tape at the beginning of the fifties, it was the only available device to record sound and its use continued more or less until 1960, depending in the countries. The principle of the record was a metal central layer, on top of which a thin layer of wax was deposited. This wax could be engraved with a special recording device with a chisel that would create groves transforming the audio signal in transversal engravings, which could be read by another device.
There is a great number of this records still around in archives, mainly in the broadcast domain, and as seen in the photography below, many of them crackle through time which makes it impossible for a normal needle to read the records.

Fig 3: a crackled record

A technology was then need to extract the mechanical information on the grove and to transform it in audio information, while sticking back the interrupted groves.
During the PrestoSpace FP6 project[15], the first developments were done, based on a discovery of a researcher and a patent associated to the particular method. The underlying principle is to exploit the specular reflection on the groove walls. In the system, an area of the disk is illuminated by a distribution of coloured rays, the colour depending on the direction of arrival. Through reflection on the groove walls, at each point of the lighted area, one direction (one colour) is selected (reflected) towards a camera. The obtained images are direct colour-coded sections of the audio signal. Several grooves are captured in one rotation of the disk with multiple images, forming a ring. These angles can be measured on an extended area of the record, typically 2x2mm, making groove tracking superfluous in the digitisation phase.
The method can be applied to almost any kind of record like clear lacquer disks, Metallophon, Thorens or coloured records. Its main characteristic is that it can read any record without any physical contact between the player and the surface of the record.

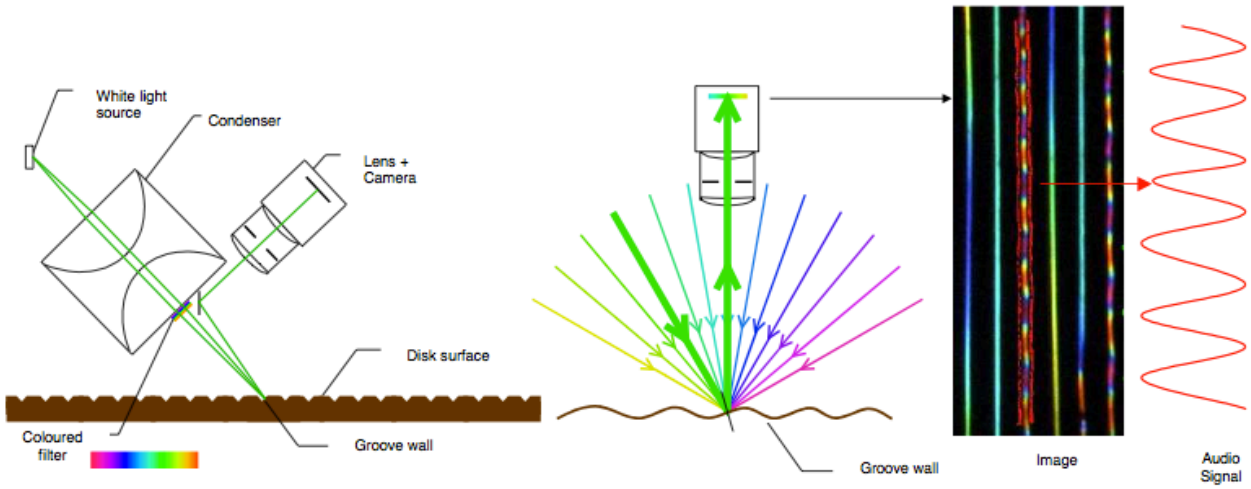

Fig 4: visual description of the applied method

After the end of the PrestoSpace project, a first prototype had been built, which could read quite slowly a record with an excellent signal/noise ratio[16]. The development of the system, which was baptised as the “Clareety” record player, was a strong priority for Ina because the process of digitizing the 225.000 records existing in the archives was a very long one and the amount of crackled records was about 60% of the whole. A potential industrial partner had already been identified through a French company based in southern France, not really specialized in record-playing but with a high competence in electronic devices. Ina and this company presented the project to a French national call for industrial take-up in 2011, which succeeded in obtaining a good financing mainly for the industrial partner, who would build the machine while the recognition software and the crackled reconstruction software. The objective was to industrialize a machine within 3 years; however after one year the company went into liquidation and the industrialization of the project had to be interrupted.

[15] www.prestospace.eu
[16] For a complete description of the process, see: http://www.prestospace.org/project/deliverables/D4.3.pdf

Ina continued on its own side the development of the software and built a new compact prototype much faster than the first one and with a higher precision in the light to sound transformation. The results were very good, however slower than real time. The emphasis was put in solving the crackled record reading for which no technology exists today. Since July 2014 good results have been obtained and in the first quarter of 2015 the first crackled records will be processed while continuing the developments in order to accelerate the acquisition and processing times in order to advance towards an industrial machine in the future.

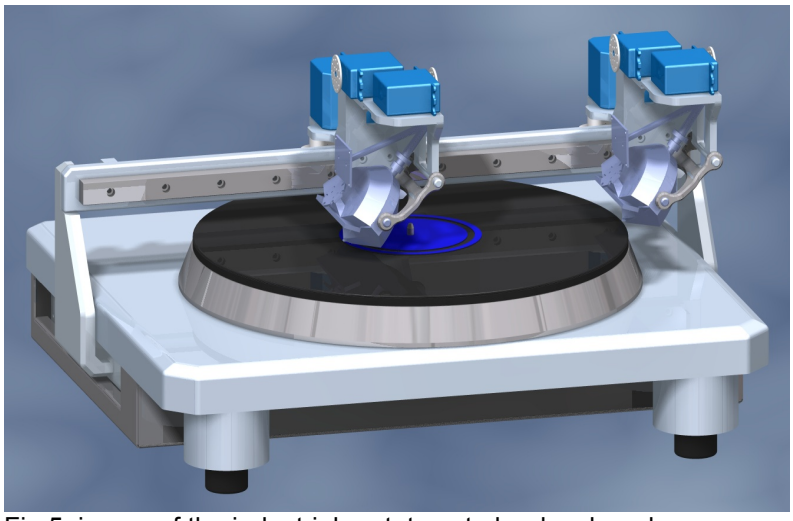

Fig 5: image of the industrial prototype to be developed

2.2.4 Use-case 4: WAISDA

WAISDA[17] was a very interesting experimentation done within the PrestoPRIME project in order to analyse the impact of user-based annotation in the quality of the generated information. The principle of the experiment was to propose a series of 1600 videos randomly presented on the website and to have users tag what they were seeing in different scenes of the presented videos. When the tags coincided with tags done by other users, the tagger would obtain points. This experimentation had a great success among users, which is already a difficult task, and produced large number or results.

Waisda? (which translates to *What's that?*), is a video labelling game launched in May 2009. It invites users to tag what they see and hear and receive points for a tag if it matches a tag that their opponent has entered. The underlying assumption is that tags are most probably valid if there's mutual agreement. Over 2,000 people played the project and within six months, over 340k tags have been added to over 600 items from the archive.

Results were very encouraging in terms of coincidence of tags on the same video and were a good way of finding how a tagging game, whose intention is to enrich the description of videos, can generate interest among user and produce useful results.

[17] http://research.imagesforthefuture.org/index.php/waisda-video-labeling-game-evaluation-report/

2.2.5 Other examples of impact from research results

Many are the successful results of projects, with also some failures as for example the wonderful audiotape reader developed within the PrestoSpace project. However efficient it was, the market for tape-recorders is still so saturated with existing machines, that the economic impact was considered too low by the industrial in order to launch a large-scale production.

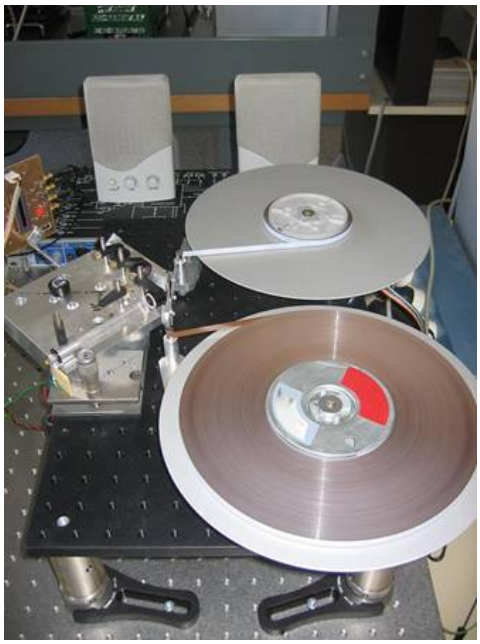

Fig 6: image of the tape-reader prototype developed within the PrestoSpace project

It was successful result of technology transfer from one domain: digital tape head to another: audio-tapes, the resulting technology proposed several improvements since it could read tapes at a higher speed than normal speed (up to 8 times) and could automatically adjust the azimuth of the heads in order to obtain the best signal quality with very little pressure on the tape so to diminish damage. However as expressed already, the market was saturated with existing analogue devices and the industrial development stopped at this point.

Another very successful results was the Samma system developed by the American company Media Matters in the PrestoSpace project[18]. This company developed highly specialised tools for analogue to digital migration, which, after the company was bought by a big actor of the domain; Front Porch Digital, became a reference for manual and automatic transfer, with hundreds of machines sold around the world in two versions: an industrial version capable of managing dozens of video cassettes in an automated way, a solo system, permitting highly reliable one by one transfer of videos or any kind of digital carrier, with an efficient quality control, cleaning, generation of technical metadata.

Finally, in the audiovisual restoration domain, important projects like DIAMANT, AURORA or BRAVA and also PrestoSpace, have opened the way to new tools (DIAMAND and S&W machines) that have been long established and are daily providing efficient and increasing

[18] http://www.fpdigital.com/solutions/samma

services to all content producers and holders. It is evidently easier to develop software as plug-ins for existing systems since they can be easily integrated in existing environments.

2.3 Impact in Cost for the different use-cases

Each of these developments has had an important effect on the general economy. Either because a machine didn't exist before, or it permitted to achieve what couldn't be done previously; either by permitting and action, which reduces the workflow's cost and has an important impact on human resources or in the general economy of a process.

As explained in chapter 1 of this document, the cost of digitization and preservation is divided in a series of small costs some of them easily calculable (as the price of a machine) others very complex to calculate (as the maintenance cost through time). Therefore the impact of research results may be analysed from a cost reduction point of view; either as from a process optimization, where with not directly reducing costs a global improvement in performance was obtained.

The sources of cost-reduction can also be of a completely different nature. Non-acting when preservation is needed, also has a cost which can and has been calculated (see Chris Lacinak, *The cost of non-action* at the FIAT/IFTA conference in Amsterdam 2014)[19]. This gives a strong perspective and the fact that non-action has an important impact in the functioning of a repository; however action is often delayed through a lack of knowledge on possible roads to follow or through uncertainty on financing over the medium and long term.

2.4 Knowledge extracted from these experiences

These experiences show the complexity of changing existing sceneries, however they show the slow but steady impact these actions have had on different sectors of the activity. If digital preservation were easy and inexpensive, it would have been long ago achieved. However it's intrinsic complexity and the diversity of components has needed a strong external pressure and action in order to address the different technical and management components of it. The different Framework Projects brought this external pressure for Research, which have increased the societal importance of Cultural Heritage, and have promoted an understanding, the development of tools and services and the widest possible diffusion of results. Initiatives like the PrestoCentre were an output of these results within the audiovisual community; the Presto4U project goes forward and contributes to a wider dissemination of knowledge, existing tools and global understanding of the impact among different communities.

2.5 Conclusions

There are still many actions to be undertaken to approach the global ambition of providing strong preservation infrastructures, tools and knowledge to all those willing to preserve their audiovisual contents. Global funding will still be one of the major lacks preventing many well-informed and intentioned content-holders and owners to advance in this direction. New tools appear and are available; workflows are better understood and can be

[19] http://www.beeldengeluid.nl/en/blogs/research-amp-development-en/201410/fiatifta-keynote-high-price-inaction

brought down to a step-by-step analysis of costs and implications. Global economies can be achieved by adopting an industrial approach to the whole processing of preservation. However if the available funding is not there, not much will be achieved. At some point European and National funding institutions will have to take this under account in order to significantly advance in the results.

This analysis also shows the complexity and diversity of Research take-ups and the different forms it make take. Probably the strongest drawback comes from the fact that the digital preservation domain has not a strong economical impact in our society and the developments are slow and medium to small sized. The investment of large industrial companies in tools or services for digital preservation has thus been slow and small sized due to the lack of structural funding to preserve the audiovisual heritage.

3 What is the cost for each CoP to undertake active preservation of their assets

3.1 Analysis of the actions and tasks related to digital preservation

Both born-digital and digitized content present a large number of challenges in relation to preservation: the increasing size of data requiring preservation, the selection of content when the totality cannot be preserved and the selection of modes of both content storage and migration to ensure long-term preservation are the most important. Despite our best intentions, it has become increasingly difficult to adequately preserve valuable digital content because of a complex set of interrelated technological, financial, and organizational pressures, including:

- Growing financial issues as increased costs and decreasing resources
- Lack of adequate staff, in numbers and expertise
- Increased complexity and volume of data
- Rapidly accelerating technological change

Despite differing interests among the various communities, we need to dramatically improve cross-organizational cooperation to increase the impact and make the most of the investments undertaken by institutions. If each institution does not have the resources to fully fund digital preservation activities on its own, spending the budget for specific assets and sharing capabilities with other institutions might lead to an optimization of resources and make a bigger impact. It's also crucial to identify and subdivide preservation functions that could be outsourced from all those that can be safely handled within the structure.
It then becomes necessary to work together to raise the profile of digital preservation, and to highlight the real costs of ensuring long-term access. We must also coordinate to develop comprehensive coverage on standards bodies, and promote specific community monitoring of technology changes relevant to digital preservation.
Specific priorities identified include:

- File format interoperability
- Effectiveness and affordability in storage solutions
- Ensuring content integrity with quality control procedures

There is a clear need for the concerned communities to share their assessments of preservation risk and their plans for mitigating those risks; to develop use-case driven best practices for specific storage solutions and to acknowledge the need for adoptable, portable and interoperable standards and protocols. Additionally, there is a need to identify more cost-efficient methods of preservation. The research must be conducted on sustainable funding and effective models. Basically, digital preservation communities need to develop a strong-shared evidence base to face all those challenges.

3.2 Economy in digital preservation systems

Techniques for reducing the cost of systems are always valuable, but they are especially valuable for digital preservation systems. Few if any institutions have an adequate budget for digital preservation; all of them must practice some form of economic restrictions. They preserve less content than they should, or take greater risks with it, in order to meet the budget constraints. On the contrary, reducing costs in acquiring and operating the system could lead directly into some combination of more content being preserved or lower risk for the preserved content.
We will discuss cost reduction at each of the stages of digital preservation, ingesting the content, preserving it, and disseminating to the eventual final users. Where possible, we try to identify a set of cost components, not all of them commonly applicable to the different communities.

3.2.1 Economy in submission and ingest phase

The cost of ingesting the content has three components: the cost of obtaining permission to preserve the content, the cost of actually ingesting the content, and the cost of creating and ingesting any associated metadata.

3.2.1.1 Obtaining permission

Permission from the copyright owner is required to make and preserve copies of copyright material. Although obtaining explicit confirmation for each item ingested is not a task to be underestimated, it might be worth the cost in a long-term perspective. Some digital preservation systems are intended to preserve content whose copyright the host institution owns. On the other hand, some archives may collect and preserve copyright content without obtaining exploitation permission. The downside of this approach is that if anyone claims to be the copyright owner, the material in question must be immediately removed. This is not a viable policy for systems preserving important curated collections.
Negotiating and obtaining permission can be difficult, time-consuming and expensive. Copyright deposit systems must operate in established legal framework that vary from country to country[20]. Other systems must negotiate individually with each publisher. The costs of doing so have been definitely identified as a major impediment to preservation. Wider adoption of the Creative Commons license, which provides the permission needed for preservation and thus eliminates negotiation, could greatly reduce the cost of preservation, however this is not really doable with recent contents still under exploitation.

3.2.1.2 Ingesting content

Just as with obtaining permission, if the ingestion of content and any necessary verification to establish its authenticity can be automated, the per-item cost of ingestion will normally be very low. To the extent to which humans are involved in the ingest process, the cost of the process can be very significant, and the conservation quality of the contents may induced unexpected costs.

20 http://www.nyu.edu/tisch/preservation/program/student_work/2012spring/12s_3049_Gronsbell_a2a.doc

3.2.1.3 Ingesting metadata

Much of the discussion of digital preservation has focused on metadata, for example which kind of metadata should be preserved along with the content, and what descriptors for representing it, but there has been less focus on the impact that metadata authoring and validation costs can have on the overall economics of the system.
All digital repositories should carefully evaluate the obligations associated with proper object documentation, particularly the amounts of metadata required to ensure longevity to the collection and the expected costs to cope with them. Repositories that manage acquisition, archiving and delivery services, for example, will have higher operational costs than the ones solely dedicated to object management. First-time digitizers typically overestimate the production costs and underestimate the intellectual costs associated with making the right selections and providing the most helpful metadata.
In the case which metadata can be supplied by the original creator of the content, or extracted automatically from the content itself, the cost impact will generally be low. To the extent to which content must be elaborated by hand with metadata, the cost impact will be significant.

3.2.2 Economy in the archival and preservation phase

The costs of preserving the content and its associated metadata can be various: the cost of acquiring and continually replacing the necessary hardware and software; operational costs such as power, cooling, bandwidth and staff time; and the cost of the necessary format migrations. Systems with few replicas have to be very careful with each of them, using very reliable storage hardware and expensive off-line backup procedures.
On the other hand, multiple back-up systems can be less careful with each of them, for example using consumer-grade hardware and depending on other replicas to repair damage rather than using off-line backups. Usually the per-replica cost can in this way be reduced enough to outweigh the increased number of replicas.

3.2.2.1 Analogue to digital transfer

Nowadays nobody longer questions the usefulness of a proper digital transfer. The variety of old analogue and digital carriers and the speed at which they are getting unusable, forces us to move towards a file-based platform, and to do it quickly. Equipment to play them all back is getting numerically insufficient and so it is very likely that a great deal of the existing bulk will never be transferred. In addition, each kind of support is subject to specific problems, related to the recording techniques of the period and the characteristics of the material on which the image has been impressed. The exact identification of the type of support and the state of deterioration are the prerequisites for the planning of subsequent interventions and to establish priorities. Minimizing costs in this area consists basically in allocating matched combination of carriers (dividing them by type and state of health) to one or more transfer sessions, in order to treat efficaciously same kind of items with similar processing. The major factors that affect the transfer cost are the degradation level of the object, the technical obsolescence of players, the lack of spare parts and the need of skilled operators inside the facility.

Restoration operations, when needed, require an artistic *savoir-faire*, scientific training and a good understanding of the ethical issues involved, such as the full respect for the originality and authenticity of every work. Aside from the latter but fundamental issue, restoration is often beyond the skills and internal infrastructure of most of the archives, especially the smaller companies. The choice in this case is to outsource this process, especially where equipment, facilities or expertise is not already in place.
We naturally tend to focus on the literal costs of moving documents into digital form, but this neglects other crucial and expensive parts of the process, especially preparing and selecting the materials to be digitized and assembling information about the materials.

3.2.2.2 Format Choice

The sustainability of digital file formats and the risks of file format obsolescence persist as significant challenges. Now that organizations are accumulating large collections of digital materials, it's important to develop actionable strategies for monitoring information about the digital files heterogeneity the organizations may be managing.
Research and mass adoption of a standard format is not an easy task at all. We are currently in a period of transition in which no one has defined standards, neither in law nor *de facto*. We can talk about evaluations essentially based on the level of diffusion of codecs, always bearing in mind that probably the "standard" format will only come in a while, and it will certainly be compatible with already widespread formats. It was held that the master copies should be, wherever possible, digitized in an uncompressed mode or with lossless compression, while access may be converted to "lighter" formats, even with lossy compression.
There are quite a number of different video file formats that are purposely created for various delivery and distribution and there is no single universally standardized format for audio-visual preservation, which makes digital archiving still difficult and quite confusing. The importance of choosing the acceptable format for long-term archiving is approximately based on the following criteria:

- *Usage* of the digital media – whether used as digital masters or access copies
- *Sustainability* of the format – based from recommendations of standard-giving bodies and established archiving institutions
- *Data size* – with obvious implications in terms of storage requirements and, consequently, costs

Digital Master

The identification of the format for archive copies must take account of all the features that characterize the conservation actions of a master: a lower frequency of use than a fruition copy, the need to ensure long-term storage, the need to check it periodically with integrity verifications, the extraction of additional derivatives, still maintaining the highest possible quality according to the availability of the institution, both in terms of cost, infrastructure, and management.
The processing of moving images presents a series of concern related to the nature of the image, and to the quality with which we make it usable. There are a lot of elements to be taken into account before drawing up a digitization plan, as well as the essential requirements of identity and integrity.

The table below gives an overview and description of some commonly used video file formats in digital archiving.

Format	Full Name	File Extension	Description	Developer	Type	Codec
AVI	Audio Video Interleaved	Avi	File format for moving image content that wraps a video bitstream with other data chunks, e.g., audio	Microsoft and IBM	Wrapper	Uncompressed
MXF_GC_UNC	MXF Generic Container Mapped to Uncompressed Images		Essence container for the MXF file format that carries uncompressed video	SMPTE	Wrapper	Uncompressed
MXF	Material Exchange Format	Mxf	Object-based file format that wraps video, audio, and other "essences", optimized for content interchange or archiving by creators and/or distributors	SMPTE	Wrapper	MPEG4/H264 DV-based/DV-based50 DV (MiniDV)
QuickTime	Quick time file format	Mov	This is a general-purpose media format that can contain multiple video, audio, text, and other tracks	Apple	Wrapper	Apple ProRes MPEG-4 (Part 2) DV MPEG-2 H.264
MPEG4 Part 10/ H.264	ISO/IEC 14496-10:2003		Block-oriented motion-compensation[21]-based video compression standard	Video Coding Experts Group (VCEG) together with the ISO/IEC JTC1 Moving Picture Experts Group (MPEG)	Encoder	Lossy
DV Family	DV, DVCAM, DVCPRO (family of related formats)	(Unwrapped raw files) dv dif	DV Stream files multiplex audio and video together digitally on a DV videotape	IEC, SONY, Panasonic, SMPTE	Encoder	DV 25 DV 50
MPEG2 / H.262	ISO/IEC 13818	Mpg mpeg	MPEG-2 is a video standard used for modern digital video format, including digital television broadcast and DVD	Video Coding Experts Group (VCEG) together with the ISO/IEC JTC1 Moving Picture Experts Group (MPEG)	Essence-only format[22]	Lossy
MPEG-4 Part 2	ISO/IEC 14496-14:2003	Mp4 M4a	MPEG-4 is an open standard video format intended for cross-platform, Internet, and multimedia delivery of video and audio content.	Video Coding Experts Group (VCEG) together with the ISO/IEC JTC1 Moving Picture Experts Group (MPEG)	Essence-only format	Lossy

[21] http://en.wikipedia.org/wiki/Motion_compensation

[22] An essence-only format is defined as basic A/V or graphic data, compressed or not. For example, JPEG data is graphic essence and PCM audio is raw sound essence. Metadata and closed caption subtitles, on the other hand, are not strictly essence but are fundamental data types.

Sustainability of video file formats

The following sustainability factors, devised by the *Library of Congress*[23], are created to aid in the decision-making of which formats will be suitable for the long-term preservation of digital information. When choosing which output format to use in digitisation strategies, apart from the overall quality attributes of a specific format, it is important to evaluate the characteristics of each format for the use of future generations.

- **Disclosure** – the degree to which complete specifications and tools for validating technical integrity exist and are accessible to those creating and sustaining digital content.
- **Adoption** – the degree to which the format is already used by the primary creators, disseminators, or users of information resources.
- **Transparency** – the degree to which the digital representation is open to direct analysis with basic tools, including human readability using a text-only editor.
- **Self-documentation** – digital object that contains basic descriptive metadata (similar to the analogue to the title page of a book) and incorporates technical and administrative metadata relating to its creation and early stages of its life cycle will be easier to manage and monitor for integrity and usability and to transfer reliably from one archival system to its successor system.
- **External dependencies** – the degree to which a particular format depends on particular hardware, operating system, or software for rendering or use and the predicted complexity of dealing with those dependencies in future technical environments.
- **Impact of Patents** – the existence of patents may slow the development of open source encoders and decoders and prices for commercial software for transcoding content in obsolescent formats may incorporate high license fees.
- **Technical protection mechanisms** – content for which a trusted repository takes long-term responsibility must not be protected by technical mechanisms such as encryption, implemented in ways that prevent custodians from taking appropriate steps to preserve the digital content and make it accessible to future generations.

The file format in which to store data is a factor of primary importance for the re-use of these data in the future. As technology continually changes, researchers should consider both hardware and software obsolescence. How will our data be read if the software used to produce them becomes unavailable? Formats more likely to be accessible in the future should be:

- Non-proprietary
- Open, documented standard
- Common usage by research community
- Unencrypted
- Uncompressed

It is therefore necessary to keep in mind all of these elements during the development of the audio-visual digitization plan, proceeding primarily to the choice of a suitable format

[23] http://www.digitalpreservation.gov/formats/sustain/sustain.shtml

that can ensure an adequate level of interoperability and compatibility with existing software, as well as a large degree of diffusion.
Regarding audio-only file formats issues, they are significantly easier to address. The standards for digitization of audio are well established, and fully documented in *IASA-TC 04*; it is then important to concentrate on new issues regarding the evolution of the community and the implications of this evolution from a preservation point of view.

3.2.2.3 Storage Solution

Storage costs are only one element of the total cost of digital preservation, but it's a factor to be taken particularly into account, as it is a large part of the continuing cost. Usually in expenditure forecasts for storage solution plan it's implicitly assumed that Kryder's Law[24] will continue in the future as it has in the past.
If this assumption will not occur, it would have two significant effects:

- The proportion of digital preservation overall costs represented by storage costs would greatly increase, since the cost of storing any individual object would no longer rapidly become insignificant.
- The projected total future cost of digital preservation would rise significantly.

Unfortunately, there is a growing body of evidence[25] suggesting that future improvements in storage cost per bit will be much slower than in the past. This applies to disk, tape and the various forms of solid-state storage.

Linear Tape-Open

Long-term preservation is the ultimate goal for virtually all objects of historic, cultural or public significance. Tape remains one of the best storage solutions of digital video that is worth considering: at scale, i.e. in large tape robots, it has an affordable cost, low power consumption and relatively high reliability for long-term access. The recording technology used by tape lags about 8 years compared to disks, but it is approximately on the same cost-per-bit curve as disks.
The digital information is stored using magnetic tapes (*LTO* – Linear Tape Open[26]), which have high capacity storage volume for backups and archives with 2.5TB of native capacity (uncompressed format) per cartridge[27] and life expectancy of 30 years[28] (in proper storage conditions). Although quite reliable for long-term archiving, this type of storage may be considered too expensive for non-professional archive[29] (one LTO-6 cartridge costs 60-80 Euros) and will require dedicated equipment (external LTO drive that may be priced between 1,500 to 17,000 Euros). Nonetheless on a choice of performance over price, LTO offer superlative points: LTO technology is continuously developing under the LTO Consortium, it has a widely adopted open standard and, with LTO-5 and the latest format LTO-6, it uses a file system - Linear Tape File System (LTFS) for ease of file access and

24 Mark Kryder, a distinguished scientist in electrical engineering and physics, in a 2005 *Scientific American* article titled "Kryder's Law" stated that areal density of bits on hard drives roughly doubles every two years. Strictly, Kryder's Law is not about cost. But until very recently, the disk drive business was highly competitive, so increases in areal density resulted in corresponding decreases in cost per bit. In practice, consumers got double the capacity at approximately the same price every two years.
25 http://www.emc.com/collateral/analyst-reports/expanding-digital-idc-white-paper.pdf
26 http://www.lto.org/index.html
27 http://en.wikipedia.org/wiki/Linear_Tape-Open
28 http://www.fujifilmusa.com/shared/bin/LTO_Data_Tape_Seminar_2012.pdf
29 http://www.tapeandmedia.com/lto-6-tape-media-tapes.asp

data management - a function similar to hard disk and flash media. More importantly, data is substantially more secured, because the data stored in tapes can be considered non-volatile. Although physical damage may be occur due to mishandling, state-of-the-art tape library housing solves this issue - some LTO systems implement robotic arms in ingest and media management operations.
Given the ideal advantage of LTO storage, an archive may well consider the usage of LTO for their long-term archival of audio-visual media, particularly those that will not be regularly accessed once the digitization of the entire collection is completed.

Hard Disk Drive

Short-term preservation aims to preserve and maintain an object economically yet as carefully as possible, allowing an institution to make a final decision regarding the future destination of an object. Immediately after digitization, the files will have to be submitted to the archive. The question however is: what other uses of the digital media will the Archive have after digitisation and how will the files be read? Accessibility is one issue to consider in this case especially for media that needs to be regularly accessed.
External Hard Drives (HDD) are the common storage media used for quickly exchanging files because of its portability, ease of use, high storage capacity reaching up to 4TB and relatively manageable cost per gigabyte (a 4TB external HDD would be priced at 150-250 Euros). HDD are simply hard disks in robust plastic or metal casing with either Firewire, USB and Thunderbolt interfaces. This gives more flexibility on the amount of storage and allocation of the physical disks and can likewise be used as back up drives. Most HDDs have about 6 years of average life expectancy[30] with 3 to 5-year warranty.
NAS is a file storage system arranged in a Redundant Array of Independent Disks[31] or RAID. The primary function of a network-attached storage (NAS) is to serve files and operates at the file level that is accessible to authorized users. NASes are basically bigger versions of HDDs with larger storage capacities (12-36 TB) that are directly attached to the system using higher-end and faster connectors such as gigabit Ethernet, Serial Attached SCASI (SAS) and also Thunderbolt. Most NASes will have supporting applications for data management. NAS appliances will have a price range between 3,000 to 10,000 Euros that includes 3-5 year warranty and technical support.
While HDDs seem more practical, these media are considered to be more fragile than tape-based storage medium. Portable hard drives are more vulnerable to damage. One known problem of hard disk drives is that the data tend to freeze up to the point of becoming unreadable because the HDD has been untouched for a prolonged period of time. This is a typical scenario that can be expected in passive archiving, where not all data is regularly accessed or updated frequently and was left to reside on a shelf. On the other hand, this type of backup system is ideal when data needs be randomly accessed, when the amount of data to be stored can justify the cost of the equipment and when there's a need for quick, economic user-initiated archive. As with any backup storage considerations, reliability, quality and performance are important. Media should be regularly checked and periodically migrated to the latest storage device.

[30] http://blog.backblaze.com/2013/11/12/how-long-do-disk-drives-last/
[31] http://www.seagate.com/internal-hard-drives/nas-drives/nas-hdd/

Archiving in the Cloud

Online backup storage is an attractive option, especially since a) data should be backed up in different geographic locations anyway and b) with online backup we don't have to worry about periodic hardware upgrades.
Newcomers to online cloud backup have a lot to learn and many conditions to consider. What about the initial transfer when we upload all of our files to the cloud? Depending on web traffic, the speed of our network and the reliability and bandwidth of our cable provider and other factors, it might take a long time to upload even just a few TB of files to the cloud. What if the connection gets dropped during the file upload? Do I have to start over again?
Moreover: are there limits to my file sizes? An hour of home video, a single uncompressed file, can be about 12 GB. Some services limit the size of the files they will accept to less than half of that. Does that mean I have to split all my video files in half?
Costs and pricing plans vary. Some services charge per volume of data. For example, one service charges $10 a month for 1 TB of storage or $100 per 10 TB per month, but the maximum per-file size is 8 GB. Other services may have flat – and much, much lower – rates with no data-size limitations. So the latter seems to be a more reasonable approach.
It certainly takes a lot of research and feature comparison to find a reasonably priced online service to fit the needs. It might also take some trial and error, so it is important not to get locked into a long-term plan before we are certain to be satisfied with the service. For example, some services allow paying month-by-month, but they require yearly subscriptions.
The two essential elements we should research before deciding on an online backup service are 1) cost and 2) ease of use. Consumer cloud storage is still in its infancy and business models are still being invented, therefore it is mandatory to take an informed and well-thought decision before starting uploading contents.
From the foregoing considerations, we can draw some conclusions:

- Optimistically, for the rest of this decade the rapid decrease in cost-per-bit of storage that has been a constant of the last three decades will be much slower.
- This will make the costs estimates much bigger and much harder to predict than would be expected.
- In a period of economic stringency, this increases the importance of developing accurate, predictive models of storage and other preservation costs.
- For much of this decade, tape is likely to maintain or improve its existing cost advantage over disk for long-term perspective preservation.
- The pricing models of current commercial cloud storage services are not yet suitable for long-term professional storage.

3.2.2.4 Operation

As with any activity involving humans, system administration is expensive and error-prone. Yet digital preservation requires very low rates of system administration error over very long periods of time. The obvious technique is to assign each replica to its own administrative domain, so that a single administrative error can affect at most one replica. In a peer-to-peer system, such as LOCKSS[32], this is naturally the case; other distributed

[32] http://www.lockss.org/

architectures may require more costly measures to achieve separate administrative control of each replica.
The only approach to reducing operational costs while maintaining low rates of operator error is to eliminate, as far as possible, the system's need for operator intervention. The large number of replicas envisaged for the LOCKSS system goes in this direction and has been successful in making the per-replica cost of administration affordable.

3.2.2.5 Checking digital content: how and when to check fixity?

In digital preservation, a key part is establishing and checking the "fixity" or stability of digital content. Many in the preservation community know they should be checking the fixity of their content, but how and how often?
Fixity in this context is the property of a digital file or object being fixed or unchanged. In this sense, it is synonymous with bit-level integrity. The *PREMIS Data Dictionary*[33] defines fixity check as "the process of verifying that a file or a bitstream has not been changed in a given period". The most widely used tools for creating fixity information are checksums (like CRCs[34]) and cryptographic hash functions[35] (like *MD5*[36]), but there are other methods such as expected file size and file count that provide basic fixity information.
The general approach in fixity checks, whether a scripts-automated process or a manual workflow, consists in:

- **Generating/Checking Fixity Information on Ingest**: whenever possible, it's ideal to encourage content providers to submit fixity information along with content objects or at least create fixity information once received the materials, because any future checking of is going to require these initial values.
- **Checking Fixity Information on Transfer**: Transferring data from one storage system to another is a potential point at which digital content could be damaged.
- **Checking Fixity on a Segment or Portion of a File** when that segment is to be provided to an end user, or when other portions of the file are to be changed. Examples of this implementation include the checksum of the encoded audio data within an audio file or individual frame-level checksums for video files.

Resource constraints may largely prevent from making fixity information for all content at fixed intervals. Different resource requirements emerge as the scale of digital files and objects increases both in number and size. Furthermore, fixity checks typically increase the usage of the media and of the mechanical devices that read and handle the media, which are factors contributing to the overall reliability[37].
The state of the art in fixity checking involves distributed fixity checking and frequent, robust repair to intentional or unintentional corruption[38]. This is done by replacing corrupted data with the distributed, replicated, and verified data held at "mirroring" partner repositories in multi-institutional, collaborative distributed networks. Digital preservation consortia like MetaArchive[39] and Data-PASS[40] use LOCKSS for this kind of distributed

[33] http://www.loc.gov/standards/premis/v2/premis-2-2.pdf
[34] For more info on cyclic redundancy check: http://computer.howstuffworks.com/encryption7.htm
[35] A common method is to compute a Message Digest ("hash") at one point and recalculate the Message Digest at a later point; if the digests are identical, the object has not been altered. For more info on cryptographic hashes: http://id.loc.gov/vocabulary/preservation/cryptographicHashFunctions.html
[36] http://en.wikipedia.org/wiki/MD5
[37] NDSA Infrastructure & Standards Working Groups, "Checking Your Digital Content" Draft Fact Sheet
[38] Jefferson Bailey, *File fixity and digital preservation storage*
[39] http://www.metaarchive.org/

fixity checking and repair. As well, many institutions use a self-maintained distributed repository system that allows them to replace damaged content with a verified, uncorrupted copy.
Built-in functionality like automated fixity checking and repair is probably the most desired feature in future preservation storage systems and is a perfect example of the complex interplay between access, performance, preservation requirements, storage infrastructure, and institutional resources.

3.2.2.6 Outsourcing vs. In-house approach

Digitization process requires expertise and modern technology. It can be built up and established in-house, but it can be allocated to a large number of specialist providers. Specific requirements, project scalability and resource availability (staff, infrastructure, etc.) will determine whether and to what extent we should make use of external services.
The primary argument for in-house approach is that it gives the institution close control over all procedures, handling of materials, and quality of products. But there are other reasons why organizations may choose to complete all digitization components in-house, including the ability to define requirements incrementally, making changes to specifications based on gradual experience, to retain direct control and to ensure security and proper handling.
Digitizing in-house may be more appropriate in case we already have the skills and facilities, or if we would like to develop them, perhaps for future endeavours.
On the other hand, outsourcing may make more sense for those projects that do not have a great amount of material to digitize. Handing over part or even the entire process to a specialized company with the experience and expertise, an established workspace, professional equipment and facilities, and tried and tested workflows certainly has its advantages. There are other compelling reasons to consider taking advantage of an outsourced service, among which we can include:

- Organizations can budget and rely upon a set cost per digitized item
- Limited risk
- Obsolescence costs are borne by vendor/partner
- Problems related to staff expertise and staff retention are incurred by the vendor/partner

3.2.2.7 Format migration

Format migration involves both engineering costs, in implementing the necessary format converters, and operational costs, in applying them to the preserved content. The engineering costs will be equivalent whatever approach is taken, but the operational costs will vary. The operational cost of migration may be large and will be incurred at unpredictable intervals, making it difficult to budget. This can lead to a drastic solution of having to choose what to keep and what not, discarding material whose migration cost exceeds its perceived value.
All this is an indication of a clear need for standards and the development of model plans for ensuring data integrity in these migrations. Much of the current practice is developed

[40] http://www.data-pass.org/

on *ad hoc*, one-off basis. Given that the forward cycle of migration will clearly be a continual part of digital preservation work, it is essential to develop clear guidance on how to plan for and manage these changes. This kind of guidance would inevitably point to issues that require further development of protocols and standards for interoperability and evaluation to help continuity.

3.2.3 Economy in dissemination phase

The cost of disseminating the content has various components, spanning from providing a strong access platform for the general public, complying with any access restrictions imposed by the agreement under which the content is being preserved (see paragraph 3.2.1.1), rights management, to the required cost for supplying copies to authorized users.

3.2.3.1 Online publishing and commercial exploitation

Cost structures are changing with the transition from analogue to digital world, as well as the access to archives has become increasingly valuable over the last few years. Digital archives are way more accessible than analogue ones and the new models of commercial exploitation have appeared to be even more profitable. The benefits of faster search and asset retrieval, easier local and remote access via networks, lower cost of replication and distribution, have pushed almost all the companies to exploit these new markets through different distribution channels.
Actual dissemination costs such as the cost of operating a web server and the bandwidth it uses are likely to be relatively low, given the archival nature of the preserved content. Content that is expected to be popular will typically be disseminated as an access copy on an industrial-level web server.

3.2.3.2 Access Copy

Derivative files are created for access or for reproduction. These are also called service, delivery, viewing or output files that are generally unsuitable for long-term digital preservation or not regarded as permanent parts of an archival collection. Unlike the digital master, the access files may be of lower resolution and are specifically created for distribution (i.e. streaming copies for the web or upload/download files for the database).
If the digitization objective is based exclusively around the final delivery format (i.e. online upload), an access copy is oftentimes more prioritized or favoured than creating a state-of-the-art digital master, because it is less time-consuming and it requires lower storage space than an uncompressed file. Deciding right from the start which will be the intended delivery format would be cost and performance effective, because it reduces the data size early in the workflow. The disadvantages are the consequent loss of quality, the difficulty in accommodating future standards and the lack of useful archival copies. Another disadvantage is that when a long-term archiving copy is needed, the whole process of digitisation is then repeated because the derivative file is not a higher quality format. This is both time-consuming and expensive, especially when the digitization is outsourced.
The quality of the derivative files will depend on the encoding adopted in the creation of the file format. When video is compressed the data rate is lower and therefore the size of a

file is smaller but as mentioned, ultimately lowers the image quality (though in some cases, not visually noticeable).

3.2.3.3 Access restrictions and Rights Management

Complying with the access restrictions typically involves an authentication system[41]. This allows sites to make informed authorization decisions for individual access of protected online resources in a privacy-preserving manner. Developing such system can be a complexity level comparable to the digital preservation system itself, because it needs IT-skilled staff for its adoption and must be maintained, verified and replaced with a newer system as it becomes obsolete. There are administrative costs involved too, as users are introduced to and removed from the system, and as the publishers with whom the agreements were made need reassurance that they are being observed.
Internet is also changing the nature of distribution of digital media from a passive Business-to-consumer flow to a much more interactive cycle where creations are re-used, combined and extended ad infinitum. At all stages, the rights need to be managed and honoured with trusted services. For this reason it is important, at least for all those communities where the sale of exploitation licenses is a core business, to invest in Intellectual Property Rights management systems in order to protect the contents from unauthorized uses. IPR metadata embedded in files facilitate the tracking of unauthorized use by third parties and are also used to secure the payment of licensing fees and royalties associated with the represented footage.

3.3 Results of the enquire

The following discussion will focus on an analysis of each community of practice, trying to highlight specific characteristics of their domain and all those barriers – economic, practical, logistical – that could prevent from setting up an effective preservation activity.

3.3.1 TV, Radio and New Media CoP

Public broadcasters have been working together for years in finding common answers to the digital era preservation challenges. It's a community with high-skilled people involved and, due to great re-use and constant re-proposition of archive material, it has been always concerned to old contents preservation and re-access over the years. At the same time, it is a community with a high level of variation: many broadcast institutions are large corporations with solid engineering teams, capable of solving highly complex issues in short time, while on the other hand we may find very small and local players with all the limitations of a tight budget. This has obvious implications in the way these two extremes manage audio-visual archives.
Broadcasters act over the entire lifecycle, from creation to play-out: ingest, preservation (short-term storage requirements and long-term planning) and delivery (access for reuse on various platforms and networks). Many of them have already put in practice a fully digital production and preservation workflow and are merely looking for better ways of performing and optimizing tasks. Others, especially those who have an important legacy of

[41] Shibboleth is a current example. Website: http://shibboleth.net/about/

analogue material or recently-born small companies, may yet have to start with setting up a solid preservation structure and typically do not spend much money on their archives because there is no short-term benefit. Generally the state of the art is represented by the national public broadcasters, which with their extensive archives have the mission of enhancing old content values and making sure that there is a genuine process of preservation for a future secure access.
Format quality has always needed to be consistent: there's a strong tradition in standardization and quality control. Format choices are quite stabilised as well, and tend to be more homogeneous than in other communities, although the variety among stakeholders could lead to non-standardized exchange protocols, often related to "local" practices and personal experience. With the ever-changing models of digital recording devices with their associated video file formats, it is difficult to keep up with all the technical changes in the market. A related issue is the adoption of a standardized "wrapper", which can greatly facilitate the successful long-term preservation of digital files, particularly to support future file migration and interoperability.
In this regard, the MXF "Material Exchange Format" was designed to achieve total interoperability in file-based media production processes and it is able to support multiple audio and video streams processed by different codecs, together with a set of metadata that describes the content material. MXF was intended as a platform-agnostic stable standard to carry a subset of the Advanced Authoring Format (AAF) data model for exchanging finished media products. It is the predominant media container format in video broadcasting and professional media, but practical MXF encoder implementations still face a huge variety of standard conformance and interoperability issues.
The task of digitising is never an insignificant issue, because even small broadcasters have huge archives. Converting all of the films and tapes into digital formats requires specialised equipment and huge amounts of effort. Despite the numerous variables to take into account, defining and determining costs is an important step in planning for a long-term repository for broadcast content. On the most basic level, it is necessary to know whether there is or will be enough money to perform the tasks laid out.
A professional preservation storage system may consists in a subset of servers, for several hundreds terabytes of raw storage, with a software system that supports shared collections that can be distributed across multiple organizations and heterogeneous storage systems. The costs for hardware, software, bandwidth, and back up are significant. In addition it's an on-going cost depending on volume: cost-per-terabyte is expected to decrease in time, but the increasing amount of data produced and ingested make sure that overall costs will not be easily lower in the near future.
More and more tools and methods for improving metadata creation are emerging. Automatic tools can be used to extract technical metadata from the files and populate databases and catalogue records. Metadata documentation is considered a crucial task to complete in order to facilitate information retrieval and provides the services to populate, maintain, and access descriptive information about the archive holdings and the administrative data to manage the archive. Costs are nominal compared to other items of expenditure, as often the software is open-source or developed in-house by staff Programmer/Analysts and most of the costs depend on volume, and relate to hardware maintenance and electricity for running the servers.
The rights situation in broadcasting is pretty clear: the broadcaster pays for the right to broadcast production, traditionally a main broadcast and one repeat. All other uses require a new negotiation with the rights holders (actors, musicians, composers, writers, producers) because they own the rights to their efforts, and in general they haven't signed for anything beyond the original contract. Much of the current effort in reducing the

obstacle of rights clearance involves codifying and automating (so far as is possible) the clearance process.
Producers and other rights holders can receive revenue from video sales, licensing and royalties. Users pay the rights holder and/or the repository to access preserved digital content at the point-of-use. This kind of model requires good search indexing and a correct pricing scheme, but all the revenues could be used to support the organization, including its preservation activities.
For concluding, the big problem of the community is not at a technological level, but it's basically economic: many public service broadcasters, in particular in Eastern Europe, are seriously under-funded. Some can barely cover their day-to-day running costs and capital investment in new equipment has often been a low priority for many years, resulting in obsolescence of equipment. In a high-technology industry such as broadcasting, failure to invest in infrastructure will result in increased costs, lack of competiveness and, ultimately, failure. Saving broadcast contents requires a new approach that brings preservation principles into emerging all-digital production and delivery environments, storage systems, and media asset management.

3.3.2 Sound and Music archives CoP

The community of Music and Sound Archives is very large; it involves any collection of sound recordings or media, where the main concern is the audio information. The strong link among all music and sound archives and sound collection owners is the nature and importance of the information conveyed by the collections, and the quality of the sound representing that information. The preservation quality level of this community is quite advanced on multiple levels: its storage capacity requirement is much lower than the required amount for video or film, because in the early days of the digital age an uncompressed format has been quickly established as a standard (PCM codec; sampling rate and bit depth, etc.). This has led over the years a considerably lower waste of energy and economic resources than other communities, allowing music and sound archives to ingest, preserve and make publicly accessible their contents.
Minimum standard storage systems are usually two physical back-ups geographically separated on:

- Optical media with digital preservation files written in their native format (i.e. ".wav")
- An open source tape standard such as Linear Tape Open (LTO), using TAR software to write the tape

A professional company's storage configuration has more redundancy requirements and it's ideally composed of:

- Redundant Array of Independent Disks (RAID 1 or greater) or an Automated Mixed-Media Library (AML), including continuous check sum monitoring
- Physical back-ups on optical disks or, preferably, a data tape standard such as Linear Tape Open (LTO)

Despite good news, this community shares most of the common challenging issues of the other communities, like short life of storage devices, deterioration of old carriers, the loss of data, and the obsolescence of formats and systems.

Sound Archives

Archives that hold audio materials are mostly local and regional archives with mixed collections, but the range is so large that they can span from large dedicated institutions and national public libraries to smaller archives specialising in a particular type of genre, radio material, oral history documents, interviews, and local folk music recordings.
As we said above, these types of archives implement a well-structured workflow and have been fixing the vast majority of their problems for at least 20 years, both in terms of format choice, storage systems and migration procedures; common technical choices and procedures are widely spread for traditional material.
However, lack of funds has always been one of the remarkable problems to be faced: in this context, considerable amounts of resources are required for the digital transfer process of all the old sound carriers. Digitize the entire audio-visual heritage of a great institution such as a national archive means having to deal with hundreds of thousands of audio-visual media (78 rpm records, wax cylinders, lacquer discs and 33 rpm, 45 rpm records, audio tapes, audio CDs, VHS, Betacam, etc.), often in a precarious state of health.
Aside from the deterioration of the carriers, what has truly become a commonly perceived threat, perhaps even worse for the future of information retrieval, is the obsolescence and the relative lack of availability of original equipment for playback. Maintaining the availability of playback equipment will become an increasingly serious problem in the near future, as manufacturing companies may shut down and spare parts are increasingly difficult to find.
A further complication is the requirement of having qualified staff being able to work with old carriers and playback equipment. Now that production has largely turned digital, these kinds of technicians tend to disappear, leaving the place to those who are digital natives and consequently have little if any training in the analogue world. However, this kind of expertise is badly needed, especially in institutions engaged in audio-visual preservation, as digitisation requires constant calibration (in particular tape recorders) for an optimal signal extraction as well as repairing and tuning of the old equipment.
In short, requirements in terms of equipment and technical expertise, particularly on analogue carriers and playback equipment, are such that many institutions are struggling with this. It is obvious that the transfer to a digital format is seen as a good solution, but due to the inadequacy of the infrastructure, some archives may find more effective outsourcing the transfer process to external experts, becoming even easier to make an estimation of the costs involved.

Musical Archives

They typically share exactly the same problems and issues of the Sound community, except for those archives directly related to music production activities, which present a general preservation problem due to the complexity of the production environment and the economical implications this may have for their activity.
Concerned actors are record labels/publishing companies, music recording facilities and contemporary music production centres. They produce music in complex digital environments, which implies the management of hundreds and even thousands of sound files inside specific software for musical recordings, editing, and post-production.
Preservation is extremely important in the music industry domain, since old productions are often “remastered” or “remixed”, which means that:

- The 2-track master is recovered in order to make a brand new (digital) version of the mastering process.
- A new version of the same work may be issued in the future with a rearrangement of the sound-files or even an inclusion or replacement of existing instruments by different ones.

In case of “remasters”, a company needs to preserve all the open-reel master tapes in a climatically controlled environment. Since there are no specific deadlines for a remastered album reissue, the preservation process must continue for an indefinite period of time.
The case of “remixes” is even more complex, because it’s crucial to keep the integrity of the original mix and to be able to access the software session within the fast-changing technology environments; and it is equally important to be able to understand how the mix was done and structured in order to easily access the numerous sound files and change any component or aspect of it. The preservation process for these production environments is the same as with any audio content, however files contain not only sound-files, but also many other different types of files, strictly dependent on proprietary software that may evolve or even disappear. The main problem in such a complex environment is that the slightest software incompatibility (at any level: OS, DAW, plugins, etc.) may result in a loss of the session for further reuse. If no effective action will be undertaken, the major production centres for classic, contemporary, or popular music risk of losing most of their collections due to complex preservation environments and lack of descriptive information permitting the understanding of the preserved contents.

3.3.3 Video production and postproduction CoP

This community is almost completely based on commercial activities. Its typical involved actors are small to medium size production houses, audio and video post-production facilities, advertisers and marketing companies and media services for broadcast. All of them act on the free market and are hired on specific niche tasks, basing their income on production activities, buying and sales of programs and rights, advertisement, post-production, etc.
This CoP thus creates and manipulates content for others, tending to a strong natural overlap towards neighbouring communities, where mutual interests and issues exist. In particular his main interlocutors are Footage libraries, Broadcast and Film and Filmmakers CoPs.
Even though film studios and broadcasters themselves are media producers, a large part of all the production chain is outsourced and handled in one way or another by the activities of this community. The key point to emphasize is that they do not own the media they work on: video production companies are usually hired to originate content for a customer, and post production companies always work on other people's content. Nevertheless there is an established tradition of holding master material even after edit, as a service to customers in case of further changes, and this is where comes into play the interest in all activities related to the preservation.
It is however difficult to generalize given the heterogeneity of users and business sectors (we are dealing with at least 10.000 companies in Europe) and the consequent different interests that each one brings forward. The gaps between the different sub-communities, their communication across business-related borders, and awareness to cost-efficient solutions more than on rational intelligent use of shared assets and needs, affect the

cohesion and thus the ability and willingness to cooperate and solve common problems. In this domain there is an abundant technology development for media asset management, file formats and wrappers, file transfer methods, digital archiving systems, but given the still insufficient standardization among stakeholders, problems like interoperability and compatibility are key issues likely to occur. Because this sector works on hiring for many customers, it absolutely needs technology that can move successfully between different companies and steps of the production lifecycle.

The opportunities presented by file-based working are focused on reducing cost and time and increasing flexibility. In practical terms this means that the transition to a fully digital workflow (which in large part is already done) is a definite advantage. For production and post-production companies this means a more efficient production chain, while reducing costs, and easier access and delivery to more customers.

In this context, one of the core operating decisions will be the format and resolution requirement of the video files to be produced and preserved. Creators will want to submit the highest quality program files possible to the customer, which then become the preservation master files. The format, encoding, and resolution of the preservation master files can have a significant impact on preservation and access needs in the future. Consequently, they should be of high enough quality that they could suit the needs of each user group, now and in the future. Most typically used formats are those connected to camcorders manufacturing brands and to the most common post-production software. This is in part promotes uniformity, but at the same time the different ways in which these tools are used give fragmentary outcomes.

An extremely important point introduced since the advent of digital is the possibility of lowering classic logistics expenses by using online delivery. High definition or forthcoming 4k content requires high-speed Internet access for uploading. Currently, most production companies (small ones in particular) do not have access to high bandwidth networks required (the order is a genuine 100 Mb/s) to distribute file-based content over the Internet, because professional fibre-based services would be too expensive and most of the time heavily underutilised. Then the final stage of the workflow has remained largely unchanged for production companies, and tape remains the standard delivery format. File-based delivery is however on the increase and, it’s a matter of a few years, will become the preferred medium for final delivery.

Speaking about storage solutions, smaller companies use consumer removable HD drives as primary option as well as cloud-based services. With regard to the latter, many users are still concerned with the degree of security in cloud-based storage solutions and they will require better and in-depth information, if we want to see some chance of mass adoption. Medium and larger companies use largely LTO tapes to back-up their productions for long-term archiving, but only the largest company have professional proper storage solution composed by multi-HDD redundant devices, shared over a network. There is definitely an increased and more confident interest in using cloud-based facilities, not only to store digital media in rational and cost-efficient ways, but also to access more dynamic tools for interacting and interlacing with other users or customers. However, cloud-based storage is only really applicable (for the moment) for “temporal” storage needs, in case content needs to be accessed by multiple parties or it is used as a short-term backup. In other use cases, a combination of removable drive and LTO is a lower cost and effective solution, already in use by the industry.

There is also a strong demand for tools or methods that would make it simple and flexible (even more automatic) to create metadata documentation in ingestion, in order to identify, search and retrieve the media across different user and platforms.

In general terms, people want to focus more and more on making content and having the job done, rather than spending their days in time-consuming activities that aren't directly connected to some form of income. This is perfectly understandable as, we must not forget, they are purely commercial activities and the search for new market areas is their primary objective. Being centred in the search for new clients and new productions, they don't consider old content to be a re-exploitable value and therefore archiving is not yet a primary mission. The use of new cloud storage services, which could greatly simplify the whole system of archiving, might change the idea for which the lack of a short-term value is not worth the costs to support long-term preservation measures. But for the moment this chance really misses the reason to exist.

3.3.4 Film collections and filmmakers CoP

Among all communities, the most homogeneous is probably the one composed by public and commercial film archives. Every stakeholder here has similar problems and issues with digital content preservation, but nobody yet has a long-term strategy or solution that does not require significant and on-going capital investment and operational expense. Even if feature film production nowadays shifted almost completely to digital, thus lowering significantly the production costs, digital cinema archives will not become proportionately cheaper, as they are not going to (immediately) eliminate the old costs of film archiving, and at the same time requiring new investments to support digital preservation activities. The longevity of this type of collections is primarily determined by digital migration or emulation, as well as the physical conservation of master tapes. So digital archives will require recurring expenditures to support the regular procedures of data quality control, formats migration and/or emulation required for long-term preservation of digital assets.
When a film arrives at the archive is catalogued in databases, repacked if necessary, and placed on a shelf in climate-controlled archive. Prior to archiving, legally deposited films are carefully checked for scratches, imperfections or omissions. Cataloguing the existing collection is an on-going task and hand-documentation is typically a serious priority.
Protecting the original by creating new film masters is the gold standard in film preservation, but the process is time-consuming, exacting, and expensive. As previously mentioned, environmental storage conditions are very important for extending the life of analogue and digital videotape as well as film and at least some level of cold Storage and controlled temperature are mandatory for both their production and master-level content. This cost represents certainly one of the most important items of expenditure for an institute.
Digitization is typically based on project funds and there aren't always enough funds for digital transfer without a commission from the clients. So, even though is a crucial task, only a fraction of the collection is digitally converted.
High-resolution digitising is done in professional scanning lab. Except for a few exceptions, public and non-profit archives contract with commercial film laboratories for preservation copying. Costs vary within a range given by labor and time required for the task and cannot be reduced to a uniform price: the more complex the work, the higher the final price. Once a film is scanned, the digital information is considered as the digital master and it's is stored either on a server or on LTO tapes placed at different geographical locations.
The film community has not reached a consensus on a standard digital preservation file format and associated codec for moving images. In pursuit of such standards, archives and standards-setting bodies should not simply aim for a single format and its related

technical details, but consider a range of formats that may address to the wider interests of the community. Among these, JPEG 2000 is a promising format offering compression performance and multiple resolution representation, but it hasn't been widely adopted by the users and it is not even part of the most common software. Other formats (like Apple ProRes SD or HD format) are proprietary and may have their limitations, but they are widely adopted by stakeholders and that's what people is basically asking for.
A major common problem in this community is that archives have to deal with not a lot of elements-per-collection, but the size of each one is huge. Uncompressed digital master of a feature film at 2k is in the order of 1-2 terabytes; in the near future scans will be done regularly at 4k, with the consequence that the storage capacity requirements will exactly quadruple and costs accordingly.
For these reasons digital archiving has lower delivery, access and utilization asset costs compared to film-only analogue archives, but it requires higher levels of investment to support the on-going digital preservation process, which may include processes like data migration or emulation every few years. The latter can be motivated by a variety of factors such as physical media decay, media or media drive obsolescence, even prior to complete system obsolescence. Older media drives may face escalating maintenance costs, there may be new user service requirements, or new media formats and/or more compatible file formats are introduced in users technology and applications. In order to face these issues, a fully automated solution for videotape migration is often implemented on a professional level with digital migration and Content Storage Management technology like *Front Porch Digital*.
Digital archivists can use emulation strategies to reduce or even (theoretically) eliminate data migration. However, a serious drawback to emulation is the cost and complexity of developing and maintaining emulation tools. To avoid the risk that old emulation tools will not work on future computer platforms, software engineers must keep adapting and updating them.
In conclusion, film archiving is not just a matter of putting digital storage media (magnetic hard drive, magnetic data tape or optical disk) on a shelf next to original analogue film. It is important to note that storing hard drives on a shelf in a cool, dry room is not a preservation strategy. Hard drives must be occasionally powered on to maintain the mechanical in good condition, and data integrity must be periodically validated (quality control is always prioritized lower than getting the job done, often based on trusted delivery from lab/facility). The long-term accessibility of digital assets on magnetic data tape or magnetic hard drives or optical disks cannot be reliably protected for the long term just by keeping the humidity and temperature of the archiving environment within an acceptable range. The annual preservation costs for a complete set of digital motion picture source materials are substantially higher than those for film, and all digital asset storage requires significant and perpetual spending to future maintain accessibility.

3.3.5 Footage sales libraries CoP

This CoP gathers those footage sales companies that exploit audio-visual holdings by conceding the temporary use of clips and footage to third parties (like broadcasters, documentarians or filmmakers) seeking out particular subject matter. The archive, being the copyright holder, will sell the "license" for the specific footage or, in the case of public domain material, they charge the access to quality prints or transfers of the material.
Footage sales are commercial activity: providers of such footage charge a per-second fee for their services, often on a sliding scale depending on the commercial purpose of the

project and the scope and method of its transmission to the public. They play thus a dual function: they primarily sell clips, but at the same time they act as an archive, being therefore obliged to practice some kind of preservation activities, if they want to continue to exploit content in the future.
Footage libraries community spans from big broadcaster sales division to large, medium and small stock footage companies that could be specialized in specific market sectors.
Because of this variety of sources, business models can vary strongly and two basic sorts of footage sales entity can be roughly identified:

- A first group is identified in all those stakeholders that act as memory institutions. They are mostly public-funded companies and for them footage sales is a side activities, rather than core business; *Istituto Luce*, *British Pathé*, audio-visual archives like *INA* or *B&G*, the *Imperial War Museum*, they all do footage sales, but not as their central activity. They are public service companies and their mission is both to enhance the audio-visual heritage and to ensure access for future generations. The long-term heritage preservation could not be necessarily a legal obligation, but it is at least mentioned in their statutes and, as institutes supported by public funding, they have a “moral” duty of dealing with it. A problem with the overall institutional business model is connecting the profit from footage sales back to the archive whose existence creates the possibility of footage sales. In large institutions, from the *BBC* to film institutes and museums, the footage revenues are not necessarily fed back to the archive, because 'archive' and 'sales' may be in very separate departments of the institution. This situation works to the disadvantage of the archives, particularly in public service and non-commercial entities.
- The second group (and the larger one) is represented by purely commercial footage collections. Sales are their core business and they exist only because they can sell their footage. Companies such as *Getty Images*, *ITN Source, Sky News, Reuters* and *Associated Press* are completely different from the previous group. Keywords here are cataloguing, fast access and availability, whereas everything has to be functional for the daily news production. Not having a public function, they are basically market-driven with a very pragmatic approach to the content preservation policy.

Despite this variety of players, all the professionals involved in this community share common problems such as necessity of digitization, effective storage solutions, delivery and preservation format choices and so on.
Such organizations must dedicate substantial resources not only to transfer and storage technology to hold the moving images, but to cataloguing and describing the footage so it can be easily searched, and providing customers with a system for preview and selecting material. The material assets of this “industry” must be properly inventoried and catalogued in order to face the daily requirements of customers in a very short time. Usually cataloguing and inventory are done together as part of a unique task, and immediately after hand-documentation is performed as well, which takes much longer time. Of course these activities could be iterated and repeated to describe constantly incoming AV content in a daily growing archive (like a current news archive) or just done one time to catalogue a more static historical archive (like the one of *Luce*). Usually archives could have access to EU funding in program like ICT-PSP to enrich their metadata and catalogue. National funds could be available as well, but nowadays are increasingly rare. Manual detailed cataloguing could be accelerated using semi-automatic tools to extract metadata from AV content.

Analogue formats on films and videotape are still commonly used, and that's why most of the community concerns are greatly shared both with Film archives and Broadcast CoP. The digitization process is at an advanced stage, but more for access or selling reasons than for optimal preservation purposes. All this affects the quality of the final copy itself: in this domain digitization is really a customer-driven activity and often quality requirements are just those sufficient for the distribution and sale of the content. This does not mean that it is the best possible quality. Even in case digitization is done for access and online publishing, a crystal clear quality is not the priority: it is rather unlikely using a 4k quality scan just to load contents on lossy-compression platforms such as You Tube and the result is often a watered-down compromises solution.
It turns out that for the most part of the archives the preferred preservation master is still considered the 35 mm: whereas most of the archives come from analogue world, film preservation, compared to digital technology, it's considered safer and cheaper. In addition, it still persists among professionals, including many of Film and Filmmakers CoP, the belief of the superior quality of film over digital.
There is however a digitization need for all those obsolescent formats that are in a decay state or in danger of becoming unplayable due to the disappearance of playback machines. Considering the great variety of old carriers to take into account, implementing a well-structured digitization plan can be a difficult task. Costs may vary depending on the source material and on the purposes (for preservation or for access), and in all cases digitisation will represent one of the biggest items of expenditure for an archive. The most common file-based format is *ProRes*, but the truth is that many customers are always asking for *Digital Betacam*. Restoration cost should not be underestimated either, as it heavily depends on the type of support, its state of preservation and the amount of restoration desired. This latter choice has ethical as well as economical implications.
In terms of storage system, the common sense is to go for data redundancy; therefore more "static" archives opt for a choice like LTO-4 or LTO-5 in separate places. On the other hand all the archives that are subject to frequent or daily increase of their contents might need more dynamic management systems, both at the level of ingestion and access.
A special attention has also given to media rights management tools and to the proper use of standard and descriptive metadata related to Intellectual Property Rights (IPR). They are key assets for content exploitation, as footage sales operations only exist in case content has clear rights ownership. IPR metadata act as fingerprinting technology and are designed to protect the contents against misuse, by tracking and detecting unauthorized use of content.
Archives recognize in particular the value of adopting systematic descriptive metadata and optimizing their audio-visual resources for search and discovery. Making content easily searchable and findable it's crucial, otherwise the whole exploitation process becomes lagging. Developing an internal information retrieval system, which index the hand-made descriptions made by cataloguers, may be a good investment of resources.
In recent years, audio-visual archives, large and small, have been making their holdings accessible online; many have decided to open the archives to popular UGC platforms such as You Tube or Dailymotion, others prefer to develop proprietary access portals. The fact is that Internet has become the predominant medium of retrieval and sales and nowadays an investment in an online showcase is considered essential. Content that is being produced digitally today is finding its way online and the business of footage sales will have finally to deal with these new commercial exploitation systems.

3.3.6 Research and scientific collections CoP

Research across virtually all knowledge domains is being transformed by digital technologies. Scientific instruments and information technologies extend powers of observation, documentation and simulation, at the same time that we are able to discover, analyse, and visualize or otherwise represent more data than we could using only analogue recording media. As a direct result, we have seen a new focus on the accessibility, integrity, and stewardship of data, and significant changes in archiving practice.

Research and scientific collections have specific requirements for professional staff in order to fulfil preservation functions. However, full-time curators who oversee a collection are something of a rarity. Care of collections is often in the hands of extremely dedicated and motivated research personnel who perform preservation-based activities in a quasi-independent way. Furthermore there is often a lack of professional support that cannot be replaced by voluntary support, especially as orientation training, accompanying advices, and guidelines; rules or standards for dealing with collections are generally not available or at least there is a low level of awareness. This is due to the fact that it has not existed until now a community really conscious of their requirements, but rather individuals involved in very different research areas (linguistic, math, environmental forecasts, medicine, etc.) that had no identity as such.

There are really very few archives with a genuine mission of audio-visual material preservation; in most cases, the researchers keep their data on a single computer, or through server systems in RAID configuration. Moreover, this community hasn't yet adopted cloud-based storage systems.

Even with widespread recognition of long-term value and well-defined selection criteria, data creators may lack strong incentives to preserve. At the economic level, there is a scarcity of resources. Generally, funds for a given project are invested in staff and in the search for new data, not in the preservation of the old ones.

Metadata documentation and indexing of a collection are basic requirements for making its scientific potential visible and enabling adequate access for research. It's not unusual however that even the most basic inventory – the digital or analogue documentation of a collection in inventories that are accessible and comprehensible to third parties – is not even done. Even though the advantages of digital documentation and indexing of objects are no longer in dispute, there is often a lack of resources (staff, server capacities, equipment, software) precisely in this area, thereby preventing visibility, availability and easier access.

Regarding digitization phase, the transfer from analogue to digital is minimal and is limited to audio carriers (open reels and cassette tapes). The scientific use of the video is mainly born in the digital age, so there isn't an urgent need for digitization. It must be said that, although this process of digitization is somewhat limited in quantity, in all those cases where it is necessary to transfer content it turns out to be both time consuming and very expensive, given the absence of the instruments and the skills to do so. Moreover the matter of which would be the best archival format in a long-term perspective is not even considered, as the acquisition process is simply limited to acquire and preserve the native output format of the camera.

More generally, this community has a very limited preservation planning, as their primary mission does not consist in preserving itself. Cost calculation is highly dependent on each specific research funding, inside which an item of expenditure may be dedicated (or not) to preservation. The resources devoted to maintaining, giving access and preserving audio-visual contents are definitely very limited and typically temporary, as part of more general

research projects and activities. The follow-up expense for digitisation, the on-going costs for server capacities, data maintenance, updates, conversion, etc., which need to be met, should be taken into account as infrastructure costs, but at the present the financing is in many ways unsuited to maintaining a fundamental infrastructure and the core tasks are often not sufficiently covered.
In order to help researchers in disseminating outcomes, to protect the intellectual property of the research, to reproduce research results produced by other research groups and to extend previous research work, this community still needs to understand the importance of tracking the records and keeping all the documentation produced, and then ultimately considering audio-video preservation as a final turning point.

3.3.7 Learning and teaching repositories CoP

AV technology has a growing demand in this domain, due to the increasing pressure on innovation and education enhancement. Audio-visual content is used to support distance learning courses, and capture other activities like conferences and seminars.. Audio, video and films are also used in classroom to support teaching and many universities have started to act as producers of content by recording lectures, making podcasts and sharing videos on line.
Nevertheless, unless they are expressly prioritised, university and academies collections often do not receive sufficient consideration in matters of financing, and they have to work hard to secure appropriate resources. Smaller collections in particular are often affected by long and complicated administrative procedures. Collections are often little known, do not have appropriate resources and are underused.
Overall, university-based collections have critical staff shortages. Only in a very few cases do university collections have adequate staff resources that meet preservation requirements. Some collections have to completely do without budgeted staff, while many are subject to high fluctuation. Not infrequently, the lack of staff means that research enquiries cannot be dealt with. Catalogues currently used by the majority of university libraries have been mainly set up for cataloguing books and don't include many fields for the description of AV digital objects. For a large number of collections, even the first step of inventory taking – the digital or analogue documentation of a collection in inventories that are accessible and comprehensible to third parties – raises problems. Often only parts of the collection are documented, and any further documentation is either impossible or can only happen very slowly due to a lack of resources.
Some CoP members use the central catalogue established within their national library network (e.g. OPAC in Italy) or integrated library management system like Voyager by Ex Libris. When national infrastructures are used costs are often proportional to the amount of records added into the library and include administration costs. More established archives in larger institutions have developed open-source solution (based on Fedora). Others use CATDV and Filemaker, which in terms of costs are considered quite high. It could also be a project-by-project scenario and some funds might be available within that particular project.
At the moment preservation in many institutions is done as a good willing because they don't have a mission to preserve so not all the preservation actions are undertaken. A full preservation plan is very difficult to consider, because of the lack of funds and the way collections have grown within universities, often scattered across different departments and without a full inventory. Some CoP members have only recently started to take some

preservation actions and it is still not very clear how preservation should be organised, making it very difficult to calculate expenditures.
Costs are quiet varied across universities, but there is often a perception that storage represents one of the highest costs for many institutions. The current tendency in many institutions is to focus more attention on keeping access formats (often compressed) because storage is very expensive and there is a need for immediate availability of content. Large-scale spinning disk systems are considered as ideal, but not all the collections come to possess advanced quality servers in RAID configuration.
Automation of metadata extraction is strategic in this field and these types of tools are really needed, but they are still very expensive and most institutions have no funds available. Among the missing tools, there is a particular demand for automatic scene detection tools to capture time code metadata, and software for capturing bibliographic metadata from existing database and import in new cataloguing system.
Quality control is currently done with visual checks and checksums. The CoP might be interested in automated tools to facilitate the correct ingestion of files and monitor quality during digitisation. With the right tools technical difficult goes down, but costs are still considered very high and not a priority for most universities.
Developing a medium and long-term preservation strategy, which captures the requirements of the various users in charge of the individual collections, requires synergies relating to administration and technical assets (server capacities, digitisation, restoration, technical personnel) and therefore an investment of considerable resources. Besides software acquisitions are often controlled by IT departments, which have little knowledge of the complicated nature and management of AV materials. Some institutions have received funding because they have taken part in European projects. Occasionally funding is released for a particular department to support research or a new course. In general we have to deal with very limited or no funds at all for digitization, migration, restoration and security environment. At present the financing of university collections – in terms of both structure and volume – is in many ways unsuited to maintaining them as a long-term infrastructure or facilitating their use in line with demands. Basic tasks relating to collections (care, preservation, indexing) are often not sufficiently covered by core funding. Moreover, collections are often at a disadvantage due to internal criteria for resource allocation that tend not to be suited to AV collections. Consequently if universities experience a budget cut, collections are the first to suffer, at the expense of the effectiveness in content preservation.

3.3.8 Art & museum object, artists and their representatives CoP

Since early 1970s, as the new technology became commercially available outside corporate broadcasting, video has established itself as a tool for art creation. Nowadays any contemporary museum around the world is likely to host a wide range of video works, spanning from small monitor projections to complex multi-channel installations. With this new artistic and technical development, new specific preservation challenges came as well, initially linked to videotape deterioration and playback machines obsolescence and then with all the problems involving a file-based workflow.
The main issue about video art preservation is that the collecting institution has no control over the technology used to create a work. Heterogeneity (of formats and forms) is thus the rule, and with diversity comes obsolescence: if only one kind of video capture and projection were required, video works could be preserved as needed. When literally

anything can come into a collection, the requirement to preserve technology becomes enormous.
Collecting institutions that acquire works are ultimately responsible for their longevity. They own a relatively small number of artworks (in the order of a few hundreds), but all have a high market value. In any case their documentation can be surprisingly complicated. Numerous standards have been set by librarians and archivists to keep catalogue records consistent between various institutions and databases. Usually there's a huge amount of descriptive information around those artworks, regarding the aesthetic intentions, physical and technical attributes of the work and the parameters of possible contexts for its presentation (aspect ratio, resolution, number and display specifications, etc.), and an outline of concerns for the on-going life of the artwork. Acquiring as much information as available, usually granted by the artist at the time of the acquisition agreement, is crucial for managing art collections and their re-proposition in the future. It is important to say that the amount of ingested information in the form of metadata is minimal compared to the generality of the overall required information, preserved as a normal "instructions" text file.
Most of the old content is still on analogue and digital tapes, which are either deteriorating or quickly getting obsolete and useless for the lack of playback equipment. Nevertheless museums and collections need to keep video art accessible; it is therefore vital to work as soon as possible on a file-based mass digitization process. Because of their age and the difficulty in finding functional playback equipment, 1/2" open reel and 2-inch *quadruplex* videotape are the two most at-risk formats. Formats like *U-matic* and 1" Open-reel Type C, have a lower but increasingly significant risk factor. Contemporary digital cassette formats, including Hi8, *DV* and *MiniDV*, have a relatively high risk factor due to the extremely small size of the tape and cassette mechanisms. In this scenario, if we also consider *VHS*, *S-VHS*, *Betacam* and 8mm formats, it's easily understandable that the time available for having this process done is never enough.
Although preparing the budget for a videotape digitization project can be a challenging task, the costs are actually determined by a relatively small number of variables, like the original format, its condition and the program length, as well as the desired destination format and level of restoration. In creating a budget, an often-underestimated item of expenditure is the cost of quality control. Once restored and digitized, each new file should be reviewed for a final quality check, checking things such running time, audio and video errors or drop-outs, etc. which can turn out to be time-consuming activities. The choice of outsourcing the process can be considered, even if typically there is a tendency towards in-house post-production facility. It is nevertheless true that, without appropriate technical equipment, very obsolete formats can be played back only by specialized labs, consequently adding additional expenses for digitizing.
There is still very little consensus within the community regarding a digital file format that is appropriate for the long-term preservation and they are always struggling with compression issues, codecs, and file compatibility. Although these issues have yet to be resolved among archivists, considering the limited number of works and the moderate space storage requirements, there is almost always a tendency to keep an uncompressed (or lossless compressed), 10bit video in either an *.avi* or *QuickTime* wrapper.
In any case, from now on, the artists are deeply aware of the importance of quality in their work and, unless there are specific aesthetic reasons behind a decision, they maintain the production chain on the highest quality standards. Among lossy compressions codecs, *ProRes* is a largely adopted solution due to its use in *Final Cut*.
In planning a preservation project, the cost of the storage system and its maintenance play an important chapter. Museums and art collections, for the high value of art works, must

necessarily focus on security and data redundancy, even if this means the most expensive option. They are usually in possession of redundant independent hard disks system (such as RAID-6) or mirrored servers in different geographical places. It is also good practice in most cases the use of a further back up on LTO, which will be placed off-site; storing data on different media prevents from being tied to a single technology and keep content safe from any eventuality.
As part of the preservation, a collection must also take account of migration, in order to keep the content accessible in the long term. If nowadays for most of the cases this still means transferring an old physical carrier to a file format, in a near full-digital future the process will imply a transfer of content every few years to a non-obsolete file-based archival format supported in the industry.
Emulation is another preservation strategy and it's even more complex than simply preserving digital files, but it can be an effective solution in case of computer-based works: it consists in re-creating the technical environment required for viewing an object by running a program from an out-of-date system on a contemporary one. Operating systems, software and plugins are heavily subjected to fast obsolescence processes, and often within just a few years a new software version is responsible for creating interruptions in backward compatibility. This approach aims to maintain all the information about the hierarchy and the overall infrastructure of the work, so that the system can be reengineered by future systems to emulate the original environment. In some cases this is the only solution to be adopted, but it's definitely high-priced considering the need for very qualified personnel for the engineering work, which are often not at the disposal of the museum.
In conclusion, video work raises issues that are absolutely specific compared to other communities and its preservation needs are wide ranging. Because of their complex nature, they might require a broad range of preservation expertise. Unlike other visual arts disciplines, video art involves the preservation of multiple components, such as sound and/or visual media, display equipment, sculptural elements, and any process that is specifically contingent to the interaction with the observer. In the future, it is likely preservation master will be transcribed, re-recorded, or cloned to a different format as they evolve. Regardless of the future kind of system, the quality of the image and its technical specifications must be preserved as they were originally conceived, even if the transfer is capable of achieving higher resolution or enhanced details, and the real big challenge of this CoP will be ultimately to maintain the functionality of the artwork as closely as possible to what the artist originally intended.

3.3.9 Personal Collections CoP

Digital preservation is a familiar issue among the world's leading cultural institutions. But despite experts and professionals can claim good successes in preserving digital files and collections, most of the general public – the largest group of digital file stakeholders in the world – are unaware of what digital preservation or personal digital archiving is or why they should care.
Today, most people have some sort of digital files to preserve, among which photographs are probably in greater numbers than audio-visual materials.
Personal collections community typically has to deal with two kinds of scenarios, each belonging to different historical periods.

Collectively produced Web content

Social networking and User-generated content (UGC) sites, whose origins date back to a decade ago if not less, play important social, cultural, and political roles in contemporary life. A new form of communication and knowledge production not created by a single individual or group, but by the collective efforts (coordinated or uncoordinated) of many individuals. User-generated content is generally created outside of professional routines and practices. They often do not have an institutional or commercial market context and can be produced by non-professionals without the expectation of a profit or remuneration.
This particular context brings out questions that are still unanswered: what should be preserved of this content? Whose responsibility is it? Who owns the content once uploaded on host sites? Who preserves the content? And who should pay for it?
Many of the ownership and rights issues are unclear or we take for granted things that in fact are not. Furthermore, and that's the basic point of this group, UGC sites do not pursue the goal of content preservation, but they are just concerned in stocking content on-line, without long-term guarantees.
A sustainable preservation strategy for collectively produced Web content may looks like nothing that exists for other digital materials. Previous questions can only be answered over time and empirically. Today, it is important to formulate the right questions to pursue, and to test and model different strategies for preservation. Some actions should be taken now, such as building partnerships between content creators and owners, and the roles and responsibilities among those who collect and preserve these materials.
For all the above reasons, the following discussion and its assumptions will focus only on the next group.

Homemade film movies

Old films don't hold private meanings just to the people who shot and to those who appear in them; local historical societies, researchers, cultural enthusiasts, documentarians and regional archives may give a value to homemade movies.
Typically, they are business-oriented activities, where a content producer (or the inheritor) donates some audio-visual material to an archive, which takes charge all or part of the digitization costs. We may encounter film transfer services that offer free transfer work in exchange of the rights to incorporate footage in their own documentary projects or in their stock footage libraries. Content producer then will be asked to sign over a legal agreement with the archive, so that they can put clips from films online on a dedicated platform. The purpose is, given a monetary value of the movie, to sell contents and share the revenues between the owner and the host site.
Often, the saved expense of having the films transferred will make this sort of arrangement worthwhile. However, there are a few important questions from the point of view of the donor, like:

- What is the real value of the film?
- Who keeps the original carriers?
- Do I know the legal agreement with the archive in terms of content exploitation?

In case stakeholders exclude the intention of contacting an archive or a transfer facility, little attention is dedicated to long-term preservation, not even having the awareness of the basic issues. This creates an urgent risk, and a crucial need to ensure the long-term preservation of high-value materials that have no natural steward or are otherwise in

danger. Beyond the specific content of their repositories, stakeholders face common issues, such as lack of core skills, or the impression that preservation tasks would go beyond their means. There is no preservation planning, if not at a rudimentary level, and there isn't even knowledge about risks associated with typical media.
Content documentation is a major issue: people immediately appreciate the value of storing home movies along with some documentation for the benefit of those who come later, but it is unlikely to go beyond mere names and events and include crucial information that might seem obvious to the immediate family. If for analogue media a basic documentation often occurs, digital files are often left with the progressive output file number of the camera. Metadata is largely an obscure concept to the general public.
Storage systems are typically domestic, with the bad trend of storing all the digital collection in only one place. Cloud storage is gradually taking place as a primary or back-up storage solution, although the reliability of archiving services is not considered as a major issue. Commercial organizations like cloud storage providers, in case of data loss, usually deny any responsibility, but in people's mind continuity and preservation over time is taken for granted, even if unfavourable conditions are already present (and explicitly accepted) under the terms of the contract.
Unlike many professional organizations, people at home typically do not have special services or knowledge on how to guard their digital data from loss or corruption. Individuals and families often don't have enough skills to develop an archiving project and without education and training most of them are in the position of simply doing the best they can.
Cost of a perpetual storage can be extremely expensive for individuals without a profit-making activity and it might be hard justifying investments without simple cost calculation models (like a "pay once/store forever" solution, knowing the cost per terabyte, etc.), which makes possible to decide what we can afford to save and what we must give up.

3.4 Conclusions

Viewing preservation through an economic lens shows that there are structural and systemic problems in providing sustainable digital preservation. These structural problems can, however, be offset by certain choices and remedies that reduce barriers to sustainability. The preservation of digital objects involves a wider range of processes to be put in place since before the file is created: from the selection of the archival format, to the prediction of the space required for storage of archival copies and backups, up to a proper metadata implementation to ensure long-term retrieval and access.
In most cases, the suitability and sustainability of the strategies adopted emerge from a combination of the conditions prevailing in the environment and the choices made by various actors. The big problem in this case, and it is transversal to all the CoPs, is that these choices are influenced more by restricted economic contingencies than from long-term requirement. Capturing more and more information to an increasingly high quality makes preservation expensive; it takes longer to gather and transmit more complete information, and it costs more to store it. Not all companies have the same mission, and not all have the same financial means to face the task. Contemplating a digitization project, involves considering seriously the costs and what they might mean for the organization.

Final Conclusion

A proper method of work, for setting up a widespread and well-targeted policy of protection of the Audio-visual Heritage, should start from the knowledge of a reality that's now extremely rich and varied, not easily governable according to a single criterion, much less reducible to the schematization.
Of course, it is always desirable to reduce preservation costs over the long term.

Document information

Delivery Type	Report
Deliverable Number	6.3
Deliverable Title	Economic CoP Impact Analysis
Due Date	June 30th 2014
Submission Date	June 29th 2014 (first release); Dec 11th 2014 (revised final version)
Work Package	WP6 – Evaluation of emerging needs, project impact and sustainability of outputs
Partners	INA, B&G, DFI, TATE, TV2, LUCE, KCL, CNR
Author(s)	Luca Bagnoli; Daniel Teruggi
Reviewer(s)	Walter Allasia
Keywords	Preservation costs; economic impact; RTD
Document Identifier	Deliverable_D6.3_presto4u_11_12_2014 version 1.1
Dissemination level	PU
Document Status	Released

Project Acronym	Presto4U
Project Full Title	European Technology for Digital Audiovisual Media Preservation
Grant Agreement	600845
Project Coordinator	Stichting Nederlands Instituut voor Beeld en Geluid
Contact Details	Sumatralaan 45, 1217GP Hilversum, The Netherlands. msnyders@beeldengeluid.nl

Document Status Sheet

Version	Delivery Date	Comment	Author
0.1	11/6/14	Preparation draft	Daniel Teruggi
0.9	26/6/14	Second version	Luca Bagnoli
1.0	29/6/14	Third version, released	Daniel Teruggi, Luca Bagnoli
1.1	11/12/14	Update	Daniel Teruggi