

2019 EDITION

DIGITAL PRESERVATION

SOUND AND VISION

policy, standards and procedures



November 2019

Author:

Annemieke de Jong, Digital Preservation Officer, Archives

Photography:

Aad van der Valk, Daria Scagliola en Stijn Brakkee

Graphic design:

Tice Grafisch Ontwerp, Weesp



CONTENTS

<i>Management summary</i>	6
1. Introduction	8
Revision of the 2015 document	10
2. Mission and organization	12
2.1 Mission statement	13
2.2 Public task	13
2.3 Organizational model	14
2.4 Knowledge, research & development	15
3. Strategic policy framework	16
3.1 Multi-year policy	17
3.2 Collection policy	17
3.2.1 Collection typology	18
3.3 Agreements on access	20
4. Preservation principles	22
4.1 Digital collection management	23
4.2 Core definitions	23
4.3 Scope	24
4.4 Archive formats	24
4.5 Preservable formats	25
4.5.1 Preservation master, mezzanine and proxy files	25
4.6 Non-preservable formats	26
5. Preservation strategies	28
5.1 Guaranteeing authenticity	29
5.2 Guaranteeing integrity	30
5.3 Migration	31
5.4 Scenario's for preservation : 'just-in-case' and 'just-in-time'	33
5.5 Preservation of new media objects	34
6. The preservation workflow	36
6.1 Information model	37
6.2 Generic workflow stages	37
6.3 Process details and technical metadata	38
7. Preservation levels	40
7.1 Full preservation and bit preservation	41
7.2 Preservation menu	41
7.3 Preservation actions	43

8. Preservation planning and control	46
8.1 General risk management at the NISV	47
8.2 Risk management in the digital archive	47
8.2.1 Organization and finance	47
8.2.2 Liability	47
8.2.3 ICT	48
8.2.4 Information management	48
8.2.5 Staff competences	48
8.3 Preservation watch	49
8.3.1 Designated Communities	49
8.3.2 Technological developments	50
8.4 Preservation planning	51
8.4.1 Preservation Plans	51
8.4.2 Migration Plans	51
9. Access, discovery, copyright and licensing	52
9.1 Portals and platforms	53
9.1.1 Media professionals, the general public, the educational domain and researchers	53
9.2 Search and discovery	54
9.2.1 Metadata	54
9.2.2 Thesaurus and identifiers	55
9.3 Copyright and licensing	55
9.3.1 Contracts and clearance	55
9.3.2 Licensing and blocking rules	56
9.3.3 Reuse conditions	57
9.3.4 Access agreements and codes of conduct	57
10. Technical Infrastructure	58
10.1 The NISV IT environment	59
10.2 Workflows and media asset management	60
10.3 Storage management	60
10.4 Backup and disaster recovery	62
10.5 Security	63
10.6 Maintenance	65
Appendix 1	66
NISV Formats / Preservation strategies	
Appendix 2	72
Staff Competences	



The purpose of this document is to further contribute to the development and promotion of the Netherlands Institute for Sound and Vision (in short Sound and Vision or NISV) as a leading media archive that has identified sustainable digital preservation as one of its leading business processes. 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 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

In 2016, the first TDR (Trustworthy Digital Repository) certificate has been acquired: the Data Seal of Approval¹. The NISV will apply for a second certificate, the Core Trust Seal (CTS)² in 2019-2020. 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 certification requirements will in turn be used to further improve and implement the policy.

¹ <https://www.datasealofapproval.org/en/assessment/>

² <https://www.coretrustseal.org/>

MANAGEMENT SUMMARY



INTRODUCTION

1

The domain of digital preservation extends to include the complete lifecycle of digital objects, in other words all processes relating to their ingestion, storage and 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 data and metadata objects and the workflows within which they are managed. But digital preservation does not stop with storage and granting access in formalized processes; it also relates to contacts with the outside world. Agreements must be reached between the depositors of collections and those who lay down procedures. Which materials can be accepted and how do the depositors want that material to be preserved by Sound and Vision? There are also various user groups for the Digital Archive. How do these users – 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 for 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 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 continuity, financial and otherwise, the sustainability of the formats and the quality of the way in which the collections

are made accessible, must 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 (Open Archival Information System, ISO 14721). 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 methods are described for each element within this framework, with reference to the accompanying documentation. This documentation may be a standard-related policy document, a users' manual, a model, an elaborated procedure or a standard itself.

REVISION OF THE 2015 DOCUMENT

This document is a revised and fully updated version of the first Preservation Policy Plan, that was published in 2015. The new edition represents the situation at Sound and Vision as it is today (2019). The adapted and extended sections in the plan demonstrate that the work on the construction of the controlled preservation environment has progressed substantially, both in terms of policy and knowledge, and in respect to actual implementation.

- In 2018 the new MAM system DAAN was launched. The implementation of this system and its modules for import, workflow management, quality control and IPR management, affects the execution of preservation operations in many ways, as is described in detail in Chapter 4, 5, 6, 7, 10.
- A number of recent changes at the (internal) organizational level that are associated with digital preservation policies and procedures, is added to the original information in the Chapters 2 and 3.
- In 2017 the Collection Policy was recalibrated to include new collection areas. This necessitated the formulation of policies for the preservation of new media types. The newly developed approach is explicitly described in the adapted Chapter 5 on Preservation Strategies and is referred to in other chapters (Ch. 4, 7).
- A new Collection Policy plan was published in 2019. Under this Policy, there is no hierarchy within the collection components. This new version of the Preservation Policy discusses when and where - based on the concept of designated usage- differences can and must be made in the choice of formats, sustainability guarantees, preservation levels, workflows and metadata, (Ch. 3, 4, 7, 8, 9).
- The subject of preservation risk management is extended and included as part of the re-designed Chapter 8 on Preservation Planning & Control.
- Chapter 8 contains recently developed procedures for the management of the preservation planning process and references new and updated policy documents in this area.
- Formalized access conditions are an important requirements of a controlled and reliable preservation environment. The way in which the institute handles access provisions, manages licenses and protects copyright and privacy, is therefore presented in a new, additional Chapter 9.
- All IT policies and procedures associated with digital preservation management are updated and gathered in one and the same Chapter 10; a section on (information) security policy was added.
- A list of tape formats, file formats and metadata formats currently in use, as well as an inventory of the type of the Digital Archive's staff required knowledge and competences, can in this version be found in Appendices I and II.

The document is now structured as follows: following an outline of the organization, its mission and its strategic policy for 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 in the domain of digital preservation, and the way in which Sound and Vision has translated these operating principles into its own preservation strategies, the format choices, the internal workflows, the preservation services and the preservation planning and control 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.

These sections are followed by a chapter on Sound and Vision conditions and arrangements regarding access, discovery, copyright and licences. The design and implementation of the complex Sound and Vision technical infrastructure is then clarified. IT policies for the Digital Archive are described in the form of an account of the procedures for making backups, disaster recovery, migration and security.



MISSION AND ORGANIZATION

2

artoren

2.1 MISSION STATEMENT

The mission and the vision of the Netherlands Institute for Sound and Vision (NISV) read:

Sound and Vision wants to improve everyone's life in and through media by archiving, exploring and clarifying that media. Paramount to this is the freedom of thought and expression in text, image and sound.

Media is an integral part of the world we live in. Everyone has their own media personality. In 2025, Sound and Vision will be the guardian of and leading institute for journalistic and media archiving and interpretation. Paramount to this are the key democratic value of freedom of expression in text, image and sound, and its creation and communication. To this end, Sound and Vision will actively promote its extensive collection and open it up as much as possible, for everyone – from professionals to individuals. To inspire them and to help them excel. In this way, Sound and Vision will help to build a more media-savvy world.

To achieve this, it will seek as much collaboration as possible and take the lead where necessary. Through its building, its collection and its museum, it will do everything in its power to create free thinking in media. Sound and Vision sees the broadcasters, scientists, government, politicians, educators, heritage organisations, public libraries, business community and creative industry as key partners in making this a success.

The mission and vision imply the task and assignment of Sound and Vision: the collection, storage and provision of access to media heritage and audio-visual collections 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 the mission. After all, by preserving this heritage for the future, Sound and Vision

is able to maintain the digital collections entrusted to it, with a view to ensuring permanent access for various user groups.

2.2 PUBLIC TASK

The Institute for Sound and Vision was founded in 1997 from a merger between the central public broadcasting archive (AVAC), the Broadcast Museum, the film archive of the Government Information Service (RVD-FA) and the Foundation for Film & Science (SFW). Since then the NISV has been formally responsible for collecting, preserving and presenting the national audio-visual heritage of the Netherlands. The public task is twofold: the organization acts as the corporate AV-programme archive for the Netherlands Public Broadcasting Organization (*Nederlandse Publieke Omroep* – NPO, the administrative organization of Dutch public broadcasters) and as a national cultural and historical media institute for education and academic research, and for the general public. The NISV is structurally funded from the government media budget, supplemented by funds, sponsors, EU projects, ticket sales and exploitation of the Sound and Vision building.

The Institute has been mandated to preserve the collections by the government. Furthermore, the NPO and the representative organizations of copyright holders (associations of independent television producers and record producers) have granted permission for preservation and for the cultural and educational usage of the collections. The Institute preserves its collections at various curation levels (i.e. from additional basic metadata to detailed additional description and contextualization, from bit preservation to full preservation), depending on the cultural value of the material, the amount of metadata supplied by the depositor, technical limitations and/or the contracts with the data producers.

The NISV collections consist of digital-born radio and television programmes and analogue and

digitized legacy collections (film, audio, video). More recently, selections of web videos, websites, games, text materials and specific object types have been added. As a (sometimes paid) service, digital collections of external partners from the cultural heritage and media domain may be ingested and preserved.

The collection amounts to over 1 million hours of AV materials. The NISV's digital repository is referred to as the Digital Archive. Its contents (2019) amount to over 26 petabytes, including backups. Currently a backup of the Digital Archive is housed by a professional commercial partner at the Media Park (SLA available on request). A second backup has been implemented at the Royal Library.

The NISV identifies several external user groups or Designated Communities: Media Professionals, Teachers and Students (at all educational levels), Heritage Professionals, the General Public and Researchers. Depending on copyrights, licenses and authorizations, these user groups may consult, order and/or re-use material from the collections, each via their own platform.

2.3 ORGANIZATIONAL MODEL

Sound and Vision's Digital Archive has been inspired by the OAIS reference model³ in the structuring of its organization, technical infrastructure and functions. The functions Ingestion, Access, Storage and Data (Information) Management are implicitly embedded in the Institution's organizational structure. The Administration functions are currently a split responsibility between the NISV's Information Governance Board (IGB)⁴, the ICT department and preservation management staff. Preservation planning is a shared task of preservation management and the ICT department. Preservation watch is being performed by ICT staff, by the NISV controller, by Exploration staff and by the Customer Service Centre and the media managers.

In total, approximately 30 FTEs work on various digital preservation processes, in different departments throughout the organization. Selection, ingestion and transfer workflows are conducted by the media ingestion managers and by ICT staff. Storage processes and application management are the responsibility of ICT. Information managers and metadata managers control the workflows and the metadata. Providing access and customer contact is taken care of by the media access managers and by the Customer Contact Centre. All the required skills of the employees that work on digital preservation have been defined (see Appendix II). Skills and competences are regularly updated by in-house and external training and knowledge sessions.

In 2018, Sound and Vision restructured the organization to completely focus its products and services on specific target groups. The target group oriented model was selected to enable the activities to be guided by the requirements of the Institute's customers. It involves employees with similar interests from different departments, and employees with different knowledge and competences, working together in multidisciplinary teams at projects or activities, set up for different target groups. The model aims to create strong connections between collections, customer requirements and the available resources (both equipment and personnel) so as to prevent compartmentalization of the organization.

'Target group managers', each dedicated to their own Designated Community (i.e. Media Professionals, Academic Researchers, Heritage Professionals, the General Public, Educational User Groups) are ultimately responsible for NISV products and services to meet the needs of their target group. A larger group of 'product managers' form the multidisciplinary teams, to develop and optimize specific products and services, thus achieving the associated targets with Key Performance Indicators (KPIs).

Department managers manage the staff and provide the appropriate processes and working methods within their teams. These managers are

³ <http://www.oais.info/>

⁴ *Assignment and Structure Information Governance Board NISV (2019)*

⁵ <https://www.beeldengeluid.nl/kennis>

responsible for the operationalization of certain high-level policies, as is the case with 'digital preservation', the explicit duty of the manager of the Archive department. Furthermore, curators play a key role in the development of the Sound and Vision collection and in the provision and use of that collection in the various products and services. A chief editor is ultimately responsible for the scope, quality, relevance and reliability of the content of all products and services.

2.4 KNOWLEDGE, RESEARCH & DEVELOPMENT

As an institute for media, consisting of both an archive and a museum, the NISV plays a key role in documenting, interpreting and presenting Dutch socio-cultural history, as recorded in media. The NISV leads and takes part in several national and international research and development projects that focus on innovation in archival and access functions.⁵ Based on practical work and research, the Institute develops, collects and actively disseminates knowledge and information on media history and on all aspects of digital audio-visual archiving via a series of national and international platforms.

Much knowledge has been recorded in standard policy documents about data, metadata, processes and standards, drawn up by preservation policy staff in preparation for the implementation of the controlled preservation environment. The relevant process and procedure descriptions are now publicly available⁶, based on resultant best practice. The development of further normative and practical digital preservation knowledge has meanwhile been made a structural responsibility of information- and preservation management staff. Annual activity programmes involve quality management, preservation planning, techwatch, preservation metadata development, certification trajectories, and the development of guidelines for the ingestion and preservation of collections of (national) non-broadcast organizations. Legal staff have been appointed to handle privacy protection

and digital rights management. In collaboration with the ICT department and the Archive department, all the necessary tools, documentation and knowledge are being built for the further implementation of the OAIS-compliant preservation environment.

The availability of preservation knowledge throughout and outside the organization is ensured by the knowledge workers of the NISV. They are responsible for structuring and disseminating the information on the organization's public publication site, in a Knowledge bank⁷, and at various internal and external workshops and conferences. Activities and priorities of the knowledge workers are documented in a Knowledge Policy Plan. The subject of digital preservation also figures permanently on the agenda of the Exploration department. Via this department, Sound and Vision is involved in national and international knowledge and research projects in the field of preservation. Research and development themes have been formulated in respect of preservation metadata, web archiving, game archiving, linked data, data analysis and alternative preservation formats. At national level, preservation knowledge is generated and disseminated by the participation of Sound and Vision in the Sustainability Work Package (*Werkpakket Houdbaar*) within the two-year project cycles set up by the national Network Digital Heritage (NDE).⁸ Sound and Vision functions as the national AV node in this network. The Institute actively takes part in the planning and realization of various sub-projects and initiatives concerning the promotion of expertise on digital preservation, the setting-up of common preservation services and the development of national preservation research programmes. Furthermore, the institute is a founding member of AVA_Net, the national audiovisual heritage network. Sound and Vision performs a key role in the execution of the core task of AVA_Net: bundling and exchanging knowledge on all aspects of audiovisual archiving, within and between the larger and smaller organizations and archives that hold audiovisual materials.

⁶ <http://publications.beeldengeluid.nl/> (Governance section)

⁷ <https://www.avanet.nl/kennisbank/>

⁸ <https://www.netwerkdigitaalervoed.nl/activiteiten/digitaal-erfgoed-houdbaar>



**STRATEGIC POLICY
FRAMEWORK**

3

The institute has been mandated to preserve the collections by the government (Media Act 2008, Section 2.1)⁹ Furthermore, the NPO and the representative organizations of copyright holders (associations of independent television producers and 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 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 the collections (or parts of them) must remain accessible in the long term.

3.1 MULTI-YEAR POLICY

The Multi-Year Policy Plan¹⁰ sets out Sound and Vision's strategic ambitions in respect of collection development, accessibility, long-term storage and preservation and the national audio-visual hub function of the Institute. One of the ambitions in the plan is to establish a controlled preservation environment guaranteeing the future-proof storage of the collections for the depositors, and permanent access for users, formally certified as a Trustworthy Digital Repository (TDR). This transition to a professionally managed Digital Archive has taken on solid form in the current policy period 2016-2020. The NISV was awarded the Data Seal of Approval for Trustworthy Digital Repositories¹¹ in the spring of 2016 and will be preparing an application for its successor, the Core Trust Seal (CTS)¹².

3.2 COLLECTION POLICY

The Sound and Vision Collection Policy Plan 2019¹³ 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 sets out a collection profile and explicit clarification of the policy principles for selection, preservation and access to the collections. All the user groups are identified.

In 2017 the Collection Policy was recalibrated to include new collection areas such as online media, games, websites and written press products. A new division of the collection was created as a basis for a different design of the selection and acquisition process for works and groups of works that meet a certain stylistic or substantive criterion. The underlying idea is that in the current media landscape – from a cultural, economic and political point of view – the content and functions of media productions are more significant from the perspective of collection building, interpretation and storytelling, than their technical form (medium or platform). The collection is split up into four main categories, each of which is subdivided: 1) News, Information and Current Affairs, 2) Culture and Entertainment, 3) Amateur and Business Productions, and 4) the Media Landscape (production, distribution and reception of media).

⁹ Act 2008, art. 2.1: <https://wetten.overheid.nl/BWBR0025028/2018-05-30>

¹⁰ Multiyear Policy Plan 2016-2020: <http://files.beeldengeluid.nl/beleidsplan/2016-2020/>

¹¹ <https://www.datasealofapproval.org/en/assessment/>

¹² <https://www.coretrustseal.org/>

¹³ Collection Policy Plan 2019

The NISV applies the following criteria to the appraisal and selection of the collection:

1. Intrinsic features:

- a. Historical (including cultural history and media history): special and important people, events, places, activities in the past, time, processes, lifestyles.
 - Artistic: works by individual film, radio and TV makers, draftsmen and journalists who are of exceptional quality and/or characteristic of a particular innovation, style, movement, oeuvre, etc.
 - Information value: important for academic and scientific research and study.
- b. Social and societal
 - Social: current relevance to individuals and groups in Dutch society.
 - Experience (collective memory): events and media moments that are recognizable to large parts of society and with which many people identify themselves.

2. External features

- a. Usage
 - Museum value: suitability for use in presentation, education and research in an organization's museum, on site and online.
 - Media value: suitability for re-use by media professionals, amateurs and the public.
- b. Special characteristics (that enhance final value/valuation)
 - Origin in a special source or collection.
 - Ensemble value: complete collections, addition to/completion of existing collections.
 - Rarity (uniqueness).

A brand new Collection Policy was published in 2019. This document describes the broadened collection profile of Sound and Vision. Under the Collection Policy, in principle all incoming and acquired digital media forms are subject to preservation. There is therefore no hierarchy: the various collections (i.c. audio-visual materials, digitized paper archives, context collections, games, websites,

web videos, written press materials) are no longer divided into 'core collections' on the one, and 'supporting collections' on the other hand. This is a logical approach when it comes to the cultural-historical appraisal of the collections. In practice (preservation), however, differences can and must be made, whereby (based on usage by the Designated Communities) collections or collection items are linked to suitable formats, different sustainability guarantees and preservation levels, and therefore to particular workflows and metadata. For this reason, collection policy and preservation policy are inextricably linked. They cannot develop separately from each other.

3.2.1 Collection typology

From the perspective of preservation and provenance, the collection can be classified as follows:

A. Digital-born collections: daily ingestion of radio and television

Radio and television productions from public (and commercial) broadcasters. To some extent these productions have intrinsic cultural and historical value. The material is selected for long-term preservation because of its cultural and historical value, re-use value and/or research value.



This material includes:

- 100% of 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).

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

- A selection of broadcasts from commercial stations: radio 1,000 hours and television 500 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, video and audio material mainly consisting of productions from public broadcasters through to the year 2007 (i.e. the start of digital broadcasting production). A smaller category consists of acquired films, e.g. films produced by amateur filmmakers. Most of the material has been digitized in the framework of the *Beelden voor de Toekomst* project (Images for the Future 2007-2014). 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. Games

In 2015 Sound and Vision took its first steps towards a games collection. This collection currently comprises Dutch games mostly from the 1980s and 1990s. The games have been acquired as physical carriers, such as tape, floppy disk or CD-ROM. In most cases these carriers were donated to Sound and Vision by the producers of the games, which allowed the Institute to make arrangements with them about the way in which, and the legal conditions under which, Sound and Vision can exploit particular games.



D. Web videos

Web videos on rising YouTube stars, trends, parodies, political campaigns, branded content and social criticism. Sound and Vision's web collection so far consists of approximately 8,400 videos selected from 2006 onwards. Formal agreements with professional producers of web video have been set up in order for these materials to be legally acquired, and to be accessed on site. For non-professional content an opt-out letter is being sent to the maker of the video.

E. Websites

Websites on media culture are collected on a regular basis since 2014. Currently, Sound and Vision holds a collection of approximately 300 archived sites. The aim is to establish a webarchive that contains a representative selection from each of the collection domains defined by Sound and Vision, insofar as these domains have an online presence on broadcasters' websites, fan pages, forums and blogs and in interactive online documentaries. The selection process will take the contents of other Dutch web archives, such as the Royal Library, the National Archive and the regional archives into account. Sound and Vision has pragmatically decided on an opt-out agreement with website owners. Due to copyright restraints, the web archive can only be accessed on site.

F. Digital media collections from third parties

Collections from Dutch media and heritage organizations with a deposit agreement that make use of selected ingestion, storage, preservation and access options within the infrastructure of the Digital Archive. At present these include parts of the AV collections of Amsterdam City Archives, the backup of the Royal Dutch Library collection and material from the Dutch Premier League/ ECV. The Digital Archive also preserves 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 solely a technical hosting relationship is maintained, the material is not made available for discovery or re-use.

G. Press collection

This collection provides an overview of the Dutch press since 1630. It includes more than 40,000 Dutch political cartoons and press illustrations, and over 20,000 unique Dutch newspaper and magazine titles. Most of the material is donated by news organizations, journalists, cartoonists and illustrators and private collectors. Sound and Vision's press collection is currently - for the larger part - being preserved at the International Institute of Social History (IISG) in Amsterdam. The collection is accessible to researchers and other interested visitors in the study hall of the IISG. It's also searchable and partly accessible through the catalogue of this institute.

The objects acquired are mainly in analogue form but are increasingly digital-born. In order to sustainably preserve the creation process of contemporary cartoonists, more research will be needed into the various digital recording techniques these artists use for their work-in-progress files.

3.3 AGREEMENTS ON ACCESS

Sound and Vision identifies several types of external user groups or Designated Communities. These are Media Professionals, Heritage Professionals, Teachers and Students (at all educational levels), the General Public and Researchers. There are also internal users: employees of the ingestion department, ICT staff, access, information and preservation staff, the museum staff and employees working on knowledge and research projects. All undertakings and agreements relating to the consultation and use of the collections by 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 the NPO, the public broadcasters 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 the CLARIAH consortium (with Surfnet and the Stichting Kennisnet as authenticators), 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, the University of Amsterdam, Vrije Universiteit Amsterdam, the University of Twente, Delft University of Technology, TNO, Utrecht University, Maastricht University and the University of Groningen).





PRESERVATION PRINCIPLES

4

4.1 DIGITAL COLLECTION MANAGEMENT

The common thread in structuring all elements of the preservation workflows and environments at Sound and Vision is the Open Archival Information System (OAIS), ISO Standard 14721.¹⁴ This tried and tested reference model for digital archives offers a framework within which the processes for ingestion, storage, access, migration and delivery are interlinked 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).¹⁵ This document serves as the framework for 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 a national archive, as a corporate AV archive for the public broadcasters, as an 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 the components of this preservation policy are based on the requirements described in this document

4.2 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 uses the following definitions:

Integrity

The object is demonstrably unchanged at bit configuration level from the instance it was stored or migrated. This can be demonstrated by a checksum.

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 its typical characteristics have been preserved to the highest extent possible.

¹⁴ <http://www.oais.info/>

¹⁵ *Quality Criteria Digital Archive NISV V1.1*: <http://publications.beeldengeluid.nl/pub/403>

4.3 SCOPE

The preservation policy extends to include both permanent and temporary preservation of all the files stored in the Digital Archive. This relates to the collections of Sound and Vision proper and to other collections. This latter category consists of materials stored by the Digital Archive as a service to other organizations (for a short period or for the long term, with or without access).

All the components of the collections come within the preservation scenarios elaborated in this document (see Chapter 7: Preservation levels). Agreements on these scenarios with the collective or individual depositors may relate to quantities, quality, storage period, mode of access and user rights, and to the conditions for 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).¹⁶ A standard template has been developed for the actual contract and the SLA's (Service Level Agreements) with depositors.¹⁷

4.4 ARCHIVE FORMATS

As a national institute, Sound and Vision wishes to exert as much influence as possible on the choice of its archive format. The ideal starting point for all objects in the Digital Archive could be a *lossless* format, from which the appropriate delivery formats for the various Designated Communities can be derived, time and again. However, the possibilities for choosing lossless as a standard are in fact determined by practical, political and/or financial circumstances and aspects. An obvious example is the immense collection of NPO broadcast

programmes (amounting to 80-90% of the entire Sound and Vision collection), for which the current broadcast production standard is followed – as both archive and delivery format (which coincide in this case).

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 a national cultural institute and a media archive. This dual role means that, in an increasing number of cases, the choice of preservation format may differ, depending on the cultural historical importance of the materials and/or the requirements of depositors. The focus of effort is therefore to bring the choice of preservation format for these parts of the collection more into line with the archive's own national responsibility and competences.

Although the use of a lossless format for the sustainable preservation of compressed files may be desirable, the advantages (no loss of information) should always be weighed against the disadvantages (storage capacity, processing speed etc.). In the case of large collections, when the risk of information loss is small and negligible from the viewpoint of the associated Designated Community, the NISV will opt for direct migration. Lossless format will in fact only be relevant if the Digital Archive is no longer able to produce a suitable delivery format from the current professional broadcast format. The chance of this happening is small.

At the moment it is not decided which lossless format Sound and Vision should prefer. There are still many fundamental and technical questions to be resolved in the ongoing research into the use of an open, lossless format, intended as a preser-

¹⁶ *Manual for Creating a Submission and Order Agreement* : <http://publications.beeldengeluid.nl/pub/400>

¹⁷ *Template Digital Archiving Service agreement*: <http://publications.beeldengeluid.nl/pub/463>

vation master for collections of non-broadcasters. To this end, a number of pilots are currently being carried out, to provide an understanding of the technical aspects, costs and sustainability as well as the requirements of the various (intended) depositors and user groups.

4.5 PRESERVABLE FORMATS

Sound and Vision has thus far selected a limited number of formats for long-term preservation. The Digital Archive is able to issue detailed guarantees for the sustainability of these formats. The organization applies 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 Sound and Vision's internal technical and catalogue infrastructure, 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, because these are considered to be essential properties for the NISV broadcast collection. In addition, it must be possible to transcode the format to other common formats, using current transcoding software. It must be possible to carry out quality analyses of the format using standard analysis software.

Tentative preservable formats

For the selected preservable formats, Sound and Vision is in principle able to give solid guarantees of sustainability. Meanwhile, the organization has started collecting non-AV media types and other new formats. This category currently contains games, websites and web videos. In order to guarantee sustainable access to these new forms of media, much remains to be investigated and

tested. Formats of this type are qualified for the time being as 'tentative preservable formats'. The aim is to keep these formats permanently playable and usable and in doing so, eventually label them as objects that are guaranteed to be preservable.

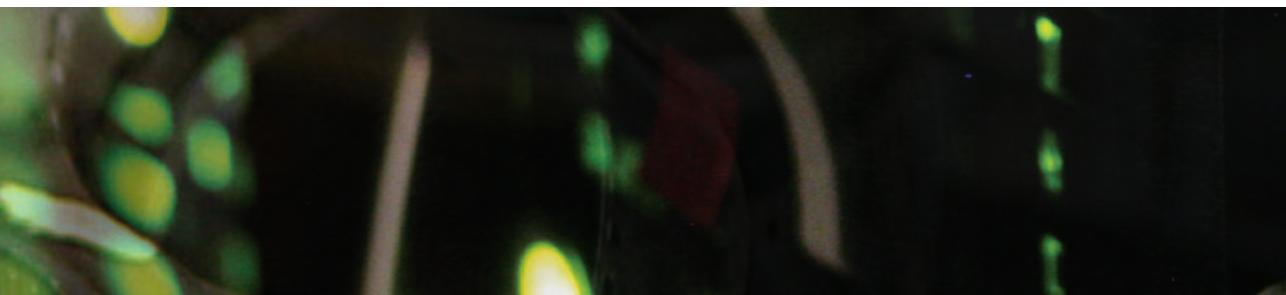
4.5.1 Preservation master, mezzanine and proxy files

Preservation practice at Sound and Vision is based on three types of file, each with a particular 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 institute's own collection or on behalf of a depositor. The choice of the MXF format traditionally goes hand in hand with Sound and Vision's position in the production process of the public broadcasting organizations, in this case its 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 current production format used by the public broadcaster, which must be suitable for immediate delivery to this largest Designated Community of the NISV.

Because this format is submitted by the public broadcasters as a source format, and because, being an industry standard, it is highly suitable for professional re-use, at Sound and Vision, MXF serves simultaneously as the preservation master and the standard delivery format. If certain specified conditions are met, guarantees can be issued for a permanently 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 shown on the various Sound and Vision search interfaces. When users order excerpts, a '*partial restore*' can be made from the MXF (a partial restore is a sequence of an AV file with the accompanying technical metadata, which is delivered separately as a component), and the material can be transcoded.

For a large subcollection of scanned films – in genres representing particular national value – a high-resolution standard has been chosen, namely DPX. This format is employed because film requires a high-quality digital archive format, since the intrinsic quality of film is higher than that of television material.

Another example is the ongoing research into the use of an open, lossless format intended as a preservation master for specific collections of non-broadcast depositors. To this end, a number of pilots are currently being carried out to provide an understanding of the technical aspects, costs and sustainability.

Mezzanine

Due to their size, preservation masters in DPX format cannot be rapidly transported. Furthermore, the content cannot be processed easily by video editors in the broadcasting production environment. As a consequence, the Digital Archive does not supply DPX directly. Instead, it produces a mezzanine, an intermediate format derived from the preservation master. This mezzanine is produced in XDCAM/MXF, the same format as used for HD-TV. 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 under the same conditions as all other MXF files.

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 Sound and Vision's catalogue infra-

structure, i.e. the internal catalogue, the online catalogue for use at the Media Park and the search interfaces for the general public and academic researchers. A number of other organizations (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 Sound and Vision's catalogue infrastructure: in the case of distribution via the internet, it is not possible to continue to offer payout guarantees. Storage, format specifications and the affected software components are no longer under the control of the Digital Archive in that case. Long-term preservation in line with 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. The proxyfiles themselves will in their current form not be preserved for the long term, even though the concept of having a proxy format for easy payout will remain a long term access strategy, set up to serve the NISV Designated Communities.

4.6 NON-PRESERVABLE FORMATS

The main operating principle is that the depositor delivers the file in the preservable format specified by the Digital Archive. The depositor is thus responsible for the quality and correctness of the submission. A second principle states that Sound and Vision can reject the material 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 transcoding or arranging transcoding of the submitted format to the preservable format. This is

only possible if the cost of the conversion is 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. For such time as transcoding can be done using current standard software, new preservation masters can be produced from the original. Once the original format can no longer be played out, the most recent preservation master will acquire the status of original. This also happens in case it is agreed with the depositor to replace the original. The scenario applies solely if a good-quality preservation master of the original *can* actually be produced. In other words, quality requirements also play a role, in addition to the financial conditions. The same standards apply to converting as to regular 'internal' transcoding within the Digital Archive: it must be clear in advance which characteristics of the original file must be retained during transcoding and which 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. A transcode will be made of the files for use on the portals of Sound and Vision. The original files – which may be in 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 will not be possible, or only to a very limited extent, hence the transcode of the files.

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 SLA's.





PRESERVATION STRATEGIES

5

5.1 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 has not been altered unintentionally or without authorization. For this purpose the lifecycle of the object is recorded.
3. 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.

Ad 1. *Quality control*

All incoming objects are checked and validated in line with a number of strict ingestion procedures, whereby a number of quality checks are structurally carried out in separate system modules for import, workflow management and quality control. Automated controls and monitoring mechanisms, together with analysis and extraction tools, guarantee the quality and completeness of the files and the metadata, during every ingestion update and after each update. Sound and Vision's MAM system DAAN also defines the rejection criteria for metadata, as well as a workflow for repairing, logging and documenting errors.

Recently, a quality check was done on *all* legacy MXF files previously stored in the archive, to capture technical metadata and generate quality reports in order to facilitate preservation planning.

Ad 2. *Life cycle management*

Recording all events in the lifecycle preserves the '*chain of custody*' of each object, i.e. the details in the context within which the object is created, ingested, stored and used. These data form part of the category of *provenance metadata* within the 'preservation metadata'. Together, these data form the evidence for the 'credibility' of an object as it has been maintained over time. These data describe where the object came from and who

has processed and used it. Creating a record of this lifecycle is achieved by means of automatically generated metadata, triggered by events such as movements of the object. This information is subsequently added to the object. In this way, provenance metadata demonstrates that the object has not been unintentionally altered. Monitoring and recording all predefined steps in the lifecycle of each individual incoming object in preservation metadata both guarantees and demonstrates the authenticity of the object. 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'.

With the launch of the new Media Asset Management system DAAN in 2018, the formal identification and registration of a large amount of process data is a fact. All files and metadata, from whatever source, go into the same system through standardized ingestion workflows. Additional provenance metadata based on events in the DIVArchive storage system will be captured in the Business Intelligence (BI) tool that is currently being developed. The output of all these processes is logged and structured and can be used as an overview of the life cycle of the files from ingestion through to storage and access.

Ad 3. *Essential properties*

The third aspect of authenticity is sustainable and meaningful access: the object is available in a usable format, and understandable to the users. To be able to guarantee this form of authenticity, the 'essential properties' of an object must be determined. These properties are the technical, aesthetic and intellectual characteristics of files, that must be preserved over time and throughout the various technological changes. The preservation of this authenticity can be measured by the degree to which – following a migration or other transformation – the specified essential properties of the original object (the preservation master) have been preserved in the new form of the object.



The Preservation Metadata Dictionary 2.0¹⁸ contains a list of *all* the technical properties of the currently specified preservable file formats at Sound and Vision. Which of the properties of an original file must be preserved is decided on each individual migration or transformation of objects, formats or collections. The choices and motivation are recorded in separate preservation action plans that are drawn up for each migration. In the case of a format migration, the Digital Archive is able to explain which of these technical properties relate to the more abstract and conceptual characteristics that make an object authentic and significant in the eyes of the user or depositor.

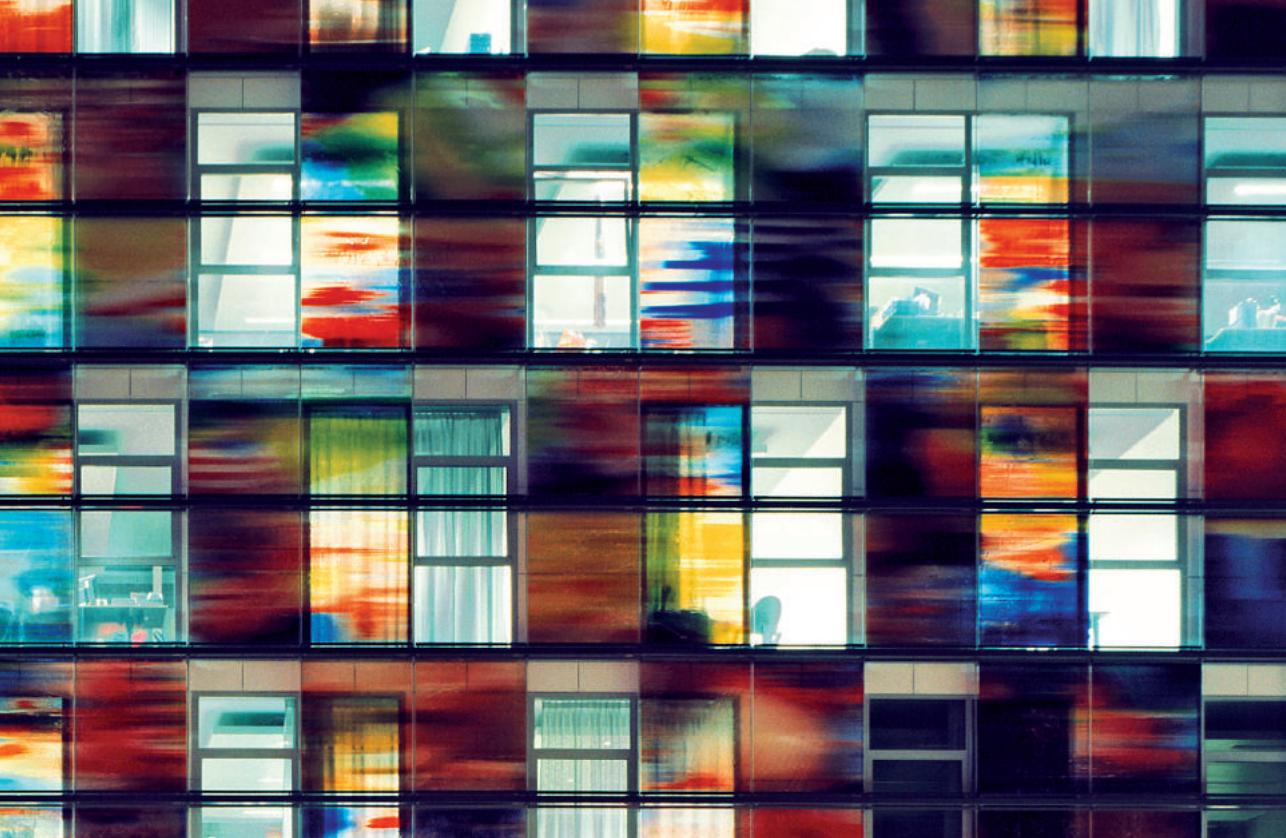
The concept of 'essential properties' is relative and context-dependent. Decisions on what is and is not 'essential' are taken in regard to the intended usage and perception (i.e. for what purpose does the user or depositor want to use the object? In what cases does a user perceive an object as

an authentic object?), Sound and Vision's own Collection Policy (i.e. in what form should the objects be kept for which kind of usage?) and the technical possibilities and limitations of the source format and the target format. Practical issues, such as processing time and cost, may also play an important role.

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 identical bit-for-bit. This data integrity is secured by Sound and Vision through the use of checksums. To create a string that is unique to the object, all the bits in a particular amount of data are processed using a certain algorithm. A checksum is calculated prior to ingestion,

¹⁸ Preservation Metadata Dictionary 2.0: <http://publications.beeldengeluid.nl/pub/615>



or when the AV-object is stored in the Archive Management system. During every processing or movement of a version of the object for storage/restorage, copying, migration or delivery to users, a fixity check is used to calculate the checksum again and compare it to the checksum previously stored. This is done in order to ensure that the object is identical to the data that was transferred in an earlier phase.

For heritage collections and commercial services, checksums are produced by the depositor, prior to ingestion. For the collections of public broadcasters, checksums are not submitted by the depositor. In these cases, the quality check done by the software program Baton¹⁹ is used to check whether the file contains a header and a footer. 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, as it will detect

errors in the container format and the audio and video data. A header-and-footer check can only be carried out on files that actually include a footer. As concerns the preservation formats, this is only the case for MXF at present.

5.3 MIGRATION

The main current preservation strategy operated by Sound and Vision for its own collections and for the materials from other depositors that must be preserved is *migration*. This first of all means that LTO carriers on which the AV productions are stored, are transferred to current carriers at set times, depending on the state of technology. Every five to seven years a complete migration of all the tapes in the Digital Archive takes place, when the objects are once again read through and rewritten. If errors appear during the daily use of the

¹⁹ <http://www.interrasystems.com/file-based-qc.php>



materials (e.g. if it turns out that a tape cannot be read), Sound and Vision operates documented procedures to restore the file from the backup. This preservation strategy arises from the fact that the Digital Archive houses a very large and rapidly growing digital broadcast collection that has to be available for permanent and intensive re-use within the professional media production environment.

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 MXF formats in the large collection of public broadcasting productions will be 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 they were established in 2007.

5.4 SCENARIO'S FOR PRESERVATION : 'JUST-IN-CASE' AND 'JUST-IN-TIME'

The choice of digital AV-formats ingested and stored the Digital Archive has thus far been dictated primarily by the institute's role as the corporate AV archive of the public broadcasters. The broadcast format therefore traditionally functions as an archive format. Sound and Vision only ever made one substantial exception: the format choice for a legacy collection of highly valued 16mm films, as part of an externally funded digitization project (Images for the Future 2007-2014).

Because of the traditional focus on the standards of the broadcast production environment, Sound and Vision has always been able to keep the number of preservable formats limited. Also, non-public broadcasting material, if not already produced or delivered in a preservable format, was generally normalized before ingestion into the Digital Archive. Due to statutory obligations, the format for materials produced by the Dutch public broadcasters will change only if a new format is selected by the broadcasters. In the event of such a format change, all the productions stored in the Digital Archive to date will in principle be transferred integrally, based on a 'just-in-case' scenario.

As a result of the increasing emphasis on Sound and Vision's role as an institute for media culture, its collection policy has recently been expanded substantially to include objects other than mere broadcast AV. The Institute will have to be able to support the ingestion and preservation of more and different types of formats. These formats do not only differ in terms of type and quality, they often have a shorter and more unstable lifespan. Regardless of the use of these new objects now and in the future, however, sustainable preservation and access is as desirable here as it is for the archive broadcast formats that have mainly dominated the Digital Archive until now.

The just-in-case scenario is neither feasible nor desirable for the new media materials that fall within the expanded collection profile. A different preservation strategy will be implemented here: the *just-in-time* approach. This scenario means that the objects will initially be stored 'as is', with playability checked in advance. In a fixed monitoring cycle (for example, every four or five years) the objects will be checked to ascertain whether they are still playable within the current playing environment, taking into account the outcomes of technology watch procedures and the requirements of the user groups, which may have changed. If the object cannot be played out any longer, it will be examined whether another environment offers an alternative, or whether the original format will have to be converted ('just in time') to a

preservation format. In the Preservation Metadata Dictionary it will be recorded which technical properties of the formats are sustainable in the long term and which playback environments are considered to be working. The Sound and Vision systems will be equipped to ingest and store the new media objects acquired: the files, the preservation metadata and the descriptive metadata.

5.5 PRESERVATION OF NEW MEDIA OBJECTS

a. Games

Currently, the games collection consists mostly out of games from the 80s and 90s, games that were distributed on physical carriers. Preservation focuses on preserving the disk images. A disk image is produced by making a 'sector-by-sector' copy of a storage medium. The original physical carriers (tape, floppy disk or CD-ROM) are stored as objects for exhibition purposes. The main preservation method is emulation, since this allows games to be preserved in their playable form and is both platform-independent and fit for scalable and – potentially – online access. This means that, apart from describing the digital files themselves, the software environments on which they depend in order for them to be recreated are extensively documented. The PREMIS²⁰ notion of 'environment' is used to structure the preservation metadata that apply to these environments. Gameplay videos are recorded with the purpose of evaluating the emulation outcomes in terms of 'look and feel'.

Future acquisitions might bring additional types of games, for example distributed online or in app stores. For each of these types, additional container files and executables will have to be checked in terms of their eligibility for emulation as a preservation method. Where games depend largely on online capabilities, or where emulation software is not (yet) available, documentation will play a more prominent role as a preservation method.

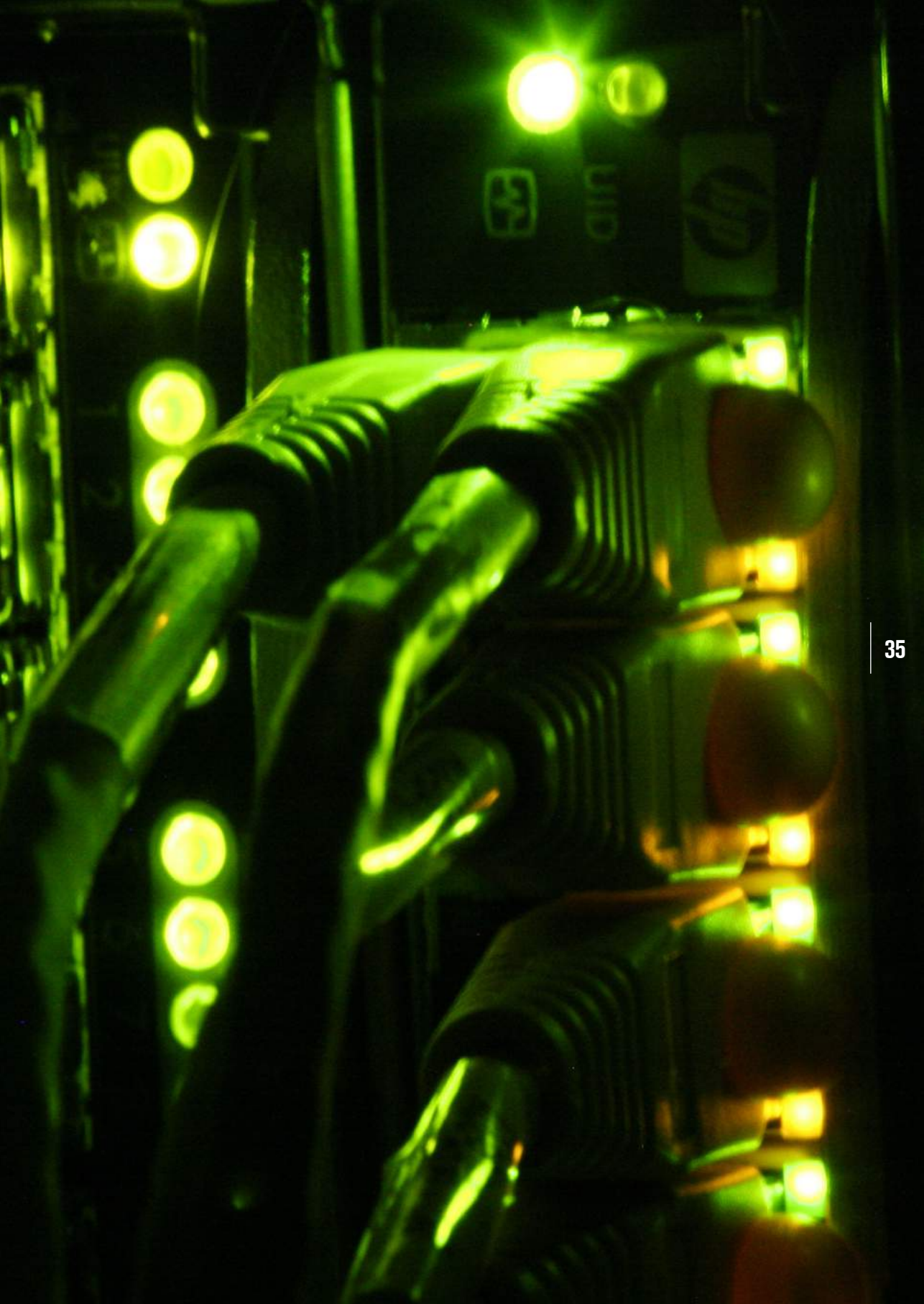
b. Web video

The web videos are collected from the web in various formats. A lossless format to be used for sustainable preservation is under research. Conversion only takes place if the initial format is in danger of becoming obsolete. Until then the files are archived 'as is'. This means minimal loss of information, efficient and sustainable storage, and the expansion of the technology watch cycle by five years. The preservation master is standardized to the format to be selected, with the intention of preserving the original technical characteristics of the file. If the conversion is reversible, it may be decided not to keep the source format. Agreements are set up on retention of the source format, the tooling to be used for the conversion and the technical properties of the files (source format and preservation format) to be recorded. These properties will be documented in the Preservation Metadata Dictionary. They also function as input to the software profile for quality control.

c. Websites

An external supplier, Archiefweb, has been chosen for the Web Archives Dashboard (WAD). Using state of the art crawling software, the selected websites are crawled and saved as a WARC file. In case the WAD tool proves inadequate (for archiving certain types of dynamic content) and the website is of high value, Sound and Vision uses the Webrecorder software, in which a website can be traversed manually. The results of this crawl are also saved as a WARC file. Archiefweb manages the access copies of the entire Sound and Vision web archive. Once a year the institute receives a dump from the web archive and stores this as a backup. In the long run it may be desirable and/or necessary to emulate the old browsers in order to guarantee authentic access to the archived websites. This requirement has been communicated to the supplier as a feature request.

²⁰ <https://www.loc.gov/standards/premis/>





**THE PRESERVATION
WORKFLOW**

6

6.1 INFORMATION MODEL

Sound and Vision has developed an information model²¹ that acts as a reference when 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.

The information model gives a generic description of which workflows are used for the ingestion, storage and access functions. The model records all the 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 enabling verification that all the events in the lifecycle of an object tie in with the preservation policy of the archive. A comparison of provenance metadata and the events in the information model establishes that – if the workflow was completed correctly – no unexpected actions have been carried out on the object. The lifecycle of a digital object thus develops in a controlled and verifiable manner.

To enable the files and accompanying metadata to be identified and managed 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, with added metadata). And what is delivered to users is often only 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 (i.e. Submission Agreements, contracts, SLA's) in which all the agreements are laid down on issues such 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 guarantees that the file was submitted complete and correct by the depositor. The fixity check cannot be carried out if a depositor is not able to supply a checksum. 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. In that case, the Digital Archive itself will generate a checksum, *after* the material has been ingested, in order to enable the integrity of the file to be checked following any future actions.

During ingestion, the format of the object is first determined and technical metadata are extracted. These data relate to the material properties of the AV file (for example, *aspect ratio*, *colour space*, *codecs* and *bitrate* used). Technical metadata extraction is necessary primarily 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 qualify for a just-in-time procedure) can be detected in time and taken into account when planning migration actions. The extraction of technical metadata is also important to enable verification that the formats comply with the quality agreements drawn up with the depositors.

²¹ Information Model NISV: <http://publications.beeldengeluid.nl/pub/389>

An optional component (i.e. depending on the agreements with the depositor) of the workflow is the automated Quality Assurance process, whereby the master files are checked for quality with regard to a certain profile and for some generic features (such as completeness of the footer). During this process technical metadata are also extracted, and the outcome of quality checks is recorded. This metadata is gathered in a separate report that is accessible via the MAM-system.

All the ingested packages, with content and metadata, are then assigned an *identifier*, a unique label that forms a permanent reference to the object stored in the Digital Archive. This stage represents the end of the ingestion phase, and the ingestion packages, referred to as 'SIPs' (*Submission Information Packages*), are ready for final storage. Together with additional files (including subtitling and context information about AV production) plus the technical metadata extracted and any descriptive metadata added manually, 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 Sound and Vision's information model remains limited. The AIP need only be allocated a definitive storage location, itself stored in the metadata.

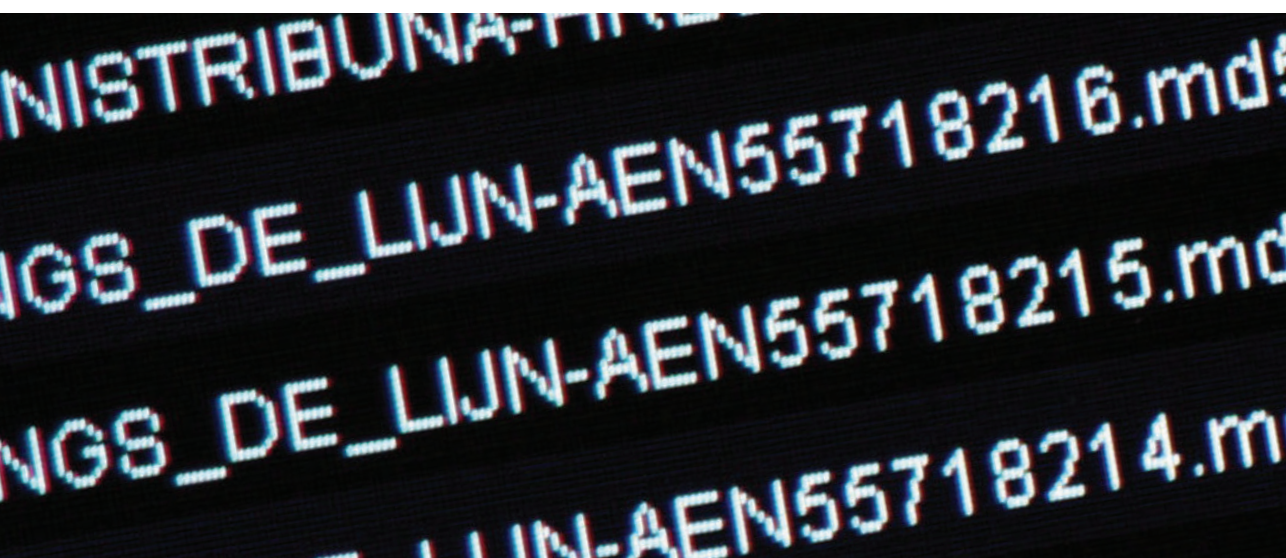
The last element of the workflow describes the actions relating to the *Dissemination Information*

Package or DIP. The DIP workflow normally starts with user authentication, whereby a determination is made of the authorizations of the user registering 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 successful 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. The access workflow is also generically created within the information model, so that a 'request' for access can relate either to simply a 'search based on metadata' or an order for all or 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 automatically generated and logged. 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 this way, both



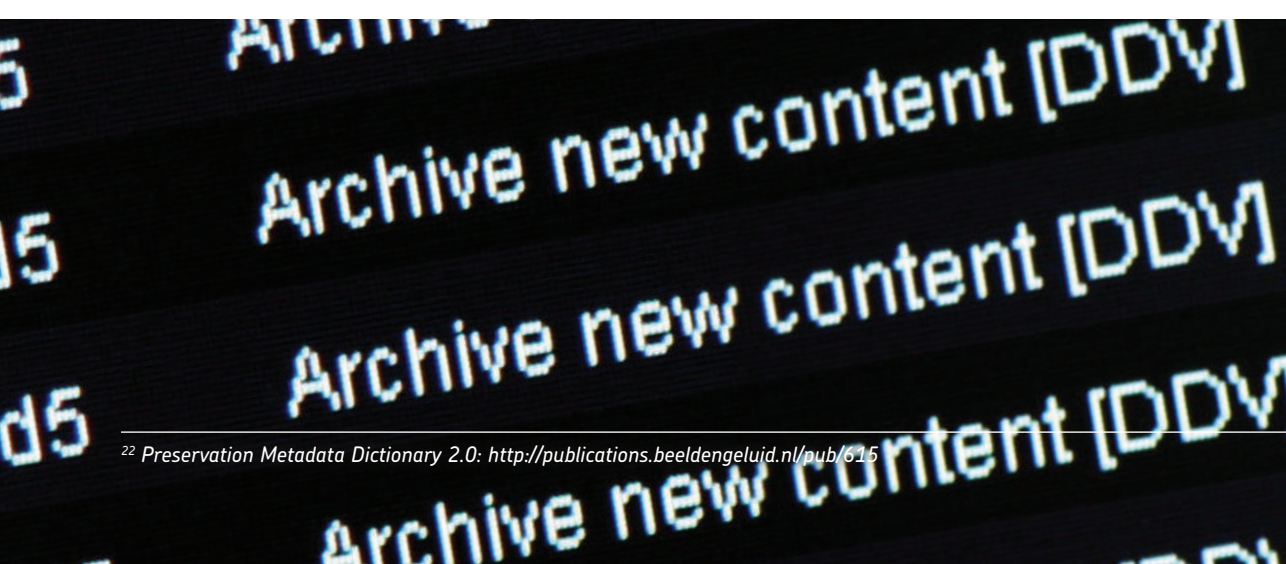
the process data and part of the technical and administrative data (in preservation metadata terms – the events and provenance metadata) about the files are recorded.

The implementation of the new Media Asset Management system DAAN in 2018 has considerably improved the creation and structuring of both process data and technical properties. The output in terms of process data is structured as tasks in a workflow management system, thus making it available as an *audit trail* of the ingestion of one file or a group of files. A new analysis and extraction tool improves checks on the quality and completeness of the files, enabling more technical data about the files to be generated. The new MAM system has established a central facility in which more 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 has been mapped with the form and content of the attributes in the Sound and Vision Preservation Metadata Dictionary 2.0²², the overview of all the technical metadata that may be involved in audio-visual preservation. For each preservation format the dictionary defines what technical metadata, managed in the MAM (as metadata or gathered in text files that themselves are additional files to the master files) or even outside the MAM (in the storage management system), must be available as

preservation metadata.

Finally, a determination will be made of how all this information can be exported from the source systems. This output, the preservation metadata, can then be accessed and manipulated in a structured and standardized manner, for the purposes of accountability and the planning and implementation of the preservation activities.



²² Preservation Metadata Dictionary 2.0: <http://publications.beeldengeluid.nl/pub/615>



PRESERVATION LEVELS

7

7.1 FULL PRESERVATION AND BIT PRESERVATION

The generic preservation workflow previously described 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 (i.e. the applicable collection policy), the requirements of the Designated Communities, the technical facilities, the cost and the copyrights.

There are two main levels for the preservation of files:

1. Bit preservation ('passive preservation')

The file is stored as it is received. Usability for the short and long term is 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. There are two possible preservation strategies to achieve this level:

a: Just in case: the material is stored in a 'preservable' format. The risk of this format becoming obsolete is very small.

b: Just in time: the material is stored as is; the accessibility of the format is monitored; the file will be transcoded to a generic, lossless format only if and when accessibility becomes a substantial risk.

ments relating to the level of preservation are set out in detail in the contracts and the Service Level Agreements with the depositors.

The preservation services Sound and Vision offers always have *full preservation* as a starting point, for the institute wants to be able to make materials available via the recipients of the service and to connect collections. Mere bit preservation offers no possibility to do so. This approach may be followed nevertheless, but is generally seen as a temporary solution, accompanied by customization and extra costs.

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. In the case of all levels the depositor remains responsible for the material, and he 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 kept as backups, at two different locations. Depending on further agreements, these may be offline copies or copies in the tape robot.

Within 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 relate, for example, to the returnability of the original objects, retrievability, the quality of storage and migration, the link with metadata, the extent to which the standards are fulfilled, guarantees of no unintentional changing of objects and sustainable playability.

7.2 PRESERVATION MENU

A 'preservation menu' explains the different preservation scenarios or 'levels' of delivery, storage and access. The menu describes the type of storage service provided, what metadata to submit, what preservation actions are undertaken, and what form of preservation applies. All the agree-

Level 1 scenarios by definition relate to short-term solutions, as bit preservation means that playability is *not* guaranteed in the long term. The level 1 scenarios are broken down into levels 1a to 1c. Depositors can opt for level 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 that is acceptable for full preservation. Levels 1a and 1b are therefore generally not scenarios for material that forms part of Sound and Vision's collections. Levels 1a and 1b 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 NISV collections, so that this material is stored rapidly and safely before being converted to a format for full preservation. This level applies, for example, 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.

The level 2 scenarios (2a to 2c) were developed for the collections stored and made accessible by Sound and Vision in the long term. This refers to material for which the Institute accepts permanent care, either as as part of its own collections or as a service for a smaller or larger heritage organization. The collections concerned can be accessed via the Sound and Vision catalogue. They are indexed and 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 carried out.

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

7.3 PRESERVATION ACTIONS

The guarantees for each level are based on specific actions that are undertaken during the ingestion and storage of the objects. Some actions do not apply to all levels and this way differentiated service levels can be provided.

Explanatory notes

Integrity calculation

A checksum is calculated for the file so that it can be used for integrity checks any time the file is read by the system, for instance during a full restore or during media refreshment actions.

Backup

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

Filename check

Before the import starts, filenames are checked for compliance with agreements. Sometimes the archive creates the filename itself, sometimes

the supplied filename needs to be augmented by the archive, and sometimes the filenames are supplied in full by the depositor and need only be checked.

Carrier replacement

A decision is taken periodically on whether carriers need to be replaced. In the event of such replacement, files are checked for integrity during 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. The backup process 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. This check is performed at collection level, based on lists of content that ought to have been delivered.

Preservation actions per level

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

Integrity check

A checksum (of some audio flows) is carried out prior to ingestion to determine whether the file has been received complete. N.B. Checksum is always the preferred control mechanism because this check provides the most reliable 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.

- *Just in case: material must be in a 'preservable' format. These formats are defined with a specific profile: the technical properties of the file are to be checked during format analysis. The probability of the need for a format migration is very small.*
- *Just in time: if the file is accepted by the sanity check, regular monitoring will check the accessibility of the file. If the risk of obsolescence becomes manifest for this format or this specific version of the format, a migration will be planned.*

Format analysis

The format is analysed and the basic properties are recorded in the MAM system, based on the metadata in the file header. Discrepancies may cause the file to be rejected.

Format and Quality check

A check to determine whether the file has the correct fields and the correct structure in the header. Profile and technical properties (bitrate, aspect ratio etc.) are checked. A limited number of checks are also carried out on the quality of the essence. In the case of MXF this is done by Baton QC. Discrepancies will generate an error regarding technical properties and warnings of quality issues in the log files.

- *Just in case: the formats are checked against a specific profile. The technical properties of the file must be within specs.*
- *Just in time: the files will undergo a sanity check to determine whether the file will be processed by our standard environment. This*

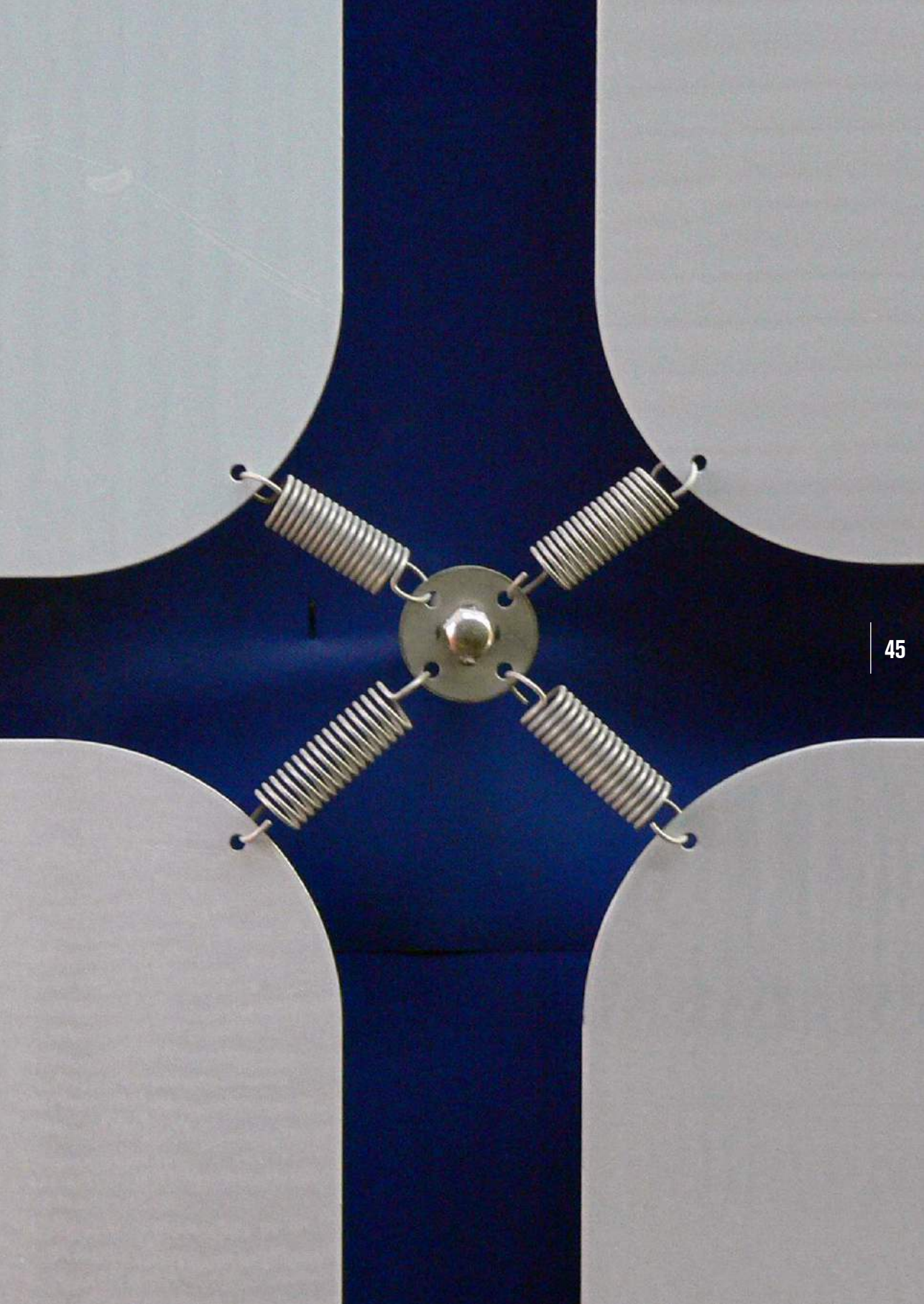
means that a proxy of the essence can be created and that the file will be able to be ingested in our systems.

Metadata content check

A check to determine whether the metadata file (XML) complies with agreed metadata rules for that specific ingestion flow.

Metadata technical check

A check to determine whether the metadata file (XML) complies with the agreed format in technical terms (well-formedness). This check is necessary in order to process the XML.



A collage of film-related items. In the foreground, a white marker with blue text "DIXON CHINA MARKING" lies horizontally. Behind it, a blue folder with a white label featuring the number "2" is visible. In the background, several reels of film are stacked, with a prominent one showing a large circular hole. The overall scene is lit with a warm, golden light, suggesting an indoor setting with film equipment.

**PRESERVATION PLANNING
AND CONTROL**

8

8.1 GENERAL RISK MANAGEMENT AT THE NISV

Risk management at the NISV's enterprise level currently focuses primarily on the financial processes, conducted by the Sound and Vision department of Planning and Control. The Institute's Supervisory Board (*Raad van Toezicht*) includes an audit committee that – in collaboration with the institute's controller – conducts annual tests of the most important financial risks. A number of scenarios are then elaborated, indicating to what extent Sound and Vision's financial stability may be at risk. If necessary, control measures will be implemented.

The NISV's intention is to further centralize and direct risk management from 2019 onwards and to promote greater continuity in its execution. Not only financial risks, but also strategic, operational and ICT risks will be managed more formally. To achieve this, the organization plans to implement a governance and compliance programme based on the COSO-ERM standard. The idea is to involve Sound and Vision's various departments and disciplines at all levels of risk management. One of the disciplines identified is digital preservation.

8.2 RISK MANAGEMENT IN THE DIGITAL ARCHIVE

In terms of risks associated with the continuity of the Digital Archive, many relationships and interdependencies exist between the various organizational disciplines. To identify and control all the digital preservation risks, risk management initiatives have been launched in collaboration with the ICT department and the Department of Planning and Control. The first preservation risk inventory report was issued in 2013, based on the OAIS standard²³, the SPOT model for Risk

Assessment²⁵, DRAMBORA²⁶ and the normative Quality Requirements for Sound and Vision's Digital Archive V1.1²⁷. All the relevant management domains that relate to potential risks were defined. The report functioned as an initial outline inventory, and represented the starting point for the expansion and improvement of the digital preservation management mechanisms. So far this inventory has led to a series of approaches and solutions in the key risk areas associated with the continuity of the Digital Archive.

8.2.1 Organization and finance

For the Digital Archive, risk management in this area implies that all changes to the organization must be anticipated. These may relate, for example, 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 includes an increase in the cost of digital storage, for example, or specific changes to current or new software and hardware contracts. Changes of this type are identified during the preparation of the annual budget. Sound and Vision then carries out a risk inventory, with an estimate of the possible influence on operations, and the potential financial consequences. On that basis a stress test is organized, in which the risks for the multi-year 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.

8.2.2 Liability

Complying with Copyright Law and the General Data Protection Regulation (GDPR) has become one of Sound and Vision's main priorities. For contract management, copyright policy and data confidentiality issues the NISV has dedicated legal staff (i.e. a copyright lawyer and a privacy officer). The staff have an internal advisory role and draft or review all copyright-related contracts entered into by the NISV. Relevant information is provided on a regular basis in order to sensi-

²³ <https://www.coso.org/Pages/erm-integratedframework.aspx>

²⁴ <http://www.oais.info/>

²⁵ SPOT model: <http://www.dlib.org/dlib/september12/vermaaten/09vermaaten.html>

²⁶ DRAMBORA: <https://www.repositoryaudit.eu/>

²⁷ Quality Criteria Digital Archive NISV, V1.1 <http://publications.beeldengeluid.nl/pub/403>

tize all employees to copyright matters and an understanding of and compliance with contractual agreements and data confidentiality. The NISV has fixed protocols in place regarding privacy protection, the anonymization of data, blocking of materials and online publication. The organization performs standardized procedures regarding requests for the blocking of broadcast material and the settlements thereof.

8.2.3 ICT

The ICT risk management strategy is informed by regularly performing a security audit and by monitoring technical, organizational and financial developments within the digital storage domain, such as the process of evolving LTO generations, or new functionalities and services offered by suppliers of storage software. Error logs are checked to make sure that all components are able to carry out the procedures in the infrastructure, according to the specifications. Risk management is carried out by using all this information to plan ahead, in order to ensure future storage volumes can be handled, techniques will remain affordable, and so on.

8.2.4 Information management

Information quality-related risks are reduced by the implementation of a formal structure for determining, controlling and guarding the framework of integral information management: the Information Governance Board (IGB). The IGB²⁸ is responsible for establishing central Information Policy and

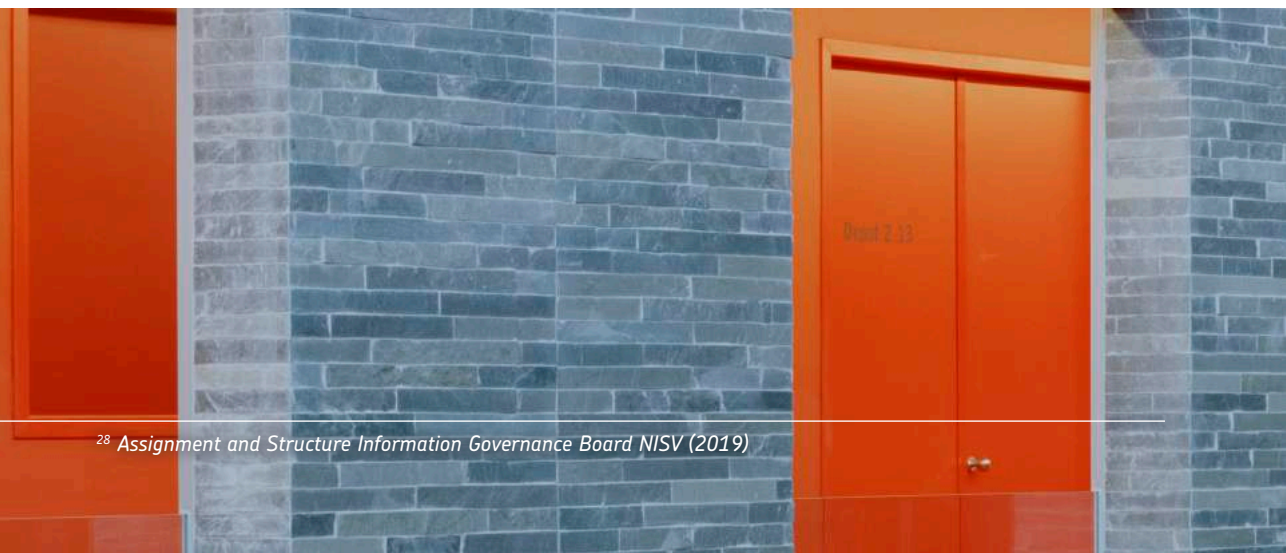
supervises its implementation. Its scope is Sound and Vision's entire information landscape, i.e. all the digital systems, applications, processes, data and metadata that are used to achieve or support the organization's objectives.

The Board's focus points are:

- a) Matching policies, projects, activities and plans in the field of information management with the strategic objectives of the organization;
- b) Ensuring consistency, coherence and links between the execution of tasks and disciplines;
- c) Promoting and monitoring the quality and efficiency of information management based on a shared vision and common standards.

8.2.5 Staff competences

Sound and Vision works in a number of different ways to promote and maintain the knowledge and competences necessary for working within the processes and systems of the Digital Archive. A fixed list has been drawn up, setting out the required skills and competences of all employees working on preservation tasks, storage, access and data management. Single points of failure are prevented by distributing critical knowledge amongst multiple employees. Managers are responsible for identifying and developing the necessary knowledge and competences within their units and departments. Budgets for education and



²⁸ Assignment and Structure Information Governance Board NISV (2019)

training at individual or group level are allocated by the unit managers in collaboration with the HR department, on an ad hoc basis. At present, the management decides what expertise can normally be insourced and what expertise must be developed and/or retained within the organization. Certain SLA's with contract parties include specifications of the knowledge and competences that the party in question must guarantee to have in house in order to provide the service adequately.

8.3 PRESERVATION WATCH

Different types of change and risks can influence the planning and execution of preservation activities. To control and 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 monitored and managed. Detection, monitoring and documentation of the various developments that could affect the preservation activities is part of Sound and Vision's risk management approach. Specific monitoring mechanisms fall under the heading of *preservation watch*, which helps to ensure the critical preservation functions of the Digital Archive.

8.3.1 Designated Communities

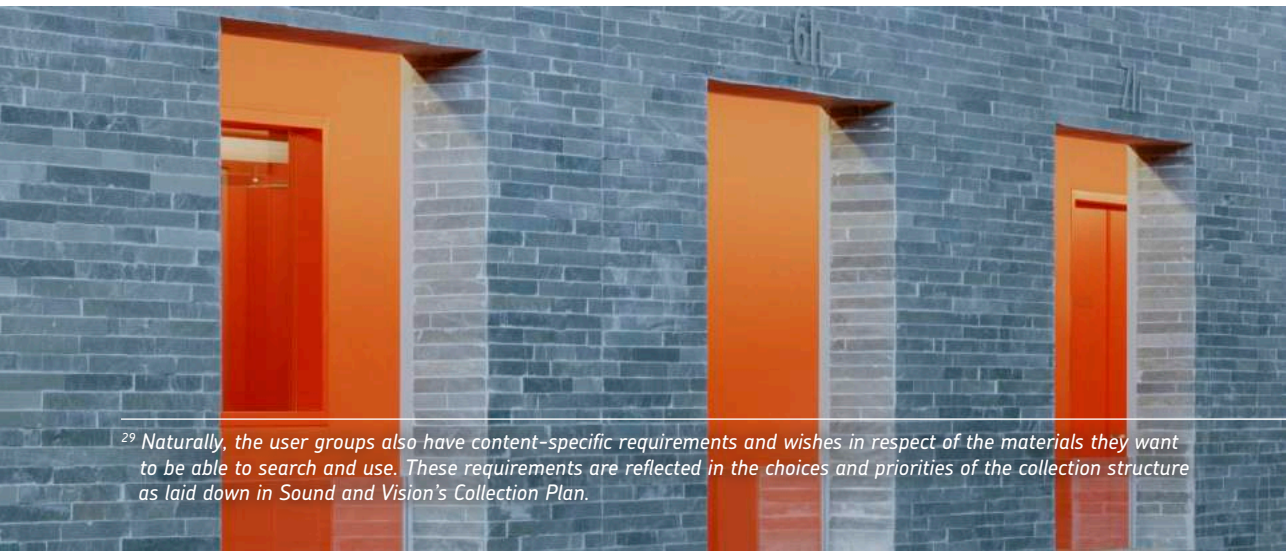
An essential influence on the development of strategies and the preparation of preservation

plans is user demand. The collections must remain permanently accessible to the various external and internal Designated Communities. In technical terms this means that the Digital Archive must simultaneously serve different user groups, each with its own specific requirements in respect of video quality, granularity, browser and navigation facilities, metadata and interoperability.²⁹ 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. on format, storage method, metadata that have to be stored). The result is a direct relationship between the preservation choices and the way in which users wish to use the media content.

As a consequence, it is essential to know precisely how each of the various user groups will want to request material from the collections. For 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 meet for re-use? 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 desired format and metadata and the intended use of the material. Delivery requirements will change accordingly.

²⁹ Naturally, the user groups also have content-specific requirements and wishes in respect of the materials they want to be able to search and use. These requirements are reflected in the choices and priorities of the collection structure as laid down in Sound and Vision's Collection Plan.



Sound and Vision controls these changes by regularly monitoring its Designated Communities. Using both formal and informal methods, 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 assessments, Sound and Vision can adjust its preservation and alter its service provisions.

The NISV's services to its Designated Communities are assessed in different ways, at different points in time and at different intervals.

- For the broadcast *media professionals* Designated Community this is done by annually reviewing the service contracts containing provisions on formats and metadata. Another way to monitor and survey users' needs are the regular training and information meetings between archive employees (i.e. the Customer Contact Centre and access media managers) and representatives of these user groups on subjects such as system training, licensing constructions and general conditions.
- The requirements of *film and documentary makers* as to preservation and access formats are part of the agreements between these user groups and the Archive. Their requirements are registered at the point of acquisition or donation of their collections.
- The institute also negotiates service conditions with the *user groups* in the *educational domain*, taking into account feedback on the outcome of access training sessions and user meetings at fixed intervals and surveys of the needs of these user groups and depositors or their representatives.
- For the Designated Community '*general public*' the consumer delivery formats are monitored by Sound and Vision's Exploration staff, as are requirements regarding streaming functionalities, online presentation formats, search facilities and descriptive metadata.

8.3.2 Technological developments

Format progression and innovations in storage media and playout software also have an independent influence on ability to issue guarantees for adequate ingestion, storage and sustainable access. Also, new media formats may come into play that have different features and require new strategies to ensure their longevity. Structured processes are therefore needed to monitor and anticipate technological developments of this kind.

By systematically addressing these developments, they can be foreseen, and sufficient time can be set aside to make cost estimates and carry out research into technical viability. In other words, technological innovations can be evaluated as they relate to current preservation in good time.

Another underlying principle when laying down preservation strategies and plans is the outcome of 'Technology Watch', a procedural mechanism that enables Sound and Vision to consciously approach technology changes and innovations in a controlled manner. This procedure has recently been prolonged by the Information Governance Board.³⁰ Technology at Sound and Vision watch is also shaped by participating in national and international (research) projects in the field of digital preservation, by membership of and active participation in international partnerships and through the regular attendance of conferences and seminars.

Trends and developments in respect of formats (container or encoding format) and software and hardware for digital AV archiving are regularly discussed by experts. The agendas for these meetings are set by a small team, with a mixed technical/organizational background, but each with a special preservation assignment. Different specialists may be asked to join, depending on the topics. The outcome of these meetings varies. Some topics will address a far-away future, others will provide direct input to preservation plans or even migration plans that are currently topical.

³⁰ Policy Memo Preservation Planning:(2018)

8.4 PRESERVATION PLANNING

Risk management in general and Preservation Watch in particular, are part of the Preservation Planning functional entity in the OAIS model.³¹ This entity provides risk assessments and recommendations to mitigate those risks. It does so at the operational level by evaluating the contents of the Archive and periodically recommending archival information updates, laid down in detailed Migration plans, and by developing recommendations for Archive standards and policies, presented to the Information Governance Board in Preservation Plans.

8.4.1 Preservation Plans

Recommendations for Archive standards and policies are documented in preservation plans. New file formats require new strategies, such as the 'just in time' strategy for web video formats. The templates for technical meta/data also need to be extended and sometimes redesigned. A preservation plan clarifies the context of these alterations or add-ons, along with notes on risks that are mitigated, specific goals and the foreseen impact on digital assets already in the Archive. The collection itself is then explained: which Designated Community is leading, and what will be the designated use; the nature and scale of the expected ingest, the 'significant properties' of the material, and notes on selection criteria or demarcation in agreement with other archives in the Netherlands. In order to make an informed decision, special attention is paid to preconditions or assumptions regarding technical issues, planning, internal users (availability, competences), and procedures to be redesigned, implemented or just applied.

At the heart of the preservation plan are the scenarios, followed by a recommendation. The scenarios may differ in the outline of the preservation strategy, chosen preservation formats, implications for the metadata dictionary, technical requirements, and so on.

These preservation plans are discussed by the Information Governance Board and as a result lead to assignments to implement tooling, prepare

specific upgrades to IT infrastructure or start prototyping a new format. Also, the outcome may be the formulation of add-ons to the preservation metadata dictionary, or even to current preservation policies themselves.

8.4.2 Migration Plans

When it comes to actual changes to assets that have already been ingested, a migration plan is needed. In the strict sense, migration means changing a file to a new format. But in terms of preservation any bulk change in the AIP is considered a migration. So whether a migration plan is needed will depend primarily on the scope of the AIP. So far Sound and Vision has defined the AIP as the package of essence, descriptive metadata and administrative metadata (such as technical, provenance and rights metadata). A migration plan is needed when the content of these packages is affected by a change in handling, storing or access, or even when there is a risk that they could be affected. For instance, when a new storage facility is implemented, tapes are phased out, or a quality check on a specific cross-section of the Archive is planned.

A good example of a planned migration is the repair flow for MXF files. This is to be introduced because a range of flaws and exceptions have surfaced as a result of a newly implemented QC procedure for all archived MXF files. These need to be analysed and categorized. Many will be fixed automatically, but some will need a specific transcode or even customized repair. Once this analysis has been completed, taking random samples, a migration plan will be drafted to plan the resources and timeframe needed, and to define the various categories and preferred solutions.

³¹ NISV Preservation Planning Manual following OAIS: <http://publications.beeldengeluid.nl/pub/399/>



**ACCESS, DISCOVERY,
COPYRIGHT AND LICENSING**

9

9.1 PORTALS AND PLATFORMS

The NISV makes as much material as legally possible available to its Designated Communities: media professionals, user groups in the educational domain, academic researchers and the general public. All access and re-use conditions are communicated via the NISV portals. Depending on copyrights, licences and authorization, the user groups may consult, order and/or re-use material from the collections, each via its own platform. The institute has appropriate licences in place for each community. These consist of customized licences and international access standards such as Creative Commons. In order to use the portal, users will have to agree to the applicable End User Licence Agreements (EULAs) that specify their rights and obligations.

The MAM system facilitates a series of search portals to the catalogue and the collections stored in the Digital Archive: the Media Professionals Portal (MPP), the General Public Portal (GPP), the internal NISV search interface Studio, and the educational platform (Sound and Vision At School)³². In addition to these portals, separate platforms offer extensive context information to the collections. Current examples of two of these platforms are the NISV Collection wiki³³ and the Music Encyclopedia³⁴. The website Open Images³⁵ is an example of an open media platform that offers online access provided under the CC licensing model. Collection material that falls within the public domain or is openly licensed can be browsed through and downloaded from this platform. Via the Open Images platform these materials are also uploaded to Wikimedia Commons³⁶, the media database of Wikipedia.

9.1.1 Media professionals, the general public, the educational domain and researchers

The Media Professionals Portal³⁷ (MPP) provides exclusive online access to the Digital Archive (consultation and viewing in low-resolution quality

and keyframes; online ordering and retrieving of high-resolution material) for professional users in the media domain. A distinction is made between rights holders and users who only retrieve material. Before the connection can be set up, the organization needs to sign a contract with terms and conditions. The organization will then be added to the NISV's firewall rules.

The General Public portal³⁸ (GPP) is an online portal targeted at the general public, where users can search the collections without having to log in, and play out selected audio and video sequences, if copyright regulations allow this. The GPP consists of metadata from AV assets in the Digital Archive that have been released for this Designated Community. The material can be previewed online (if IPR regulations allow this) and on site in the NISV's Media Museum.

Material from the Digital Archive was previously made available to users in the educational domain by two main dedicated streaming platforms: Teleblik and Academia. Teleblik was the NISV's portal to AV material meant for use by students and teachers at primary, lower and middle and senior secondary schools. The platform provided both editing and presentation tools. Academia was the NISV's outlet to AV resources for higher education (students and teachers at universities) that also provided editing and presentation tools.

A fully renewed integrated Educational Media Platform (Sound and Vision At School) was launched in the fall of 2019, offering central integrated access for all levels of education, from primary to university. This educational platform provides source material for end users (teachers and students) that can be edited, structured and contextualized in educational settings. The NISV clusters and thematizes the material on offer. Metadata for the entire collection is searchable, and a large selection (approximately 100,000 assets at go-live) can be retrieved and used.

³² <https://beeldengeluidopschool.nl/#/home>

³³ <https://wiki.beeldengeluid.nl/index.php/Hoofdpagina>

³⁴ <http://www.muziekencyclopedie.nl/>

³⁵ <http://www.openbeelden.nl/en>

³⁶ <https://commons.wikimedia.org/>

³⁷ <https://www.beeldengeluid.nl/collectie/collectie-voor-makers-en-professionals>

³⁸ <https://zoeken.beeldengeluid.nl/>

Videos can only be played out through a login, and authentication will be taken care of by the Dutch educational organizations SURFnet and Kennisnet. Material from the collections can be shared via this platform and embedded in external educational environments (educational catalogues and applications). The records can as metadata be included in a catalogue, using the OAI-PMH protocol. The content and users of the former platforms Teleblik and Academia are migrated to the new platform.

Sound and Vision is one of the partners in the CLARIAH (Common Lab Research Infrastructure for the Arts and Humanities) consortium.³⁹ This academic consortium consists of more than 40 partners: in addition to all the Dutch humanities research institutions, university libraries, heritage institutions, public organizations and companies are also affiliated. The primary goal of CLARIAH is to develop a digital infrastructure that enables researchers in the humanities to conduct innovative and data-intensive research across collections from various institutions.

The institute further presents selected AV material on international (educational or academic) platforms such as EUScreen⁴⁰ and Europeana⁴¹. For the conditions and terms relating to the actual use of the moving image materials visitors are referred back to the main portal.

9.2 SEARCH AND DISCOVERY

All the portals mentioned above are connected to the same multimedia catalogue through APIs (Application Programming Interfaces). A search engine is provided, as well as ordering

and delivery services. The MAM system logs all search behaviour on the portals in order to enable accessibility and searchability for the Designated Communities to be analysed and enhanced.

9.2.1 Metadata

The Digital Archive's catalogue is the starting point for all collection access.⁴² The underlying metadata model was inspired by the IFLA's (International Federation of Library Associations) FRBR⁴³ (Functional Requirements for Bibliographic Records) model, and allows for the layered identification of content. It links different levels of digital productions (programme, series, season, item, time-coded metadata) to catalogue descriptions and information on the analogue and/or digital carrier.

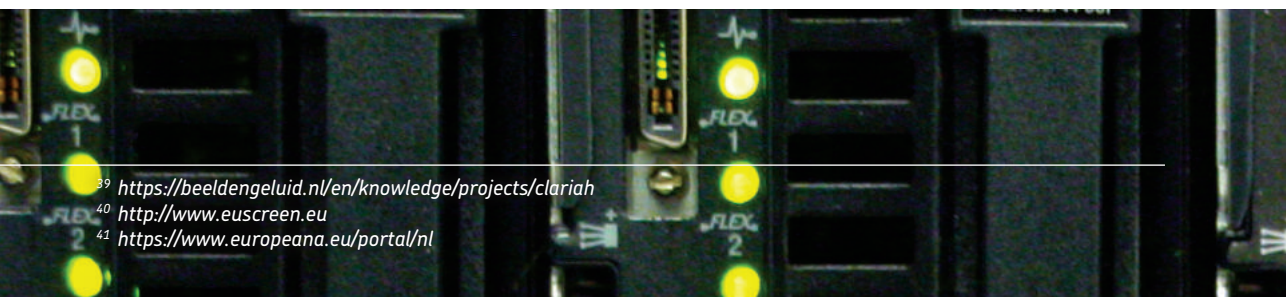
Descriptive and administrative metadata from the public broadcast production environment are pulled from the source systems to flow directly into the Sound and Vision catalogue. The metadata associated with materials from non-broadcast producers is imported via standardized ingestion workflows. Automated checks and monitoring mechanisms, together with analysis and extraction tools, guarantee the quality and completeness of the metadata and files during every ingest and after each update. The specifications of the new MAM system also define the rejection criteria for metadata, as well as a workflow for repairing, logging and documenting errors.

Tagging techniques such as speaker labelling and keyword extraction based on Teletext 888 automatically add to the descriptive metadata in the catalogue. In selected cases the ingested metadata are added to, enriched and contextualized manually.

³⁹ <https://beeldengeluid.nl/en/knowledge/projects/clariah>

⁴⁰ <http://www.euscreen.eu>

⁴¹ <https://www.europeana.eu/portal/nl>



9.2.2 Thesaurus and identifiers

The common thesaurus for audio-visual archives, the GTAA⁴⁴ (*Gemeenschappelijke Thesaurus Audiovisuele Archieven*), is an essential tool for indexing content. The GTAA structures the relationship between words and concepts and connects disparate parts of the collection. It supports both media managers and users in creating consistency between different spellings of keywords, locations and names while assigning subject headings and keywords or while searching. It also enables the NISV to enhance the automatic indexing applications, thus ensuring the quality and consistency of descriptions throughout the catalogue. The GTAA is available to other audio-visual archives via an Open Database Licence. The NISV's collection can be linked to other data resources via the GTAA, using linked data principles.

OAI harvesting is permissible, and will in 2019 be enabled for the entire Digital Archive. The Digital Archive provides internal identifiers, which are maintained to be internally unique and persistent. Persistent Identifiers, a formal defined attribute category in the NISV's Preservation Metadata Dictionary 2.0, will be implemented by the end of 2019. The NISV has selected the URN:NBN system for these PIDs that guarantees that handles are globally unique and can be resolved on the Internet.

9.3 COPYRIGHT AND LICENSING

The NISV ensures, to the extent possible, that data are created, curated, accessed and used in compliance with legislation and ethical norms. Complying with Copyright Law has become one

of the NISV's main priorities in the past few years. Specialized staff are appointed to focus on collection access policies and copyright policy issues and to draft or review all copyright-related contracts entered into by the NISV. In order to sensitize the employees to copyright matters and an understanding of, and compliance with, contractual agreements, the staff provide internal copyright instructions. Information on the conditions for the use and re-use of the collections, on copyright issues and on the valid codes of user conduct is available on all NISV portals. The institute's general terms and conditions do not provide for measures if they are not complied with: in such cases the conditions refer to Dutch law and the Dutch civil courts. If a user fails to comply, the original rights holder will be contacted by the NISV. The rights holder will be informed and will have the final say.

9.3.1 Contracts and clearance

There are three types of contracts with data producers: the service agreement with the collective public broadcasters, which is updated and renewed annually, service agreements with non-public broadcasting organizations, which are drawn up with each data producer, and donation agreements. Data producers can choose whether or not to transfer copyrights to the NISV. In the case of donations this happens regularly; in that of service agreements generally no IPRs are transferred.

Most of the NISV's audio-visual collections are thus copyrighted, and the rights should always be cleared before the use or re-use of the materials. Sequences and whole programmes can only be published, viewed or re-used according to what is specified in a licence agreement, unless one of a number of exceptions applies, as detailed in the

⁴² <https://zoeken.beeldengeluid.nl/>

⁴³ IFLA FRBR model: <https://www.ifla.org/publications/functional-requirements-for-bibliographic-records>

⁴⁴ Common thesaurus AV-archives: <http://gtaa.beeldengeluid.nl/>

End User Licence Agreements (EULAs).

9.3.2 Licensing and blocking rules

The NISV's MAM system is equipped with an integrated licensing module that stores all contracts (License agreements) as records in the catalogue. Updates of information on makers, contracts and copyrights are preserved. Copyright information is also made accessible in the catalogued metadata for each individual asset (at collection, programme, sequence and item level).

Sound and Vision plays a coordinating role in the licensing process. Positioned as an intermediary between users and rights holders, the Institute functions as a gateway, where all transactions are registered. By providing the appropriate intermediate services, the NISV aims both to enable the use of materials and secure the proper financial benefits for rights holders. The conditions for the use and re-use of specified material for educational and cultural purposes are included in a collective licensing agreement (the so called 'Archive Agreement') between the NISV and the public broadcasting organizations and associations of independent television and record producers.

When a third party requests content from a data producer for use or re-use, the NISV will function as intermediary in the licensing process. The licence is signed by the requester and by the NISV as the data producer's representative. The intermediary makes sure both parties adhere to agree on costs and other conditions in the license agreement. Public broadcasters, the largest group of data producers/rights holders and data consumers of the Digital Archive, provide licences for the use or re-use of all programme materials they own the rights to. General and specific information on licences, copyright, regulations and exceptions in the professional media domain can be found on the Media Professionals Portal. This contains special sections on the terms and conditions for providing licences for archival materials and the rules applicable to the use of sequences, and provides various licence forms that can be downloaded, including a release-of-liability form. All arrangements concerning rights and licensing are

laid down in the Service Agreements between the NISV and the NPO.

The NISV applies a protocol regarding requests for the blocking of broadcast materials and the settlements thereof. General rules on blocking are included in the agreements with the public broadcasters. Rights holders can request blocking via a special form on the Media Professionals Platform (MPP). Blocked material cannot be ordered or re-used, only the metadata can be viewed in the Sound and Vision catalogue. If requested (because of e.g. portrait rights/privacy law) it is also possible to anonymize personal data and – in exceptional cases – to block the carrier itself, so that the programme cannot be viewed any more. The Customer Contact Centre is the NISV department that processes these requests. A form is sent to the metadata managers for verification. In case of doubt, the specialized legal staff may be approached. If individuals object to the accessibility of certain data or mistakes are detected, they have the option to contact the NISV in order to have the metadata changed or corrected. The essence remains unaltered. Confidential data is not available to non-authorized users outside the Digital Archive environment. Users may request special permission to access this content via an internal workflow. Staff will inform users about the restrictions on use and re-use.

The NISV applies a protocol regarding requests for the blocking of broadcast materials and the settlements thereof. General rules on blocking are included in the agreements with the public broadcasters. Rights holders can request blocking via a special form on the Media Professionals Platform (MPP). Blocked material cannot be ordered or re-used, only the metadata can be viewed in the Sound and Vision catalogue. If requested (because of e.g. portrait rights/privacy law) it is also possible to anonymize personal data and – in exceptional cases – to block the carrier itself, so that the programme cannot be viewed any more. The Customer Contact Centre is the NISV department that processes these requests. A form is sent to the metadata managers for verification. In case of doubt, the specialized legal staff may

⁴⁵ <https://www.beeldengeluid.nl/collectie>

⁴⁶ <http://www.beeldengeluid.nl/auteursrechten>

⁴⁷ <https://zoeken.beeldengeluid.nl/licenses>

be approached. If individuals object to the accessibility of certain data or mistakes are detected, they have the option to contact the NISV in order to have the metadata changed or corrected. The essence remains unaltered. Confidential data is not available to non-authorized users outside the Digital Archive environment. Users may request special permission to access this content via an internal workflow. Staff will inform users about the restrictions on use and re-use.

9.3.3 Reuse conditions

The general conditions for the use and re-use of specified material for educational and cultural purposes are included in a collective licensing agreement (the 'Archive Agreement') between the NISV and the public broadcasting organizations and associations of independent television and record producers.

The public has access to metadata of the collections via the online catalogue on the General Public Portal (GPP), but they are only permitted to view moving images that have become public domain, or collections that have been licensed under a Creative Commons licence by the rights holders. The GPP has a section containing general information on the conditions and instructions for the usage of audio-visual material, including how to contact the Customer Contact Centre regarding more complicated copyright issues and requests.⁴⁵ Information on copyright holders and restrictions can be viewed in the metadata at collection, programme, sequence or item level. End-user organizations have the option to create a tailor-made contract, by uploading conditions and restrictions on re-use. Sound and Vision also makes use of standard licensing models, such as Creative Commons. A link to the conditions will be shown when material on one of the NISV web platforms is licensed under CC. For all other material the standard conditions of use apply.

Specific information on AV copyright regulation is made available to users in the AV heritage domain and the creative industry, to enable these Designated Communities to determine whether their intended usage is appropriate and permitted.⁴⁶

Re-use conditions are also stated in the conditions for delivery and use based on the relevant laws of the Netherlands and the EU.⁴⁷ If a private user wants to purchase material, he/she – having agreed to the terms and conditions – can create an account. Before the actual purchase, the customer has to agree to the conditions, regulations and restrictions on online purchase under Dutch law.⁴⁸

For media professionals, the general access and use regulations included in the terms and conditions for licensing and rights are available on the Media Professionals Platform (MPP). This has sections on rules regarding the usage of archival material by this Designated Community. Any individual media professional may request a personal account with specific rights. Based on these rights, the user can browse, preview, request a licence and download material. An account request needs to be made via a form signed by the mother organization. Individual contracts are available in a standard format and contain specific information regarding the purpose of re-use. With every individual contract the user needs to agree to the restrictions set by the licence holder and in general needs to agree to the terms and conditions set by the NISV.

9.3.4 Access agreements and codes of conduct

For the browsing, viewing and online editing of the low-resolution AV files available on the NISV's educational platforms, the original rights holders, i.e. the organizations that represent the individual media producers, film makers and record producers, have granted general permission in the Archive Agreement with the NISV. The conditions for the consultation and use of the NISV's collections by users in the educational domain are formally laid down in agreements with the CLARIAH Consortium and the Archive Agreement.

As regards Sound and Vision At School, NISV's new, integrated audio-visual portal for primary, lower, middle and senior secondary schools and for universities, information on all the aspects of access and usage and the codes of conduct, can be found in the General Conditions on the site.⁴⁹

⁴⁸ <https://www.beeldengeluid.nl/collectie/aanvragen-voor-privegebruik>

⁴⁹ <https://beeldengeluidopschool.nl/#/algemene-voorwaarden>



TECHNICAL INFRASTRUCTURE

10

10.1 THE NISV IT ENVIRONMENT

At the core of the NISV's primary and secondary business processes lies an advanced technological infrastructure that enables the ingestion and sustainable storage of, access to and exchange of all the NISV's digital materials and all supporting tasks and secondary activities. Sound and Vision is fully responsible for maintaining the infrastructure 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.

Rules and regulations regarding IT systems, standards and procedures are clearly defined and apply to the entire IT environment.⁵⁰ Policies and procedures within the NISV's IT department comply with ITIL methods and with Sound and Vision's Quality Criteria for the Digital Archive. Because of the scope and complexity of the IT infrastructure and the fact that it is constantly changing – as a result of developments inside and outside the organization – all IT activities are centralized.

The NISV distinguishes four application groups:

1. Business-critical applications, considered critical to the Digital Archive's core business (media asset management, catalogue management, thesaurus management, storage management and metadata and file import applications).
2. Dissemination applications, used by NISV staff to provide access to archive content and by external users to access content (search and order processing systems, educational and museum portals, access portals for researchers and access portals for the general public).
3. Office applications, used by NISV staff (office applications, ticketing systems, personnel management, email and document management, internet).
4. External applications (social media to provide access to archival content, not managed by the NISV (YouTube).

The institute has divided its technical infrastructure, so that the business-critical processes are managed in a separate environment, referred to as the 'Production environment'. New IT concepts are created in the Sound and Vision Development Environment. These products have an alpha or beta status and cannot simply be put into use in the Production Environment. They are therefore tested for functionality and correct operation in the so called Acceptance or Staging Environment, as are products that are provided for by external suppliers.

Another environment, the Museum environment, manages the group of systems that facilitate exhibitions and activities in the NISV museum. The Office environment, with its office applications and facility systems, supports all these processes. The NISV websites and portals are managed in the Hosting environment.

The various IT environments serve both employees and groups of external users (Designated Communities). The services are based on different types of agreements, both with external users and with the suppliers of the systems and applications. The NISV's network infrastructure and the connected network equipment are essential to make the IT services available inside and outside the organization. The network consists of the following components:

- The back-end (i.e. the core network components, such as the routers and firewall);
- The cabling between the back and front end;
- The front-end (switches) and WiFi access points. The NISV provides the possibility of a VPN connection to allow internal and external users to access the internal network.

⁵⁰ *Operational IT Policy Plan NISV (2017-2020)*

10.2 WORKFLOWS AND MEDIA ASSET MANAGEMENT

At the heart of the Production environment is the MAM system DAAN (Digital Audiovisual Archive of the Netherlands), a proven VizOne product supplied by VizRT. The standard functionalities of VizOne have been substantially supplemented with a number of functions tailored to the NISV's preservation functions. Additional modules for import, workflow management, IPR management and quality control have been installed. Storage and access to the source files is controlled by the archive management system, which acts as a 'go-between' between the MAM system and the actual storage units, including the tape libraries.

All files and metadata – from any source – are ingested into the MAM system through standardized workflows, managed by the workflow system MAYAM Tasks. Automated checks and monitoring mechanisms, together with advanced analysis and extraction software (BATON) guarantee both the quality and the completeness of the files and the metadata, during every ingestion and after every update. The integrity of metadata is detected during ingestion and after updates. In the case of MXF files, a check on file integrity by verifying header and footer metadata is accepted. The MAM system has rejection criteria for metadata built in, as well as a workflow for repairing, logging and documenting errors. The output of the quality control actions for all incoming files can be mapped with the pre-defined standardized preservation metadata in the NISV's Preservation Metadata Dictionary 2.0.

In addition, there is a Physical Asset Management System (PAM) in place that holds information about the analogue AV carriers in the collection (film, video, audio). An interface between PAM and DAAN has been built to record updates (i.e. the output from digitization actions). The systems are directly connected to the Digital Provision (DDV), the central infrastructure of the broadcast production environment. Separate import facilities for files and metadata are in place for non-broadcast

material. The MAM system directs all ingestion workflows and is responsible for the management of the metadata. A thesaurus system supports the index and search processes.

The provenance of all ingested items, along with the descriptive metadata and technical metadata on the files/carriers is captured, along with references to the source system and a source ID. The date and method of ingestion are recorded. The archive management system DivArchive logs and stores basic provenance data with the AV-file. By way of a naming convention, data producers are identified in the various ingestion workflows, and an internal persistent identifier is generated, establishing a link between metadata and files. The process of extracting and recording technical file characteristics will ensure that different versions can be distinguished whenever a new version of a file enters the Digital Archive. Different data producers will deliver files to different watch folders, in order to create yet another way to keep track of their identity.

The output of all processes is logged and structured and can be used as an overview of the life cycle of the files from ingestion through to storage and access. A business intelligence tool is currently being developed to map, integrate and structure pre-defined preservation metadata recorded in different applications for the purpose of digital collection management and preservation action planning.

10.3 STORAGE MANAGEMENT

The NISV employs DivArchive (an Oracle product) as an archive management system. Digital archive masters are stored primarily on LTO tape. A disk cache has been created to enable recently ingested material to be delivered faster (on a temporary basis) than is possible from tape. Tapes are loaded into a Storagetek tape robot. The administration of master files takes place in DivArchive, which is responsible for ingestion, integrity monitoring, copying actions, access, logging, allocation to tape

27-005

27-004

27-003

groups, the allocation of resources, partial restores and checks on the location of the stored files. The storage management system is owned and managed by Sound and Vision.

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 onto its own tape group. This is done to ensure maximum control over the storage policy for each collection. Using the MAM system, staff are also able to check whether ingestion has succeeded and is complete. If manual intervention is required, tasks can be created in the system to correct errors.

In addition to the integrity checks that take place in DIVA, the software tool 'Tape Analytics' from Oracle is used. This tool monitors the quality of the data on the tape media. By means of a process that takes place in the background, tapes are scanned for signal quality and degradation in order to prevent the development of 'bit rot'. The application notifies the storage manager if it is necessary to replace a medium. DIVA can signal that a file no longer matches the checksum, but in the exceptional case that tape copies are not readable, a data loss is inevitable. The tape analytics process proactively guarantees that this cannot occur.

Tape migration

Every five to seven years all data tapes are re-written to a new version of LTO. A migration plan is drawn up for this process, which is harmonized with the Information Governance Board and the Management Team of Sound and Vision. The plan also includes an estimate of future volumes. This planning 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 archive management system, with checks carried out, based on checksums, that files have been correctly copied or that all the files

eligible for migration have actually been included and the correct backup policy has been implemented. These actions are logged by the archive management system. This process metadata (category: 'events') is preserved in a separate data warehouse. The Digital Archive has already performed multiple migrations of the OAIS type 'refreshment' on the LTO tape storage media and file servers.

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.

10.4 BACKUP AND DISASTER RECOVERY

The frequency of backups and recovery times are determined for each IT environment. All vital datasets within the various IT environments have been identified and located. These are either irreplaceable components of the digital collections, or data and metadata that are indispensable for the daily continuity of the archival processes and services. Extra back-up mechanisms have been set up for these key archival datasets.

Files that are stored on tape are managed by the storage management system DiVArchive, which is responsible for creating backup copies and ensuring that these are identical to the primary archival copy. Files arrive on a disk cache and can only be removed after at least one successful copy has been made on tape. This process includes a check that 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. Recov-

ery is necessary whenever a tape fails to work (it may emerge that a tape is no longer readable during a migration action, for example, or when an order is received). The backup can then be retrieved for copying. Browse files of the preservation masters are also backed up on disk and tape. The aim of these backups is to reduce costs if, in the event of the 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 with the preservation masters.

Files that are kept on file servers are mirrored and a tape backup is created every night. This process is strictly monitored. The nightly backup includes all the database files that contain the catalogued metadata and the database of the storage management system. In the event of data loss or database defects, backup copies can be restored from tape. Policy prescribes that backup restore tests are performed regularly, to ascertain that the Production environment can be fully restored in case of emergency. The stored checksum information is used to check that backup copies are identical and files have been copied properly. More customized options are available if higher levels of accessibility and/or security are required. Backup guarantees depend on the preservation level agreed with the depositor/data producer and/or the appraisal of an item or collection that requires 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 pre-determined level of backup policy is implemented. Detailed backup procedures are laid down in the collective or individual contracts and SLA's with the data producers.

A disaster recovery approach is used to spread the geographical risk. One full backup of the Digital Archive is placed with a professional commercial partner at the Media Park. A second full backup (including power feed) has been implemented at the Royal Library in The Hague. To avoid tape wear, this second backup is not to be used for the delivery of content; it functions

primarily as a disaster recovery facility for the NISV's Production environment. A third backup of selected, key parts of the collection is also stored in this facility.

Periodically, during migrations, all the data in the Digital Archive, including all the backup services, are tested and the checksums are compared. These recovery tests take place in a separate Acceptance environment, in order not to disrupt the ingestion, storage and access processes in the Production environment.

10.5 SECURITY

Sound and Vision's IT security policy⁵¹ is a summary of the general security rules and procedures to secure the entire digital infrastructure against all (deliberate or unintentional) internal and external threats, human error and incompetence, and deliberate sabotage. It is an important part of IT Risk Management and complies with OAIS security policies and the NISV's Quality Criteria for the Digital Archive V1.1.

The NISV's security policy document divides security measures into several groups:

- Physical Security
- Identification and Authentication
- Access control
- Data integrity
- Data confidentiality

To secure internal access to the workplace and the applications, two layers of security are deployed in the relevant environments: a password and two-way authentication. User accounts have limited rights, but temporary exceptions may be granted under strict regulations. Data security procedures to guarantee the protection of data under the Personal Data Protection Act are the NISV's responsibility. Strict rules for authorization, storage and security risk management apply. Any exchange of data between information systems is executed as a rule using an automated method. Data stored in the Google Drive environment is

⁵¹ Security Policy Plan NISV V1.0 (2019)

not to be shared with external persons.

The MAM system provides the option to grant access to specified parts of the collection as part of a preservation service to data producers in the cultural heritage and media domain. After ingestion, data producers or 'tenants' may access and/or download their own materials from this MAM, based on user and password authentication. Access is not permitted to anyone else. The internal MAM system remains accessible solely to NISV administrators. Tenant access level, rights and restrictions are specified in the contracts.

A monitoring system performs regular predefined checks of the critical business systems, i.e. the databases, the production server devices, selected office and support services and the metadata base of the core collection. Alarms and events that entail significant risks automatically flow into a ticket registration system. These alerts may be sent to duty managers, who will carry out the appropriate actions. Any incidents and disruptions occurring within the IT environments are recorded. The construction of this knowledge database reduces the chance of future disruptions.

The implementation of all security solutions is regularly tested in the various IT environments (Production, Acceptance, Development, Museum) and on the network and the connected equipment. An independent organization conducts audits on the available documentation (procedures) and the enforcement (and enforcement policy). Penetration tests are also performed, based on a quick scan (i.e. a search for known vulnerabilities using various automatic tools). Other penetration tests detect less common vulnerabilities or combinations of vulnerabilities. Security improvements are made on the outcome of these tests. Deterioration of storage media is handled by pro-actively migrating to new LTO versions every five to seven years. During migration, checksums ensure that new copies of archival objects are consistent with the original versions. Deterioration of disk storage is taken care of by regularly replacing servers that are older than five years.

A strict level of security is needed for a large part

of the collection, due to copyright considerations. Further security measures are guaranteed by having a separate network and portal in place, accessible only to connections from IP addresses on the NISV whitelist. The most important security measures for both on-site and online access are described in the Preservation Policy document.

The following general measures have been introduced to secure the infrastructure and personal details of users against data loss, damage and abuse. Specific supplementary agreements are made with each department.

- At the data centre, a theft and fire alarm system has been fitted at two locations (Sound and Vision and the company Ericsson). This is linked to a security organization and/or the police and fire brigade.
- The data centre has a high-quality power supply, UPS, an emergency cooling system and an emergency power supply to ensure that access to materials can continue for some time in the event of a power failure.
- To prevent physical access by unauthorized persons, the servers that form part of the infrastructure are at a secure location in the data centre.
- 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 names and passwords.
- Users are not granted any rights to which they are not entitled under 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 planned backup procedures have been correctly implemented.
- Installing security updates for standard package software bundled with equipment in consultation with suppliers. Purpose: to retain functionality.
- Implementing recovery from incidents within configuration management.
- Following correct delivery (in line with the agreement with the data producer/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. This provides material redundancy. This tape can then be stored outside the tape robot, at the discretion of Sound and Vision.

10.6 MAINTENANCE

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 infrastructure status.
- 'Problem, change and release management', i.e. timely identification of possible shortcomings and implementation of technical application management tasks such as configuring the infrastructure.
- Identifying trends, frequent incidents and mutual links and causes; if possible reproducing recorded 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: to retain functionality.
- Implementing (and ordering) essential maintenance on the infrastructure, and if necessary replacing equipment or components.
- Incident recovery in the infrastructure.



APPENDIX

1

NISV FORMATS, PRESERVATION STRATEGIES AND STANDARDS

A. Data tapes

Linear Tape Open (LTO)

For long-term storage data tapes of the LTO type are used (Linear Tape Open). LTO is a powerful, scalable and adaptable tape format that helps address the growing demands of data protection. It's also an open format, licensed by some of the most prominent names in the storage industry to ensure a broad range of compatible tape drives and cartridges.

Preservation strategy : just-in-case scenario

<https://www.lto.org/technology/what-is-lto-technology/>

AXF

The NISV tapes are written in AXF format, an open-standard object-container for file-based assets and their associated metadata. AXF allows any type, size and number of files to be stored, transported and preserved as a collection while maintaining full independence of file or operating system and underlying storage. AXF addresses the need for a long-term open storage format while overcoming the limitations of legacy file and container offerings. AXF was first published as an International Standard by the Society of Motion Picture and Television Engineers (SMPTE) in 2014 and later by ISO/IEC in 2017.

Preservation strategy : just-in-case scenario

<http://www.axf.io/>

B. File formats

Video

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 soft-

ware 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 HD422/50 Mbps.

Preservation strategy : preservable format, just-in-case scenario.

Audio

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

Preservation strategy : preservable format, just-in-case scenario.

Film

The chosen format for the digitization of 16 and 35mm film is DPX in three different resolutions : 4K and 8K (used for AV-productions with an outstanding cultural, historical and/or esthetic value) and 2k for other highly valued programmes.

For the mezzanine: XDCAM/HD422/MXF.

Preservation strategy : preservable format, just-in-case scenario.

Text

For written archives (born digital) the institute maintains PDF as our standard. Text files accompanying programmes are also stored in PDF format.

Preservation strategy : tentative preservable format, just-in-case scenario.

Subtitles

For additional subtitling files accompanying the MXF, .890 (proprietary, from supplier Cavena) and .STL (used in EBU format).

Preservation strategy : tentative preservable format, just-in-case scenario.

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

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

Photographs

Photographs and paper objects are preserved as TIFF. In the case of cartoons, illustrations and magazines belonging to the NISV press collection, NISV also maintains TIFF as the standard for both digitized and born digital materials.

Preservation strategy : *preservable format, just-in-case scenario.*

Games

The acquired physical carriers (thus far: tape, floppy or CD-ROM) are converted into digital disk images: ISO 9660 (nb. this file format can also contain other optical carriers), TAP (file format of a rough tape copy) and FLP (floppy) files. The original carriers are stored as objects for exhibition purposes.

Preservation strategy: *emulation, tentative preservable format*

Web video

Web videos are downloaded from the web. In some cases the depositor submits the file him/herself.

Currently most web videos are published as a H.264 file in a MP4 container. Web videos that were produced in earlier periods, were stored in a variety of file types, common to the particular production period. For the online and onsite publishing of the content, the MAM system DAAN creates its own proxy upon the file entering the archive, in an MP4 / AVC format. Certain levels of this file (such as a Baseline@L2.1 profile) are standardized. This standardization is useful to ensure playability on different platforms. Web videos that cannot be transcoded to this proxy format are to be migrated to a lossless preservable format.

Preservation strategy: *preservable format under investigation, just-in-time scenario.*

Websites

Websites are archived in the WARC-format (Web ARChive file format), the widely used ISO standard for archiving websites (ISO 28500:2017) strongly advocated by the IIPC (International Internet Preservation Consortium). The web archive as it was built up until 2016 also contains ARC-files,

the predecessor of WARC, these will be converted to the WARC-format in 2019.

Preservation strategy: *tentative preservable format, emulation.*

C. Metadata

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.

Metadatamodel

The metadata model covers all descriptive metadata, as well as most technical metadata of all types of assets. The hierarchy is set-up similar to the FRBR-model, except for the *Work* level; this is not modelled as a separate level, although assets belonging to the same work can be grouped via the title thesaurus attribute which provides for grouping different assets under a uniform program title. Programmes (together with series and seasons, as well as scene descriptions) describe the *Expression* of some artistic content. Items describe the physical entity or *Manifestation* of the program.

An item in the NISV MAM-system, together with the package when needed, represent the files that are kept for playout or viewing, and for preserving the content in our archive. At a more concrete level these files are managed in a filesystem, where several copies are kept. These copies or instances can be interpreted as the *Items* of the FRBR. When it comes to digitized material, the analogue carriers will also be described as items in the hierarchy, with a pointer to our physical asset management system. The same holds for analogue material that is not yet digitised.

Metadatamodel NISV MAM system (2018)⁵²:

FRBR IFLA: <https://www.ifla.org/publications/functional-requirements-for-bibliographic-records>

⁵² <https://publications.beeldengeluid.nl/pub/671>

Descriptive metadata

For the *daily intake* and acquisitions, in principle the externally delivered descriptive metadata that are received are considered sufficient. One-off acquisitions must be handed over with a minimum set of structured metadata. 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, rare
- Value B: core collection
- Value C: no major value
- Value D: little to no value or irrelevant

All descriptive metadata within the Sound and Vision information system is structured according to the internal metadata model. 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.

Thesaurus

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 speak-

er labelling. The GTAA is also used to add already structured keywords upon ingestion of materials into the broadcasting production systems. All concept-schemes (the ones used internally and the ones that are published for external use) are managed in OpenSkos, both that are used internally and 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.

Common Thesaurus AV-Archives GTAA:

<http://gtaa.beeldengeluid.nl/>

Preservation Metadata

The goal of the Sound and Vision Preservation Metadata Dictionary (PMD) is to describe the digital objects in the Digital Archive as well as how said objects are to be processed and managed. In the PMD, the technical metadata attributes are defined that can be allocated to each digital object (audio, video, film, text, photograph) ingested in the Digital Archive. The Dictionary also contains a section for the Persistent Identifiers.

The Sound and Vision PMD has been developed to conform to PREMIS (Preservation Metadata Implementation Strategy) an international metadata standard designed to support the preservation and long-term sustainability of digital objects. The current conformance level is 1: the NISV preservation metadata can be mapped to PREMIS. In a later stage it will be decided if and how to implement PREMIS level 2 (export from the NISV systems in PREMIS metadata).

The PMD will be further developed to include attributes describing actions ('events'), results of those actions ('outcomes') and their associated 'agents' (responsible organization, software or person). Also the object category for intellectual items will be extended with environments.

The PMD documents the preservation metadata with which Sound and Vision guarantees sustainable access to digital objects. Assurances are found in different ways and on different levels.

The preservation metadata is used to:

- a. Group digital objects in order to enable:
Specific preservation planning and actions
Data management; data quality, efficient and consistent management
Collection forming, application of retention policy and access policy
- b. Manage the lifecycle of digital objects
- c. Control the migration of assets to new formats

The PMD helps document which information must be minimally available in order to carry out these tasks. The PMD thus documents the basis description for OAIS conformant Information Packages

Preservation Metadata Dictionary V2.0:

<http://publications.beeldengeluid.nl/pub/615>

PREMIS Data Dictionary for Preservation Metadata, version 3.0, June 2015: <http://www.loc.gov/standards/premis/v3/premis-3-0-final.pdf>

APPENDIX

2

REQUIRED KNOWLEDGE AND SKILLS OF DIGITAL ARCHIVE STAFF

Four years ago new job profiles were prepared for the staff in the various units at Sound and Vision. 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, user experts, preservation management staff, information specialists, specialists in collection policy, legal specialists, ICT control and management staff, metadata management and ICT development.

The following areas of knowledge are important for staff within the digital preservation environment. Depending on their position in that environment they hold knowledge in a number of the following 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

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, SLA's, 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/layout requirements and requirements of Designated Communities and methods for monitoring and user survey in relation to selection/retention and preservation planning.

8. General ICT knowledge and ICT knowledge in the archival 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.

