White paper

OAIS Compliant Preservation Workflows in an AV Archive

A requirements project

Netherlands Institute for Sound and Vision

Annemieke de Jong

Beth Delaney

Daniel Steinmeier

5th of September 2013

Introduction

Since its establishment in 1997, the Netherlands Institute for Sound and Vision (Beeld en Geluid) has always served two functions. First, as the central production archive for all public broadcasters (20) in the Netherlands. In this role the archive is responsible for storing and providing access to broadcasted television and radio programs. At the same time Sound and Vision functions as the Dutch national audiovisual archive. Thus, its collection also includes Dutch cultural heritage audiovisual material, including scientific material, documentary film collections, photos and objects, amateur film and AV collections from businesses and social organizations. In addition, Sound and Vision plays a central coordinating role in the Dutch broadcasting and AV cultural heritage landscape, gathering and disseminating knowledge in the digital storage and exhibition domains. Currently Sound and Vision manages a collection that includes more than 800,000 hours of radio, television and film programming. Yearly, thousands of hours of digital born material, created in the broadcast domain, is added to the archive.

During the recent past, Sound and Vision has become best known for its analogue digitization work. The huge project ”Images for the Future” has seen the creation of digital files for more than half of its analogue film, video and audio collection. Another important development was the establishment of a direct IT interface between the archive and the digital broadcast production systems. Since 2007, all the broadcasted radio and TV programming as well as its related metadata has been ingested into the archive via an automated digital workflow system. In Sound and Vision’s depot, approximately 6 petabytes of digital born and digitized material is stored in its digital repository system. Every day, sound and image fragments from the repository are delivered to hundreds of users in the professional broadcast domain as well as at home, in educational institutions, in businesses and as visitors to Sound and Vision’s museum, The Experience.

Preservation

Now that Sound and Vision has digitized a great deal of its collections and thousands of new digital files are ingested daily, is faces an important new challenge: the management of a fast growing digital storage repository and the increasingly related complexity that brings with it. The chance of failure that the storage of digital material brings with it is great; it is necessary to maintain firm control over the lifecycle of digital files. In order to achieve long-term preservation, more processes, procedures and metadata need to be incorporated into the system. The role and responsibility of all parties, producers, archivists and users, needs clear definition.

The primary question is: how do we ensure that stored digital material stays accessible for the users? The different formats, size and location of digital files, delivered daily to a larger and larger client base is huge and growing still. How do we make sure that we are able to deliver up-to-date formats that the Designated Communities can actually use?

When the first IT integration with the broadcast production environment was designed, long-term preservation was not yet a major consideration. Thus, during the initial digital ingest and storage infrastructure design, the processes needed to ensure long-term preservation were not explicitly incorporated. Yet, in both roles, as a broadcast production archive and as a national AV heritage coordinator, guarantees of trustworthiness and long-term preservation must be met.

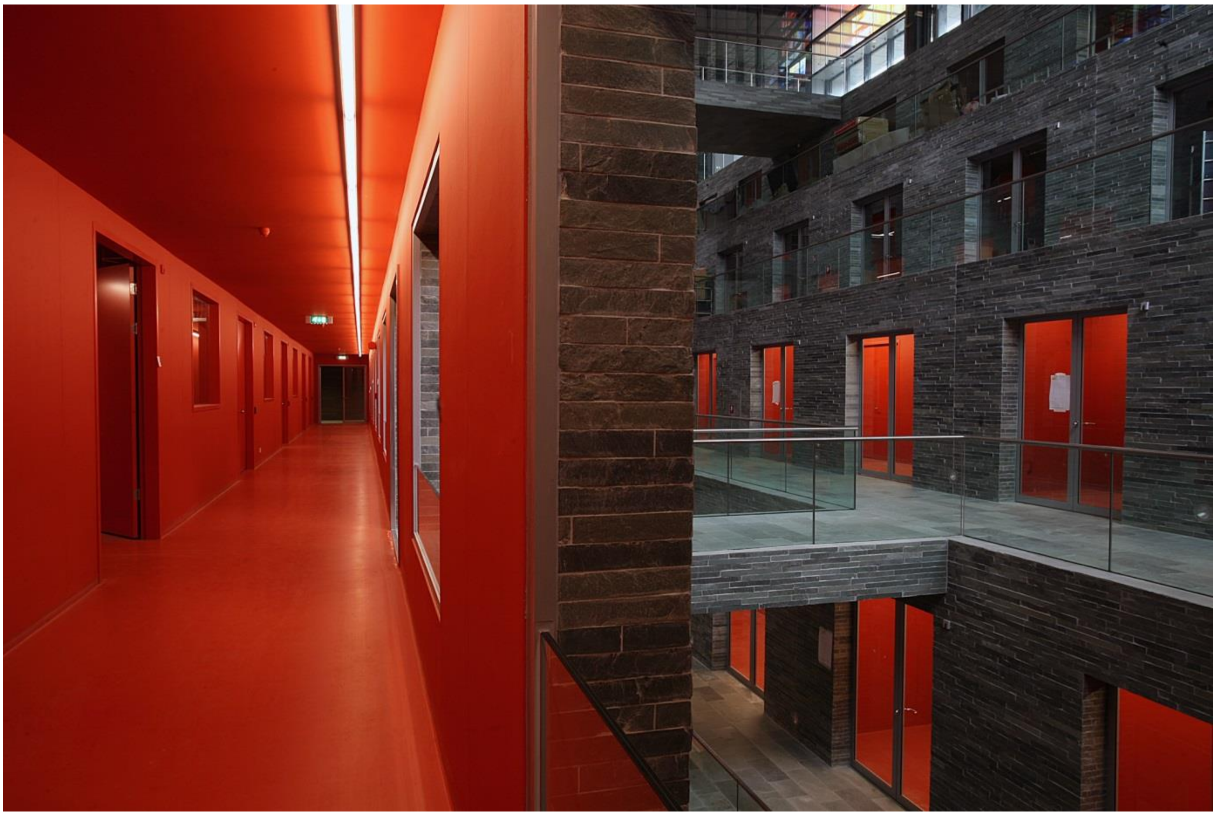

Fig. 1 *In Sound and Vision's vaults, approximately 6 petabytes of digital born and digitalized material is stored to date. How to keep it safe and accessible for the long term?*

Sound and Vision has taken this new challenge head on and has made it its most important goal for the coming period. Sound and Vision wants to become a "trustworthy digital repository" for Dutch audiovisual cultural heritage collections. Whether it is public radio or television programs, or other audiovisual material, it must be safely stored and made permanently available for whomever wants to use it. In 2012 this new strategic goal lead to the establishment of a project wherein the requirements for a trusted AV archive repository were to be defined. What was to be delivered was a set of normative policy documents that could guide the establishment of an OAIS compliant audiovisual archive environment. Its primary reference framework would be the "Quality Requirements Digital Archive Sound and Vision".

In addition to a set of policy documents outlining OAIS compliant organizational and administrative requirements, the central driver of the project was the development of Digital Object Management requirements.

This consisted of the development of a workflow model for the ingest, storage and dissemination of digital files and metadata. Data description would be enhanced through the development of a preservation metadata dictionary, composed of technical, administrative and provenance attributes. Together, the workflow and preservation metadata dictionary form the Information Model for Sound and Vision's digital archive. To date, all the project documents have been completed.

OAIS

The Information Model's workflow and metadata development was mostly inspired by the OAIS reference model. The ISO-standard OAIS[1] defines the functions that ensure "trusted" long-term preservation and access to digital information objects. Establishing a common language to describe the objects and processes, it offers a conceptual approach upon which actual business processes can be defined , allowing the archive to fulfill the preservation goals a TDR promises. Its central theme is the guarantee, offered to those whose material is deposited in such a repository, to long-term accessibility.

The model provides guidelines for defining and formalizing processes and data throughout the archive chain: from ingest, to storage to dissemination. By tracking and registering certain defined events in the lifecycle of each individual digital object in "preservation metadata", the authenticity of the ingested object can be demonstrated, and thus the basic requirement of 'trustworthiness' fulfilled. The Archive or repository is able to demonstrate its responsibility to its depositors as well as its users.

OAIS is a worldwide reference model that to date has seen wide take-up primarily in digital libraries and traditional archives. This can also be said for PREMIS[2], the most important preservation metadata standard. The applicability and implementation of this standard in the media archive domain, where the emphasis has always been on access and re-use – is still scarce. How to manage and preserve fast growing volumes of digital material in a rational and responsible way, is nevertheless a question that increasingly confronts broadcast and other large audiovisual collection holders.

The Information Model Digital Archive Sound and Vision 1.0 formulates Sound and Vision's first, normative answer to this question. In the model, OAIS and PREMIS based processes and metadata have been modified to fit the specific situations and needs of AV files managed in a dynamic production environment. Its development particularly reflects the domains in which Sound and Vision operates, as a production archive for Dutch public broadcasters (NPO), as a national AV cultural heritage repository and as an online service provider of AV content to various user groups.

[1] ISO 14721:2003. The Open Archival Information System Reference Model
[2] http://www.loc.gov/standards/premis/

Authenticity and integrity

To date, Sound and Vision could not provide guarantees concerning two essential aspects found in a TDR: the integrity and authenticity of the objects in the repository. Upon what do such guarantees lie? In general an OAIS compliant repository must be able *to demonstrate* the integrity and authenticity of digital files/objects held in the repository through some sort of audit trail. Such a trail is created over time and documents preservation events that have occurred throughout a defined preservation workflow.

For example, demonstrating the integrity of a file requires *documenting* that data streams have not been corrupted during transport or while in storage. Validation and error checking mechanisms allow the repository to track and verify that no corruption has occurred. Fixity information demonstrates file/object integrity over time. The degree of authenticity is judged on the basis of evidence.[3] Demonstrating the authenticity of a file/object over time requires *documenting* its provenance, that is the creation, chain of custody and change history[4] over time. Thus, two essential things are required: defined "preservation" business processes that ensure that "preservation" related events take place; and a mechanism by which an *audit trail* can be generated and maintained, allowing the archive to *demonstrate* the outcomes of these events. The TDR project set out to define these processes as well as a robust set of technical and preservation metadata, wherein essential lifecycle management information could be managed.

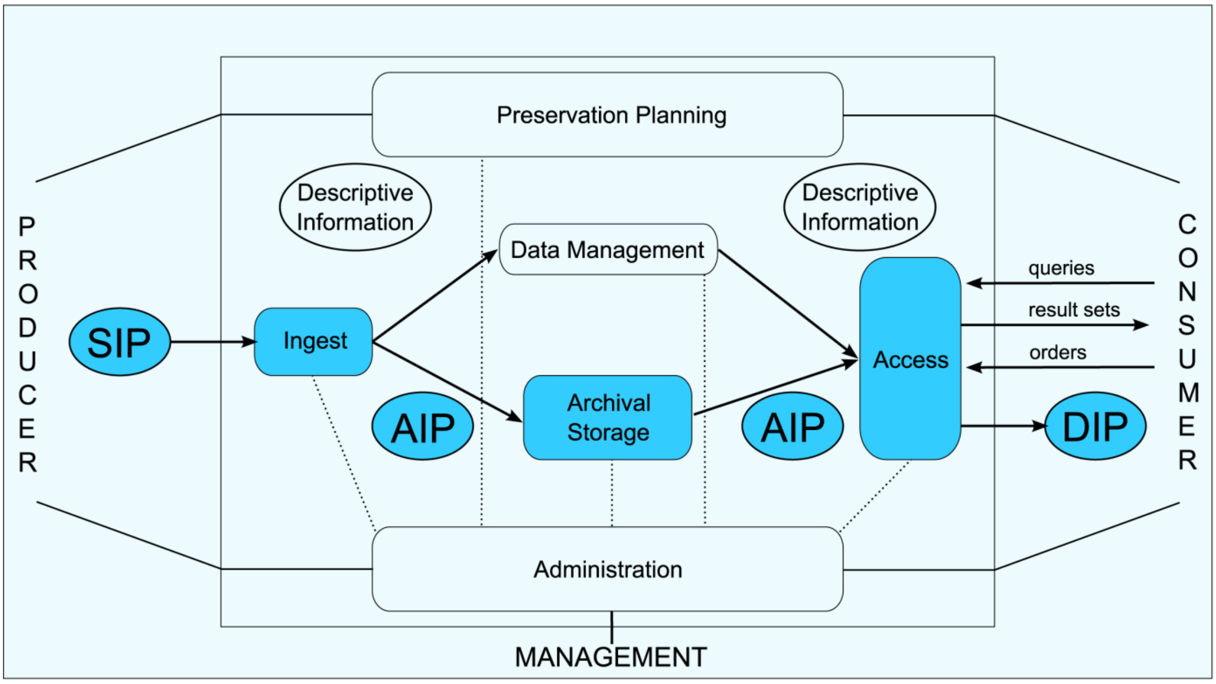

Fig.2 *OAIS Schema with marked workflow of the Information packages SIP, AIP and DIP : ingest, storage and access.*

[3] CCSDS, *Reference Model for an Open Archival Information System, Pink Book*, May 2009, pg.1-8
[4] Caplan, Priscilla. *Understanding Premis*. 2009, pg. 6

Information packages

Different types of digital objects exist in an OAIS compliant repository; the nature of these objects depend on where in the process the object exists. These “packages” are formed during defined workflow processes. The Producer/Donor submits files according to a Submission Agreement that consists of a set of terms between a producer and archive over the delivery and preservation of digital material. The files are packaged and sent to a staging area where a variety of auditing and metadata generation steps are carried out. All original files as well as all the additional files generated by the repository are then packaged and ingested into Archival Storage. Defined strategies, processes and required workflows lead to the creation of these manageable objects.

The OAIS defined digital archive is split up into different functional areas, each with associated processes. Three primary functional areas are ingest, storage and access. The digital objects and their associated metadata form an Information Package (IP). The IP is delivered by the producer and is called the Submission Information Package (SIP). The version of IP stored is the Archival Information Package (AIP). The package delivered to end-users is the Dissemination Information Package (DIP). The Information Model describes the distinction between ingest, storage and access workflows. It defines all the actions an object undergoes, thereby describing the entire lifecycle of an object, from ingest to dissemination. The object’s lifecycle history is documented in metadata, generated during certain defined events the object undergoes during its journey from ingest to dissemination. Metadata that documents what has occurred with the object over time makes up the “digital provenance” part of the preservation metadata schema.

The Information Model lays out when these “events” are to take place in the workflow and where it is documented in the preservation metadata. This digital provenance trail then provides the reference framework by which the actions an object has undergone can be verified for compliancy with an archive’s lifecycle management policy. By comparing the resulting provenance metadata with the defined workflow events, an archive can be assured that an object has not undergone any unexpected processing. In the end this provides the end-user with the needed authenticity guarantees.

Workflow schemas

The model describes not only the actions an object undergoes during various workflows, but the properties of the objects themselves as well, represented by two different schemas. The workflow schemas describe the object’s lifecycle from a linear perspective, while the object schema illustrates which metadata is present during different stages in the object’s lifecycle.

The different object types are defined as sets of essence, associated support files and metadata files. In the SIP phase, the accompanying metadata is limited to what the Producer has delivered with the file.

It contains a minimal set of descriptive and rights metadata and possibly additional information about the source of the file or other technical metadata. The AIP contains the most metadata of all three object types. It includes not only descriptive metadata but the complete set of preservation metadata, made up of technical information, provenance information, source metadata and rights. The provenance metadata documents all the events that have played a role in the lifecycle of the object. Generally, the DIP contains only a sub-set of the AIP metadata. If requested in the Submission or Order Agreement, all the preservation metadata can be included in the DIP.

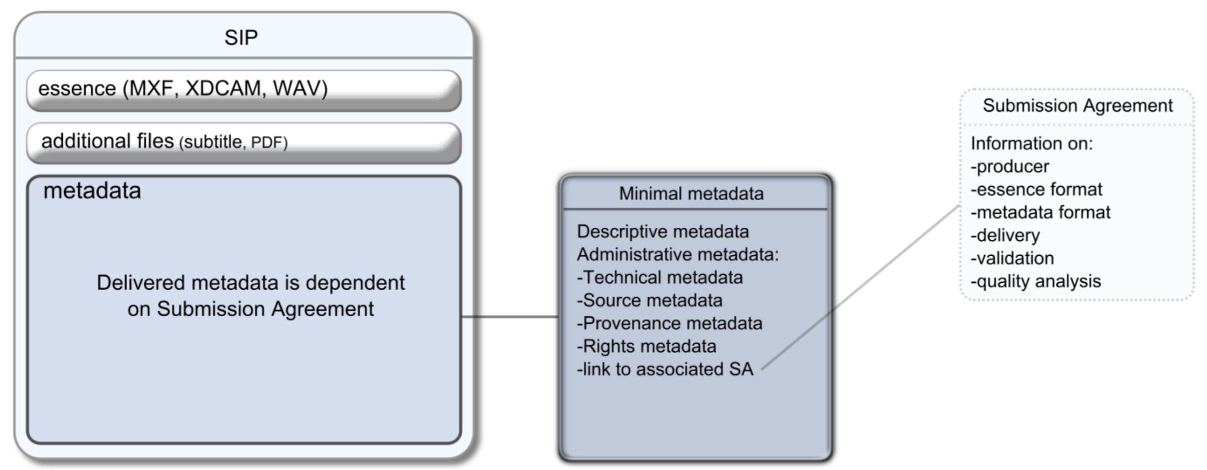

Fig. 3 *The contents of the Submission Information Package (SIP)*

A basic set of events to ensure the long-term preservation of the digital object can be found in various requirement and best practice documents. The first version of the Information Model focuses on workflows reflecting the type of audio and video material Sound and Vision provides to broadcast and cultural institutions

The first step in the preservation process actually occurs before ingest. This consists of the negotiation and documentation of terms between the owners of the material and the archive and results in the creation of a Submission Agreement. Such contractual documents explicitly define all the actions that are to take place during the archiving process and includes requirements such as formats to be delivered, rights to preserve the material, when and what kinds of validation the objects will undergo and how error reporting is to be handled. Upon ingest a virus check takes place in order to prevent any possible damage to the entire digital repository. The “fixity” check follows, which involves the comparison of two checksums in order to verify that the repository has received a file that has not been damaged during transport.

This check provides the Producer and the Archive with guarantees that the ‘integrity’ of the file is still intact upon arrival in the repository system. If Producers are not able to deliver checksums with their files, the Submission Agreement should spell out the risks involved: the Archive can only “fix” the state of the file after it is received into the repository.

Thus, the “integrity” guarantee begins from that moment on. The repository cannot be held accountable for any corruption they may have taken place during transport to the system.

The next step is file characterization, done by extracting technical metadata. Such characteristics include things such as aspect ratio, color space, codec identification and bit rate. The extraction of technical metadata provides the repository with a detailed picture of the technical nature of the files it stores. Only by exposing these characteristics can the repository identify potential risks associated with the files. Identifying a particular risk may lead for example, to the migration of a format. This characterization step also allows the archive to verify that the files submitted actually comply with those agreed to in the Submission Agreement

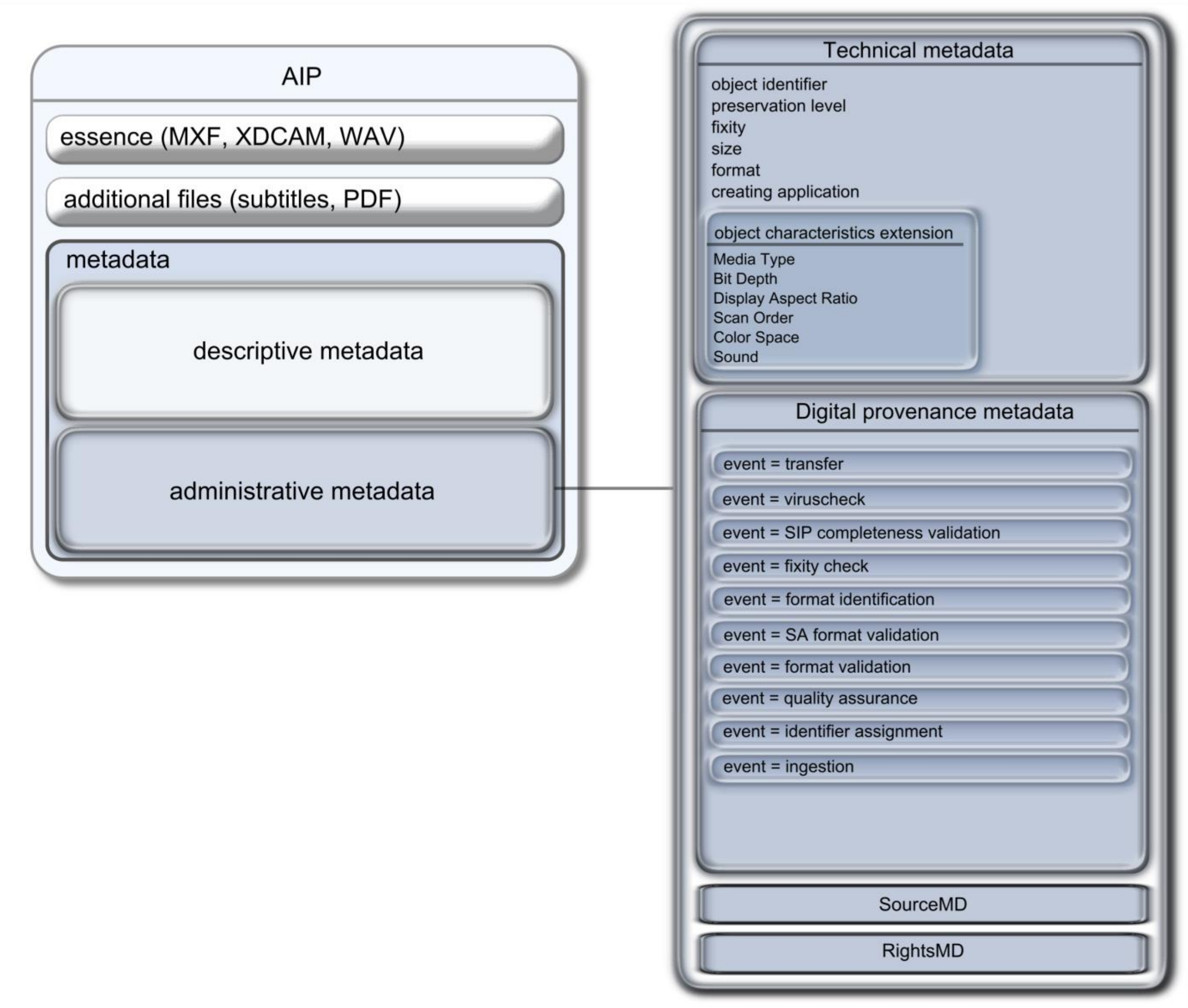

Fig 4. *The contents of the Archival Information Package (AIP).*

The workflow includes an optional “Quality Assurance” event whereby the essence is run through an automatic and/or manual visual/audio quality assessment process. Another possibility is the transcoding of a file format to what the archive itself considers its “archival” format, if agreed to with the Producer.

Finally, the object is assigned a persistent identifier and is then ready to be sent to storage. This ends the ingest phase and the AIP is ready for long-term storage. The object receives a definite location that is stored in its metadata.

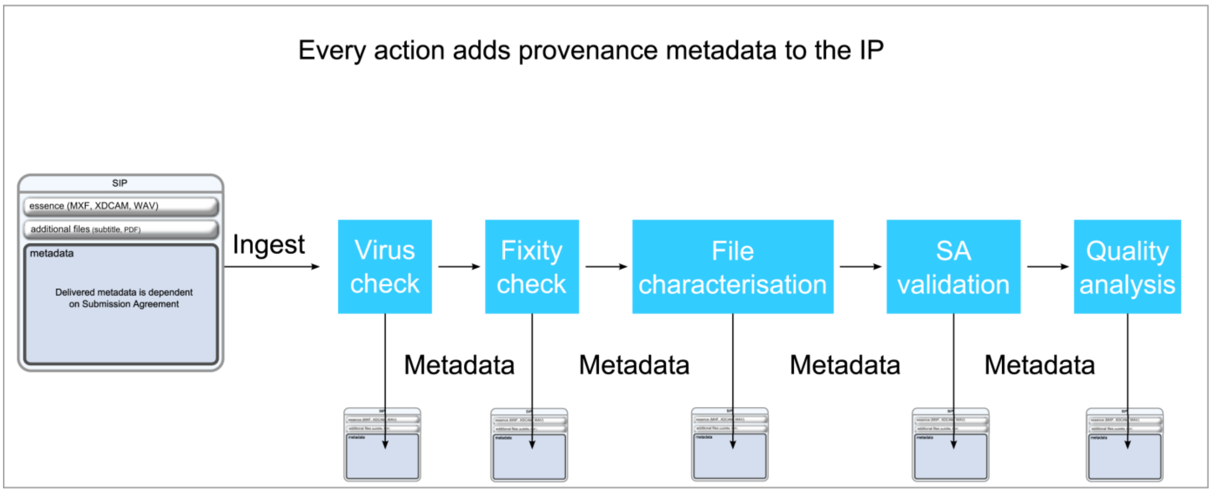

Fig. 5 *The INGEST workflow of the SIP and the various actions the package undergoes.*

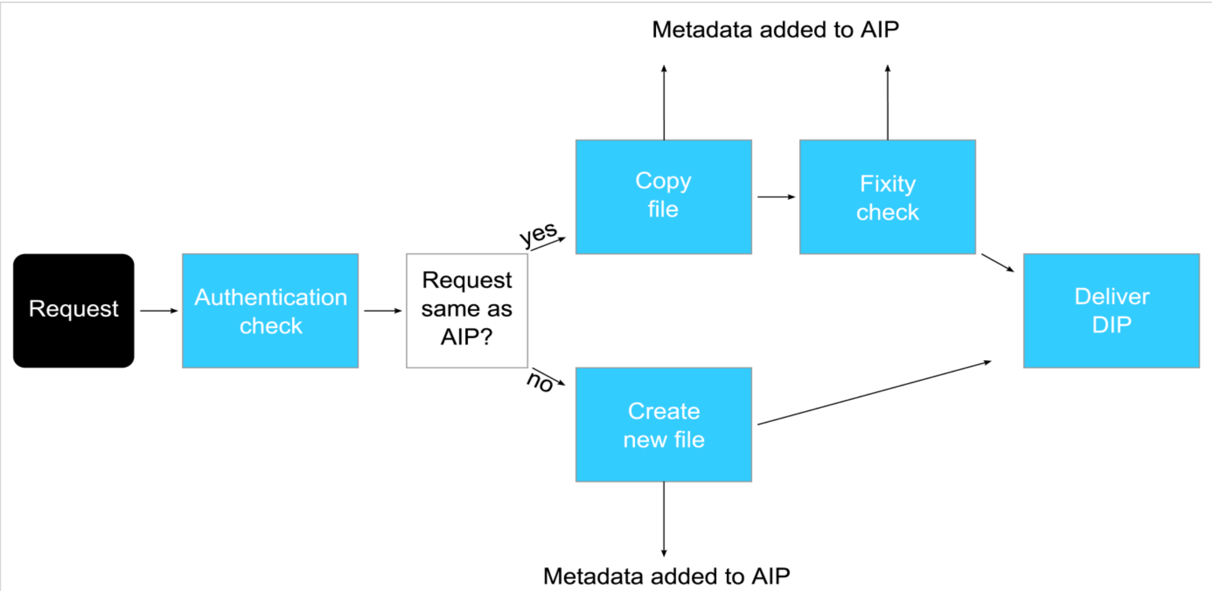

Fig. 6 *The ACCESS workflow of the DIP*.

The last section of the workflow describes how DIPs are created. The DIP workflow always begins by authenticating the user. This is followed by a search for certain types of material and a determination of the use case. When the system has determined that the material can be made available for this purpose, the DIP is generated. If the DIP is equal to the AIP, a checksum is generated and a fixity check performed to ensure the integrity of the DIP.

If a DIP is requested that requires transcoding to another format, or part of a file is requested (a partial-restore), a checksum is generated for this new version and delivered along with the DIP. This workflow is relatively generic in that a ‘request’ could include a request simply to search in the catalog as well as a ‘request’ for the delivery of an MXF file.

Developing a Preservation Metadata Dictionary

Although some technical metadata is present in various applications in Sound and Vision’s current IT architecture, it is not structurally defined or managed as a component of a preservation system. The task was to develop a metadata dictionary that includes both essential technical characteristics of audiovisual files as well as preservation metadata, which focuses primarily on digital provenance: that is, metadata documenting the creation, chain of custody and change history over time. Rights related metadata, strictly referring to the rights to preserve, also form an essential element.

For the technical metadata, a study was made of a variety of AV specific metadata schemas (PBCore, EBUcore, AES, LC VideoMd, AudioMD and NARA’s reVTMD). Metadata schemas developed by digital repositories operating primarily in the academic library domain (Rutgers University and University of Texas) were also studied primarily because of the access they provide to well-defined dictionaries based on existing standards. And, they offer examples of actual, implemented metadata systems in institutions that consider audiovisual collections an essential part of their digital preservation responsibility. PREMIS was chosen as the standard to use for the digital provenance and (preservation related) rights metadata[5].

The dictionary consists of technical metadata for audio, video and still images. It also includes PREMIS digital provenance attributes and a complete defined set of ‘events’ to be carried out as defined in the Information Model workflow. The biggest challenge in compiling the dictionary was trying to reconcile a set of diverse metadata schemas.

The schemas differed not only in their definitions of which attributes were relevant at what level (file/stream), they also had different attribute names and slight differences in definitions that made comparison difficult. In addition, differing opinions in the field as to which attributes were essential and which were ‘nice to have’ added complexity to the decision making process.

Ultimately a balance had to be found between including every possible characteristic and including what was considered essential at this point in time, with the underlying assumption that future versions of the dictionary may contain additional attributes.

The current dictionary does not yet include extensive metadata that reflects documentation of re-formatting related processes, that is, for files created as a result of a digitization project.

Although PREMIS has a place for ‘creating application’ and some creation related information, it does not include the full level of reformatting metadata that NARA’s reVTMD

[5] PREMIS was also chosen as the standard in the PrestoPRIME project whose focus was on the preservation of digital audiovisual collections. www.prestoprime.eu

offers (in fact that is why NARA created the schema); or that offered by the AES 57-2011 schema.

Attribute of	Name	Definition	Value type Text/Numeric/Date/Binary	Obligation	Repeatable	Data Constraint	Values allowed, or link to CV-list.
Moving Image Audio	duration	The elapsed time of the entire item or track in playback	Text	M	NR	Structured form.	
Moving Image Audio	dataRate	Also known as bit rate; the rate at which data is presented within the codec. Data rate of the compressed data over time expressed in bytes per second.	Numeric	M	NR	None	
Moving Image Audio	dataRateMode	Indicates whether the stream data has been processed to achieve a fixed (constant) or variable bit rate.	Binary	M	NR	CV	Allowed values (LC): Fixed; Variable.
Moving Image	timecodeInitialValue	Starting value for timecode.	Text	M	NR	Structured form.	
Moving Image	timecodeRecordMethod	Method for recording timecode on the video source item					See also: http://rucore.libraries.rutgers.edu/open/projects/openmic/index.php?sec=guides&sub=metadata&pg=t_time-code
Moving Image	timecodeRecordType	Type of timecode recorded on video source item, e.g., SMPTE dropframe, SMPTE nondropframe, etc..	Text	M	NR	CV	longitudinal (LTC); vertical interval (VITC); Other

Fig.7 *Excerpt from Sound and Vision's Preservation Metadata Dictionary V1.0 : examples of technical metadata for AV files.*

These schemas offer much more granularity and detail in areas such as what transfer machine was used, calibration, needles employed in disc transfers, etc.

This extensive digital provenance information, along with a clear link to the technical characteristics of the original analogue source, are considered by some experts to be essential in ensuring that the resulting digital file can be considered authentic by users.[6]

Finally, a more extensive set of technical metadata for analogue carriers needs development and inclusion in the dictionary as well. Here again, both reVTMD and AES 57-2011 offer solutions. In fact, a metadata expert at Rutgers makes a good case to extend the AES 57-2011 standard to cover all multi-media carriers and is working on such an implementation at Rutgers.[7]

[6] Otto, Jane Johnson(2010) 'A Sound Strategy for Preservation: Adapting Audio Engineering Society Technical Metadata for Use in Multimedia Repositories', Cataloging & Classification Quarterly, 48: 5, 403 — 422
[7] Ibid.

Attribute of	Name	Definition	Value type Text/Numeric/Date/Binary	Obligation	Repeatable	Data Constraint	Values allowed, or link to CV-list.
Moving Image Audio Photo	**1.3.1 preservationLevelValue**	A value indicating the set of preservation functions expected to be applied to the object.	Text	M	NR	CV	Example values for R: bit-level; full; 0; 1; 2. For File: bit-level; full; 0; fully supported with future migrations.
Moving Image Audio Photo	**1.3.2 preservationLevelRole**	A value indicating the context in which a set of preservation options is applicable. Repositories may assign preservationLevelValues in different contexts which must be differentiated, and may need to record more than one context	Text	O	NR	CV	Example values: requirement; intention; capability.
Moving Image Audio Photo	**1.3.3 preservationLevelRationale**	The reason a particular preservationLevelValue was applied to the object.	Text	O	R	None	
Moving Image Audio Photo	**1.3.4 preservationLevelDateAssigned**	The date, or date and time, when a particular preservationLevelValue was assigned to the object.	Date	O	NR	Structured form.	
Moving Image Audio Photo	**1.5.2 Fixity**	Container element for holding information used to verify whether an object has been altered in an undocumented or unauthorized way.	Text	O	R	Container	
Moving Image Audio Photo	**1.5.2.1 messageDigestAlgorithm**	The specific algorithm used to construct the message digest for the digital object.	Text	M	NR	CV	Example values for F: MD5; Adler-32; HAVAL; SHA-1; SHA-256; SHA-384; SHA-512; TIGER; WHIRLPOOL.
Moving Image Audio Photo	**1.5.2.2 messageDigest**	The output of the message digest algorithm.	Text	M	NR	None	
Moving Image Audio Photo	**1.5.2.3 messageDigestOriginator**	The agent that created the original message digest that is compared in a fixity check.	Text	O	NR	None	
Moving Image Audio Photo	**1.5.3 Size**	The size in bytes of the file or bitstream stored in the repository. to indicate the storage requirements or file size of a digital media item. As a standard, express the file size in bytes. - use attribute 'unit of measurement'	Numeric	O	NR	Integer	

Fig.8 *Excerpt from Sound and Vision's Preservation Metadata Dictionary V1.0 with examples of preservation metadata based on PREMIS.*

Important Findings

1. *Synchronizing Theory with Practice*

The OAIS model is an abstract one; it needs precise definition when applying it to a specific archive environment. It is not an "all or nothing" concept. Its requirements are actually high level principles and provide plenty of room for interpretation. This makes the model on the one hand, a nuanced and flexible one. Yet on the other hand, it does not provide ready-made implementation solutions. In the end, what is desirable and can actually be realized in a running archive will determine what specific solutions are chosen.

At Sound and Vision, the TDR project was running at the same time that the process to acquire a new MAM-system began. In order to enable that a new MAM would contribute to OAIS compliancy goals, steps in the Information Model workflows were quickly translated into detailed and concrete MAM requirements. This meant that some requirements outside of the IP workflow needed to be analyzed. It was at this point that it became clear that OAIS compliancy requirements themselves offer little concrete or technical solutions. It was not always clear if requirements were to be met by way of workflow implementations or through technical requirements on an application level. Each workflow requirement was then identified as one to be met by either a MAM application or elsewhere.

In the end, those OAIS compliant lifecycle management requirements that became part of the overall MAM requirements actually represents a sub-set of the overall OAIS requirements.

2. *Preservation from an Enterprise IT Architecture Viewpoint*

The need to identify which OAIS requirements were relevant in a MAM requirements list clearly illustrated that not all preservation requirements could be met by one application. It is recommended that different requirements be fulfilled by different components within the total Enterprise IT architecture. To realize this, much more insight is needed into the differences as well as the underlying relationship between the types of (meta)data, the different systems, the workflows and the functions of the Enterprise IT architecture, in relation to OAIS compliant digital life cycle management. For Sound and Vision there's still a lot of work to do.

When moving from the analogue preservation world to the digital IT domain it helps to consider preservation from an enterprise IT architecture perspective. Such a perspective requires a clear understanding of the difference between general IT processes and those processes associated with the "business" of preservation. Confusion may stem from the fact that there is overlap in some OAIS functional areas (take the functional OAIS area Administration for example, where issues about security and risk are covered, issues that form a part of regular IT management processes.) IT processes for example, focus on whether specific applications within the IT infrastructure are working properly - is it secure, are there errors, is it managing capacity properly? An Enterprise IT architecture viewpoint focuses on *the lifecycle management of the data/content:* the business activities that need to be carried out on the data; in the TDR context, the business processes that ensure the long-term preservation of the object. One illustration of this is mistakenly equating 'storage' with preservation.

In fact, storage is simply one of a set of IT applications within an Enterprise IT architecture system. It is concerned with such things as "writing/recording the data, retaining the data for an undetermined length of time, reading and accessing the data. More advanced storage may include replication of files through techniques such as RAID, erasure coding, or the simple generation of multiple copies on different media or at different sites (geo replication), or the optimization of data placement, e.g. in hierarchical storage management (HSM) systems. A preservation *system* on the other hand involves processes that need to take place during ingest and access *to content* to support preservation; it includes carrying out preservation actions performed on content. Preservation is a process that includes the storage, but is not limited to it."[8]

From an enterprise IT architecture viewpoint, a preservation strategy must be defined and translated into business processes that are then associated with IT functional areas. Applications are then found to ensure that such functional processes are carried out. The TDR project set out to define these strategies and processes.

[8] Stephen C. Phillips. PrestoPRIME D2.3.1. Service Oriented Models for Audiovisual Content Storage, pg. 10.

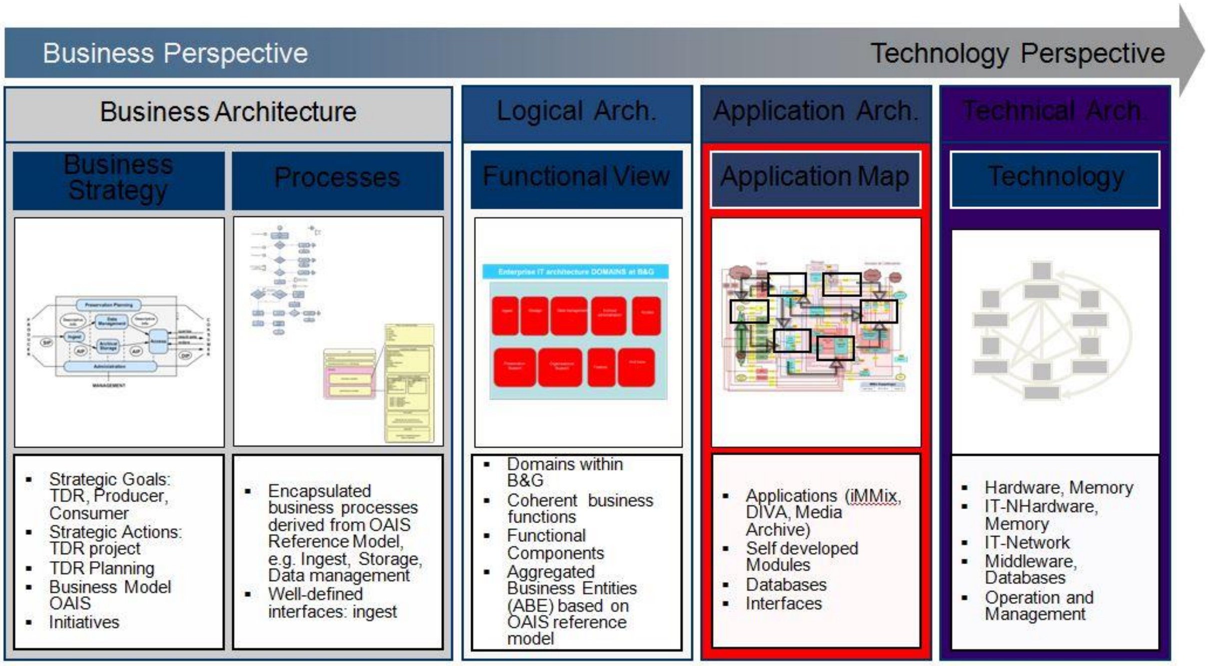

Fig. 9 *Distinctive levels within the Enterprise Architecture of Sound and Vision and their corresponding functions and domains.*

3. *Learning, communication and collaboration*

It was clear at the start of the TDR project that digital preservation processes are not confined to what the IT department manages; implementing OAIS compliant lifecycle management involves the entire archive institution: at ingest, during storage, from description to dissemination. This means that the entire organization play a role in providing long-term preservation and access guarantees: the acquisition and selection staff, the catalogers as well as the access services staff. OAIS compliancy makes more explicit the important relationship the archive has with its producers/donors as well with its Designated Communities: all incoming and dissemination activities play an integral role in the archival lifecycle of the material.

Therefore, one of the most important goals of the TDR project was to try and involve the entire organization in digital lifecycle management decision making and to share knowledge about digital preservation as broadly as possible. A project structure was chosen as the best approach, even though the work primarily focused on policy development. It was thought that by involving different staff members (catalogue, system application and access services staff as well as policy advisors) everyone would better understand their own role which would in turn, strengthen the responsibility they feel in the digital lifecycle management process. This goal was only partially achieved. It was too difficult for most project members, alongside their daily work responsibilities, to spend time researching and translating theoretical processes into archive policy. In the end, the work was carried out by Sound and Vision's IT Information Policy division, who had originally been assigned this task.

The project approach did succeed in raising the level of consciousness about different aspects of digital preservation and digital lifecycle management: what these concepts mean and how important it is to collection management. The OAIS standard, its terminology and its process framework are no longer foreign to the organization. In addition, certain technical processes such as validation and fixity checking have gained prominence.

It's important that the knowledge gained during the OAIS requirement trajectory remains the reference framework within Sound and Vision's IT department, who in the meantime, needs to start finding solutions to support the newly developed business processes. It is hoped that the defined lifecycle management requirements demonstrate to IT staff that preservation processes and workflows first need mapping to functional areas in the Enterprise IT architecture system; only then can applications be identified that may be able to carry out the needed functionality.

To Conclude

There are many essential questions and issues that remain after having defined the OAIS compliant audiovisual archive requirements. Besides working out where within the Enterprise IT architecture the preservation functions need to land, budget and cost modeling needs to start. What are the cost implications of this workflow and information scenario? Who will pay for which preservation service within the digital lifecycle management process? The archive itself as "Trusted Repository"? The producer/donor as deliverer of digital collections or the Designated Communities for whom the files need to be kept technically up to date forever?

Another burning question concerns all the previously ingested files and metadata: how, for example, to ensure that what is now essentially 'dark metadata', generated during earlier digitization processes, is brought into the preservation workflow and data management system? What do you do with all the files that were ingested earlier, into a non-OAIS compliant repository, and thus never underwent the fixity checking and validation processes an OAIS compliant repository requires, especially given the nature of the amount – 400.000 hours?

Further, the preservation business processes themselves need further development. Are we going to apply the normative workflow and metadata processes to *all* the collections that are ingested or are we going to apply them in only some cases? If so, how: are we going to define different preservation levels for broadcast production material and cultural heritage material considered to be of national importance? And are we going to apply to *all* the different types of content files (metadata, photos, contextual documents, etc) the same preservation level or will this only apply to Sound and Vision's core collection?

Value	Fixity Check		
Definition	The process of verifying that an object has not been changed in a given period.		Note: This is the comparison of two message digest calculations.
Semantic unit	Semantic component	Sample value(s)	Notes
eventIdentifier	eventIdentifierType	NIBGPS0.3	domain within which id is unique
eventIdentifier	eventIdentifierValue	E634721158	
eventType	none	fixity check	
eventDateTime	none	2013-01-16T020:20:32+01:00	
eventDetail	none		
eventOutcomeInformation	eventOutcome	{pass;fail}	
eventOutcomeDetail	eventOutcomeDetailNote		link to report if failed
linkingAgentIdentifier	linkingAgentIdentifierType	NIBGPS 0.3	domain within which id is unique
linkingAgentIdentifier	linkingAgentIdentifierValue	A282179	id of agent entity which gives name, version and type of agent performing fixity check; agentName="MD5Deep", version=3.6 ; agentType=sw program
linkingAgentRole	none	Executing Program	
linkingObjectIdentifier	linkingObjectIdentifierType	NIBGPS0.3	domain within which id is unique
linkingObjectIdentifier	linkingObjectIdentifierValue	549-274io0	id of file being verified

Fig. 10 *A burning question*: *will ALL of the materials of Sound and Vision that were ingested in previous years, still have to undergo the full process of OAIS-compliant fixity checking?*

All of these questions must eventually be fully answered and translated into policy to ensure a rational and cost effective business operation. However, the Information Model and the other normative policy documents produced in this project provide an important reference framework against which the AV archive can measure how far its current operations consciously reflect preservation lifecycle management, in the way it approaches the ingest, storage and dissemination of the collections, whether broadcast production or cultural heritage material. There lies a solid, theoretical basis for establishing a technical and organizational preservation structure in the audiovisual broadcast domain that can be considered OAIS compliant. All those who play a role in the preservation process have been identified. Sound and Vision now knows how to operate according to the standards and can thus – by implementing the systems – demonstrate how the standards have been implemented and why. All this allows the organization to prove to its producers/donors and users, how, by having implemented the basic requirements of a 'trusted' repository, it operates responsibly.

Based on the project's achievements, Sound and Vision plans to seek a sort of certification for digital archives, the so-called Data Seal of Approval[9]. The normative documents that have been delivered can serve as sound, documentary proof that some of the DSA requirements have been fulfilled. Thus the first step towards the status Trustworthy Digital Repository has been realized.

References

- Kwaliteitseisen Digitaal Archief Beeld en Geluid 1.0
- Informationmodel Digitaal Archief Beeld en Geluid 1.0
- Preservation Metadata Dictionary Beeld en Geluid 1.0
- Designated Communities Beeld en Geluid, typering en uitlevereisen, 1.0
- Handleiding Submission Agreements, 1.0

[9] http://datasealofapproval.org/

About the authors

Annemieke de Jong (r) is sr. Policy Advisor Digital Preservation at the Netherlands Institute for Sound and Vision. She is responsible for setting out the Institute's strategic policies for digital AV-collection management and AV-preservation. De Jong is also involved in the development and international dissemination of knowledge and expertise on digital AV-archiving and has several publications to her name. Annemieke de Jong is a long standing member of the Media Management Commission of FIAT/IFTA . She has a master's degree in New Media and Digital Culture and Information Management.

Beth Delaney (l) has been defining collection management system needs, implementing metadata standards and developing policy in American and European audiovisual archives for 25 years. She is experienced in both analog and digital collection management, in broadcast as well as cultural heritage institutions. As a consultant, she specializes in digital preservation processes, preservation related business requirements and preservation metadata standards in the AV-domain. Delaney has a master's in Library and Information Science with a specialization in archives.

Daniel Steinmeier is a technical specialist in the application management department at Sound and Vision. Up till now he has been involved in making AV-material accessible for educational purposes via various web platforms. Metadata has always played an important role in his work and he has worked for years with metadata standards such as IEEE-LOM and protocols for metadata exchange such as OAI-PMH. Steinmeier has a master's degree in Language and Culture studies with a specialisation in media.

De Jong, Delaney and Steinmeier together form the core of **Sound and Vision's Information Management Team** that to date has been defining the requirements for OAIS-compliant business processes and models for Sound and Vision, including the central quality criteria for Sound and Vision's digital archive , the Information Model, the Preservation Metadata Dictionary, the Guidelines for Submission Agreements, the definition of the Institute's Designated Communities and various policy documents on preservation planning and digital storage.