

# DIGITAL PRESERVATION

SOUND AND VISION

*policy, standards and procedures*

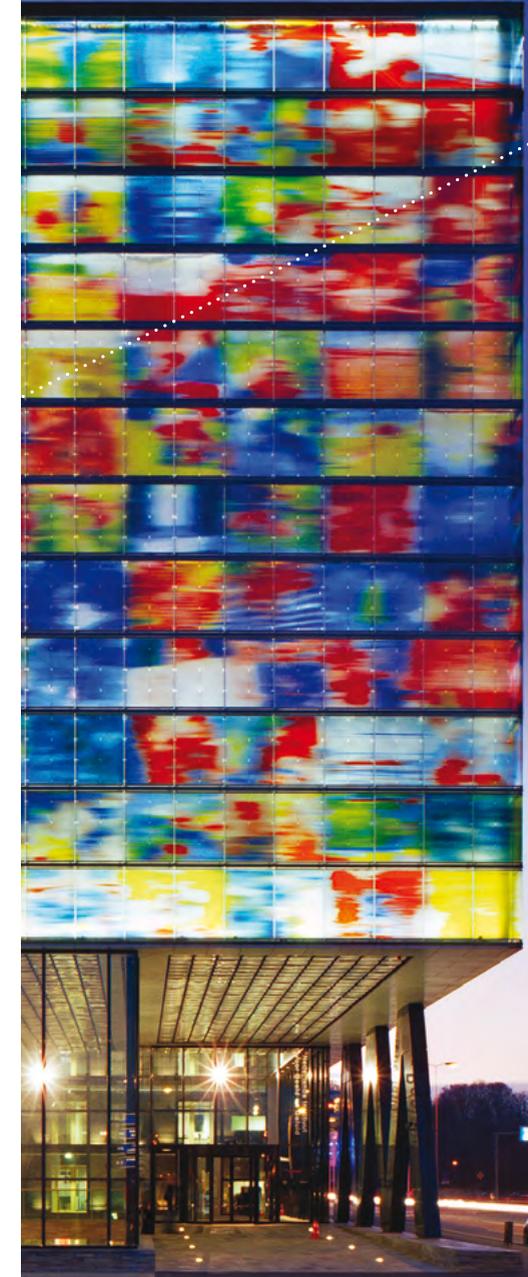


January 2016

Author: Annemieke de Jong, Digital Preservation Officer, Archives  
Contributions from: Daniel Steinmeier, ICT Dptmnt, Archives  
Photography: Aad van der Valk, Daria Scagliola en Stijn Brakkee  
Graphic design: Tice Grafisch Ontwerp, Amsterdam

# CONTENTS

|  |    |  |  |
|--|----|--|--|
| <i>Management summary</i>                            | 6  |  |  |
| <b>1. Introduction</b>                               | 8  |  |  |
| <b>2. Mission and task outline</b>                   | 12 |  |  |
| <b>3. Strategic policy framework</b>                 | 14 |  |  |
| 3.1 Multiyear policy                                 | 15 |  |  |
| 3.2 Collection policy                                | 15 |  |  |
| 3.2.1 Collection description (typology)              | 15 |  |  |
| 3.3 Agreements on access                             | 16 |  |  |
| <b>4. Preservation principles</b>                    | 18 |  |  |
| 4.1 Core definitions                                 | 19 |  |  |
| 4.2 Scope  | 19 |  |  |
| 4.3 Preservable formats                              | 19 |  |  |
| 4.3.1 Preservation master, mezzanine and proxy files | 20 |  |  |
| 4.4 Non-preservable formats                          | 21 |  |  |
| <b>5. Preservation strategy</b>                      | 24 |  |  |
| 5.1 Migration  | 25 |  |  |
| 5.2 Guaranteeing integrity                           | 26 |  |  |
| 5.3 Guaranteeing authenticity                        | 27 |  |  |
| <b>6. The preservation workflow</b>                  | 30 |  |  |
| 6.1 Information model                                | 31 |  |  |
| 6.2 Generic workflow stages                          | 31 |  |  |
| 6.3 Process details and technical metadata           | 32 |  |  |
| <b>7. Preservation levels</b>                        | 34 |  |  |
| 7.1 Full preservation and bit preservation           | 35 |  |  |
| 7.2 Preservation menu                                | 35 |  |  |
| 7.3 Guarantees and checks                            | 37 |  |  |
| 7.4 Preservation actions                             | 37 |  |  |
| <b>8. Preservation planning</b>                      | 38 |  |  |
| 8.1 Preservation Watch                               | 39 |  |  |
| 8.2 Organization changes                             | 39 |  |  |
| 8.3 Designated Communities                           | 39 |  |  |
| 8.4 Technological developments                       | 40 |  |  |
| <b>9. Standards and formats</b>                      | 42 |  |  |
| 9.1. Digital collection management                   | 43 |  |  |
| 9.2 Metadata formats                                 | 43 |  |  |
| 9.2.1 Descriptive metadata                           | 43 |  |  |
| 9.2.2 Thesaurus                                      | 44 |  |  |
| 9.2.3 Preservation metadata                          | 44 |  |  |
| 9.3 File formats                                     | 45 |  |  |
| 9.3.1 Video  | 45 |  |  |
| 9.3.2 Audio  | 45 |  |  |
| 9.3.3 Film   | 45 |  |  |
| 9.3.4 Text, subtitles and photographs                | 46 |  |  |
| 9.3.5 Websites and Internet video                    | 46 |  |  |
| 9.3.6 Computer games                                 | 46 |  |  |
| <b>10. Storage policy</b>                            | 48 |  |  |
| 10.1 Backup and recovery                             | 49 |  |  |
| 10.2 Migration                                       | 49 |  |  |
| 10.3 Retention                                       | 50 |  |  |
| <b>11. Technical infrastructure</b>                  | 52 |  |  |
| 11.1 Maintenance and security                        | 53 |  |  |
| 11.2 Continuity                                      | 54 |  |  |
| <b>12. Training and knowledge development</b>        | 56 |  |  |
| <b>13. Risk management</b>                           | 60 |  |  |



The purpose of this document is to contribute to the development and promotion of the Netherlands Institute for Sound and Vision (*Nederlands Instituut voor Beeld en Geluid*) as a leading audio-visual archive that has identified sustainable digital preservation as its primary *business*. With that in mind, this document explicitly outlines all principles and choices that form the basis for execution of this business. The digital objects and their lifecycle are defined, services and guarantees are described in detail and a record is made of how Sound and Vision (*Beeld en Geluid*) complies with technical and staff quality requirements. By documenting the current policy and the standards employed, it is possible to account to all parties that entrust their digital collections to Sound and Vision, to the users of those collections, and to subsidy-awarding bodies. The document also offers the staff of Sound and Vision transparency and clarity on the rules and procedures that apply.

Laying down this policy in its entirety offers an insight into the many implications of digital preservation. It becomes clear which business components are involved, which activities are necessary and which requirements apply, at all levels of the operation. It also becomes clear how the various tasks, processes and procedures relate to one another. An integrated preservation policy lays the foundations for ensuring governance of the Digital Archive in the context within which it operates: it provides an insight into the degree of management, control and standardization required for the orderly ingestion of large digital volumes from multiple sources, for the secure and reliable storage of materials and ensuring their access for users now, and in the future. Management of the preservation processes on the basis of a well defined policy not only promotes the effectiveness of those processes, but also helps manage both costs of the digital service provision and all the risks run by the Digital Archive at organizational, financial and technological level. In other words, the document will make a major contribution to rationalizing the operation

Over the coming policy period (2016-2020), the aim is to acquire three TDR (Trustworthy Digital Repository) certificates: the Data Seal of Approval, a DIN 31664 certificate and a DSA-WDS Certification. To ensure a solid technical infrastructure, Sound and Vision also aims to obtain ISO certification in the 27001 series (Information Security), 9000/ITIL (Quality Management), ISO 22399 (Incident preparedness and operational continuity management) and ISO 31000 (ICT Risk Management).

The link between these certification programmes and this preservation policy works on both sides: laying down policy is an important certification requirement in all cases, while the individual requirements will in turn be used to further improve and implement the policy.

# INTRODUCTION

# 1

The domain of digital preservation extends to include the complete lifecycle of digital objects, in other words all processes relating to the ingestion, storage, management and providing access to the object in question. A scope of this kind calls for the establishment of an environment in which the processes are able to take place in a controlled manner, in relation to one another. The result is a well-organized Digital Archive.

The requirements imposed on such an environment are to a considerable degree met through the standardization, formalization and documentation of both the (meta)data objects and the workflows within which they are managed. But digital preservation does not stop with the storage of and granting of access in formalized processes. It also relates to the contacts with the outside world. Agreements must be reached with the depositors of collections and procedures laid down. Which materials can be accepted and how do the depositors want that material to be preserved by Sound and Vision? There also user groups for the Digital Archive. How do these parties - each in their own specific case - wish to be granted access to the collections, and in what form should the materials consequently be stored by Sound and Vision? The conditions and guarantees in the provision of these services from the Digital Archive are all defined in the preservation policy.

The processes in and around the Digital Archive take place in a complex network of hardware, software and links with the outside world. Digital preservation is clearly closely related to the ICT environment. The state of affairs in respect of storage facilities and the protection of the technical infrastructure are integral to the preservation policy. The policy also includes securing the necessary knowledge and the competences of the staff working in such a technological environment.

A reliable preservation environment is capable of withstanding all possible threats from inside or outside the organization. The (financial) continuity, sustainability of the formats and the quality of the way in which the collections are made accessible may not be placed at risk. Preservation policy

therefore also provides mechanisms for managing the various risks run by the Digital Archive.

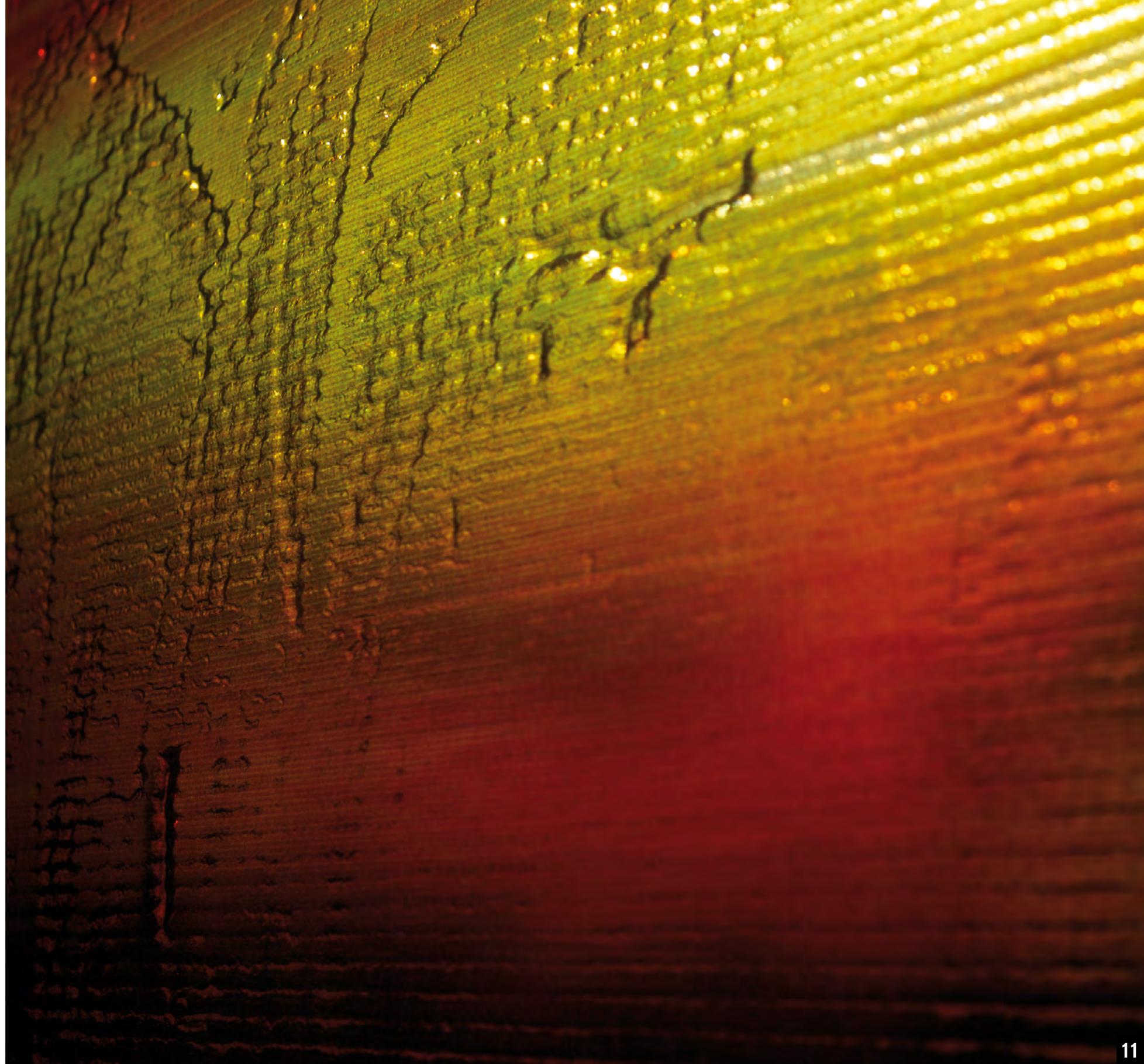
This document lays down how all these elements of digital preservation have been structured at Sound and Vision. The policy, the standards and the procedures described in this policy document tie in with the requirements for a *Trustworthy Digital Repository* as laid down in international standards, the most important of which is OAIS. On the basis of the *why* (with what aim), the *what* (which materials) and the *for whom* (the user groups), this document discusses the *how* of sustainable preservation of the collections. In this way, the document offers a framework within which the various aspects of preservation are placed in a logical, coherent context. The resultant policy and the methods are described for each element within this framework, with reference to the accompanying documentation. The documentation may be a standard-related policy document, a users' manual, a model, an elaborated procedure or a standard itself.

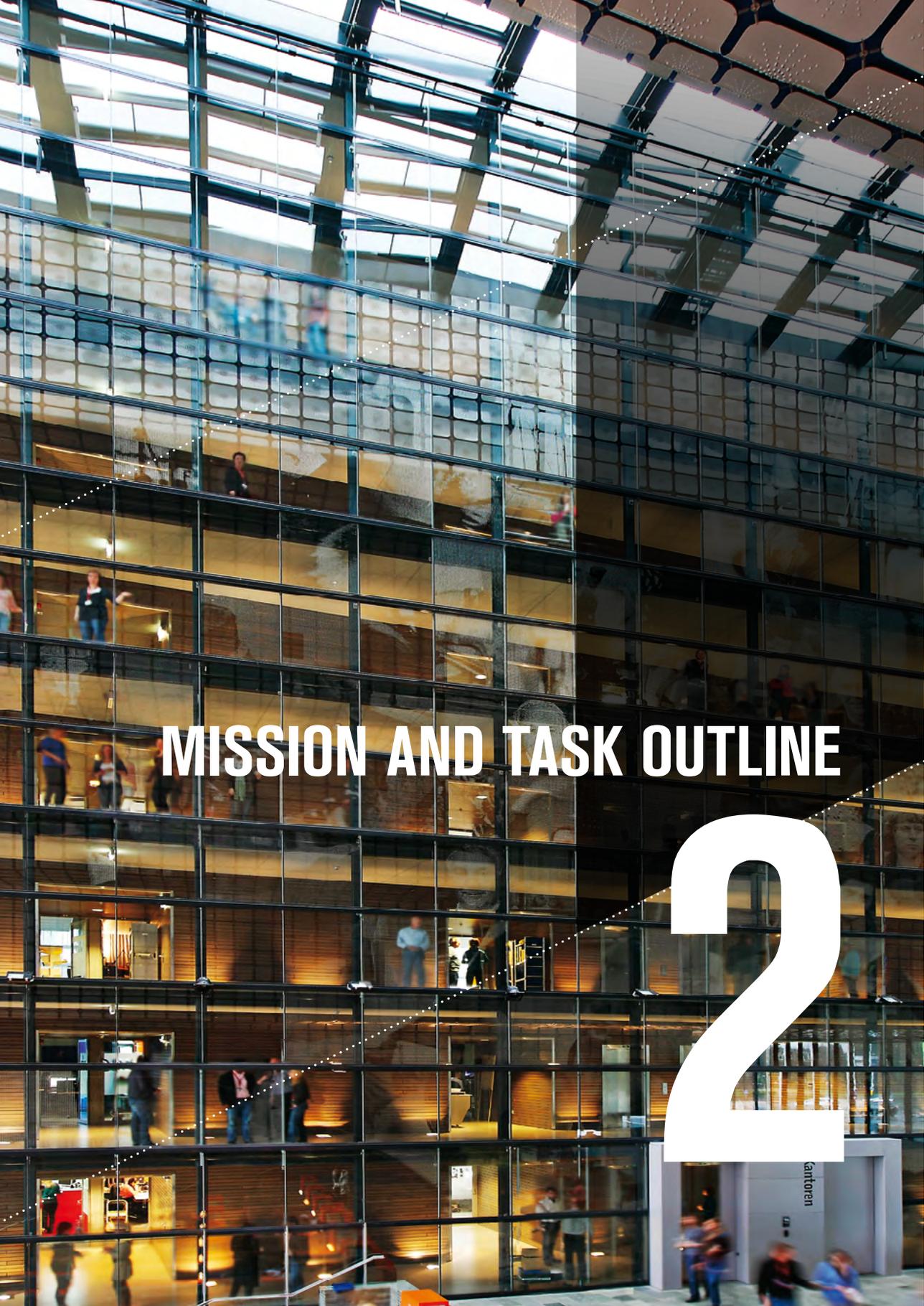
The document is structured as follows: following an outline of the mission and the strategic policy for the collection and the procedures for providing access to the collection, the preservation assignment of the institute is elaborated. This elaboration takes on its fullest form in the core of the document: the explanation of the key concepts from the domain of digital preservation, and the way in which Sound and Vision has translated these operating principles into its own preservation strategy, the format choices, the internal workflows, the preservation services and the planning mechanisms. The various forms of temporary and sustainable preservation of the materials are described, as are the conditions and guarantees imposed by Sound and Vision on those processes.

This is followed by a summary of all standards and file formats and metadata as employed in the Digital Archive. The storage policy and preservation of the technical infrastructure are then clarified in the form of a reproduction of the procedures for making backups, disaster recovery, migration and security. A separate chapter discusses the

type of knowledge required for working in a digital audio-visual preservation environment, and the way in which Sound and Vision ensures that that knowledge is developed among its staff. The document ends with a description of the policy initiatives taken in developing mechanisms for risk management.

This document represents the situation in December 2015. Work on the construction of the controlled preservation environment is anything but complete. Constant efforts are made to develop and improve the environment, both in terms of policy and knowledge, and in respect of actual implementation. At present, a large proportion of the policy governing the preservation task of Sound and Vision has been formulated. This document in fact demonstrates that many elements of this policy - but not all - are now visible in practice. To the extent that policy has not (yet) been laid down or implemented, these facts are clearly recorded. To maintain pace with the constant development of the Digital Archive, this document will be regularly updated.





# MISSION AND TASK OUTLINE

# 2

The mission statement of Sound and Vision is:

*As the guardian of Dutch audio-visual cultural heritage we keep Dutch history in sound and vision alive. We make it possible for everyone to learn, experience and create with the Dutch audio-visual history.*

This mission summarizes the task and mission of Sound and Vision: the collection, storage and provision of access to audio-visual heritage of national importance for the media industry, the creative industry, the cultural heritage sector, for education and academic research and for society as a whole. The obligation of preservation is the automatic consequence of mission, assignment and task. After all, by preserving this heritage for the future, Sound and Vision is able to maintain the digital collections entrusted to the organization, with a view to ensuring permanent access for various user groups.

# STRATEGIC POLICY FRAMEWORK

# 3

The public task of Sound and Vision is twofold: the organization acts as the corporate audio-visual programme archive for the Netherlands Public Broadcasting Organization (*Nederlandse Publieke Omroep* - NPO) and as a national cultural and historical archive for education and academic research and for the general public. The institute has been given the mandate of preserving the collections by the government (Media Act 2008, art. 2.1). Furthermore, the NPO and the representing organisations of copyright holders (associations of independent television producers and the record producers) have granted permission for preservation.

The preservation policy is a direct consequence of the overarching policy of Sound and Vision. This policy is recorded in strategic documents and in documented agreements and collaborative ventures with user groups. These documents all tie in with one or more of the roles and responsibilities of Sound and Vision outlined above. The relationship with the preservation task can be summarized as follows: the aims and ambitions laid down in the following policy plans and agreements mean both implicitly and explicitly that (parts of) the collections must remain accessible for the long term.

## 3.1 MULTIYEAR POLICY

The Multiyear policy plan 2016-2020<sup>1</sup> contains the strategic ambitions of Sound and Vision in respect of accessibility, long-term storage and the national audio-visual hub function of the institute. The transition at Sound and Vision into a professionally managed Digital Archive will take its definitive form over the coming policy period. The establishment of a controlled preservation environment, formally certified as a Trustworthy Digital Repository (TDR), will guarantee the futureproof storage of the collections for the depositors, and will guarantee permanent access to users. ISO certification of ICT management and services will complete the reliability and security of the technical infrastructure, in support of these processes.

<sup>1</sup> *Multiple year policy NISV 2016-2020*

<sup>2</sup> *Collection Policy Plan NISV, ed. M. Lauwers (2013)*

## 3.2 COLLECTION POLICY

The Sound and Vision Collection Policy Plan (2012-2016)<sup>2</sup> contains a description of the various archive roles of Sound and Vision, and of the (legal) frameworks within which those roles are fulfilled. The document offers a collection profile and an explicit clarification of the policy principles for selection, preservation and providing access to the collections. All user groups are identified.

### 3.2.1 Collection description (typology)

In December 2015, the size of the Digital Archive - including backups - amounted to more than 25 Petabytes. The collections can be divided as follows.

#### A. Digital born collections daily ingest of radio and television

Radio and television productions from public (and commercial) broadcasters preserved on assignment. To some extent, these productions have an intrinsic cultural and historical value. The material is selected for long-term preservation because of its (cultural and) historical value, its reuse value and/or its research value. This material includes:

- 100% broadcast Dutch-produced programmes from the public broadcasting organizations.
- Each year, 800 hours of unedited news and current affairs items.
- EVN material, 1,000 news items (EVN is the European news and exchange service for current affairs items).
- 100% of the programmes broadcast by the six Dutch public radio broadcasters.
- 800 hours of music recordings from public broadcasters (MOZ).
- Each year, unabridged recordings of two complete weeks of Dutch radio and television broadcasts, public and commercial (4,368 hours and 3,360 hours, respectively).

NB The public television broadcasts are stored in high resolution, the commercial broadcasts in low resolution.

- A selection of broadcasts from commercial stations: radio 500 hours and television 1000 hours, in high and low resolution.

The ingestion of digital born audio and video produced by the public broadcasters per year amounts to 8,000 hours of television and 54,000 hours of radio.

#### B. Digitized legacy collections

Digitized collections of analogue film, film, video and audio material mainly consisting of productions from public broadcasters through to the year 2007 (in this case the start of digital broadcasting production). Most of the material has been digitized in the framework of the Beelden voor de Toekomst (Images for the Future - 2007-2015) project. Now the project has come to an end, legacy materials are not being digitized on such a large scale, i.e. only on request from specific customers and as part of selection work.

#### C. Websites and interactive media

Websites, primarily produced by public broadcasters. This selection will be expanded over the coming years to include websites relating to the core collection at Sound and Vision. Twice a year, approximately 55 unique sites are collected. As well as websites, Internet videos are also collected; each year between 100 and 150. These are mainly videos that appear on YouTube (channels from 'celebrity' YouTubers and other popular videos). Finally, the ambition has been announced in the Multiyear plan 2016-2020 to reassess the collection policy by broadening the policy to include video art, interactive productions, games and virtual reality. In 2016, a number of research pilots will be initiated.

#### D. Digital (media)collections from third parties

Collections from Dutch (media) organizations with a deposit agreement that make use of (a choice from) the ingestion, storage, preservation and

access options within the infrastructure of the Digital Archive. At present these include the EYE film collection, the backup for the collection of the Royal Dutch Library, material from the Dutch Premier League/ECV and the Naturalis Museum. There are well-developed plans for preserving video registrations of the Proceedings of the Dutch Lower House of Parliament. All these collections are managed subject to the individual conditions of the depositor, and may be preserved for a short period or for the long term. In cases in which exclusively a technical hosting relationship is maintained, the material is not made generally available for discovery or re-use.

### 3.3 AGREEMENTS ON ACCESS

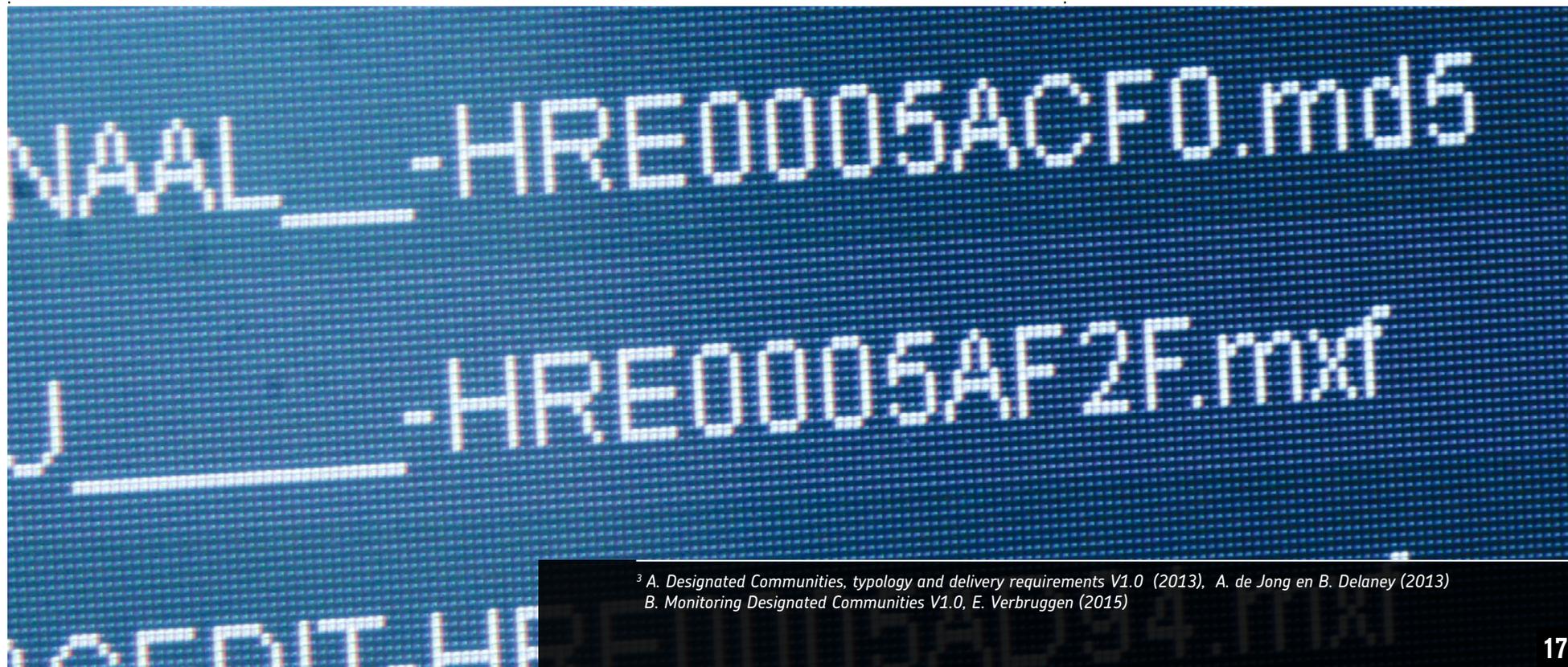
Sound and Vision identifies four types of external user groups or Designated Communities<sup>3</sup>. These are Media Professionals, users from the Educational Domain, the General Public and Researchers. There are also internal users: the Metadata Management department and the Ingestion and Access Media Managers, the Museum staff,

the staff of the Knowledge and Innovation department and the Development department.

All undertakings and agreements relating to (conditions for) consultation and use of the collections for the various external user groups are formally laid down in:

1. The service agreement with the Netherlands Public Broadcasting Organization NPO; the Chain Agreements (*Ketenafspraken*) between NPO and Sound and Vision.
2. Agreements concerning access to AV sources for primary, lower, middle and senior secondary education and higher education and research as laid down with Surfnet, the Secondary Education Council (*VO Raad*), Stichting Kennisnet, and the NPO and NTR, the CLARIAH Consortium (ACADEMIA, CLARIN.nl), and in the Archive Agreement.

3. Agreements and collaborative ventures for granting access to AV sources for scientific research (with national and European subsidy providers and for example with the University of Amsterdam, VU University Amsterdam, the University of Twente, Delft University of Technology, TNO, Utrecht University, Maastricht University and the University of Groningen).



<sup>3</sup> A. Designated Communities, typology and delivery requirements V1.0 (2013), A. de Jong en B. Delaney (2013)  
B. Monitoring Designated Communities V1.0, E. Verbruggen (2015)

# PRESERVATION PRINCIPLES

# 4

## 4.1 CORE DEFINITIONS

Within the policy framework of Sound and Vision, the term *digital preservation* should be taken to mean:

*The full range of activities and processes necessary for the intellectual and technical preservation of the digital collections over time, with the purpose of ensuring sustainable access for the user groups.*

The primary goal of digital preservation is to preserve the integrity and authenticity of the digital objects. Sound and Vision operates the following definitions:

### **Integrity**

*The object is demonstrably unchanged at bit configuration level.*

### **Authenticity**

*The object is what it purports to be; it is demonstrably unaltered since its submission or it can be demonstrated that following transformation, all typical characteristics have been preserved.*

## 4.2 SCOPE

The preservation policy extends to include both permanent and temporary preservation of all files stored in the Digital Archive. This relates to the core collections of Sound and Vision and collections that do *not* belong to the core collections. This latter category consists of AV productions that have been appraised with a lower value by Sound and Vision (see also paragraph 9.2.1) and materials stored by the Digital Archive as a service to other parties (for a short period or for the long term, with or without access). Both groups belong to what is known as the 'dynamic collections'.

All components of the core collections and the dynamic collections come within the preservation

scenarios elaborated in this document (see Chapter 7). Agreements on these scenarios with the collective or individual depositors may relate to quantities, quality, storage period, means of granting access and user rights, and to the conditions for the digital ingestion. The agreements are laid down according to the definitions in the document "*Handleiding voor het maken van een Submission en Order Agreement*" (Manual for creating a Submission and Order Agreement)<sup>4</sup>. A standard template is being developed for the actual contract and the SLAs (Service Level Agreements) with depositors<sup>5</sup>.

## 4.3 PRESERVABLE FORMATS

Sound and Vision has selected a limited number of formats for long term preservation. The Digital Archive is able to issue fully described guarantees for the sustainability of these formats. The organization operates fixed criteria for determining a preservable format: it must be a well-documented industry standard that operates on current software, as used in the audio-visual domain. The format must be able to be indexed within the internal technical and catalogue infrastructure of Sound and Vision, so that derived files can be produced, for viewing and delivery. The standard must therefore support media-related functions such as time codes, subtitling and metadata. In addition, it must be possible to transcode the format to other common formats, using current transcoding software. Using standard analysis software it must be possible to carry out quality analyses of the format.

<sup>4</sup> *Manual for making a Submission Agreement and Order Agreement V1.1*, A. de Jong, B. Delaney (2013)

<sup>5</sup> *Standard Service contract NISV* (2016)

#### 4.3.1 Preservation master, mezzanine and proxy files

Within the preservation practice at Sound and Vision, three types of file are distinguished, each with their own function within the services and tasks undertaken by the Digital Archive: the preservation master (also known as the archive master), the mezzanine and the proxy file.

##### *Preservation master*

The preservation master is the archive format permanently preserved by Sound and Vision, either as part of the core collection or on behalf of a depositor. The choice of the MXF format traditionally goes hand in hand with the position of Sound and Vision in the production process of the public broadcasting organizations, in this case the role as the corporate AV-archive. As was formerly the case for the analogue preservation master, also in the digital era, the archive format depends heavily on the production format used by the public broadcaster, which must be suitable for immediate delivery to this largest Designated Community.

Because this format is submitted by the public broadcasters as source format, and because as an industry standard it is highly suitable for professional reuse, at Sound and Vision, MXF serves simultaneously as preservation master and as standard delivery format. If certain specified conditions are met, guarantees can be issued for an enduring playable MXF preservation master. In that case it will also be possible to search through the material via metadata in the catalogue and/or via derived files (proxy files and/or key frames) that can be presented on the various Sound and Vision search interfaces. When users order fragments, a so-called *partial restore* can be made from the MXF (a partial restore is a sequence of an AV file with the accompanying metadata, that is delivered separately as a component), and the material can

be transcoded.

Despite the unavoidable influence of the broadcast production standard on the choice of the preservation format, Sound and Vision has set itself the task of introducing some differentiation in its twofold task as national cultural historical archive and as media archive. This dual role implies that in a number of cases, the choice of preservation format may deviate, depending on the cultural historical importance of the materials. The focus of effort therefore is to bring the choice of preservation format for these parts of the collection more into line with the archive's own, national responsibility. A first example of this focus was the choice of format for scanned film. For this subcollection - in respect of genres representing particular value - a standard was chosen with a high resolution; in this case DPX. This format is employed because film requires a high-quality digital archive format, since the intrinsic quality of film is higher than of television material.

##### *Mezzanine*

Due to their size, preservation masters in DPX format cannot be rapidly transported. Furthermore, the content cannot easily be processed by video editors in the broadcasting production environment. As a consequence, the Digital Archive does not supply DPX directly. Instead, a mezzanine is produced, an intermediate format that is derived from the preservation master. This mezzanine is produced in XDCAM/MXF, the same format as used for HD-TV. In terms of specifications, the mezzanine fulfils the requirements imposed on MXF files as they are ingested in the Digital Archive on a daily basis. This means that this intermediate format for digital film is effectively identical to the preservation master for television material. For that reason, the mezzanine MXF formats are also stored and preserved according

to the same conditions as all other MXF files. In practice, the MXF mezzanines therefore function as a second preservation master.

##### *Proxy files*

Proxy files are produced by Sound and Vision as a viewing copy of the audio-visual master material. These files have relatively low resolution and are not suitable for reuse. They can be accessed and viewed via the catalogue infrastructure of Sound and Vision, i.e. the internal catalogue, the online catalogue for use at the Mediapark and the search interfaces for the general public. A number of other parties (including the commercial broadcasters) can also use this function. The proxy format at Sound and Vision is MPEG-4. MPEG-4 was chosen because it is a stable and commonly used format that can be widely used on a variety of platforms.

Proxy files are not intended for use outside the catalogue infrastructure of Sound and Vision. After all, in the case of distribution via the Internet, it is not possible to continue to offer playout guarantees. Storage, format specifications and the affected software components, in that case, are no longer under the control of the Digital Archive. Long term preservation according to the policy principles outlined in this document, can therefore only be applied to digital objects over which the Digital Archive has full control and authority, during their entire lifecycle.

## 4.4 NON-PRESERVABLE FORMATS

The operating principle is that the depositor delivers the file in the preservable format specified by the Digital Archive. The depositor is thereby

responsible for the quality and correctness of the submission. A second principle states that the material can be rejected by Sound and Vision if upon submission it does not comply with the agreed specifications.

If for whatever reason a depositor is not able to deliver a preservable format, additional agreements can be made. One of those agreements may require Sound and Vision to offer support in (enabling) transcoding of the submitted format to the preservable format. This is only possible if the costs for the conversion are covered by the depositor. Financial considerations also play a role in the decision whether or not to permanently store the non-preserved original in the Digital Archive and to continue maintaining that original alongside the newly created preservation master. During the period that transcoding can be done using current standard software, new preservation masters can be produced from the original. At the moment that the original format can no longer be played out, the last-made preservation master will acquire the status of original. This scenario applies exclusively if a good-quality preservation master of the original *can* actually be produced. In other words, in addition to the financial conditions, quality requirements also play a role. The same standards apply to converting as for a regular 'internal' transcoding process within the Digital Archive: it must be clear in advance which characteristics of the original object must be retained during transcoding and what will be lost. The newly created file must further comply with the same quality requirements as material submitted directly in the preservation format.

It is also possible, at the request of depositors, to submit non-preserved formats to the Digital Archive, *without* transcoding them to a preservation master. These original files - this may relate to a

wide variety of professional and non-professional AV formats - can be ingested to be stored 'as is'. The Digital Archive offers no guarantees for this service in respect of long term playability. Access to the materials via the catalogue is not possible, or at least only to a very limited extent. This option can be selected by depositors - in consultation with Sound and Vision - if discovery and access to the material by other parties are not relevant or if the submitted material cannot be transcoded to the preservable format in advance, but that nonetheless there is a wish to have the material stored (temporarily or for the long term) in the Digital Archive.

All decisions concerning the transcoding and temporary or long term storage of non-preserved formats must be taken in advance of the ingestion phase, in consultation with the depositor, and laid down in the contracts and SLAs.



# PRESERVATION STRATEGY

# 5

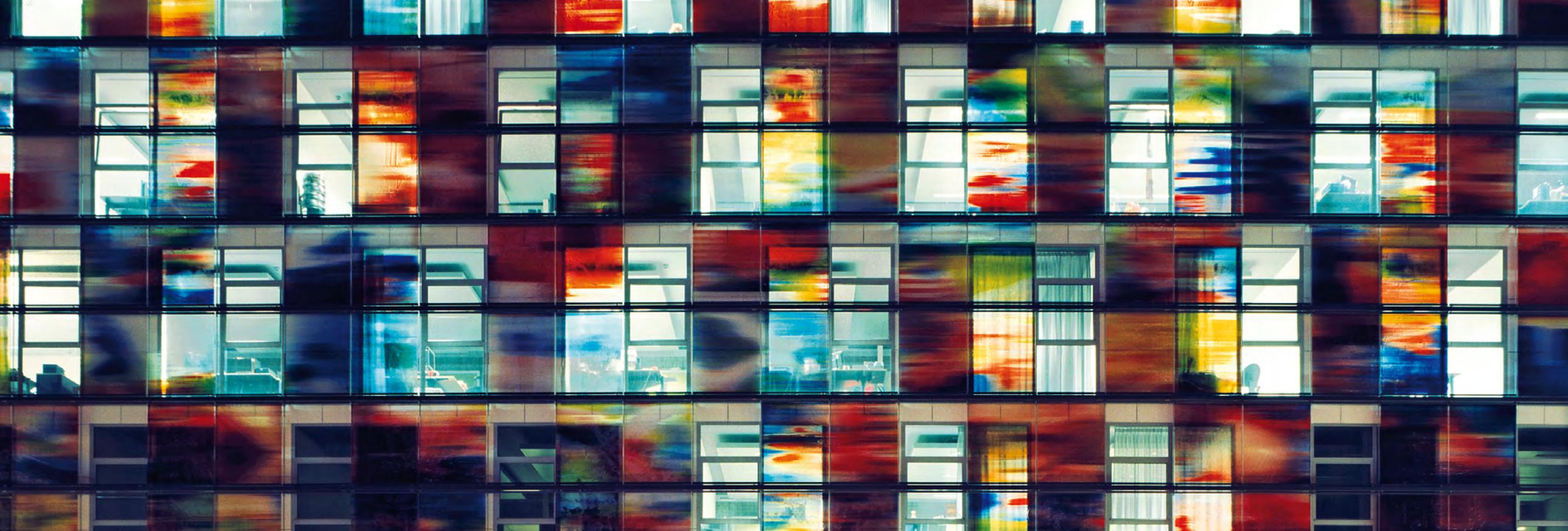
## 5.1 MIGRATION

The central preservation strategy operated by Sound and Vision for the core collection and the materials from other depositors that must be preserved, is migration. This first of all means that carriers on which the AV productions are stored, are transferred to current carriers at set times, related to the state of technology. This preservation strategy arises from the fact that the Digital Archive houses a very large and rapidly growing digital collection that is intensively reused within a professional media production environment. As a result, emulation is not an option as a preservation strategy\*. After all: in that case it would not be possible to make all the material available immediately in a usable form to the largest user group, the public broadcasters.

Migration activities also include format migration. With the exception of the migration of large collections of proxy formats, this type of migration has not yet been carried out since the first ingest of digital archive formats (in 2007). The planning and implementation for the migration of the formats in the large collection of broadcasting productions is carried out in consultation with the depositor, the NPO.

The third form of migration is the migration of the hardware such as the tape robot, the servers and the systems within which the various preservation processes take place. These innovations have been a fixed element of the annual planning, budgeting and implementation of the maintenance and permanent improvement of the infrastructure since the moment they were established in 2007.

\* Emulation will probably start to play a role with the new collection areas for games and interactive media. The enormous variety of file formats that cannot be standardized to the regular preservable format makes migration for these media difficult if not impossible as a preservation strategy.



## 5.2 GUARANTEEING INTEGRITY

Any digital archive must comply with the *requirement of persistence*. This means that the material must come out precisely as it went in, i.e. complete and in the correct data structure. This data integrity is secured by Sound and Vision through the use of *checksums*: in other words the comparison of the file against the original bit configuration, as laid down prior to ingestion, or at the moment of storage. During every processing or movement of a version of the object for (re) storage, copying, migration or for delivery to users, a *fixity check* is used to return to the checksum, to make certain that (in terms of bits) the object is identical to the content that was transferred in a previous phase. Integrity is also guaranteed by monitoring the material already stored in the Digital Archive during migration actions known as *refreshments*, and whenever the material is accessed.

Incoming objects are checked and validated according to a number of strict ingestion procedures, whereby a number of (quality) checks are structurally carried out, for example to determine the completeness of the submitted files, compliance of the format with the standard and the presence and the quality of the metadata. Many of these checks are included in the scripts of the import workflows for the various collections to be ingested.

For heritage collections and commercial services, checksums are produced by the depositor prior to ingestion. For the broadcast collections, checksums are not submitted by the depositor. In these cases, another method is used to check whether a file is complete, namely the so-called *header-and-footer check* of the file. This check guarantees that the file has been received complete and that the transport was not interrupted partway through. This check also confirms that the container format can be read. A header-and-foot-

er check can only be carried out on files that actually include a footer. As concerns the preservation formats, at present, this is only the case for MXF.

Every five to seven years, a complete migration takes place of all tapes in the Digital Archive. At these moments, the objects are once again read through and rewritten. If errors appear during the daily use of the materials (e.g. if it turns out that a tape cannot be read), Sound and Vision operates documented procedures to restore the file with the backup.

## 5.3 GUARANTEEING AUTHENTICITY

To guarantee authenticity in the context of preservation policy, three conditions must be complied with:

1. The object is what it purports to be. A quality analysis is carried out on the object to demonstrate this fact.
2. The object is usable, can be played out and has significance to the user. For this purpose, the essential properties of the object are preserved.
3. The object has not been altered unintentionally or unauthorized. For this purpose, the lifecycle of the object is recorded.



By recording all events in the lifecycle, the so-called *chain of custody* of each object is preserved: the details in the context within which the object is created, ingested, stored and is used. These data form part of the category provenance metadata within the so-called *preservation metadata* (see also paragraphs 6.3 and 9.2.3). Together, these data form the evidence for the 'credibility' of an object as it has been maintained over time. These data describe where an object came from, and who has processed and used it. Creating a record of this lifecycle is achieved by means of (automatically generated) metadata, themselves the outcome of adaptations and movements of the object. This information is subsequently added to the object. In this way, provenance metadata represent the guarantee that the object has not been unintentionally altered. By monitoring and recording all predefined steps in the lifecycle of each individual incoming object in preservation metadata, the authenticity of the object is both guaranteed and demonstrated. As a consequence, the Digital Archive is able, at all times, to account for its actions to both its depositors and its users. In this way, the basic condition is met for being 'trustworthy'. The second aspect of authenticity is sustainable access: the object is available in a usable format, and understandable to the users. To be able to guarantee this aspect of authenticity, the essential characteristics of an object must be determined, the so-called *significant properties*.

These significant properties include the technical, aesthetic and intellectual characteristics that must be preserved over time and throughout the various technological changes. The preservation of authenticity can be measured according to the degree to which - following a migration or other transformation - the significant properties of the original object (the preservation master) have been preserved in the new form of the object. The significant properties of an archival object are 'absolute', in other words, they form a fixed set that applies to all migrations and transformations. After all, in combination, they represent the characteristics that make up the essence of the object as it

entered the archive and/or as it is experienced by the user.

Another category of object properties are *not* fixed: they are relative. These are the so-called *transformational migration properties*: specified inherent characteristics of an object that are permitted to change during a migration. Which of these properties of a master file must be preserved is decided on each individual migration or transformation of objects and collections. These decisions are taken on the basis of the technical possibilities for input and/or output format, the requirements of the depositors and/or the users and the current format policy of the Digital Archive. Practical issues such as lead time and costs can also play a role. The choices and motivation are recorded in the separate preservation action plans that are drawn up for each migration.

Determination of these transformational migration properties must always be carried out in relation to the fixed significant properties. By identifying the significant properties and the transformational migration properties in advance, the successfulness of a migration can be assessed afterwards.

The Preservation Metadata Dictionary (paragraph 9.2.3) contains a list of all possible provenance metadata and technical metadata per individual object/file (among these the transformational migration properties). Which properties are considered significant, is established per archival object type (e.g. MXF, WAV, DPX, TIFF, PDF), in policy.

# THE PRESERVATION WORKFLOW

# 6

## 6.1 INFORMATION MODEL<sup>6</sup>

Sound and Vision has developed an information model that acts as reference in analysing and successively implementing the controlled preservation workflow in the processes and systems. Elements and concepts from the model are now in fact reflected in the actual ingestion, storage and access processes.

In the information model, a generic description is given of which workflows are distinguished around the functions ingestion, storage and providing access. The model records all actions or events that can take place in relation to an object and also defines the (changing) composition of the object across all of these processes. *Where* in the workflow the actions or events take place and the resultant outcome in preservation metadata, is also recorded. Predefinition is essential in order to have a reference framework to be able to verify that all events in the lifecycle of an object tie in with the preservation policy of the archive. By comparing provenance metadata and the events in the information model, it can be established that - if the workflow was completed correctly - no unexpected actions have been carried out on the object. The lifecycle of a digital object thereby develops in a controlled and verifiable manner.

To be able to identify and manage the files and accompanying metadata within the processes at a conceptual level, clear digital objects have been created within the information model, known as *information packages*. The content (essence and metadata) of the package types may differ: a submitted file may be stored in an enriched form (for example added to with metadata). And what is delivered to users is often only a part of what is stored (for example only a viewing copy of the content without the complete set of stored metadata).

## 6.2 GENERIC WORKFLOW STAGES

The first stage of the generic workflow effectively takes place before intake or ingestion and consists of the negotiation phase with the depositor of the material. This process stage results in the drawing up of contractual documents (in this case Submission Agreements, contracts, SLAs) in which all agreements are laid down on such issues as formats, preservation scenario, rights, quality controls, metadata and reports.

During the actual ingestion process, a *fixity check* is carried out to check whether the file was correctly received. This check serves to guarantee that the file was submitted complete and correct by the depositor. The fixity check cannot be carried out (or only to a limited extent, in the form of a header-and-footer check (see also paragraph 5.1) if a depositor is not able to also supply a checksum. In that case, the Digital Archive itself will generate a checksum, *after* the material has been ingested, in order to be able to check the integrity of the file following any future actions. In the absence of a submitted checksum, it is specified in advance, in the contract or SLA with the depositor, that the Digital Archive will accept no liability if material is corrupted during the transfer as part of the ingestion process.

Following this series of stages, the format of the object is determined, and technical metadata are extracted. These data relate to the material properties of the AV file, such as for example *aspect ratio*, *color space*, and *codecs* and *bitrate* used. Technical metadata extraction is primarily necessary in order to maintain a permanent overview of the various technical formats stored in the archive. In this way, future risks relating to specified file formats (for example formats that are becoming obsolete) can be detected in time and can be taken into account in planning migration actions to new formats. The extraction of technical metadata is also important to be able to verify whether the formats comply with the quality

<sup>6</sup> Information Model Digital Archive NISV V1.0, A. de Jong, D. Steinmeier, B. Delaney, Y. Hollander, P. Hoffman (2013)

agreements as drawn up with the depositors. An optional component (in this case depending on the agreements with the depositor) of the workflow is then the automated Quality Assurance process, whereby the files are checked for quality. During this process, too, technical metadata are extracted.

All ingested packages with content and metadata are then given an *identifier*, a unique label that forms a permanent reference to the object, as stored in the Digital Archive. This stage represents the end of the ingestion phase, and the ingestion packages, the so-called SIPs (*Submission Information Packages*), are ready for final storage. Together with additional files (including subtitling and context information about the AV-production) plus the extracted technical metadata and any manually added descriptive metadata, the SIP is converted into an AIP, an *Archival Information Package*. This package then enters the storage domain of the Digital Archive: the Archival Storage. The number of workflow stages for storage in the information model at Sound and Vision remains limited. The AIP need only be allocated a definitive storage location, itself stored in the metadata.

The last element of the workflow describes the actions relating to the *Dissemination Information Package* or DIP. The DIP workflow normally starts with authentication of the user, whereby a determination is made of the authorizations of the user

who registers with the system; this is followed by a request for a certain type of material intended for a certain type of use. If the system approves the request, the DIP is delivered. In a situation in which the *full* AIP is requested, a fixity check is carried out to guarantee a successfully delivery.

If a *transcode* (a different file format than the stored format) or a *partial restore* (part of an AIP and/or the accompanying metadata) is requested, the checksum of this new version is calculated by the system, and delivered with the file. Within the information model, the access workflow is also generically created so that 'a request' for access can relate either to simply 'a search according to metadata' or ordering (part of) a specific AV file.

### 6.3 PROCESS DETAILS AND TECHNICAL METADATA

A large proportion of the workflow steps and the technical properties of the files as defined in the information model are currently automatically generated and logged in a variety of forms that have grown out of practice. This happens during the ingestion, storage and access processes, as they take place across the various components of the Sound and Vision information environment. In the large workflows for radio and television collections, for example, the *end-to-end check*

of the files is a standard component of the work process. The outcomes of format checks and copy actions are monitored and logged. In the scripts of the import workflows, the checks during ingest are included. In this way, both the process data and part of the technical and administrative data (in preservation metadata terms - the events and provenance metadata) about the files are recorded, be it dispersed.

With the implementation of a new Media Asset Management system in 2016, the creation and structuring of both the process data and the technical properties will be considerably improved. All files - no matter what source they originate from - will in the future be ingested into the MAM system by the same ingest workflow. In migrating from the current systems to the new central MAM system, all legacy files will undergo the same checks, so that the same technical data will be available in relation to both legacy and newly ingested files.

The output in terms of process data will in the new situation be structured as tasks in a workflow management system, thus making it simply available as an *audit trail* of the life cycle of one or a group of files. Checking the quality and completeness of the files will subsequently be improved by using new analysis and extraction tools. As a result, more technical data can be generated about the files. The new MAM system will establish a central facility in which all data necessary for maintaining the integrity and authenticity of the

collections will be created and stored at a single location.

For a completeness check, this corpus of technical and process data will be mapped with the form and content of the attributes in the Sound and Vision Preservation Metadata Dictionary, the overview of all metadata that may be involved in audio-visual preservation. It should be made clear for how long and in what form the metadata data in question is stored in the MAM. Subsequently, further investigations will be carried out to determine the extent to which yet other metadata, managed outside the MAM (in the storage management system, in the acquisition and registration system), can be used as preservation metadata. Finally, a determination will be made of how all this information can be exported from the source systems. This output as preservation metadata can then be accessed and manipulated in a structured and standardized manner, for the purposes of accountability, planning and implementation of the preservation activities.

NISTRIBUNA-  
IGS\_DE\_LIJN-AEN55718216.md5  
IGS\_DE\_LIJN-AEN55718215.md5  
IGS\_DE\_LIJN-AEN55718214.md5

Archive new content [DDV]  
Archive new content [DDV]  
Archive new content [DDV]

# PRESERVATION LEVELS

# 7

## 7.1 FULL PRESERVATION AND BIT PRESERVATION

The previously described generic preservation workflow can be structured for each type of intake into the Digital Archive - large or small, single or structural. For this purpose, Sound and Vision has developed a differentiated preservation approach, whereby the method of ingestion, storage and accessibility will depend on (a combination of) the demands and capacities of the depositor, Sound and Vision itself (in this case the applicable collection policy), the requirements of the Designated Communities, the technical facilities, the costs and the (copy) rights.

There are two main levels<sup>7</sup>:

### 1. Bit preservation ('passive preservation')

*The file is stored as it is received. Usability for the short and long term are not guaranteed, because the file is not technically analysed.*

### 2. Full preservation ('active preservation')

*The file is first stored as received but - in order to keep it usable/playable - may be changed over time.*

*NB This form of preservation is only possible if material is stored in a preservable format.*

## 7.2 PRESERVATION MENU

A 'preservation menu' offers a choice from different scenarios or "levels" of delivery, storage and access. This menu describes the type of storage service provided, what (meta)data has to be submitted, what preservation actions will be undertaken, and what form of preservation applies. All agreements relating to the level of preservation are described in detail in the contracts and in the Service Level Agreements with the depositors.

For the depositor, each level comes with conditions for the submission of essence and metadata. Only if these conditions are met, can specific

guarantees be issued by Sound and Vision. For all levels, the depositor remains responsible for the material, and himself retains a copy as long as no feedback has been received. At that point, a formal notice must be issued that the submission has been successful, and that the Digital Archive accepts responsibility for the object(s). Furthermore, at least two versions of a file are always present as backup, at two different locations. Depending on further agreements, this may be an offline copy or a copy in the tape robot.

Within each of these two main choices - bit preservation (1) and full preservation (2) - there are rising levels of authenticity and integrity (a to c). A rising level of guarantee applies to each service type from 1a to 2c. These guarantees for example relate to the returnability of the original objects, retrievability, the (quality) of the storage and the migration, the linking with metadata, the extent to which the standards are fulfilled, guarantees on not unintentional changing of objects and (sustainable) playability.

Level 1 scenarios by definition relate to solutions for the short term. After all: bit preservation implies that playability is *not* guaranteed for the long term. The level 1 scenarios are broken down into levels 1a to 1c. Depositors can opt for levels 1a or 1b if for example, they are not in a position to transcode the submitted material in advance (or have it transcoded) into a format preservable by the Digital Archive. Levels 1a and 1b are therefore generally *not* scenarios for material that forms part of the core collections of Sound and Vision. These levels can be applied to material that is submitted with only minimal metadata (for example only a checksum) or with no metadata at all. These materials cannot be accessed, or only to a limited extent. Level 1c can be selected for material that will eventually become part of the core collection, so that this material is rapidly and safely stored before being converted to a preservable format. This for example applies when a depositor can no longer store his own collection - for whatever reason - and wishes the entire collection to be ingested at Sound and Vision, with immediate effect.

<sup>7</sup> Technical elaboration of the main preservation levels full preservation and bit preservation (i.e. Full Service archive and Managed Storage) in relation to the cost price calculation appear in the policy memorandum 'Tariefstelling en keuzes archiefdienstverlening', M. Spindel d.d. 151015.

The level 2 scenarios (2a to 2c) were developed for the collections stored and made accessible by Sound and Vision for the long term. This refers to material for which the institute accepts permanent care, either as a service for a smaller or larger (heritage) party, or as part of its own core collections. These materials can be accessed via the Sound and Vision catalogue. These collections are catalogued upon ingestion (from basic metadata through to enriched descriptive metadata); preservation metadata can then be added and a series of checks and quality controls are carried out.

## EXPLANATORY NOTES

### Integrity calculation

A checksum is calculated for the file so it can be used for integrity checks following media refreshment actions.

### Backup

Following successful ingestion, a backup is made of the essence, including an integrity check on the basis of the checksum.

### File name check

Before the import starts, file names are checked for compliance with agreements.

### Carrier replacement

Periodically, a decision is taken on whether carriers need to be replaced. In the case of such replacement, files are checked for integrity following copying.

### Restore

If media become unreadable during the storage phase, a restore will be carried out from the backup tape, and a new backup will be created. Making the backup also includes an integrity check on the basis of the checksum.

### End-to-end check

A check to determine whether or not material is missing within an agreed delivery.

### Integrity check

A checksum or header-and-footer check (only for MXF) is carried out prior to ingestion to determine whether the file has been received complete. NB The checksum is always the preferred control mechanism because this check is most reliable for detection of bit corruption.

### Format migration

If there is a serious risk that the file format will no longer be playable within the foreseeable future, the format is migrated to a format that is playable.

### Format analysis

A check to determine whether the file has the correct fields and the correct structure in the header.

### Quality analysis

A limited number of checks are carried out on the quality of the essence.

### Metadata content check

A check to determine whether the metadata file complies with current metadata policy at Sound and Vision in terms of content, before import is started.

### Metadata technical check

A check to determine whether the metadata file complies with the agreed format in technical terms.

## 7.3 GUARANTEES AND CHECKS

### Guarantees for each preservation level

|  | 1a    | 1b    | 1c    | 2a    | 2b    | 2c    |
|--|-------|-------|-------|-------|-------|-------|
| Files can be returned as they were submitted                       | Green | Green | Green |       |       |       |
| Files will not be unintentionally changed following storage        | Green | Green | Green | Green | Green | Green |
| Files are retrievable by file name                                 | Green | Green | Green | Red   | Green | Green |
| Storage media are periodically replaced                            | Green | Green | Green | Green | Green | Green |
| Files are stored well down to bit level                            | Red   | Green | Green | Green | Green | Green |
| Everything that should be stored is actually stored                | Red   | Green | Green | Green | Green | Green |
| Files are linked to descriptive metadata                           | Red   | Red   | Green | Green | Green | Green |
| Files are retrievable according to content                         | Red   | Red   | Red   | Green | Green | Green |
| Files are stored in a preservable archive format                   | Red   | Red   | Red   | Green | Green | Green |
| The metadata comply with the metadata standard of Sound and Vision | Red   | Red   | Red   | Green | Green | Green |
| Files are playable/usable for the short term                       | Red   | Red   | Red   | Green | Green | Green |
| Files comply with the valid format standard                        | Red   | Red   | Red   | Red   | Green | Green |

## 7.4 PRESERVATION ACTIONS

### Preservation actions per level

|                          | 1a    | 1b    | 1c    | 2a    | 2b    | 2c    |
|--------------------------|-------|-------|-------|-------|-------|-------|
| Integrity calculation    | Green | Green | Green | Green | Green | Green |
| Backup                   | Green | Green | Green | Green | Green | Green |
| Carrier replacement      | Green | Green | Green | Green | Green | Green |
| Restore                  | Green | Green | Green | Green | Green | Green |
| End-to-end check         | Red   | Green | Green | Green | Green | Green |
| Integrity check          | Red   | Green | Green | Red   | Green | Green |
| Format migration         | Red   | Red   | Red   | Green | Green | Green |
| Format analysis          | Red   | Red   | Red   | Red   | Green | Green |
| Quality analysis         | Red   | Red   | Red   | Red   | Red   | Green |
| Metadata content check   | Red   | Red   | Green | Green | Green | Green |
| Metadata technical check | Red   | Red   | Green | Green | Green | Green |

# PRESERVATION PLANNING

# 8

## 8.1 PRESERVATION WATCH

Preservation activities are documented in preservation strategies for the long term and preservation plans per migration for the short term. The Sound and Vision standard for laying down the strategies and plans is recorded in the document “*Handleiding Ontwikkeling Preserveringsstrategieën en Preserveringsplannen*” (Manual for the Development of Preservation Strategies and Preservation Plans) V1.0<sup>8</sup>. The Manual is intended as a reference and assessment tool for standardizing preservation planning.

Different types of change can influence the planning and executing of the preservation activities. To predict these changes - which can occur both internally or in the environment of the Digital Archive - and to facilitate a timely response if alterations become necessary, they are permanently monitored. The controlled monitoring and documentation of the various developments that could affect the planning fall under the heading preservation watch. Supported by this *preservation watch*, the critical preservation functions of the Digital Archive are ensured.

## 8.2 ORGANIZATION CHANGES

First of all changes to the organization must be anticipated. These may for example relate to general austerity measures or a substantial change to the collection profile. Changes may also relate to issues that directly affect the technical infrastructure. This for example includes an increase in costs for digital storage or specific changes to current or new software and hardware contracts. Changes of this type are chartered out during the preparation of the annual budget. Sound and Vision then carries out a risk inventory, with an estimate of the possible influence on the operation, and the potential financial consequences. On that basis, a stress test is organized in which the risks for the multiyear forecast are also calculated. On the basis of these two analyses, an assessment

is made of which risks are acceptable and in respect of which risks additional financial and/or organizational measures need to be taken.

## 8.3 DESIGNATED COMMUNITIES

A second influence on the development of strategies and the preparation of action plans relates closely to the demands of the users. The collections must remain permanently accessible to the various external and internal Designated Communities. In technical terms this implies that the Digital Archive must be able to simultaneously serve different user groups, each with their own specific requirements in respect of video quality, granularity, browser and navigation facilities, metadata and interoperability\*. Once Sound and Vision knows how the user groups wish to use the materials, this can be taken into account in making preservation choices (e.g. the format, the storage method, the metadata that have to be stored and in the preservation action plans). The result is a direct relationship between the preservation policy and the way in which users wish to use the audiovisual content.

As a consequence, it is essential to know precisely how the various user groups each individually want to request material from the collections. To what purpose does a particular Designated Community access the material? What is the specific hardware and software environment in which the material must be played out? What quality requirements must the material reach for reuse? Substantive and technical requirements of this kind from the user groups of the Digital Archive have been identified and recorded.

The situation within which Designated Communities request and use material may change, both in terms of 'receiving' systems and the desired format and the metadata and the intended use of the material. Requirements for the delivery will change along. Sound and Vision controls these changes by regularly monitoring the Designated

<sup>8</sup> Manual for developing preservation strategies and preservation action plans, D. Steinmeier (2013)

\* Naturally, the user groups also have content-specific requirements and wishes in respect of the materials they want to be able to search through and use. These requirements are reflected in the choices and priorities of the collection structure as laid down in the Collection Policy Plan at Sound and Vision.

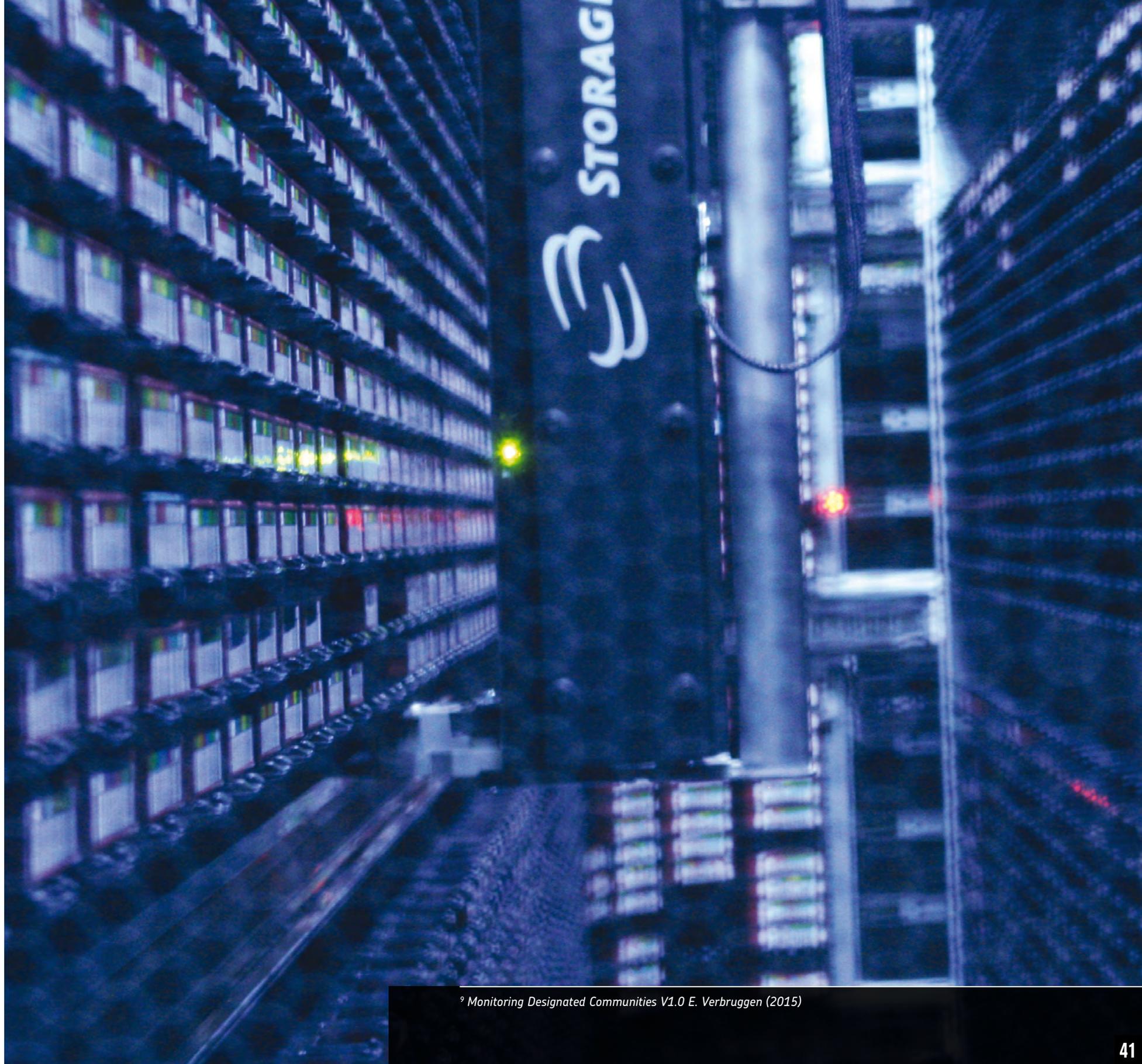
Communities. Using standardized methods (such as user research, surveys and user meetings) at fixed intervals, the departments that serve the user groups verify whether the previously recorded needs of those user groups are still up to date. Depending on the outcome of these assessment moments, Sound and Vision can adjust its preservation policy, and alter its service provisions<sup>9</sup>.

## 8.4 TECHNOLOGICAL DEVELOPMENTS

Format progression and innovations in storage media and playout software also have an autonomous influence on the ability of issuing guarantees for adequate ingestion, storage and sustainable access. Monitoring and anticipating technological developments of this kind are therefore also structured processes.

In this way, important developments can be foreseen, and sufficient time can be set aside for making cost estimates and carry out research into technical viability. In other words: technological innovations can be evaluated according to the current preservation policy, in good time.

The third underlying principle for laying down the preservation strategies and plans is therefore the outcome of the so-called *Technology Watch*, a procedural mechanism that enables Sound and Vision to consciously approach technology changes and innovations, in a controlled manner. Trends and developments in respect of formats (container or *encoding* format) and software and hardware for digital AV archiving, are carefully monitored and documented. This task is undertaken by various groups within the organization (ICT management, Application Management, Ingestion and R&D dptmnts). Knowledge is combined and filtered, and subsequently passed on to the Information Strategy Board at Sound and Vision, which board then takes decisions on the necessity and desirability of alterations, and their technical implementation.



<sup>9</sup> *Monitoring Designated Communities V1.0 E. Verbruggen (2015)*

# STANDARDS AND FORMATS

# 9

## 9.1 DIGITAL COLLECTION MANAGEMENT

The common thread in structuring all elements of the information and workflow environment at Sound and Vision is the Open Archival Information System (OAIS), ISO Standard 14721<sup>10</sup>. This tried and tested reference model for digital archives offers a framework within which the processes for ingestion, storage, access, migration and delivery are mutually linked and formally integrated with the outside world (i.e. the depositors and the Designated Communities). Sound and Vision has included the whole of the OAIS standard in a standard policy document for its own environment: “*Kwaliteitseisen Digitaal Archief Beeld en Geluid*” (Quality Requirements Digital Archive Sound and Vision) V1.1<sup>11</sup>

This document serves as the framework in organizing and creating order within the layout of the Digital Archive in respect of such aspects as:

1. The responsibilities, functions and roles of Sound and Vision (as national archive, as corporate AV-archive for the public broadcasters, as AV hub for the Netherlands);
2. The responsibilities, functions and roles of producers/depositors (broadcasters and other depositors);
3. The services and the nature of services provided to users;
4. The essence (film, video, audio) and the metadata (descriptive metadata and preservation metadata);
5. The AV formats and standards employed.

The Quality Requirements are a tool for further structuring the controlled preservation environment. The document can also be used as a self-assessment tool for the objective measurement of OAIS compliancy of the Digital

Archive. All components of this preservation policy are derived from the requirements described in this document.

## 9.2 METADATA FORMATS

The data management functions within collection management include the management of the descriptive metadata and the preservation metadata, the ability to carry out selection and appraisal, discarding and retention actions and supporting discovery of the materials.

### 9.2.1 Descriptive metadata<sup>12</sup>

For the *daily intake* and acquisitions, in principle the externally delivered descriptive metadata that are received via broadcasting information systems of the NPO are considered sufficient. All the material, upon ingestion, is allocated a value which subsequently determines the requirements with which the descriptive metadata must comply. Four value categories have been distinguished:

*Value A:*  
Irreplaceable and essential/rare (core collection)

*Value B:*  
Core collection

*Value C:*  
No major value (dynamic collection)

*Value D:*  
Little to no value or irrelevant (dynamic collection)

<sup>10</sup> <http://public.ccsds.org/publications/archive/650x0m2.pdf>

<sup>11</sup> *Quality Requirements Digital Archive NISV, V1.1*, ed. A. de Jong (2015)

<sup>12</sup> *Metadata rules NISV collections, V.Huis in 't Veld, J.Schuurman, I. Veenstra* (2015)

All descriptive metadata within the Sound and Vision information system is structured according to the internal standard based on the IFLA-FRBR model<sup>13</sup>. There is an administrative distinction between conditional metadata, basic metadata and enriched metadata:

#### Level 1:

These metadata are conditional for the information systems and processes at Sound and Vision to function. This also includes (parts of) the preservation metadata.

#### Level 2:

Following completion of ingestion, all materials come with a full basic set of metadata. This basic set will depend on the allocated value; for material of values A and B, the basic set is broader than for material of values C and D.

#### Level 3:

Following ingestion, Sound and Vision can provide part of its collection with additional descriptive context metadata.

### 9.2.2 Thesaurus<sup>14</sup>

Together with other Dutch heritage organizations, including the EYE Film Institute, Sound and Vision has developed what is known as the Common Thesaurus for Audio-visual Archives (GTAA). The GTAA is used for characterizing the content of audio-visual materials from the archive with labels drawn from a controlled and structured list of terms. At present, the thesaurus is above all used for the automatic structuring of metadata and linking sources on the basis of *linked data* principles. This is for example carried out in keyword extraction on the basis of Teletext (TT888) and in speaker labelling. The GTAA is also used to add already

structured keywords upon ingestion of materials into the broadcasting production systems. The thesaurus (which includes the GTAA and relates to the axes that are supplied externally) is also managed in OpenSKOS. The GTAA can be made available to external stakeholders as SKOS in XML/RDF under the Open Database License (OdbL). At present, the following users have access to the thesaurus: Sound and Vision, EYE, The National Archives of the Netherlands, the Cultural Heritage Agency of the Netherlands (RCE), Naturalis Museum; the Netherlands Public Broadcasting Organization (NPO); the projects CATCHPlus, Waisda, Woordentikkertje; VU University Amsterdam. The GTAA is supplied through the OAI - PMH protocol.

### 9.2.3 Preservation Metadata<sup>15</sup>

Preservation metadata include the categories of technical metadata and provenance metadata. Parts of the descriptive metadata are also included in the category preservation metadata, namely the attributes needed to identify a digital object. The fourth category of preservation metadata is the rights metadata.

The definitions of preservation metadata are summarized in the Preservation Metadata Dictionary V1.2 (PMD) developed by Sound and Vision. This dictionary contains the selection and definition of all metadata that may be used in recording the digital preservation process at Sound and Vision. In the PMD, the attributes are defined that may be allocated to each digital object (audio, video, film, text, photograph) ingested in the Digital Archive. This includes both technical metadata attributes of a file and attributes describing actions ('events'), results of those actions ('outcomes') and their associated 'agents' (responsible organization, software

or person. After all, these are the data that are required to provide the Digital Archive, its depositors and its users evidence of the digital provenance of a digital object, and hence its authenticity. The Dictionary also contains rights attributes that must be structurally related to a digital object. These rights relate not only to (re)use rights, but also preservation rights. The collection attributes in the Preservation Metadata Dictionary V1.2 are based among others on the standards PBCore, the Library of Congress VideoMD and AudioMD, PREMIS, NARA reVTMD and the ANSI/NISO Z39.87 Data Dictionary Technical Metadata for Digital Still Images. The Dictionary also contains a section for the Persistent Identifiers.

In its current form, the Preservation Metadata Dictionary V1.2 is used at Sound and Vision as a reference document for the gradual identification, standardization and implementation of this category of metadata in the actual workflow and systems. This function is expressed in two concrete applications: the newly purchased Media Asset Management System (MAM) logs and stores the outcome of the file quality control as technical preservation metadata attributes defined in the PMD. The second application is the use of the technical attributes in the PMD as an initial list of transformational migration properties. It is essential that along with the absolute significant properties as defined per archival objecttype, the properties per file that may vary per migration be identified, to be able to check after the event whether the most important properties of a digital object have remained intact following a migration, so that its authenticity can be demonstrated (see also paragraph 5.3).

The PMD is not a static document. Updates

are made when important internal or external developments (e.g. coming from the developing archival practice, or from a new release of one of the international standards the dictionary is based on) require a change or an addition to the attributes. A new 2.0 version of the PMD is currently being developed, based on a direct mapping of the new PREMIS version 3.0 and the information in the actual systems at Sound and Vision.

## 9.3 FILE FORMATS

### 9.3.1 Video<sup>16</sup>

MXF is an open standard maintained by the AV standards organization Society of Moving Pictures Engineers (SMPTE). The format is intended for professional use and is supported by a large number of different transcoders and editing software packages. Of all ingested materials in this category, standard viewing versions are produced in MPEG4. Standard Definition (SD) Material must be encoded as MXF OP1a, D10-30 or D10-50, the standard for Digital Provision of the public broadcasters, on the basis of the SMPTE guidelines. High Definition (HD) material must be encoded as XDCAM HD422/50MBbps.

### 9.3.2 Audio

The preservation format for audio is BWF. This format consists of the lossless WAV format, supplemented by additional metadata fields.

### 9.3.3 Film

The chosen format for the digitization of 16 and 35mm film is DPX.

For the mezzanine: XDCAM/MXF

<sup>13</sup> [http://www.ifla.org/files/assets/cataloguing/frbr/frbr\\_2008.pdf](http://www.ifla.org/files/assets/cataloguing/frbr/frbr_2008.pdf)

<sup>14</sup> Common Thesaurus Audiovisual Archives (GTAA): <https://sites.google.com/a/beeldengeluid.nl/gtaa/>

<sup>15</sup> Preservation Metadata Dictionary Sound and Vision V1.2 (2013)

<sup>16</sup> File specifications NISV V1.1 (2015)

#### 9.3.4 Text, subtitles and photographs

Text files accompanying programmes are stored as standard in PDF format. For additional subtitling files accompanying the MXF, .890 and .STL are used in EBU format<sup>17</sup>. Photographs and paper are scanned as TIFF.

#### 9.3.5 Websites and Internet video

The French Sound and Vision partner IMR is responsible for the crawling of the selected websites, which are stored as ARC (or: WARC) files. Investigations are still underway into a preservable format for Internet video. It is expected that an intermediate format between high and low resolution will be selected. All already collected Internet videos will at some point be transcoded to this standard still to be determined, after which the originals will still be stored for some time.

#### 9.3.6 Computer games

Over the coming policy period, Sound and Vision will be experimenting with the collection and preservation of games. If broad access is required, the most obvious strategy will be the storage of so-called disk images that can be played in an emulated environment. A disk image is an archive file that contains the structure of a storage medium (hard disk, tape drive, floppy disk, optical disk, USB card). A disk image is produced by making a 'sector-by-sector' copy of a storage medium. The required technical environment, as well as the specific preservation metadata necessary to sustain these objects under an emulation strategy, is subject to further research.

NB The websites, games, video art, internet video and the other interactive media objects are as yet not ingested, stored and made available within the controlled infrastructure of the Digital Archive. They are processed in a separate environment. Awaiting the outcome of research and preservation pilots, these object types currently own the status of candidate preservable objects.



<sup>17</sup> <http://tech.ebu.ch/docs/tech/tech3264.pdf>  
<http://tech.ebu.ch/docs/tech/tech3285.pdf>



## STORAGE POLICY<sup>18</sup>

# 10

Digital archive masters are primarily stored on LTO tape. LTO was selected because of its reliability and cost efficiency. To be able to deliver recently ingested material faster (on a temporary basis) than is possible from tape, a disk cache has been created.

Tapes are loaded into a Storagetek tape robot. The administration of master files takes place in the storage management system DivArchive, that is responsible for ingest, integrity monitoring, copying actions, access, logging, allocation to tape groups, the allocation of resources, partial restore and checks on location of the stored files. The storage management system is owned by Sound and Vision and is managed by a professional external party with whom Sound and Vision has laid down detailed SLAs, also in respect of checking and managing disruptions.

Files are recorded on tape in AXF (Archive Exchange Format), an open standard for exchange between storage systems. AXF offers space for a limited set of metadata, such as the checksum and a partial workflow history. Each collection is recorded on its own tape group. This is done to ensure maximum control over the storage policy for each collection. Metadata are stored and accessed via the Media Asset Management System (MAM) that offers a search interface, and is responsible for delivery to users. On the basis of the defined workflows, errors in the processes can be detected. Using the MAM system, staff are also able to check whether the ingest has succeeded and is complete. If manual intervention is required, tasks can be created in the system to correct errors.

### 10.1 BACKUP AND RECOVERY

Backup guarantees depend on the preservation level agreed with the depositor and/or the appraisal of an item or collection that implies a specific preservation level. As soon as the ingestion process has been successfully completed,

responsibility for the master files is accepted by the Digital Archive, and a predetermined level of backup policy is implemented. The material is stored at two different locations. The DIVA storage management system implements and monitors the backup policies and copying actions. As part of this process, it is checked if backups have been produced in time. The underlying principle is that as soon as resources become available, the backup is created as quickly as possible following ingestion. Recovery is necessary if a tape fails to work (during a migration action for example, or when an order is received, it may emerge that a tape is no longer readable). The backup can then be retrieved for copying. Each day, the system management department makes a backup of the catalogue with the descriptive metadata.

To be able to guarantee permanent and ongoing service provision, extra copies are made of the proxy files, on disk and on tape. The aim of these backups is to reduce costs if, in the event of loss of the proxy files, transcoding has to be started all over again from the MXF. In other words, the backup is not intended to guard against the possible loss of objects themselves, as is the case for the preservation masters.

### 10.2 MIGRATION

Every five to seven years, all tapes are rewritten to a new version of LTO. A migration plan is drawn up for this process, which is harmonized with the Information Strategy Board and the Management Team of Sound and Vision. The plan also includes an estimate of future volumes. This is carried out to ensure that the storage facility remains sufficient for storage in the period between migrations. Depending on the state of technology, it is possible to opt to skip an LTO version number. The migration of the tapes is carried out within the store management. Checks are carried out according to checksums to determine whether files have been correctly copied and whether files have been correctly copied or whether all files eligible

<sup>18</sup> OAIS-compliant storage policy in relation to the operation of the current storage management system at Sound and Vision is described in the document "Storage binnen OAIS, normatief model en gapanalyse" V1.0 (Storage within OAIS, reference model and gap analysis), D. Steinmeier (2013)

for migration have actually been included and the correct backup policy has been implemented. These actions are logged and recorded by the storage management system in the metadata in AXF format.

### 10.3 RETENTION

Files can only be permanently removed from the Digital Archive by deleting complete tapes. In all other cases, the files can only be discarded by not including them in a subsequent migration. Authorized persons carry out these delete operations. A rollback (i.e. retrieval from a removed tape) is possible in emergency situations, on condition the tape has not been deleted and no migration has taken place in the meantime.



# TECHNICAL INFRASTRUCTURE

# 11

## 11.1 MAINTENANCE AND SECURITY<sup>19</sup>

Sound and Vision is responsible for maintaining the infrastructure (including the Media Asset Management system) necessary for preserving the Digital Archive. The organization ensures that this infrastructure complies with professional quality requirements in respect of availability, operational reliability and security. The following measures have been introduced to secure the infrastructure and the (personal) details of users against data loss, damage and abuse.

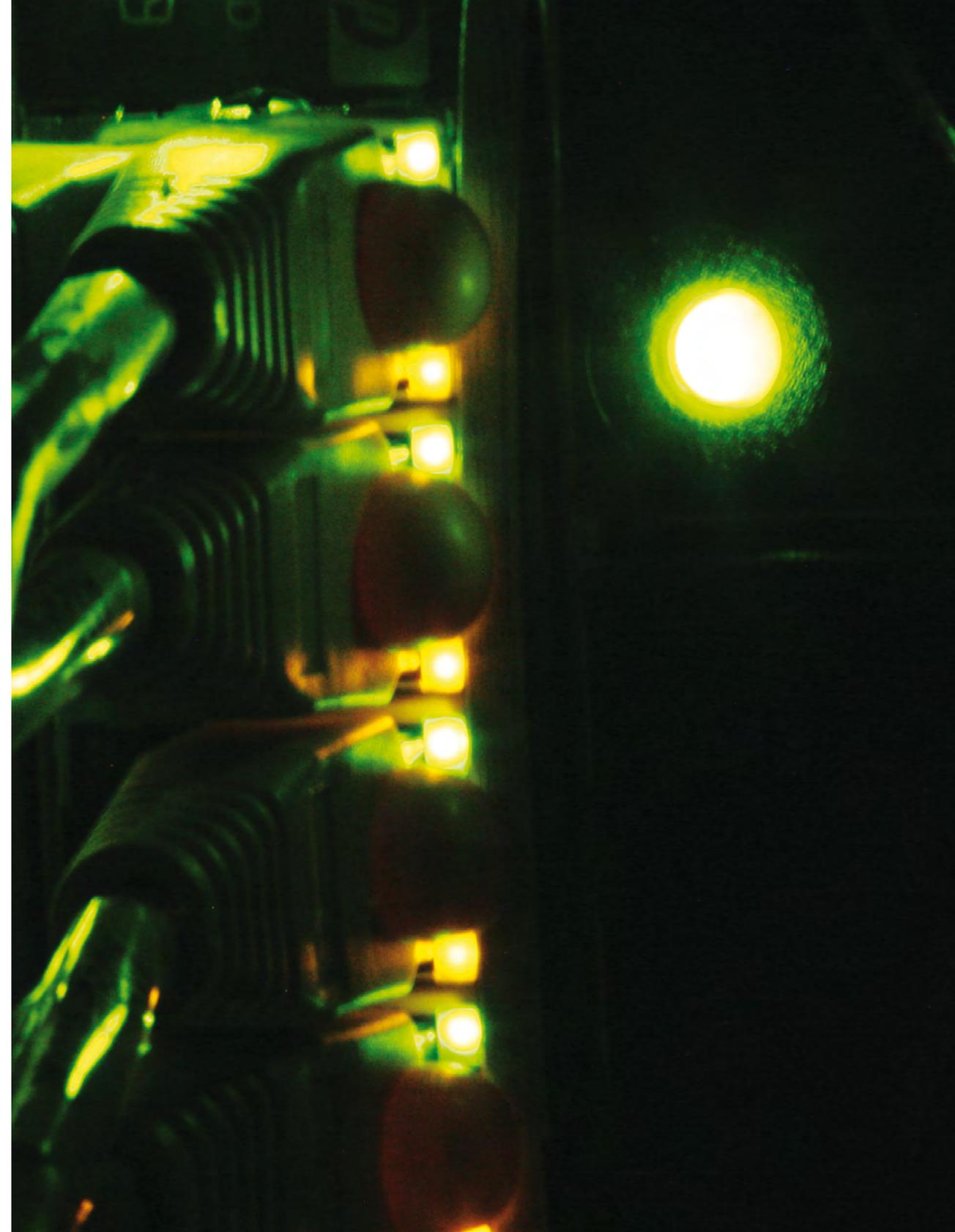
- In the datacentre, at two locations (Sound and Vision and the company Ericsson), an alarm system has been fitted for theft and fire, that is linked to a security organization and/or the police and fire brigade.
  - The datacentre has high-quality electricity, UPS, emergency cooling and emergency power supply to ensure that in the event of power failure, the access to materials can be continued for some time.
  - To prevent physical access by unauthorized persons, the servers that form part of the infrastructure are at a secure location in the datacentre.
  - Sound and Vision protects the network within which the Digital Archive is located by means of a firewall. Access to servers within the firewall is subject to IP *whitelisting* so that unauthorized persons cannot penetrate the network.
  - Authentication for access to servers in the infrastructure is based on user name and password.
  - Users are not granted any rights to which they are not entitled on the basis of the applicable agreements (for example rights to change or delete data).
- Users are not allowed to make changes to central network facilities or the software supplied.
  - Sound and Vision makes a daily backup to tape of its key systems (web applications, [meta]databases, search software and operating systems).
  - Sound and Vision checks whether the planned backup procedure was correctly implemented.
  - Following correct delivery (according to the agreement with the depositor) the material is stored on data tape in the Digital Archive. As quickly as possible thereafter, a copy is made of the material, to another tape at another secure location. In this way, material redundancy is achieved. This tape can then be stored outside the tape robot, at the discretion of Sound and Vision.

<sup>19</sup> *OAIS-compliant security policy for Sound and Vision is described in the document "Information Security in the Digital Archive at Beeld en Geluid, normative model and gapanalysis V1.0, B. Delaney, D. Steinmeier (2014)*

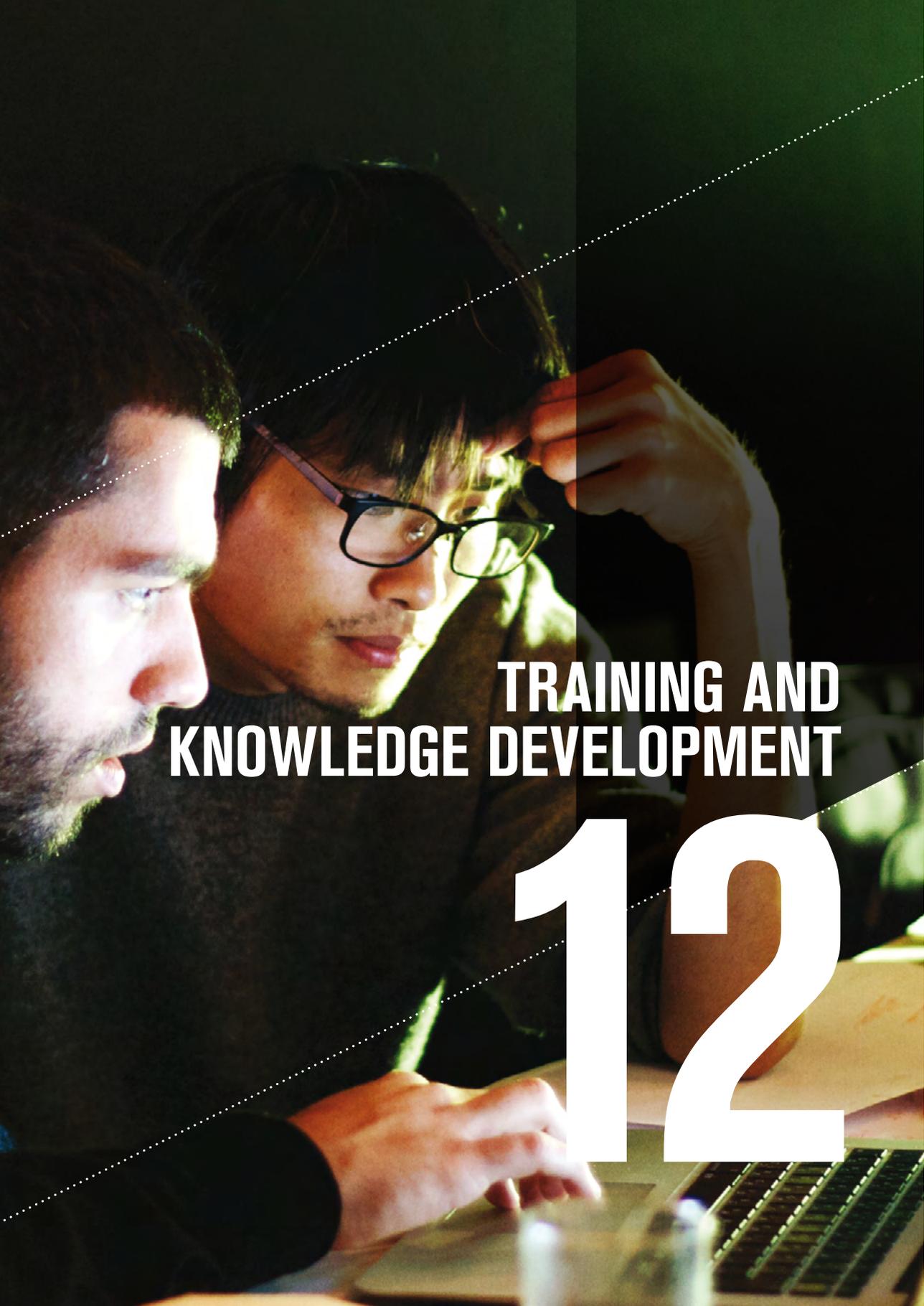
## 11.2 CONTINUITY<sup>20</sup>

To ensure minimal disruption to services and to keep the systems up-to-date and error free, staff at Application Management and the Customer Contact Centre of Sound and Vision carry out the following tasks:

- Proactive monitoring of the infrastructure status.
- “Problem, change and release management”, in this case timely identification of possible shortcomings and implementation of (technical) application management tasks such as configuring the infrastructure.
- Identifying trends, frequently occurring incidents and mutual links and causes; if possible reproducing registered incidents in a test environment.
- Monitoring the storage and processing capacity of the infrastructure and identifying capacity problems.
- Optimizing the infrastructure.
- Issuing authorizations for user connections.
- Creating daily backups of the infrastructure; in emergencies, recovering backups and the secured storage of these backups.
- Building and testing *workarounds* and/or releases, installation and testing in the test environment and installation and implementation of the workarounds and/or releases in the production environment in consultation with suppliers. Purpose: retaining functionality.
- Implementing (and ordering) essential maintenance on the infrastructure and if necessary replacing the equipment or its components.
- Incident recovery in the infrastructure.
- In consultation with suppliers, installing security updates for standard package software that is part of the equipment. Purpose: retaining functionality.
- Implementing the recovery of incidents within configuration management.



<sup>20</sup> OAI-compliant disaster recovery policy is described in the document “Disaster recovery in the Digital Archive at Beeld en Geluid, normative model and gapanalysis” V0.9, B. Delaney (2014)



# TRAINING AND KNOWLEDGE DEVELOPMENT

# 12

Sound and Vision works in a number of different ways to promote the knowledge and competences necessary for working within the processes and systems of the Digital Archive. According to the current 'integrated management' principle, every Unit Manager is responsible for identifying and developing the necessary knowledge and competences within his own unit. Budgets for education and training at individual or group level are allocated by Unit Managers, on an ad hoc basis. At present, the management decides which expertise can normally be insourced, and which expertise must be developed and/or retained within the organization. In certain SLAs with contract parties, specifications are included as to which knowledge and competences the party in question must guarantee to have in-house, in order to adequately provide the service.

Prior to the reorganization in 2015, new job profiles were prepared for staff in the various units at Sound and Vision<sup>21</sup>. Wherever relevant, in these profiles, the competences required for digital archiving have been included. In this connection, the job profiles refer in particular to Media Ingestion Managers and Coordinators, Media Access Managers, Domain Managers, Preservation Officer, specialists in collection policy and copyright, specialists in user research, ITC maintenance and management staff, metadata management and ICT development.

Those areas of knowledge that are important for staff within the digital preservation environment (depending on the job group) have all been identified and broken down into a number of categories:

1. *Knowledge of the essence* (e.g. transcoding/normalization for digital services and acquisition; input format/ output formats, QC, validation, headers, audio levels, video levels; knowledge of digital carriers [hard disks, tapes]; knowledge of technical metadata, knowledge of transport protocols for essence and the applicable fixity agreements; knowledge of old,

analogue formats).

2. *Knowledge of metadata* (e.g. descriptive metadata and preservation metadata; information management and digital lifecycle management; APIs/web interfaces, in this case knowledge of XML and mapping; knowledge of (Persistent) Identifiers).
3. *Knowledge of the workflows* (e.g. knowledge of OAIS-compliant workflow/ processes, general knowledge of the components of the hardware and software environment in which the workflows take place; data management/lifecycle management; knowledge of preservation metadata).
4. *Knowledge of copyright* (as it plays a role in all aspects of the management and use of the digital collection, knowledge of drawing up agreements with depositors and users).
5. *Knowledge of contract management* (e.g. knowledge of digital services and digital management; knowledge of the standard contracts, SLAs, Submission-Order Agreements, in this case agreements with depositors and users, including the relationship with the resultant workflows and preservation levels, knowledge of cost models; knowledge of exit agreements).
6. *Knowledge of collection policy* (e.g. knowledge of valuations, disposal policy/ selection and retention policy; knowledge of the preservation workflow, preservation levels, preservation planning).
7. *Knowledge of users* (e.g. technical/payout requirements and requirements of Designated Communities and methods for monitoring and user survey in relation to selection/retention and preservation planning).

<sup>21</sup> *Job Profiles NISV Unit Archives (2015)*

8. *ICT knowledge in general and in the AV archive context* (e.g. knowledge of AV formats, hardware, software, networks, IT standards, OAIS technology requirements, backup policy, disaster recovery, security and IT Risk Management; preservation planning and preservation levels; knowledge of storage technology and security).

Much knowledge has been recorded over the past few years in standard policy documents about data, metadata, processes and standards, all drawn up to make the Digital Archive OAIS compliant. The relevant process and procedure descriptions are now available on the basis of resultant best practices. The specific dissemination and development of the normative and practical knowledge necessary for working in an OAIS-complaint preservation environment occupies a permanent, financed position in the Digital Archive Optimization programme lead by the Preservation Officer and the Sustainable Archive Theme of the Development Dpmnt of Sound and Vision. Within these programmes, in 2016, further steps will be taken to ensure the availability throughout the organization of relevant documentation and information about training and educational possibilities including webinars, practical workshops, conferences, etc.

The subject 'digital preservation' also has a permanent position on the agenda of the Knowledge & Innovation department. Via this department, Sound and Vision is involved in national and international knowledge and research projects in the field of preservation. Research themes have been formulated in respect of OAIS-compliant workflows, Monitoring Mechanisms, Preservation Metadata, Validation and Quality Control, Significant Properties and Version Management of AV objects. Preservation knowledge was also generated during the participation of Sound and Vision in the project cycle, organized by the Netherlands Colation for Digital Preservation (NCDD), in particular by membership of the project groups certification and automatic quality control. Sound and Vision is also well represented in Theme 3

Sustainability, of the national Network Digital Heritage (NDE) programme, that has started in the autumn of 2015. The organization takes part in the projects around persistent identifiers, certification, webarchiving, preservation policy, digital life cycle management and game-archiving.



## RISK MANAGEMENT

# 13

For managing the risks relating to the operation and preservation of the Digital Archive as a whole and in its parts, a number of different initiatives have been organized. From the perspective of digital sustainability, an first inventory report was issued in 2013 based on the OAIS standards, the SPOT model for Risk Assessment, DRAMBORA, the CCSDS Audit and Certification of Trustworthy Digital Repositories Recommended Practice (TRAC) and the normative Quality Requirements for the Digital Archive at Sound and Vision V1.0.

Within this framework consisting of the most important international standards in respect of risks, the relevant management domains for the Digital Archive have been defined, where they relate to potential risks:

1. Organization: governance and viability
2. Personnel
3. Financial sustainability
4. Contracts, licences and liability
5. Technical infrastructure
6. Security risk management
7. Ingestion processes
8. Information management
9. Preservation and storage
10. Access management

Based on studies of existing documentation and procedures following interviews with MT members, the controller and the Board secretary, the policy at Sound and Vision for these ten areas was charted out, alongside the planning and monitoring mechanisms established for managing risks. For each area, the potential risks for the sustainability of the collection as they existed in the year 2013 were identified. The report functions as an initial outline inventory, and represents the starting point for expansion and improvement of the current management mechanisms.

The Customer Contact Centre has also initiated activities related to risk management. Here the focus was pragmatic, and the choice was made to start with a clearly demarcated area: identifying

the specific risks in respect of ICT in as much as they have an influence on service provision. The aim of the activities is to identify and implement the specific risk management actions, aimed at:

- The relationship between costs of services and revenue for the Digital Archive at Sound and Vision
- End-to-end service levels
- ICT single points of failure

The Customer Contact Centre has also recommended evaluation of important incidents (such as the power failure at the Mediapark in 2014), to assess lessons learned for the management of similar risks.

Based on these two initiatives, the Risk Management Mechanisms report and the activities of the Customer Contact Centre, a solid basis has now been established for further identifying the various risk areas in respect of the Digital Archive, and developing and implementing fixed control mechanisms for managing those risks.





6h

12136

7436