SOUND&
VISION

Digital preservation

Policy, standards and procedures

Colofon

2026

Author: Marjolein Steeman, Digital Preservation Officer
Graphic design: Vera Jansen, Studio Vera
Photography: Feike Faase, Jorrit Lousberg, Max Kneefel, Tomas Mutsaers and Collection Sound & Vision

Contents

Management summary

The purpose of this document is to further contribute to the development and promotion of the Netherlands Institute for Sound & Vision (in short Sound & Vision or S&V) 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)[1] certificate has been acquired by S&V: the Data Seal of Approval, followed by a second certificate, the **Core Trust Seal** (CTS)[2] in 2020. S&V will apply for renewal of this certificate in 2025. 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.coretrustseal.org/about/history/data-seal-of-approval-synopsis-2008-2018
2 https://www.coretrustseal.org

Introduction 1

The concept of digital preservation

The domain of digital preservation extends to support 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 from the Archive perspective. 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.

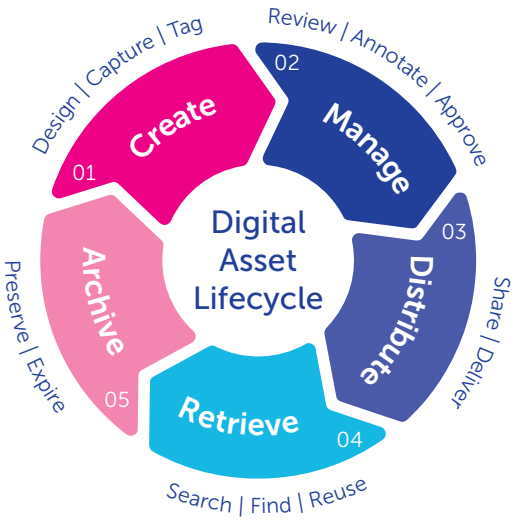


Fig. 1: Digital asset lifecycle

The requirements imposed on such an environment are to a considerable degree met through the standardization, formalization and documentation of both the data (materials) 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 the archive that sets conditions and prescribes procedures. Which materials can be accepted and how do the depositors want that material to be preserved by Sound & Vision?

There are also various user groups for the Digital Archive. How do these users – each with their own goals – wish to be granted access to the collections, and in what form should the materials and metadata consequently be presented by Sound & 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 and Information Governance. The state of affairs in respect of storage facilities, and applications that manage technical and descriptive metadata are integral to the preservation policy. The policy also includes securing the necessary

knowledge and competences of the staff working in such a technological datadriven 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 information that makes the collections accessible, must not be placed at risk. Preservation policy therefore also provides mechanisms for managing the various risks run by the Digital Archive. These mechanisms are aligned with the highly regarded ISO standard for information security.

The Preservation Policy of Sound & Vision

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 docuwmentation may be a standard-related policy document, a users' manual, a model, an elaborated procedure or a standard itself.

Revision of the 2019 document

This document is a revised and fully updated version of the Preservation Policy Plan, that was published in 2019. This new edition represents the situation at Sound & Vision as it is today (2025). It is once again a moment in time. Developments are progressing rapidly, driven by the renewal of portals, the opportunities offered by cloud solutions, and the applications of artificial intelligence. The policy will be updated at relevant times to reflect these changes.

The adapted and extended sections in the plan demonstrate that the work on the construction of the controlled preservation environment has further progressed, both in terms of policy and knowledge, and in respect to actual implementation.

- Roles concerning preservation have developed since the reorganization in 2018 and are visible in regular consultation structures. This has been elaborated on in section 2.3.

- A new Collection Policy plan will be published in 2026; and a fifth

domain will be added to the overall collection structure (see 3.2). The plan formulates the ambitions for the collection and, in doing so, issues a call to the preservation expertise for the future.

- In 2017 the Collection Policy was already recalibrated to include new collection areas, in line with the mission of a media institute. In the following years, this led to the implementation of a new workflow prior to ingestion into the archive (explained in section 6.3). The ingest of new file formats being controlled via this route, enabled an expansion of Sound & Vision's list of preservation formats (see 4.4, 4.5 and 5.3).

- The description of the generic workflow has been extended with information about the negotiation phase, which is so important for the audit trail of the acquired objects (section 6.2).

- The concept of a preservation menu has made way for an optimization approach: the different levels of preservation result from a combination of circumstances and are primarily related to file formats. This is the more realistic, new approach for chapter 7.

- Since 2021 Sound & Vision is certified to ISO 27001, for its Digital Archive. The primary method for guaranteeing information security, according to this standard, is risk management. This has led to a significant improvement in the organization in this area and hence to various adjustments in chapter 8.

- The traditional model of end-users ending up in a central portal, shifts to a more-dimensional model where you also need to bring collections to users. This is reflected in chapter 9 by adding a paragraph on user driven access.

- All IT policies and procedures associated with digital preservation management are updated. Disk management has come more to the forefront, driven by the storage of new formats on disk. This is reflected in guaranteeing integrity (5.2), storage management (10.3 and 10.4) and also in Optimal preservation (chapter 7).

Reading guide

The document is 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 & 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.

These sections are followed by a chapter on Sound & Vision conditions and arrangements regarding access, discovery, copyright and licenses. The design and implementation of the complex Sound & 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

Introducing the institute

2.1 Mission statement

The mission and the vision[3] of the Netherlands Institute for Sound & Vision (S&V) read:

Mission

A colorful, media-savvy society

We strive for a media-savvy society in which all colors are visible and everyone's voice can be heard. A free, inclusive and creative society in which everyone can participate in media.

Vision

We strengthen everyone's life in media

We are committed to enriching and strengthening everyone's life in the media. As guardian of the media, we preserve and bring the media heritage to life, so that the stories of yesterday, today and tomorrow can be told. We celebrate media in all its colours and, together with our partners, we are committed to smart, conscious and creative media use for everyone.

This is supported by five core values: connecting, surprising, progressive, reliable, colorful, and imaginative. The core value reliable directly refers to sustainable archiving:

Sound & Vision operates as an independent hub in the Dutch media landscape, where Dutch media heritage is preserved sustainably and is in safe hands. We interpret developments in the media landscape and in doing so, involve external expertise from many perspectives.

With the mission and vision Sound & Vision meets her task and assignment: 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 & 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 S&V has been formally responsible for collecting, preserving and presenting

3 https://www.beeldengeluid.nl/en/about/mission-and-vision

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. Later, the institute was further strengthened through mergers with the Museum for Communication, the Press Museum, and Muziekweb. The institute currently presents itself as:

One of the largest digital media archives in the world: Preserving and managing all types of media, from radio and television programmes to video (games), podcasts and social media.

The meeting place about media: We connect people with media and the past with the present. We do this in our colourful building, online and on location.

Reliable expert in media: As the partner and expert in media, we help people to understand and use media better, so that they become media-wise. We do this, for example, with workshops on fake news.

S&V is structurally funded from the government media budget, supplemented by funds, subsidies, sponsors, EU projects, ticket sales and exploitation of the Sound & 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.
Likewise donation agreements with private and institutional donors contain permissions for storing and preservation of the material.

Sound & Vision preserves its collections at various curation levels (as these are defined by Core Trust Seal), depending on the cultural value of the material, the file formats supplied by the depositor, technical limitations and/or priorities in planning pre-ingest workflows.

The S&V collections consist of digital-born radio and television programmes and analogue and digitized legacy collections (film, audio, video, photo). More recently, selections of web videos, podcasts, websites, games, and text materials have been added. Also, digital collections of external partners from the cultural heritage and media domain may be ingested and preserved, as a (paid) service.

The collection amounts to over 1 million hours of AV materials. The S&V's digital repository is referred to as the Digital Archive. Its contents (fall 2025) amount up to 44 petabytes

(excluding 6 petaytes as a service), including backups; 19 petabyte unique material. S&V has its own storage facility on the premises. A second storage facility is housed by a professional commercial partner at the Media Park. A third facility has been implemented on a remote location.

S&V identifies several external user groups or Designated Communities:

- Media Professionals (broadcast, creative, heritage, journalism etc);
- Teachers and Students;
- Academic Researchers;
- the General Public.

Depending on copyrights, licenses and authorizations, these user groups may consult, order and/or re-use material from the collections, each via suitable platforms.

Sound & Vision has a Supervisory Board that supervises the legality and effectiveness of the policy of Sound & Vision, also in light of the media law task of the institute. As a supervisory body, the Board pronounces on, among other things, the budget and the annual accounts, the articles of association, the board regulations and the policy plan. The Supervisory Board of Sound & Vision consists of at least seven and at most nine members and is appointed by the Board itself on the basis of a profile. The chairman of the Supervisory Board is appointed by the Minister of Education, Culture and Science.

2.3 Organizational model

Sound & Vision structured the organization to completely focus its products and services on specific target groups. To describe the organization from a preservation perspective we make use of the functions of the reference model for an Open Archival Information System (OAIS) as shown in de figure on the next page.

The functions **Ingest**, **Datamanagement** and **Access** are embedded in the department Vereeuwigen (archive management). The function **Storage** is assigned to Victorie; the ICT department. Both work together with information specialists from the department Verkennen, where dedicated digital preservation expert staff is available:

- Preservation officer for advice on operational preservation policies and practices;
- Information officers for advice on data management, interoperability and data quality analysis;
- Metadata-managers and engineers;
- Specialists user studies and polls.

In total, approximately 30 FTEs work on the four key functions, in these three departments. All the required skills of these employees have been defined (see Appendix II).

The **Administration** function is a joint effort between Victorie (ICT; systems engineering, monitoring operations, configuration management etc), *Vesting* (facilities management), *Verwennen* (customer support) and *Verhalen* (collection policy, product development, appraisal of donations and networking/negotiating with producers).

In addition, specialist staff is organised in *Vormen* (HR), *Vindingrijk* (legal), Verkennen (business intelligence), *Veroveren* (branding), *Verantwoorden* (finance) and *Verenigen* (secretariat). For preservation, especially the knowledge of *Vindingrijk* is crucial and offers expertise via:

- privacy officer
- GDPR specialists
- Legal expert on collection and access policies, contracts and copyright policy issues.

The **Preservation Planning** function has many different operators. Monitoring the environment of Sound & Vision is a task for the product managers. They are accountable to Target group owners; each of them responsible for a particular group of the Designated Community.
Note that some of the user groups do also have an important role as producer, and are monitored in that sense as well, in close coördination with *Vereeuwigen* and *Victorie*. In addition, monitoring changes in the technology environment is primarily a task for *Victorie* (ICT).

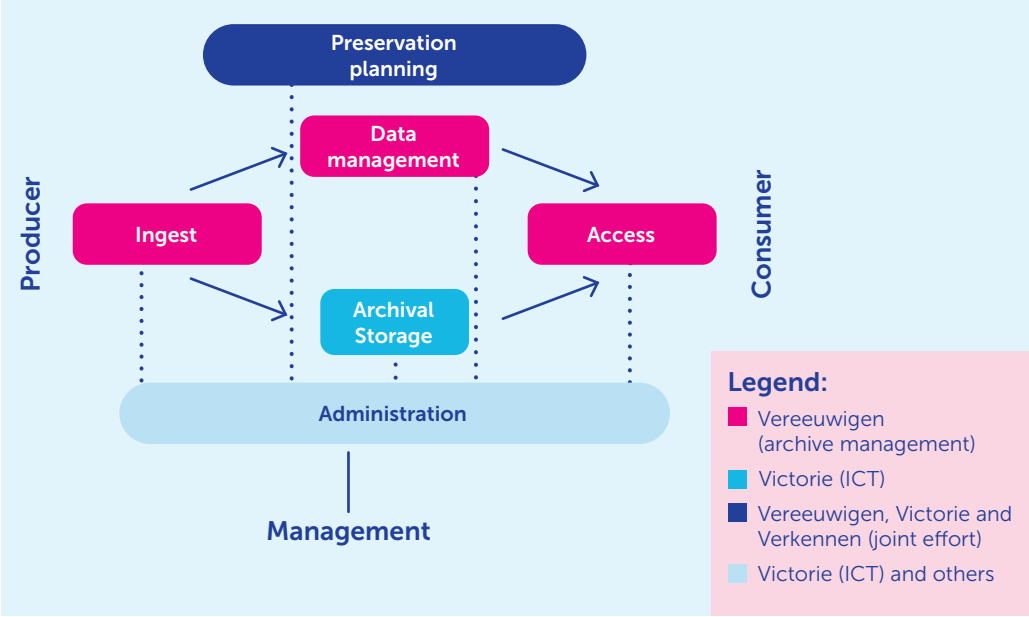


Fig. 2: OAIS functional model

However, the key function of Preservation Planning is to evaluate the file formats that are presented to the Archive and periodically recommending updates for the Archive standards and policies. This function is carried out by a designated working group of specialists from the three core departments: *Vereeuwigen*, *Victorie* and *Verkennen*, in the capacity of digitisation and ingest experts, system engineering and preservation officer.

The managers of Vereeuwigen, Victorie and Verkennen consult with each other in a preservation board (VoorAltijdEnEeuwig). They can be assisted by a specialist from their department. The board is chaired by a target group owner, who is also the escalation line; new policy guidelines will be presented to the top management via the chair.

A matrix organization

All 'department managers' manage their staff and provide the appropriate processes and working methods within their teams. These managers are responsible for the operationalization of certain high-level policies, as is the case with 'digital preservation'. The primary departments are accountable to the 'cluster manager'; staff departments (HR, legal, finance, branding and secretariat) to the director.

'Target group owners', each dedicated to their own Designated Community (i.e. Media Professionals, Teachers and Students, Heritage Professionals and Researchers, the General Public), are ultimately responsible for S&V products and services to meet the needs of their target group[4]. To reach their goals they may appeal on the commitment and expertise of the departments. The target group owners are a member of the Broad Management Team, underlining the importance of the Designated Community as a guidance for the Sound & Vision strategy and operations. This Broad Management Team is the extensive version of the so called 'Small Management Team', namely the director together with the CFO.

2.4 Knowledge, research & development

S&V initiates national and European research that makes media heritage open and searchable, monitors relevant innovations in media archiving, participates in research projects, and experiments with new technology[5]. Some research projects (EU-funded) explore specific developments in the field of preservation. An example is the current TEMS project[6] (Trusted European Media Data Space), that explores how to make media metadata consistently and reliably searchable on a European scale. One of the promising suggestions is to use an ISAN[7] number for the unique identification of audiovisual works in the European media sector.

Based on practical work and research, the Institute develops, collects and actively disseminates knowledge and information on all aspects of digital audio-visual archiving via a series of national and international platforms. Knowledge on digital preservation in particular is gathered via active participation and involvement in digital preservation expert platforms such as NDE, FIAT/IFTA, PREMIS, iPRES, and OPF. And moreover by annual visits to dedicated IT conventions (IBC and NAB).

Skills and competences are regularly updated by in house exchange sessions with other archival institutes (like MEEMOO, Rijksmuseum), in/external training (e.g. the S&V Knowledge sessions, IPR workshops, annual FRAME training INA), and international conference attendance, e.g. conferences and seminars organised by FIAT-IFTA and iPRES.

At national level, preservation knowledge is generated and disseminated by the participation of Sound & Vision in the Dutch Digital Heritage Network (NDE)[8]. Sound & Vision functions as the national AV hub 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.

Furthermore, as a founding member of AVA_Net, the national audio-visual heritage network, Sound & Vision performs a key role in the execution of the core task of AVA_Net: bundling and exchanging knowledge on all aspects of audio-visual archiving, within and between the larger and smaller organizations and archives that hold audio-visual materials.

4 the **library target group** falls outside the scope of this policy plan; the lending music-collection is still separate from the heritage collection. Research into the overlap and integration is ongoing.
5 https://www.beeldengeluid.nl/kennis
6 https://www.beeldengeluid.nl/en/research/projects/trusted-european-media-data-space
7 https://www.isan.org
8 https://netwerkdigitaalerfgoed.nl

Strategic policy framework

3

Anchoring preservation in institutional policy

Sound & Vision 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 & 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 & 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

In Media, the multi-year Policy Plan[10] sets out Sound & Vision's strategic ambitions, based on four strategic theme's, or 'circles'. From within these circles, eleven specific objectives have been set to ensure that in 2027 we are the most modern media archive in the world.

9 Act 2008, art. 2.1: https://wetten.overheid.nl/BWBR0025028/2026-01-01#Hoofdstuk2
10 In Media, Multi-year policyplan 2022-2026

The heart of the circles, being the guardian of Dutch Media heritage, is the core of our archival task and the basis for our outward-looking activities as the 'driver' of media heritage. Three objectives are set for this circle:

1. Open for archivists
2. Reliable, sustainable storage and services
3. Maximum collection accessibility.

The second circle is about keeping media heritage alive. In the third circle, S&V will celebrate the media and in the fourth, it will strengthen the media ecosystem by promoting press freedom and news literacy.

3.2 Collection policy

The Sound & Vision Collection Policy 2025 defines the archival responsibilities of the Institute, the legal frameworks in which these responsibilities are exercised, and the principles guiding collection, preservation, and access. It outlines the collection profile, specifies policy principles, and identifies all user groups.

This policy is forward-looking. It addresses the challenges of the rapid changes in the media landscape, the increasing complexity of formats, both digital-born and digitized, and the management of materials that do not easily fit into existing access portals. Like web archives or the metadata for object and museum collections, which require migration to ensure accessibility for researchers, media

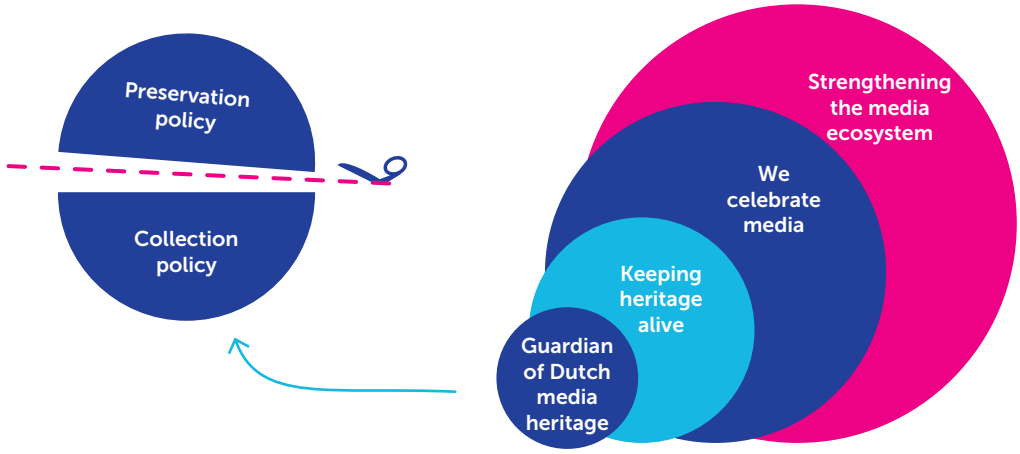


Fig. 3: Strategic goals S&V

professionals, heritage experts, and the general public.

In 2017 the Collection Policy was revised to incorporate new domains, including online media, games, websites, and the written press. A new classification system was established to support acquisition and selection processes. The shift emphasized the cultural, social, and political significance of media productions, rather than their technical medium or platform.

In 2022, the collection structure was expanded from four to five domains with the addition of Communication Culture.
The structure now consists of the following five core domains: 1) News and Society, 2) Culture and Entertainment, 3) Amateur and Business Productions, 4) the Media Landscape (production, distribution, reception) and 5) Communication Culture, documenting the history of Dutch postal and telecommunications services.

Sound & Vision will continue to expand and enrich its collections in ways that reflect the cultural, social, and technological dynamics of media. Preservation will follow a pragmatic yet ambitious approach, ensuring long-term sustainability while adapting to emerging formats and practices.

3.2.1 Appraisal and selection criteria

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

1. **Intrinsic features:**
   - Historical (cultural, media and communication history): key figures, events, places, activities, time periods, processes, and lifestyles.
     - Artistic: works by film, radio and television makers, illustrators and journalists of exceptional quality or representative of a style, movement or oeuvre.
     - Informational value: importance for academic and scientific research.
   - Social and societal
     - Social: relevance to individuals and groups in Dutch society today.
     - Collective memory: events and media moments widely recognized and with which large segments of society identify.

2. **External features**
   - Usage
     - Museum value: suitability for use in presentation, education and research, both on-site and online.
     - Media value: potential for re-use by media professionals, amateurs and the general public.

- Heritage value
  - Provenance: origin in a distinctive source or collection.
  - Ensemble value: completeness or contribution to, existing collections.
  - Rarity or uniqueness.

All acquired digital media are to be preserved in principle. No distinction is made between "core" and "supporting" collections.

Sound & Vision acquires heritage in two ways: - Passive intake: the automatic flow of materials from public broadcasters. - Active acquisition: divided into archival materials (audiovisual, image, and paper documents) and museal objects (2D and 3D).

Acquisition criteria for both areas are updated regularly in response to developments in media production, distribution and consumption. Technical infrastructure, professional expertise, and storage facilities will develop in step with the growth of the collection. While practical constraints are recognized, lack of space or technological capacity should not prevent the preservation of heritage for future generations

3.2.2 Collection typology

From the perspective of content type, the collection can be classified as follows:

A. Radio and television programs

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:

- Television: 100% of broadcast Dutch-produced programs from the public broadcasting organizations from the year 2007 (i.e. the start of digital broadcasting production).

- Radio: 100% of the radio programs broadcast by six Dutch public radio channels.

- Digitized analogue film, video and audio material consisting of productions from public broadcasters through to the year 2007.

- Each year, 800 hours of unedited news and current affairs items (clean feeds).

- Commercials have been part of Sound & Vision's collection policy for many years. However, advertisements broadcast after 2008 had not been systematically acquired. Since 2024, a partnership with Ster ensures that its rich archive, along with all new commercials broadcast, is preserved in the collection.

- Researchmaterial
  - Each year, unabridged recordings of two complete weeks[11] of Dutch radio and television broadcasts, public and commercial (3000-5000 hours each).
  - A selection of broadcasts from commercial stations: radio approximately 800 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. non-lineair AV

Productions of video or audiocontent that does not follow a fixed, scheduled timeline; online content intended for on-demand consumption.
This material includes:

- Online content NPO. More and more public broadcasters are finding us for the archiving of their online content. Some examples:
  - In consultation with NPO and various public broadcasters Sound & Vision archives non-lineair content that was published on NPO Start;
  - In recent years Sound & Vision created in coorporation with NOS (current affairs) an automated workflow for non-lineair NOS-content like for example "Jeugdjournaal Online", "Explainers" and "Examenspreekuur: 1K+ items;
  - broadcaster NTR (music registrations Pinkpop, North Sea Jazz) 3.100+ items.
- Podcasts:
  - All podcasts available in the NPO Podcastgenerator; 47K items
  - other than produced by public broadcasters (private productions); 38K items

- Approximately 27.000 webvideos have been collected since 2006, covering online YouTube and Tiktok phenomena, political campaigns, branded content, parodies, and social critique. Access is restricted to on-site consultation.

C. Material that is unpublished or published via traditional channels.

Raw material and productions for cinema, physical distribution or private use, for archival collection and subsequent research.

- Digitized analogue film, video and audio material, e.g. films produced by amateur filmmakers and media creators, corporate films, educational films. And oral history, soundscapes, sound effects, and spoken word recordings. Since 2014, digitization has continued on a selective basis, with plans to expand capacity. Future digitization

11 One of these coincides with UNESCO World Day for Audiovisual Heritage

will also include materials from the Communication Culture collection, such as the PTT audiovisual and photographic archives.

- Yearly 300+ music recordings from public broadcasters (MOZ), predominantly audio recordings for radio.

- Digitized commercial musicrecords or any other analogue carriers for private or commercial ends.

D. Games

Dutch games from the 1980s onward, acquired in digital form and, where possible, as original carriers (e.g., tape, floppy, CD-ROM). 400+ Games so far. Agreements with developers govern acquisition, preservation, and legal use. Sound & Vision also seeks to collect related production and marketing materials. Ongoing research and experiments refine selection criteria and ingestion methods.

E. Images

Of the originally analogue photo collections, that mostly refer to broadcasting productions, over 2 million TIFF files are available. The Press collection combines analogue and digital-born materials, including over 40,000 political cartoons and press illustrations. Research is underway to address the sustainable preservation of digital techniques used by contemporary cartoonists.

F. Websites

Since 2014, representative websites related to media culture and online journalism have been archived. 1K+ sites at the moment; periodic harvests still add content to over 300 of these sites. The goal is to build a balanced national web archive, aligned with domains defined by Sound & Vision and complementary to other Dutch web archives (e.g. National Library, National Archives). An opt-out approach is applied. Due to copyright, access is restricted to on-site consultation.

G. Communication culture

Future ambitions on communication culture include capturing and preserving contemporary communication such as telephone calls, pager codes, voicemails, emails, WhatsApp, and social media traffic. To achieve this, Sound & Vision will need to develop a dedicated selection strategy and technical preservation methods to ensure long-term accessibility and reuse.

H. Music

Since 2021, the music collection of Muziekweb has been incorporated. S&V is responsible for the management of both its analog and digital carriers. The collection is continuously expanded through the acquisition of new analog materials, which are subsequently digitized, as well as through additional acquisitions and the inclusion of born-digital content. A selection has been designated as cultural heritage. To ensure long-term

preservation, appropriate preservation strategies and methodologies will be identified and implemented in the near future.

3.3 Agreements on access

Sound and Vision identifies several types of external user groups or Designated Communities. These are media professionals, heritage professionals and researchers, teachers and students and the general public. 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 Archive Agreements between S&V and the NPO, and S&V and the Collective (rights) Management Organizations (CMO's)[12] concerning access to the collections for cultural, educational and (scientific) research purposes, leading to several agreements between S&V and organizations in the above-mentioned fields like CLARIAH [13] Consortium, European research projects and individual universities.

2. The Service Agreement between the Netherlands Public Broadcasting Organization NPO, individual public broadcasters and Sound & Vision (regarding the basic services provided by S&V to NPO and the individual broadcasters required by the Dutch Media law) and the Chain Agreements (Ketenafspraken) between the NPO, the public broadcasters and Sound & Vision (regarding processes within TV production all three parties are involved with).

3. The Covenant with NPO and the collective rights management organization StOPnl regarding out of commerce (AV) works, based on European and/or Dutch law[14].

4. The Covenant with Buma Stemra regarding out of commerce (music) works, based on European and/or Dutch law[15].

5. Donation agreements with individual donors, professional

12 CMO's involved are the ones for Independent TV/content producers, composers, performers, directors, script and dialogue writers and creators of (non moving) visual works

13 Common Lab Research Infrastructure for the Arts and Humanities

14 https://kennisbank.avanet.nl/out-of-commerce-convenant-audiovisuele-werken

15 https://www.kvan.nl/publicaties/convenant-muziekwerken

and private, containing licensing clauses or the complete transfer of copyrights to S&V.

6. Contractual agreements or government guidelines related to certain legacy collections. For instance the collection of the Government Information Service, Foundation Film and Sciences, and the collection for amateur film.

If no contracts are available, for instance, because copyright owners are unknown, S&V makes use of the possibilities available to Cultural Heritage Institutions under European and/ or Dutch copyright law[16] for external consultation and use of archive material, always under supervision of the S&V Legal department.

The S&V Legal department also looks into parts of the collection that might be in the public domain (PD). If and when the public domain status of a (part of a) collection has been ascertained by the S&V Legal department, consultation and use will be available to all without the need for a license. Although there are cases in which privacy or ethical concerns might lead to the S&V legal department checking requests for consultation or use.

16 For instance the European out of commerce works regulation that was transposed into Dutch law or the fact that, according to Dutch law, a cultural heritage institution is allowed to make the contents of its archives available to the public within the walls of the organization itself.

Preservation principles 4

Scope and fileformat policy

4.1 Digital collection management

The common thread in structuring all elements of the preservation workflows and environments at Sound & Vision is the Open Archival Information System (OAIS), ISO Standard 14721[17]. 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). This standard serves as the framework for organizing the layout of the Digital Archive regarding aspects such as:

1. The responsibilities, functions and roles of Sound & 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 provided to users and the safeguarding of intellectual property rights;

4. The manifestations of essence (film, video, audio, images etc) and the metadata (descriptive metadata and preservation metadata);

5. The AV formats and standards employed.

17 http://www.oais.info

4.2 Core definitions

Within the policy framework of Sound & 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.

In the context of preservation, the term **sustainable** means that the digital collections are preserved in such a way that they remain accessible and usable for their target user groups for the long term, regardless of future technological or organizational changes. This includes:

- Intellectual access:
  Ensuring that the content remains understandable, with all the necessary metadata to interpret it (context, structure, format).

- Technical access:
  Ensuring that the files can still be opened, viewed, and used on future hardware and software, if necessary through strategies like format migration or emulation.

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

- Integrity:
  The object is demonstrably unchanged at bit configuration level, from the instance it was stored or migrated.

- 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.

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 & 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 collections come within the preservation scenarios elaborated in this document (see Chapter 7: Optimal preservation). 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.

4.4 Archive formats

As a national institute, Sound & 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 would be a lossless format, preserving all information from the moment the digital file 'is born'. The appropriate delivery formats for the various Designated Communities can be derived from that file, time and again. However, the possibilities for choosing lossless as a standard are in fact determined by practical, political and/or financial circumstances. An obvious example is the immense collection of NPO broadcast programmes (amounting to 85% of the entire Sound & Vision collection), for which the lossy 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 & Vision has recently introduced some differentiation, given 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 original format of the born-digital material and the way it can be offered to or becomes available to the archive. The focus of this effort has been to bring the choice of preservation format for these parts of the collection more into line with the archive's own national responsibility and ambitions.

The use of a lossless format for the sustainable preservation of the variety of files that S&V wished to add to its digital archive seemed desirable. But weighing the advantages (no loss of information) against the disadvantages (storage capacity, processing speed etc.) yielded a different outcome. As long as the formats offered meet the requirements of the current designated community, align with the task of S&V as a national archive, and can be supported by the existing infrastructure, a certain variation was accepted and there was no compelling reason to migrate. This is on the condition of extensive documentation of technical properties and validation of the submitted files. Files that meet these requirements are accepted 'as they are'.

Files that do not meet these requirements cannot be accepted and must be migrated. For the time being however, S&V has chosen not to migrate anomalous files to a lossless format yet. Instead, the preference is migration to 'the most suitable archive format', give or take the potential loss of information. The condition

under which this policy applies is that, given the limited size and small numbers in which this currently occurs, the source format can also be preserved, to potentially fall back on in the future. At the same time, S&V remains open to developments in this area. There are still fundamental and technical questions to be resolved in the ongoing research into the use of a lossless format, intended as a preservation master.

4.5 Preservation formats

Until recently Sound & Vision only had a limited number of formats for long-term preservation. For these formats the Digital Archive is able to issue detailed guarantees for the sustainability each being a well-documented industry standard that operates on current software, as used in the audio-visual domain.

These detailed guarantees were difficult to maintain when choosing a preservation format for documents (print material) and new media (web video, pod cast). In the end it was decided that these formats (as a minimum) must be able to be indexed within Sound & Vision's internal technical and catalogue infrastructure, so that derived files can be produced for viewing and delivery. It must also be possible to carry out quality analyses of the format using the standard analysis software of the MAM-system. Within these constraints, archive formats could be added for which sufficient guarantees of sustainability apply, despite the absence of fixed, uniform industry standards.

For the selected AV-formats (including print and image), Sound & Vision is able to give solid guarantees of sustainability. Meanwhile, the organization has also started collecting materials that are non-linear, like games and websites. The sustainability of these media depends on the combination of formats, software, and even hardware. The environment in which the material operates largely determines the experience that is the subject of preservation. In order to guarantee sustainable access to these interactive forms of media, much remains to be investigated and tested. For the time being, S&V uses the most common formats in this field.

4.5.1 Preservation master, mezzanine and proxy files

Preservation practice at Sound & 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. Additional files such as subtitles, xml-files with QC outcomes etc, are part of the AIP and stored als 'auxiliary'. This type of files is not subject to

general guarantees for playout; only bit preservation is guaranteed. Unless specific service agreements with data producers are in place.

Preservation master - moving image

The preservation master is the archive format permanently preserved by Sound & 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 & 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 S&V.

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 & Vision, MXF serves simultaneously as the preservation master and the standard delivery format. If the MXF meets the specifications, S&V guarantees 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 & 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 accompanyingtechnical metadata, which is delivered separately as a component), and the material can be transcoded for delivery in other formats (mainly mp4).

For scanned films 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. An exception applies to certain legacy collections that are being scanned at high speed to make them accessible to users as MXF files. For this high-speed scanning, a solution called Fast Forward is used. It enables analogue film collections to be digitized much faster than traditional scanning processes, though with compromises in image and sound quality. Upon request, scanning to DPX can still be performed.

In addition to MXF and DPX, Sound & Vision also accepts MP4 and as preservation master for moving image, when these formats are delivered as the original source material. This approach follows the institute's preservation practice of keeping the deposited file in its original form and avoiding unnecessary transcoding that would cause additional quality loss. Although MP4 is a lossy format,

its widespread use in production and online distribution makes it a practical choice. If the MP4 meets our specifications, guarantees can be issued for permanent accessibility, and the material can be transcoded into other delivery formats when required.

Preservation master - sound

For audio the WAV format has been the lossless industry standard for broadcast. It is a widely accepted standard. For new audio collections such as podcasts, MP3 was introduced as alternative audio format. While MP3 is lossy and less durable than BWF/WAV, it is often the only source format available. Creating a WAV copy would not improve quality and would lead to unnecessary duplication.

Preservation master - still image

For still images, both TIFF and JPEG are maintained as preservation masters. TIFF, being lossless, is used for high-quality scans of photographs and documents that require long-term preservation. JPEG, although lossy, is often the native format in which material is created and delivered.

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 only supply DPX on special demand. 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 format of the preservation master for television material. To guarantee fast delivery and simplify work processes the mezzanine MXF formats are also stored and preserved under the same conditions as all other MXF files.

In anticipation of the new portal for the public, another mezzanine is made from all the MXF files to further reduce their size and optimizing access times. These are in a high bitrate H.264 MP4 format, such that they can be kept on on-line disk rather than near-line tape, so bulk upload, analysis and other mass operations are much easier to execute.

A third mezzanine is made in a streamable HLS format on a public cloud provider. This version's quality is somewhere between the MP4 mezzanine and the proxies, and intended to deliver to end user devices (TV, mobile, tablet etc.) in the appropriate quality[18].

Proxy files

Proxy files are produced by Sound & 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 & Vision's catalogue infrastructure, i.e. the internal catalogue (Studio), the online catalogue for the media professionals (MPP) 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 & Vision is Mp4. Mp4 (codec H264) 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 & Vision's catalogue infrastructure: in the case of distribution via the internet, it is not possible to continue to offer playout guarantees. Storage, format specifications and the affected software components 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 playout will remain a long term access strategy, set up to serve the S&V 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 & 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 & Vision to offer support in transcoding or arranging transcoding of the submitted format to the preservable format. This is only possible if there is agreement on the costs; S&V will factor in the value of the material to the archive. For archival material, the delivered files are always stored alongside the new master file. This means that in the future, it will always be possible to fall back on the original material, even though S&V provides no guarantees regarding its playability. The new master may have some information loss compared to the source, but its playability is guaranteed. The sometimes complex trade-off in choosing the preservation format is made by the acquisition team

18 In the future, these could replace the existing streaming formats that were specifically generated for the education portal. This involved a subset of the archive, consisting of MP4 files in various bitrates.

in consultation with specialists from the digitization and collection policy departments. In some cases, feedback is requested from the working group for preservation planning. Fig.4 shows once again the relationship between (non)preservable fileformats and the different filetypes.

For (commercial) tenants, this same policy is applicable. 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', at the tenant's expense. 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-preservable formats must be taken in advance of the ingestion phase, in consultation with the depositor, and laid down in the contracts and SLA's.

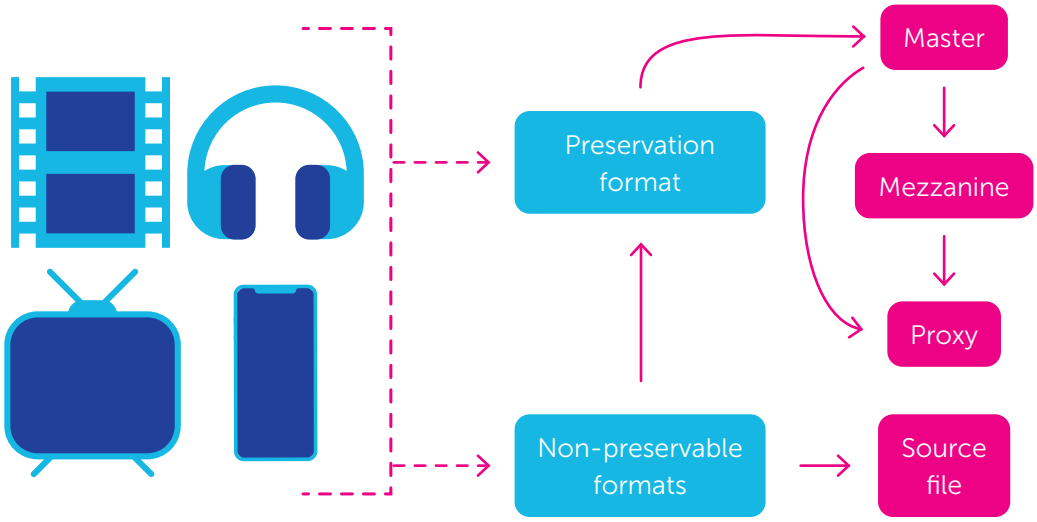
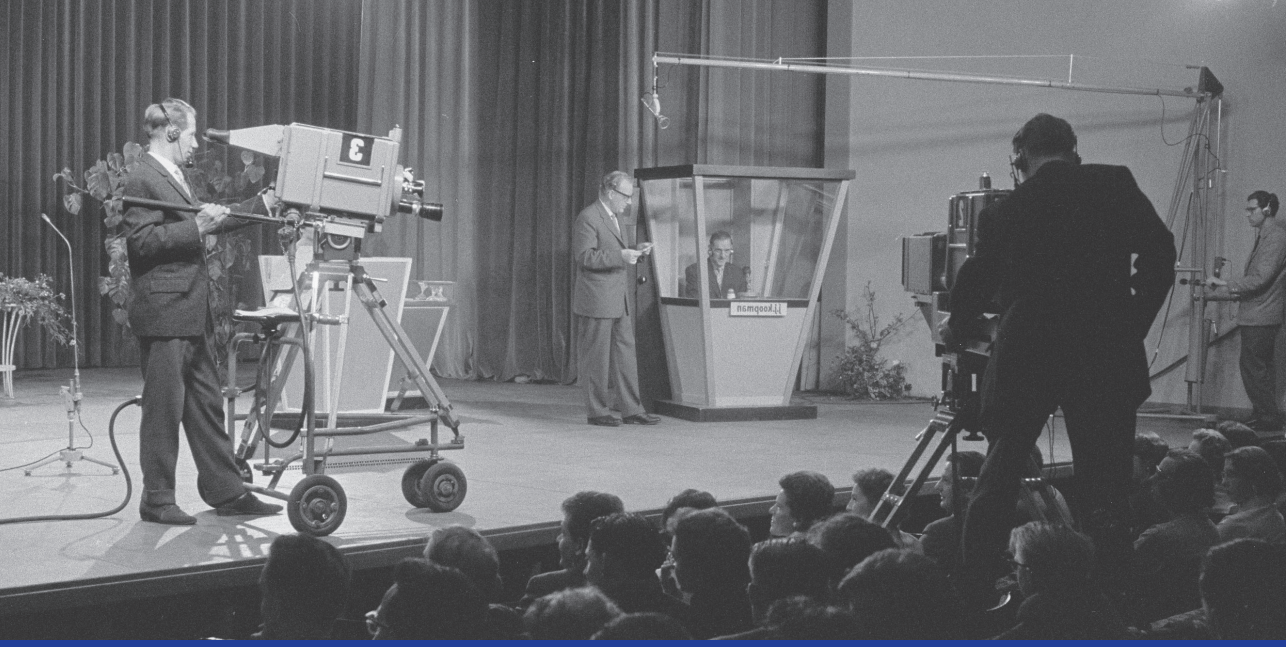


Fig. 4: Fileformats and filetypes



Preservation strategies

5

How to deliver on our promises

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 technical quality analysis is carried out on the object to demonstrate this fact. This analysis is limited to technical and archival authenticity and does not constitute a claim about the factual accuracy, truthfulness, or real-world correspondence of the content.

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 AV objects are checked by the software program Baton[19] and validated against the technical format profile. S&V has defined a quality profile for each preservation format. The check against this profile confirms that the container format can be read, and it will detect errors (deviations) in the container format as well as the audio and video codec. For AV/moving image (MXF, ProRes and mp4) this check is integrated in the MAM-system. For sound (WAV, Mp3) this check is performed during pre-ingest for all one-off custom acquisitions. For still images and documents (tiff, jpeg, pdf) such profiles are not in place yet. Instead visual checks are done, to see if the system successfully completed the ingest, including the generation of derivative files (proxy).

In the MAM system also rejection criteria for metadata are defined for several high volume ingest flows (NPO, ECV, 2K), according to agreed-upon metadata specifications. For one-off custom imports checks are built for metadata that must align with thesaurus terms. The import for each object is logged by a tasklist; alle issues and errors are flagged to draw the mediamanager's attention.

Ad 2. Life cycle management

Recording all events in the lifecycle preserves the 'chain of custody' of each object, i.e. the details in about 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

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

describe where the object came from and who has subsequently processed and used it. In this way, provenance metadata demonstrates that the object has not been unintentionally altered.

The MAM-system provides the formal identification and registration of a large amount of process data. All files and metadata, from whatever source, go into the same system through standardized ingestion workflows. After initial registration, all changes (by whom and when) on metadata are recorded in an audittrail. Once the files are stored in DIVA, additional provenance metadata based on events in the storage system is also recorded.

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 2.0[20] 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

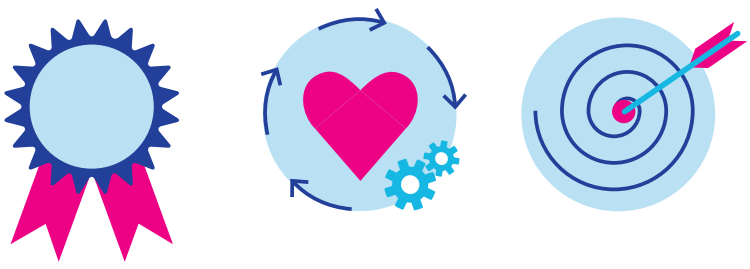


Fig. 5: Authenticity controls

be preserved is decided on each individual formatmigration. These decisions are made by the "new file formats" working group, as part of the preservation planning function. In their decision process they will take into account 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. They will weigh these factors against the more practical issues, such as processing time and organizational costs.

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 (i.e. for what purpose does the user or depositor want to use the object?) and Sound & Vision's task and mission as a media archive. Due to the volatility and temporality of context, and uncertainty regarding future possibilities, S&V always chooses to also preserve the source format. Even though S&V does not take responsibility for long-term playability, users can always revert to the most original format if desired.

5.2 Guaranteeing integrity

Any digital archive must comply with the requirement of persistence. For the files this means that they must come out precisely as it went in, i.e. complete and identical bit-for-bit. This data integrity is secured by Sound & 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 as early as possible in the chain prior to ingestion, or at least when the AV-object is stored in the Archive Management system (as is the case for daily ingest of tv-programs and so called clean feeds). If the checksum is calculated prior to ingestion it is stored in the metadata in the MAM, as well as in the Archive Management system. This is particularly the case with one-off acquisitions.

For all files that are 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, the checksum is calculated again and compared to the checksum previously stored. This fixity check is done in order to ensure that the object is identical to the data that was transferred in an earlier phase. For archival files that are stored on disk

20 Preservation Metadata Dictionary 2.0: http://publications.beeldengeluid.nl/pub/615

file, monthly, the influx per format is archived as an object in the Archive Management system. When writing the file in the Archive Management system, a checksum is created. This checksum can be used for fixity checks on any future action involving this object, thus guaranteeing the integrity of the underlying files.

LTO carriers on which the materials are stored, are transferred to new 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 migration of the materials (e.g. if it turns out that a tape cannot be read, or a fixity check fails), the operator restores the file(s) from a copy, that will exist thanks to the storage plan. This approach allows for the continuous monitoring of the integrity of the entire archive.

5.3 Format migration

The choice of digital AV-formats ingested and stored in 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 for all video and audio. The only exception to this rule concerned digitized film; this is digitized at a higher quality and made available via a mezzanine file in the broadcast format.

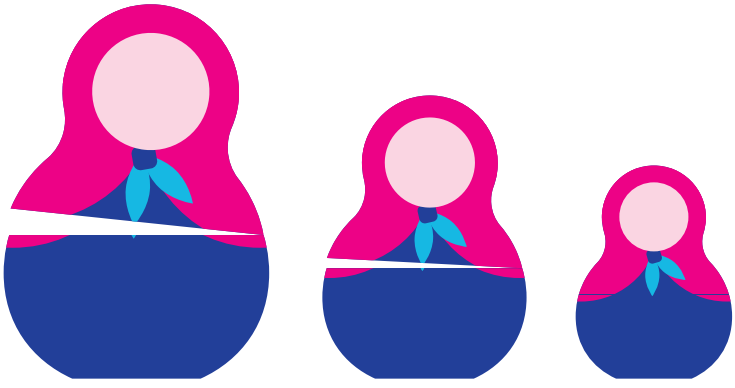


Fig. 6: Integrity controls

Because of the traditional focus on the standards of the broadcast production environment, Sound & Vision has always been able to keep the number of preservable formats limited. If in the future any of those formats would threaten to become to no longer be widely supported, Sound & Vision will consider migrating to another format. This type of migration has not been necessary 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 would always be carried out in consultation with the depositor, the NPO.

Non-public broadcasting material, if not already produced or delivered in a preservable format, would generally be normalized before ingestion into the Digital Archive. As a result of the increasing emphasis on Sound &Vision's role as an institute for media culture, its collection policy has been expanded substantially to include objects other than mere broadcast AV. The Institute supports the ingestion and preservation of more and different types of formats. These file formats (mp3, mp4, prores, pdf) are less standardised and sometimes even proprietary (prores) and appear in a wide variety of codex, resolution, bitrate and other technical attributes. Normalising would alter these files with probably loss of information and/or inefficiency in storage. Therefore these objects are initially stored 'as is', with playability checked in advance.

For each of these formats a profile has been created that defines a range of technical values, that are supported by the MAM-system; access, playability and delivery are guaranteed. Format migration only occurs when the file does not meet the profile's preconditions. Format migration takes place to an appropriate archive format. The original file is preserved as part of the audit trail.

5.4 Other strategies

Some material that must be preserved, because the content is part of the collection policy, needs a different strategy. This applies to media that, due to their interactive properties and/or dependences on hardware and software, cannot be unambiguously stored in an archive format (games and websites). It also applies to all deviating formats that cannot be migrated to an archive format because of extensive loss of quality and/or high costs. The latter are stored in the so-called Dark Archive.

a. Games

The games collection consists of a variety of games on different types of physical carriers or digital formats, like exe, apk or iso. For physical carriers (mostly games from the 80s to 10s) 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, CD-ROM, DVD, cartridge etc.) are stored as objects for exhibition purposes. Games in digital format are acquired as a DRM-free build (where possible), so with no copy-restrictions, and are also checked in terms of their eligibility for emulation as a preservation method (which comes down to the ultimate question: is an emulator for this gametype available?). The dominant preservation method for games is emulation[21], 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 PREMIS[22] notion of 'environment' is used to structure the preservation metadata that apply to these environments.

For each game, the disk image or digital format is stored together with the emulated environment and documentation in a preservation package. The environment is included in the package because the custom settings and fine tuning is unique for each individual game. Once ingested in the archive system, the files will be accompanied with a checksum to check integrity of the files in the future. Emulation is an intensive process in which settings of the emulator and emulated operating system are tried out until the game is playable. This also involves subjective judgement for example, concerning the speed at which the game runs, or the use of colors. All settings are then documented. Gameplay videos and screenshots are recorded, also with the purpose of evaluating the emulation outcomes in terms of 'look and feel' in support of the authenticity of the game.

Future acquisitions might bring additional types of games, for example VR or new portable devices. 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 (for community gaming or downloading updates and new content), specific hardware to play the game, or where emulation software is not (yet) available, documentation will play a vital role as a preservation method. In unique cases S&V even works closely together with the developer to create an archival-version of the game to make emulation possible after all.

b. Websites

Websites are harvested using state of the art crawling software, resulting in a WARC file (ISO 28500) for each domain. The harvesting is performed using the "Archive-It" collecting and accessing service from external supplier The Internet Archive. Archive-It also functions as the access portal for the entire Sound & Vision website collection.

Most websites in the collection are harvested on a recurring basis, with an annual or semi-annual frequency. The WARC files in Archive-It are automatically updated to include any newly harvested content from the domain. Every crawl is opened and reviewed; in case of deviations, an investigation follows based on the crawl report. Problems can have two causes. Firstly: incorrect or incomplete data, in which case the crawl must be done again or we have to acknowledge that the website is currently too technically complex. Secondly: it is a playback issue. The crawl itself is then correct, but the WARC viewer cannot process or display the data properly. The crawls are then accepted, with the hope that the viewer will be able to play them back in the future. After accepting the crawl, the integrity of the files are guaranteed by the external supplier. Periodically Sound & Vision receives a dump from Archive-It with all WARC files and stores this as a backup.

In case Archive-It proves unable to adequately harvest a domain (most often because of certain types of dynamic content) and the website is of high value, Sound & Vision uses the Archiveweb.page software (formerly known as Webrecorder) to traverse and crawl the website manually. The results of this process are also stored in a WARC file, which is uploaded to the Archive-It environment. In rare cases where both automated and manual crawling is not feasible, a website is documented by capturing a screen recording video. Screen recordings are ingested in DAAN. Because of the dynamic nature of the websites it will always be very hard to determine the authenticity of the archived version. The methods to ensure this are still very much in development.

c. Dark archive

Deviating formats that cannot be migrated to an archive format because of extensive loss of quality and/or high costs are stored in the so-called Dark Archive. For instance files in mpeg2 transport stream. These files are received from a third party that generates them during the broadcast process. The format is unsuitable for long-term preservation. Generating a suitable format would be complex. The files are only retained for research purposes. This concerns two full weeks of broadcast material (hourly blocks) from all nine public broadcasters, and a selection of 500 hours of programming from commercial broadcasters.

21 an-overview-of-emulation-as-a-preservation-method: https://www.clir.org/pubs/reports/an-overview-of-emulation-as-a-preservation-method/

22 Preservation Metadata: Implementation Strategies. https://www.loc.gov/standards/premis/

The preservation workflow

6

Implementing the OAIS information model for each stage

6.1 Information model

Sound and Vision has developed an information model 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 reflected in the actual ingestion, storage and access processes.

The information model gives a generic description of which workflows are used for the (pre)ingest, storage, data management 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. And to check whether necessary events have taken place. A comparison of provenance metadata with 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.

The informationmodel is based on the OAIS-referencemodel (see fig.7). To enable the files and accompanying metadata to be identified within the processes at a conceptual level, information objects have been created, known as information packages. The composition of these package types will depend on its role in the process: 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).

The workflows and the description of the packages together form a framework for information policies: guidelines on what information is needed to guarantee the integrity and authenticity of the content and to make the content searchable for S&V's audiences, given any strategic and legal constraints.

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 strategy, materialtransfer and/or temporary storage rights, quality controls, metadata and reports. See fig.8 for a schematic representation of the information package in this submission phase (SIP).

In practice, a distinction can be made between contracts that arrange a *structural influx* of (born-digital) material over a longer period, versus *one-time contracts* with depositors who transfer their material (analog or digital) all at once.

Contracts for the structural influx of programs contain extensive details regarding descriptive and administrative metadata, which will be processed according to pre-established mappings into the data schema of the MAM system. Agreements are also made concerning the quality and completeness of the data, and the procedure to be followed in case of deviations or failures. Metadata may be added in case the program meets specific formal criteria

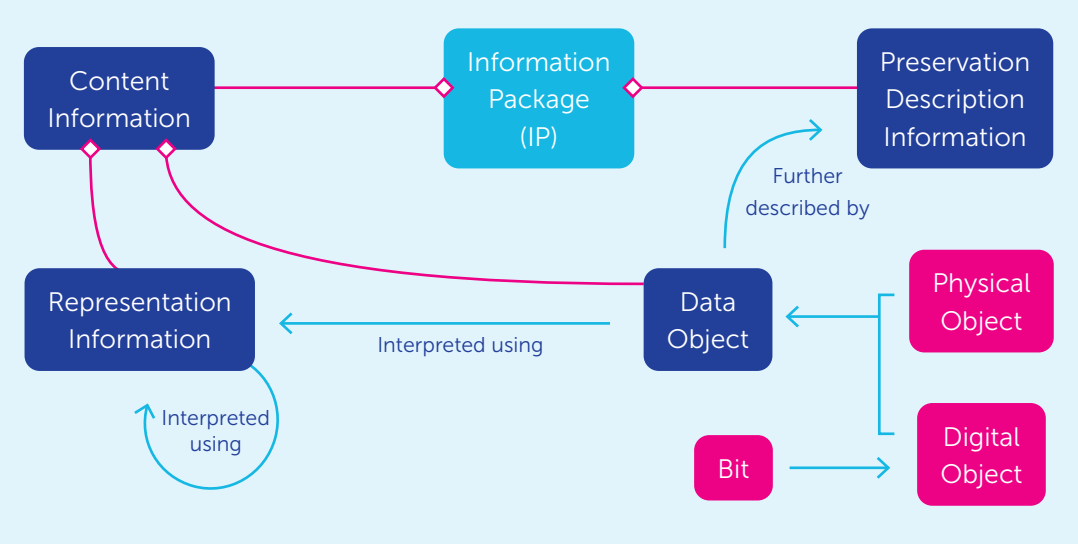


Fig. 7: OAIS information model

or for reasons of topicality (genre, category). These agreements are translated into automatic checks, to be executed during the data ingest process of that particular material.

The execution of these events is recorded in tasks and thus becomes part of the material's provenance metadata. S&V can reject and/or correct the material if upon submission it does not comply with the agreed specifications. In case of structural errors, the data producers/ depositors will be contacted by the media managers in order to improve the quality.

Curation levels are only practised for broadcast programs. In case of high valued programs a mediamanager will add a curationlevel A; default curationlevel is B. Music programs with a standard music-talk format are assigned C.

For the one-time offers, registration follows a regulated step by step procedure. Information about material that S&V rejects in advance is provided on the website. An online form asks the provider for the key details of their offering. Upon receiving an offer, an initial registration takes place and an assessment is made to determine whether the offer falls within the S&V collection policy. Subsequently, a ticket is created, in which all necessary steps such as selection, further assessment, assignment of a definitive storage location, and any digitization are planned and followed up. All correspondence with depositors, including reasons for rejection and

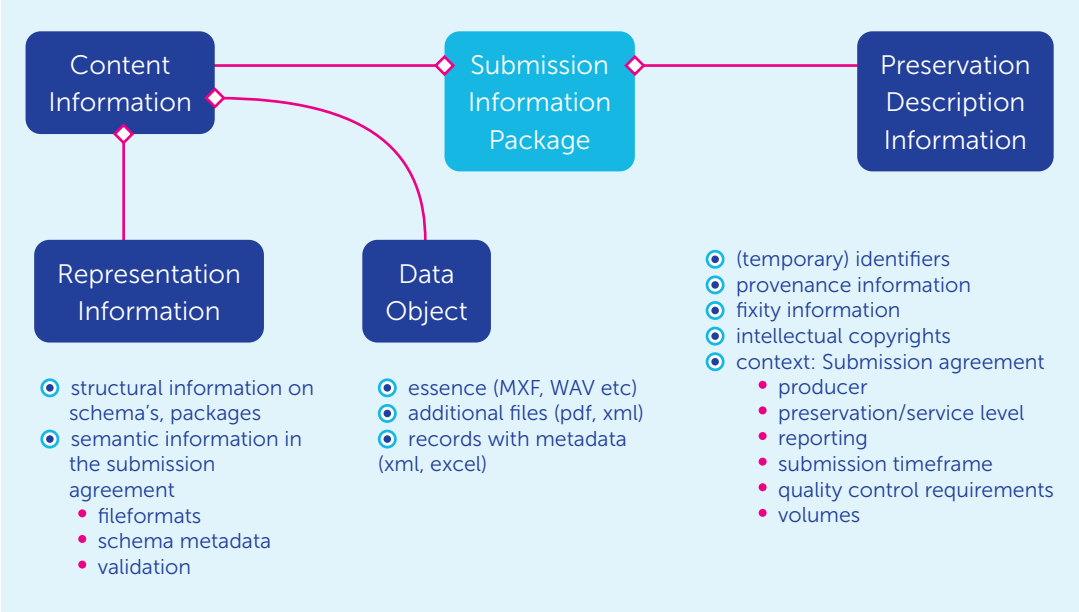


Fig. 8: Submission Information Package (SIP)

additional explanations for selection decisions, is documented within this ticket.

If the offer is accepted, a contract is concluded. The offer is registered in the depot-system. Before ingest of the digital (or digitized) material, the metadata that is available, will be assessed and mapped tot the MAM-schema on best effort-basis. Before ingest, programmes are appraised either with an A (rare, unique productions) or a B. Mediamanagers will add the curationlevel and add some categories or genres if applicable.

At ingest, standard checks are performed on values and presence of metadata, also resulting in tasks for reference. Once metadata and files are ingested, the contract is uploaded and connected to the relevant programs.

During the 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. For tenants, in the absence of a submitted checksum, it is specified in the SLA, that the Digital Archive will accept no liability if material is corrupted during the transfer as part of the ingestion process. In any case, the Digital Archive itself will generate a checksum, after the material has

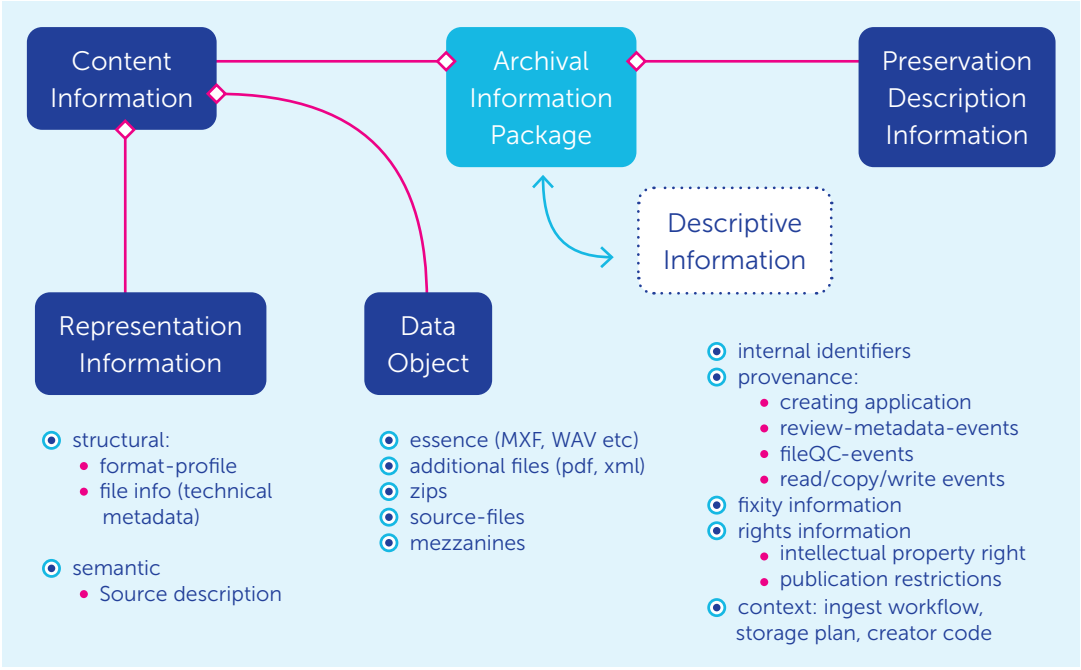


Fig. 9: Archival Information Package (AIP)

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 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, differences in technical aspects of a file format can be taken into account when planning migration actions.

An integral part for all video and audio files is the automated Quality Assurance process, whereby the master files are checked for quality with regard to the relevant profile. During this process technical metadata are also extracted, and the outcome of the 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 internal identifier, that forms a permanent reference to the object stored in the Digital Archive. This stage represents the end of the ingestion

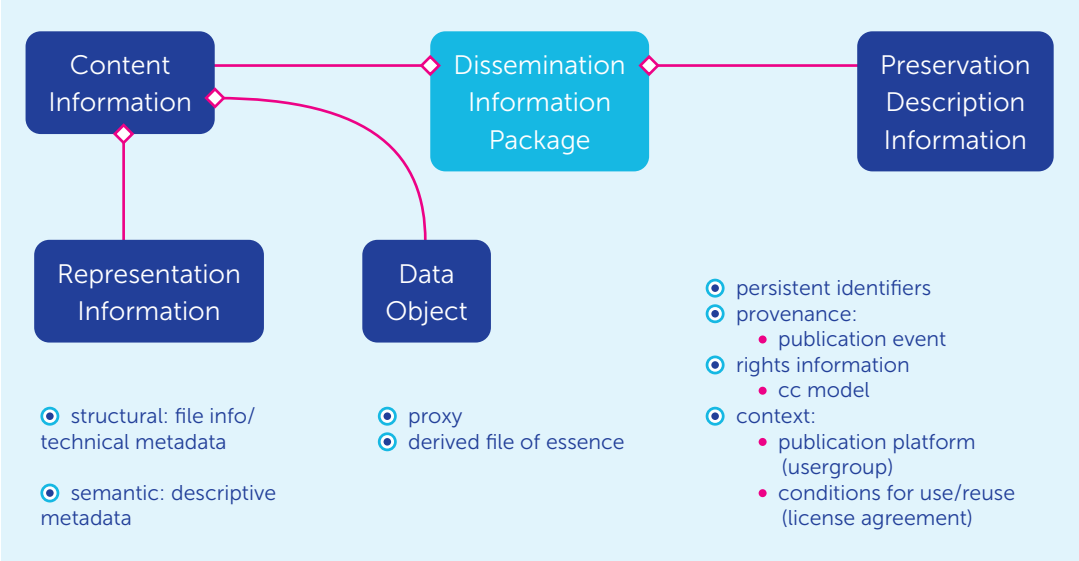


Fig. 10: Dissemination Information Package (DIP)

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, 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 & Vision's information model remains limited. The AIP need only be allocated a definitive storage location based on a set of business rules and the storage plan for those files that are stored on tape. In a parallel process, the metadata that has been mapped to the metadata schema is stored in the MAM system. All metadata management takes place there. See fig. 9 for an overview of the AIP.

In some situations it happens a better quality file becomes available, that can replace the current file. Or a file may need replacement because of a failure. In such cases the technical annotation states that the file was re-ingested. When a re-ingest is needed from a structural influx the program (and files) will be deleted first. To delete a program or to replace a file a ticket needs to be made in the servicetool; a procedure follows that ensures the removal is executed by the authorized colleague.

The last element of the workflow describes the actions relating to the Dissemination Information Package or DIP (see fig. 10). The DIP workflow may start with user authentication; depending on the service or portal being used, a determination is made of the authorizations of the user registering with the system. Access to material (the DIP) is managed in a step-by-step process. Based on rights step 1 is a prerequisite for step 2, and so on.

1. Publication (findability) of program information.

2. Releasing the playback of the proxy for viewing or listening to the content.

2+. Offering a link to a streaming version that can be embedded in another public website, possibly including restrictions on geoblocking.

3. Offering a download of the proxy or an ordering service for various derived versions of the source file, including securing consent from the rights holder.

In a situation in which the source file is requested, a fixity check is carried out to guarantee successful delivery. In case of a partial restore (part of the file) is requested, only this specific part of the file is read and copied to a new file that will be delivered. 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 (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 & 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 Media Asset Management system supports 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 each file. The built-in extraction tool enables more technical data about the files to be generated. These functions are prerequisites for file ingest: variants that are not supported by the MAM system will not be accepted. A pre-ingest workflow has been set up to allow these files to be archived nonetheless (fig. 11). To this end, checks are performed beforehand, and in case of failure, the file will be transcoded to an accepted file format. This transcoded file is then ingested as the master. Both the source file and all metadata from the QC (Quality Control) are added to the AIP (Archival Information Package) as context information.

In the metadata of the transcode, a reference to the original file will be noted (carriernumber source). Both the transcode and the source file will be ingested and in the MAM-system have a mutual projectname.

The Sound & Vision Preservation Metadata Dictionary 2.0[23], defines an 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. The Preservation Metadata Dictionary 2.0 is based on PREMIS[24] (Preservation Metadata Implementation Strategy) the international metadata standard designed to support the preservation and long-term sustainability of digital objects.

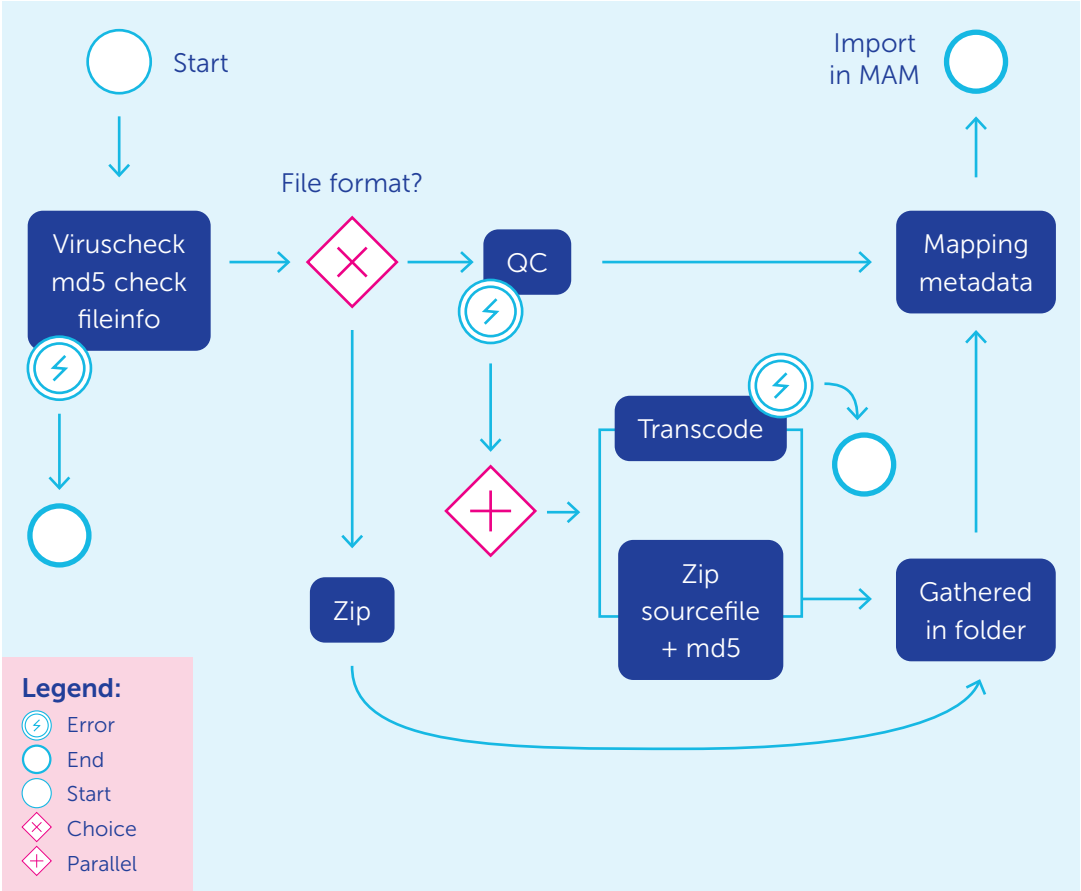


Fig. 11: Pre-ingest workflow

23 Preservation Metadata Dictionary 2.0: http://publications.beeldengeluid.nl/pub/615
24 PREMIS Data Dictionary for Preservation Metadata, Version 3.0 : https://www.loc.gov/standards/premis/v3/premis-3-0-final.pdf

Optimal preservation 7

Guarantees aligned with demand, technology, and costs

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[25]. To operationalize this, Sound & Vision offers a differentiated preservation approach, whereby the method of ingestion, storage and accessibility is optimized, based on a combination of the demands and capacities of the depositor, the technical facilities, and the costs involved.

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 and the accessibility for times to come is not monitored.

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. The material is stored in a preservation format. The accessibility of the format is monitored.

The preservation services Sound & Vision offers always have full preservation as a starting point, for the institute wants to be able to keep materials accessible for future generations. 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.

25 excluded from the general workflow: games, websites and dark archive

7.2 Preservation levels

S&V distinguishes different preservation 'levels' of delivery, storage and access. The levels are dependent on preservation actions, both concerning the files and the metadata. All the details relating to the level of preservation are set out in the contracts and the Service Level Agreements with the depositors.

For the depositor, each level comes with conditions for the submission of essence and metadata. Only if these conditions are met can specific guarantees be issued by Sound & Vision. For tenants (preservation as a service), this is formalised in the sense that 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). For structural ingests the agreement states how S&V detects errors during ingest and outlines the depositor's responsibility for resubmitting the affected content.

Furthermore, for one-time acquisitions, responsibility for the material lies with B&G at the moment of transfer. Until ingested, the files are stored separately on a server, potentially along with a provided checksum and metadata.

Within bit preservation (1) and full preservation (2) there are rising levels of authenticity and integrity. 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.

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.

Level 1a stands for the Digital Acquisition Depot. Files that have been transferred to S&V but are not ingested into the MAM-system yet. These files are kept in a folder on a disk storage. Until succesfully ingested. Actions are documented in the Acquisition module (per acquisitionnumber).

Level 1b is used for temporary storage if, for example, a depositor is not in a position to transcode the submitted material in advance (or have it transcoded) into a format that is acceptable for full preservation, but S&V is willing to accept the content in order to have it transcoded in the near future. Due to priorities content is stored at level 1b. The metadata is not ingested, but stored with the content; these materials cannot be accessed, or only to a limited extent.

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

Legend:

- – Does not happen
- O Implemented in standard procedures
- ! visual check after ingest in stead of QC control
- !! additional measures are taken >> see integrity calculation
- !!! profile for validation is not available yet

Fig. 12: Rising levels of guarantee

Level 1c refers to material that cannot be delivered in a preservation format and where migration would be relative costly. This applies for instance for files that are generated as a by-product of the television broadcasting process and are included in the collection for research purposes.

The level 2 scenarios (2a to 2c) were developed for the collections stored and made accessible by Sound and Vision for the long term. This refers to material for which the Institute accepts permanent care, either as part of its own collections or as a service for a tenant. The collections concerned can be accessed via the Sound and Vision MAM-system. They are indexed and catalogued upon ingestion (from basic metadata through to enriched descriptive metadata). Preservation metadata has been added and possible checks and quality controls have been carried out.

Level 2a (no technical quality control against a format profile) applies to pdf, jpg and tiff. This level also applies to game-packages and warc-files.
Level 2b (with technical quality control, no structural checksum control) applies to all mp4 and pm3 files that are stored on disc, not on tape.
Level 2c applies to all mp4, pro-res, dpx, mxf and wav files**!**, stored on tape.

! although: wav files from automated flows lack a technical quality control; these files are basically level 2a.

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 are provided.

Backup/Copies
Following successful ingestion, copies on tape are made of the essence based on the storage plan. Material on disks will always have a second instance based on mirror-rules.
For the influx of material on disks over a period of time, a backup will be made on tape as well.

Carrier replacement
For tapes a decision is taken periodically on whether carriers (LTO-versions) need to be replaced. In the event of such replacement, files, are checked for integrity during copying. As the influx of files on disk is being stored on tape, these are also checked during these carrier replacement actions.

Integrity check at ingest
If a checksum is delivered by the depositor[26], this is used at ingest to determine whether the file has been received complete and correct. If not, a checksum is created prior to ingest for all one-offs (during pre-ingest) and all digitisation flows.

	1a	1b	1c	2a	2b	2c
• Copies/Backup • Carrier replacement	–	○	○	○	○	○
• Integrity check at ingest	–	□	□	□	□	□
• Format migration • Format information extraction	–	–	–	○	○	○
• Format and Quality check	–	–	–	–	○	○
• Integrity calculation	–	–	–	–	–	○
• End-to-end check/ completeness • Metadata content check • Metadata technical check	□	□	□	□	□	□

– Does not happen ○ Implemented in standard procedures
□ Depending on collection and agreements with depositor

Fig. 13: Preservation actions per level

Format migration
If the delivered format is not compliant with the S&V preservation formats, it will be migrated to a preservation format. The original source-file will be kept as well, serving as contextinformation. The preservation format will be monitored for long-term accessibility.

Format information extraction
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, Mp4, and ProRes this is done by Baton QC. Discrepancies will generate an error regarding technical

26 For instance for the daily ingest radio flow

properties and warnings of quality issues in the log files.
The formats are checked against a specific profile. The technical properties of the file must be within specs.

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.

End-to-end check/completeness 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. The implementation of this check may defer per collection (daily influx vs one off donations).

Metadata content check
A check to determine whether the metadata file (XML) complies with agreed metadata rules for that specific ingestion flow (applies for ddv and radio flows and some tenants).

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. (applies also for all ingest via GMI)

Preservation planning and control

8

Riskmanagement on enterprise, digital archive and fileformat-level

8.1 General risk management at S&V

Enterprise-level risk management at S&V is the overarching approach that focuses on all risks that affect the organization's value. It focuses on the organization's objectives:

- Strategic: What events (risks/ opportunities) could occur that would require S&V to adjust its strategic goals? For example: political crisis
- Operational: What risks affect the efficiency of daily processes? For example: system failure.
- Reporting: What risks impact the reliability and integrity of the financial and KPI reporting? For example: fraud.
- Compliance: What risks does the organization face by not complying with legal and contractual obligations? For example: publication without the consent of the rights holder.

The strategic objectives are laid down in the multi-year plan along with an impression of the organization's risk appetite per objective.. In the strategic risk management cycle, the most important strategic risks, impact, ownership and the costs/ benefits (multi-year budget) of control measures are discussed by the management team, and weighed based on the organization's risk appetite. This process is facilitated by the CFO. The CEO then discusses it with the Supervisory Board. A number of scenarios are then elaborated. If necessary, control measures will be implemented. All risks are expressed in terms of impact (financially or in terms of reputation loss) and probability and listed in the annual board report.

As input for this annual risk cycle, departmentmanagers evaluate their operational risks when preparing their annual budget. Based on urgency, topics requiring attention may be added to the strategic list. An external risk analysis is updated every year as well. All risks are analyzed in detail and outcomes are discussed by the management team before a digest is presented to the board.

Compliance risks are yearly assessed by the managementteam, based on an inventory of all laws and regulations applicable to S&V. Reporting risks are managed by the financial contolmeasures and the yearly statutory audit.

8.2 Risk management in the Digital Archive

Since 2021 Sound & Vision is certified to ISO 27001, for its Digital Archive. This standard outlines the requirements for an Information Security Management System (ISMS), which is a systematic approach to managing sensitive company information so that it remains secure. The Risk Management Cycle (fig. 14) shows the ongoing activities to continuously improve the organizational resilience.

Risk management plays a central role in ISO 27001. The standard requires to identify all risks associated with the availability, integrity, and confidentiality of information and implement controls (measures) to lower the likelihood or the impact of the risk. Sound & Vision monitors all measures by regular internal audits and reviews the effectiveness of the overall risk management process each year during a management review. In order to give direction to the risk assessment process S&V identifies all risks (internal and external) that are relevant for its information security. By understanding these risks S&V is able to target the most significant threats and vulnerabilities. These factors can be identified on five risk areas associated with the continuity of the Digital Archive.

A controlling and initiating role is vested in a separate security committee, that consists of officers from various disciplines, including IT, facilities, HR, privacy, and preservation. This committee holds an independent position with respect to the organization and receives direct information on results related to information security for specific areas. In addition, this committee organizes internal audits to systematically check the effectiveness of all control measures.

Fig. 14: Risk Management Cycle

Organization and finance

For the Digital Archive, all changes to the organization are potentially relevant in terms of threats or vulnerabilities. For example, general austerity measures or a substantial change to the collection profile. Or 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. Furthermore, the potential loss of ISO-certification could have reputational consequences. Issues of this type are identified during the strategic risk management cycle, driven by the preservation watch: technical monitoring and monitoring the designated community.

Liability

Complying with Copyright Law and the General Data Protection Regulation (GDPR) has been one of Sound & Vision's high concerns, given the interests involved. For contract management, copyright policy and data confidentiality issues S&V 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 and other contracts entered into by S&V. Relevant information is provided on a regular basis in order to sensitize all employees to copyright matters and an understanding of and compliance with contractual agreements and data confidentiality. S&V has fixed protocols in place regarding privacy protection, the anonymization of data and otherwise placing restrictions on online publication. The organization performs standardized procedures regarding requests for the blocking of broadcast material and the settlements thereof. These measures are referred to in the security policy under data confidentiality.

ICT

Sound & Vision's ICT systems have high availability and contain high-quality content and information, for various usergroups. Therefore they involve a range of potential issues. Disruptions and outages can impact operational continuity, while the threats of unauthorized access, malicious intrusion through hacking, and crippling ransomware attacks pose dangers to data integrity and system availability. Finally, issues affecting system performance and overall reliability can undermine the effectiveness and trustworthiness of Sound & Vision's digital services. These issues are managed by regularly performing an external security audit (penn test) and by all other measures (physical, autorisation, access, data integrity) in the security policy of S&V.

Information management

Information for Sound & Vision is a core asset, much like its technical infrastructure and human resources and it must be properly valued, protected, and leveraged to achieve S&V's goals. Strategic decisions rely heavily on accurate and timely information. Poor information management can also lead to increased other risks (ICT, liability). The preservation board (*voor altijd en eeuwig*) is responsible for establishing a Digital Information Policy and supervises its implementation. Its scope is Sound & Vision's information landscape in the context of the Digital

Archive, i.e. all the digital systems, applications, processes, data and metadata that are relevant to the OAIS functions Ingest, Storage, Datamanagement, Access and Preservationplanning. The Board's focus points are:

a) Matching policies, projects, activities and plans in the field of information management with the strategic objectives of Digital Preservation;

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.

Recent topics include policies regarding legacy terminology that may be perceived as offensive, as well as the application of generative AI within the archive. The primary objective may be to enhance completeness, findability or inclusivity. However, additional measures are required to ensure the authenticity remains uncompromised.

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 (Appendix II), 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 training at individual or group level are allocated by the unit managers in collaboration with the HR department, on an ad hoc basis. S&V strategically engages external experts and temporary staff to acquire specialized skills and specific expertise for short-term projects. 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

Issues in each of the risk areas may have an impact on the effectiveness or continuïty of the digital archive. More specifically, certain developments can impact the effectiveness of preservation activities and can be a reason to implement changes in the preservation policy or choices in its execution. Detection of relevant changes and the conversion to

an adequate and timely response are crucial. In general two types of developments are distinguished.

8.3.1 Designated Communities

Understanding the needs and preferred access methods of the Designated Community is crucial to the archive in order to accomplish its mission of accessibility. The community of S&V contains different user groups. Each has its own interest in accessing the content, ranging from playback and (re-)experience to research and reuse. Preservation choices (e.g. on format, storageplan, metadata schema's) must maximally support this variety of interests with the right balance where they might conflict. Shifts in expectations between user groups will possibly require new strategies or lead to policy adjustments. These user-interests (and shifts) are managed collaboratively by the product managers.

A recent example of impact on infrastructure driven by a shift in expectations from Designated Communities is the National Digital Heritage Strategy[27] as developed by NDE. This strategy aims to maximize the findability of digital heritage on the Web, to increase its impact, interoperability, and use for all conceivable user groups. This has led to the development of the Digital Heritage Reference Architecture (DERA)[28]. Sound & Vision has committed to this architecture, which has a widespread impact on all its services. To manage this transition, the organization has conducted a gap analysis to identify the differences between the infrastructure for the new public portal and the DERA standard. The identified gaps will now be addressed and closed in a phased, step-by-step approach.

Of at least equal importance are the developments occurring with material producers. S&V aims for any submitted material to ingest it without information loss. New emerging production formats, adopted by important creators may directly have an impact on the policy of preservation formats. To monitor this kind of changes, the Victorie and Vereeuwigen departments maintain contacts with key producers (public broadcasters and tenants). Additionally, product managers relay important signals from their user groups as far as these groups also have a role as producer.

A topical example of emerging fileformats is the expected broadcasting of ulta-high-definition-television (Ultra HDTV, also known as 4k-resolution due to its nearly 4000 pixels in width), that contains four times the number of pixels in Full HD. These files are approximately six times larger than the current MXF files. Details (bitrate, collorscheme)

27 https://zenodo.org/records/14760720
28 https://zenodo.org/records/5562062

about the format are not yet clear; S&V is keen to play a coordinating role between different broadcasting organizations (public and commercial) to arrive at a standard and to be able to anticipate it in a timely manner. The expectation is that this format will then be added to the existing preservation formats.

Finally, when it comes to facilitating access to the material S&V has developed different outlets for different usage and needs. Using both formal and informal methods, the departments that serve the user groups verify whether the previously recorded needs of those user groups and/or producers are still up to date.

- For the broadcast media professionals Designated Community this is done via regular contact on the service contracts containing provisions on formats and metadata. In addition, performance against metrics agreed upon in the Service Level Agreement (SLA) is reported on a quarterly basis, such as how quickly programs must be available after broadcast, the permitted download time for an order, and site downtime. Another way to monitor users' needs are the regular training and information meetings[29] for representatives of these user groups on subjects such as system training, news on collections or materials, licensing constructions and general conditions for re-use. On top of that, every one or two years, a user survey is conducted to assess the appreciation of the portal and user support among all users of the media professional portal.

- For film and documentary makers sessions and meetings about archival use are organized at festivals and professional events. Contacts are maintained to provide them with input on research and ethics. And programs such as internships and trainee positions for emerging makers or researchers are offered.

- S&V actively seeks contact with trendsetters in the online domain of students and pupils: teachers who can explain exactly what they need from the archive to convey their stories. Together with them, S&V creates educational packages with archival content that are widely shared, thanks in part to the online visibility of the co-makers.

- For the Designated Community 'general public' the users behaviour on the portals is monitored by Sound & Vision's Verkennen staff. Four personas have been developed that represent the diverse users of our collection, based on this research and data. These personas are systematically used as reference for new developments on search performance and UX-design.

- For the Designated Community 'scientific research' a specific product Media Suite is developed. The creation of the Media Suite follows a co-development approach: the interface is designed in continuous interaction with users. From the start, scholars have been actively involved in formulating the requirements, prototyping and testing the interface, as well as in the evaluation of the Media Suite's features.

- Through AVA_Net, the national knowledge network for audiovisual heritage professionals, S&V organizes meetings with the audiovisual heritage field, such as the annual symposium, knowledge sessions, and webinars. On behalf of AVA_net, Sound & Vision organizes the Trendmonitor[30]; a biennial study that provides insight into the state and challenges of the field. Additionally, as the national AV hub in NDE, Sound & Vision is involved in the updates to the National Digital Heritage strategy, a process that is preceded by an inventory of the heritage sector.

8.3.2 Technological developments

Innovations in storage media and playout software also have an independent influence on the ability to issue guarantees for adequate ingestion, storage and sustainable access. The ICT strategy is informed 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.

By subscribing to industry news and updates from the current or potential vendors, and attending international industry trade shows, S&V stays well-informed about technological developments that could impact the technical environment and preservation facilities of the digital archive. Topics that need to be monitored are:

- Hardware obsolescence: tracking the lifespan of storage media (LTO-tapes, hard drives, optical discs), servers, and other infrastructure components.

- Software obsolescence: monitoring file formats (e.g., new versions, deprecation of old ones, emerging standards), operating systems, application software (word processors, image editors,

29 https://www.beeldengeluid.nl/collectie/voor-makers-en-professionals/daan/trainingen-voor-daan
30 https://kennisbank.avanet.nl/trendmonitor-audiovisuele-collecties-in-nederland-2024/

scientific analysis tools), and programming languages.

- Emerging technologies: identifying new preservation tools, techniques (e.g., validation, QC, transcoding), and storage solutions that could improve efficiency or effectiveness.

- Network and security technologies: staying informed about cyber threats, new security protocols, and network infrastructure changes that could impact access or integrity.

Insights into, and awareness of, the relevance and pace of innovations are further enhanced 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.

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. Technology alerts and requirements from users/producers are gathered and discussed in terms of impact on preservation strategies and operational implementation. See fig. 15 for activities of Preservation Planning according to OAIS.

When it comes to new file formats and legacy issues, this function is carried out by a designated working group of specialists from the three core departments: Vereeuwigen, Victorie and Verkennen, in the capacity of digitisation and ingest experts, system engineering and preservation officer. Together they define new packaging designs, given the applicable policies. And they monitor the implementation as well.

Likewise, another working group is concerned with metadata management. Issues that arise from the department of Vereeuwigen when handling metadata, for example on new acquisitions or ingest flows, are presented to this group of specialists. They have roots in datamanagement and workflowmanagement, and they give advise on handling metadata, given the schema's and businessrules, how they have been implemented.

All issues that have a potential impact on the digital archive or proposals that imply an adjustment to the existing policies, are submitted to the preservation board, by these working groups. The board is represented by three departments (Victorie, Vereeuwigen, and Verkennen), as well as a delegation from the file format and/or metadata expertise groups. Once in agreement, new policy is approved by top-management.

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. Sound & 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. Migration plans are submitted to the expert working group and/or information specialists; they are approved by the Preservation Board.

A good example of a planned migration is the repair flow for MXF files. This was introduced because

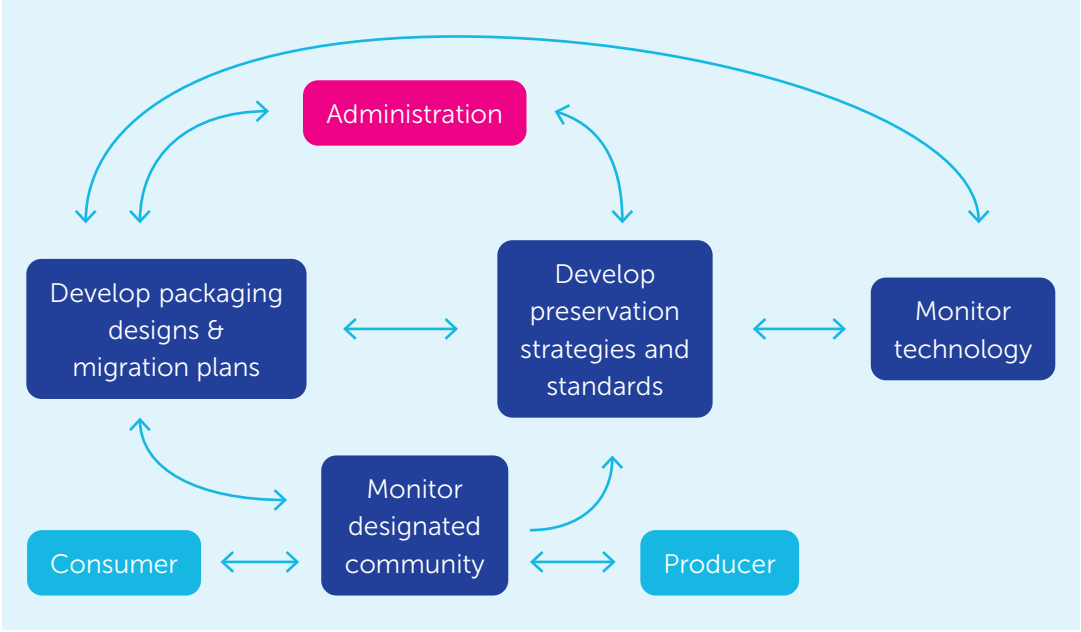


Fig. 15: Activities of Preservation Planning

a range of flaws and exceptions had surfaced as a result of a newly implemented QC procedure for all archived MXF files. These were analysed and categorized. Many could be fixed automatically, but some needed a specific transcode or even customized repair. Once this analysis had been completed, a migration plan has been drafted to plan the resources and timeframe needed, and to define the various categories and preferred solutions. After the project was finished, the plan and documentation were evaluated.

A second example is the project to isolate MP4 material that was ingested at an early stage, before the proper checks were implemented, and re-ingest it through the correct controls. This is done so that it has the same documentation as the current MP4 ingest. For this, too, a plan is made in advance so that the results can be checked and evaluated afterwards.

Yet another example concerns the music collection. Sound & Vision manages an extensive music collection, a selection of which has been designated as cultural heritage. A policy will be developed regarding the digital ingestion of this collection into the archive. This process will also be conducted in collaboration with the working group, with formal decision-making taking place via the preservation board.

Access, discovery, copyright and licensing

9

Differentiated for all designated communities

9.1 Platforms and services

Sound & Vision strives to provide its Designated Communities with access to all legally permissible material: media professionals, user groups in the educational, heritage and academic domain, and the general public.

A specialized Makers team supports producers, directors, and researchers in using archival material for documentaries, series, and other special media productions. In these projects S&V assists with searching for relevant footage and audio, arranges the necessary licenses, and manages requests for the digitization of analogue material. They remain involved throughout the entire production process, from the first request to the final delivery, and ensure that all aspects of archival use are well-coordinated, both practically and in terms of content.

To support individual search by users themselves, 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), and the educational platform (Sound and Vision At School)[31].

All access and re-use conditions are communicated via these portals. Depending on copyrights, licenses and authorization, the user groups may find, playout, order and/or re-use material from the collections, each via its own platform. The institute has appropriate terms of use in place for each portal in order to use the portals, users will have to agree to the applicable End User License Agreements (EULAs) that specify their rights and obligations.

Online collections attract more and especially other users, with new expectations and skills. The traditional model of end-users ending up in a central portal, shifts to a more-dimensional model where you also need to bring collections to users.

31 https://beeldengeluidopschool.nl/#/home

9.1.1 User-driven access

Users are already somewhere 'on the web', ready to be served: content and desired user facilities are integrated onsite. S&V has several initiatives, each aiming to empower the user. In a practical way; in social media applications, on various mobile platforms, in multimedia research environments, in educational tools and via programmable interfaces and APIs.

Researchers that want to conduct innovative and data-intensive research across collections from various institutions, make use of the Media Suite. Sound & Vision is one of the partners in the CLARIAH (Common Lab Research Infrastructure for the Arts and Humanities) consortium[32]. 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. One of the workpackages has resulted in the Media Suite for AV-material, whose services are being continuously improved. The catalogue and the collections stored in the Digital Archive are accessible via this infrastructure. The connection is made via middleware that in future possibly will be available for other outlets as well, for instance for the new public portal that is currently being developed.

Also, the Sound & Vision Data Platform[33] offers access to several datasets for reuse. A dataset, in this context, is defined as a collection of data and/or metadata, logically grouped on source or content and collectively described and published, insofar as copyright restrictions allow. The datasets are available to a wide range of users, including developers, researchers, and enthusiasts, with the goal of improving discoverability, contextualization, and engagement with these collections. The platform provides access to this information, along with documentation and examples of reuse. The datasets are offered in various formats, including linked data, data dumps, and via APIs or Wikimedia Commons. The datasets have also been published in the national dataset register[34] of NDE.

The institute further presents selected AV material on (inter) national (educational or academic) platforms such as EUScreen[35], and Europeana[36], Oorlogsbronnen[37] and Wikimedia Commons [38], with the appropriate rights statement. For the conditions and terms relating to the actual use of the moving image materials visitors are referred back to the main portal. In addition, Sound & Vision actively pursues and concludes distribution deals with both national and international partners, including Getty Images, ANP and Videoland. The material that is present on the platforms of these distribution channels is manually transferred. The metadata is being mapped to the

specified schema's, such as EDM: Europeana Data Model[39].

The website Beeld & Geluid PeerTube[40] is an open media platform that offers online access and options for re-use under the CC licensing model. Collection material that falls within the public domain or is openly licensed can be browsed through, downloaded from this platform and remixed into new works. This platform also provides an API, making it easy to develop auxiliary tools.

In addition to these portals, separate platforms offer extensive context information to the collections. An examples of such a platforms is the S&V Collection wiki[41]. Stories about the collection are told on Tilt with regular posts on topical items.

9.1.2 S&V portals

Media Professional Portal

The Media Professionals Portal[42] (MPP) provides restricted 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. Before the connection can be set up, the organization will be added to the S&V's firewall rules and as a result receive the restricted access. The MPP is also accessible to the general public who wish to search the collection on-site at S&V at dedicated terminals. Visitors receive a login code and as they are 'within the institute's premises' can preview all content, based on a specific exception in Dutch copyright law[43].

The MPP is available for tenants, to search through their own collection. They are provided with a login and based on the organisation they get access to the material that belongs to the tenant.

General Public Portal

The General Public portal[44] (GPP) is an online portal targeted at the general public, where users can search the collections without having to log in. Users can play out only selected audio and video sequences, for which the rights holder has granted permission to do so or that are in public domain. The GPP consists of metadata from

32 https://beeldengeluid.nl/en/knowledge/projects/clariah
33 https://data.beeldengeluid.nl/about
34 https://datasetregister.netwerkdigitaalerfgoed.nl/catalog.php?lang=nl&uri=http://data.beeldengeluid.nl/id/datacatalog/0001
35 http://www.euscreen.eu
36 https://www.europeana.eu/portal/nl
37 https://www.oorlogsbronnen.nl
38 https://commons.wikimedia.org/
39 https://pro.europeana.eu/page/edm-documentation
40 https://peertube.beeldengeluid.nl/
41 https://wiki.beeldengeluid.nl/index.php/Hoofdpagina
42 https://www.beeldengeluid.nl/collectie/collectie-voor-makers-en-professionals
43 Artikel 15h Auteurswet
44 https://zoeken.beeldengeluid.nl/

AV assets in the Digital Archive that have been released for this Designated Community; by default these settings are set per tennant, per ingestflow. Work is currently underway on a new public portal ("de Schatkamer"), which will allow more content to be played, based on an agreement that utilizes the possibilities of European out-of-commerce regulations, in addition to renewed agreements with broadcasters and third party rights holders regarding licensing possibilities.

Educational Media Platform

A dedicated Educational Media Platform (Sound & Vision At School)[45] offers 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. S&V clusters and thematises the material on offer. Metadata for the entire collection is searchable, and a large selection (approximately 100,000 assets) can be viewed, retrieved and used, by logged-in users. Authentication is handled through the Dutch educational organizations SURFnet (the IT cooperative of education and research) and Kennisnet (ICT support organization for schools).

Material from the collections can be shared via this platform and embedded in external educational environments (educational catalogues and applications). The metadata can be included in a local catalogue, using the OAI-PMH protocol.

9.2 Search and discovery

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 search ability for the Designated Communities to be analysed and enhanced.

9.2.1 Metadata

The Digital Archive's catalogue in the MAM-system is the starting point for all collection access. The underlying metadata model has similarities with the IFLA's (International Federation of Library Associations) FRBR[46] (Functional Requirements for Bibliographic Records) model, and allows for the layered identification of content. It links different levels of digital productions (program, series, season, item, time-coded metadata) to catalogue descriptions and information on the analogue and/or digital carrier.

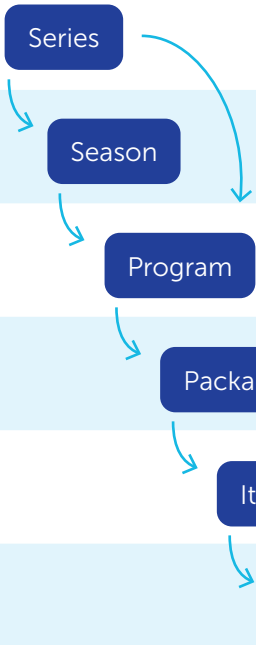
MAM Entity	Role	EBUCore Equivalent
	Editorial grouping	Collection (series)
	Editorial grouping	Collection (Season)
	Editorial object	EditorialObject (Program/Episode)
	Playout grouping	EditorialObject (Compilation Package)
	Holds media	Instantiation + Resource
	Timecoded metadata	Segment/Annotation

Fig. 16: Metadata model with reference to EBUcore

In this model the program corresponds to the expression in the FRBR-standard. The work is not modelled as an entity; programs belonging to the same work can be grouped via a thesaurus attribute. The item relates to the digital manifestation of the program. For an analogue carrier a so-called placeholder is created at item-level. The FRBR items are managed in the file management system; not in the catalogue.

The figure shows a reference to the EBUCore equivalent. EBUCore[47] is a set of metadata that has been identified as being the minimum information needed to describe radio and television content. The core set of metadata presented in EBUCore is compatible with Dublin Core[48]; the generic standard of fifteen basic elements that describe the essence of a resource. The EBUCore is recommended when describing and providing access to audiovisual content.

45 https://beeldengeluidopschool.nl/#/home
46 IFLA FRBR model: https://www.ifla.org/publications/functional-requirements-for-bibliographic-records
47 https://tech.ebu.ch/docs/tech/tech3293.pdf
48 https://www.dublincore.org/resources/metadata-basics

Descriptive and administrative metadata from the public broadcast production environment are pulled from the source systems and mapped directly into the Sound and Vision data model. The metadata associated with other structural influxes is also imported via standardized mapping workflows. Automated checks and monitoring mechanisms, guarantee the quality and completeness of the metadata during every ingest including rejection criteria for metadata, as well as a workflow for repairing, logging and documenting errors.
Next to these custom ingest flows a General Metadata Importer is used. Some new flows of structural influx are supported by this tool. Metadata that comes with digitised content or one-off acquisitions is also handled this way.

The metadata schema of the MAM system serves as the basis for publishing metadata on S&V's own search portals. It is used in a slightly simplified form to allow for the efficient configuration of the search engines. Additionally, a subset of the internal metadata model is typically mapped to other schemas, to be published on other platforms. This includes a mapping to Schema.org[49], a standard that has elaborated on the structural metadata for a creative work such as a TV-series. With this mapping in place, the metadata can be easily interpreted and programs become findable for others who use this vocabulary. The connection with additional metadata also opens up opportunities to improve search results for our users and to offer additional context.

The level of detail of descriptive metadata varies over time. As a rule, more extensive manual descriptions were provided in the past when selection was limited. With the automation of ingest, there was greater reliance on the (formal) source data. With the arrival of annotation techniques, the metadata for certain parts of the ingested material has become richer once again. Tagging techniques such as speaker labelling and facial recognition automatically add timecoded persons thesaurus labels to the descriptive metadata in the catalogue of a selection of daily ingested radio and tv programs based on genre-criteria. In selected cases the ingested metadata are enriched and contextualized manually. These metadatafractures[50] receive special attention from S&V, to manage expectations regarding findability and curation.

9.2.2 Thesaurus and identifiers

The common thesaurus for audio-visual archives, the GTAA[51] (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 S&V to enhance the automatic indexing applications, thus ensuring the quality and consistency of descriptions throughout the catalogue. The GTAA is available as linked open data to other audiovisual archives and heritage institutions via the NDE Network of Terms[52]. This is a search engine for terms based on shared thesauri, classification systems, and reference lists. This engine makes it possible for heritage institutions across the Netherlands to use the GTAA to describe their collections. It is created in collaboration within the Digital Heritage Network, and is managed and maintained by the Cultural Heritage Agency of the Netherlands.

The Digital Archive provides internal identifiers, which are maintained to be internally unique and persistent.

Persistent Identifiers, are a formal defined attribute category in S&V's Preservation Metadata Dictionary 2.0. S&V has selected the URN:NBN system for these PIDs that guarantees that handles are globally unique and can be resolved on the Internet[53]. The PID is automatically assigned to archive objects (program level) for which the metadata (not necessarily the stream) is shared on the public portal (GPP). This way each PID can refer to/be resolved to a single location, accessible to everyone. If a program has to be removed from the GPP, the PID will redirect to a standard explanation stating that the program has been depublished for specific reasons.

9.3 Copyright and licensing

S&V ensures, to the extent possible, that data are preserved, curated, accessed and used in compliance with legislation, contractual agreements and ethical norms. Specialized staff has been organised in a separate Legal department. They are appointed to focus on collection access policies (including privacy and ethical matters) and Intellectual Property Right (IPR) issues and to draft or review all copyright-related contracts entered into by S&V.

In the event that material (metadata or files) has been published on a portal incorrectly or if S&V has otherwise acted inconsistently with the rights, this is considered a data leak. S&V has a security incident procedure in place which conforms to ISO 27003

49 https://beeldengeluid.github.io/beng-lod-ontospy
50 https://data.beeldengeluid.nl/showcases/data-fractures
51 Common thesaurus AV-archives: http://gtaa.beeldengeluid.nl/
52 https://termennetwerk.netwerkdigitaalerfgoed.nl/en
53 example: URN:NBN:NL:IN:20-PGMTV2002009481 will be resolved via https://persistent-identifier.nl/ to the GPP landingspage

standards, to handle these situations. In order to sensitize the employees to copyright matters and instill an understanding of, and compliance with copyright law, contractual agreements, and ethical and privacy sensitivities, 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 portals. S&V does not actively monitor compliance by third parties and 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 S&V. The rights holder will be informed and will have the final say in how to respond to the infringement.

9.3.1 Copyright and the need for clearance

There are four types of current contracts with data producers regarding copyrights and the need for clearance:

1. The Archive Agreements with NPO and CMO's,
2. the service agreement with NPO and the collective public broadcasters, which is updated and renewed annually,
3. service agreements with tenants (commercial archiving services),
4. and donation agreements.

Preservation actions by S&V are permitted by the agreements for all materials. However, publishing content and disseminating content for reuse is only regulated (if at all) under specific circumstances, for restricted usergroups. Outside of those, these actions will need to be agreed upon with the individual data producers.

Data producers can choose to transfer their copyrights to S&V at the moment the material is transferred. In the case of donations this happens occasionally; in service agreements generally copyrights are not transferred. More often data producers broaden the permission for Sound & Vision by allowing play-out and/or re-use under circumstances. For this, they issue licenses to S&V ranging from very specific to generous release. These license clauses are documented in the metadata of the individual program(s).

In some cases data producers do not own all copyrights (or neighbouring rights) to (archive)material. If this is the case, dataproducers are asked to inform S&V of any third party rights owners they are aware of, so this information can be added to the metadata and taken into account when clearing the content. In case of transfer of rights, S&V is liable for any infringements of third-party rights that the data producer was unaware of at the time of copyright transfer.

As S&V came about as a result of several mergers of different government and other types of archives, there are a great number of legacy contracts between S&V's predecessors and data producers of those times. Specialized staff from the S&V Legal department, spend time researching these contracts in order to establish whether copyrights were transferred or licenses granted to S&V's predecessors and whether there might be any third party rights that need to be taken into account. Sometimes to solve a current issue, sometimes focused on a specific collection.

Any outcomes of this research are added to the IPR information in the catalogue. Updates of information on makers, contracts and copyrights are logged. Copyright information is made accessible in the IPR table within the catalogued metadata for each individual asset (at collection, programme, sequence and item level) and, if available, contracts containing rights metadata are uploaded to the applicable catalogue entry in the system.

Most of S&V's audio-visual collections are thus still protected by copyrighted, and these rights must be cleared before publishing or delivering the materials for reuse. Sequences and whole programmes can only be published, viewed or re-used according to what is specified in the licensing clauses of the agreement with the data producer.

A final point regarding the publication of metadata: If content publication is allowed, then its metadata can implicitly be published as well, as it is essential to be able to search for and find specific content. However, the publication of metadata becomes a subject for discussion as its relationship with content discovery weakens. For this reason, Sound & Vision has made agreements (with the NPO) on which metadata may be published as open data.

9.3.2 Licensing and blocking rules

The MPP portal is equipped with an integrated licensing module which can be used by the media professionals or persons of the S&V customer service team for the clearing on request, for re-use. The licensing module stores all contracts (License agreements) as records in the MAM-system.

Public broadcasters, the largest group of data producers/rights holders and data consumers of the Digital Archive, provide licenses to each other for the re-use of all programme materials they own the rights to, via this module. In general the requests are checked whether the material belongs to the broadcaster itself; in that case no license is needed. Also, for some programmes (that have high current value) broadcasters have agreed to share material without a license (only financial settlement afterwards). All the same, a broadcaster may indicate in advance for a certain program that

they always want to see and review any request for this material; in that case a license is always needed. Sound & Vision functions as a gateway, where all transactions are registered. By providing the appropriate intermediate services, S&V aims to both enable the use of materials and secure the proper financial benefits for rights holders, as well as recording the conditions as agreed upon.

When a third party (non public broadcaster) is involved as requestor or rights holder, S&V will function as intermediary in the licensing process. S&V makes sure both parties agree on costs and other conditions in the license agreement. Unless the rightsholder has appointed S&V as their representative for the licensing of their content, S&V will contact the rightsholder. The license is signed by the requestor and by the rightsholder (or by S&V as the data producer's representative), via the S&V licensing module and with the help of a S&V customer service.

Prohibition rules

S&V applies a protocol regarding requests for the blocking of broadcast materials and the settlements thereof. General rules are included in the agreements with the public broadcasters. Rights holders of broadcast materials can request restriction of re-use via a special form on the Media Professionals Platform. Other data producers will highlight any content that should be restricted in the donation agreement with S&V leading to the proper restriction being implemented upon entering the catalogue. These restrictions can mean that a license check is enforced (see above) or that material is blocked. Blocked material cannot be ordered or re-used in the MPP/GPP. In special circumstances also the browse preview-file can be removed from the MPP. Or, even going further, the program can be completely hidden from MPP/GPP-search.

If requested (because of e.g. portrait rights/privacy law) it is also possible to anonymize personal data. The Privacy Officer processes these requests. A workflow is established to record the outcome of the request and share this with metadata management, so they can anonymize data when needed. The essence remains unaltered. In case of doubt, the DPO (Data Protection Officer) may be approached. If the request also involves blocking content (not only anonymize metadata), S&V will contact the rights holder to see if they are willing to request for a block.

For its *legacy collections* (copyrights with S&V or unknown) that might or might not be in the public domain, S&V has created a separate registration with regard to sensitive content, based on privacy and/or ethical grounds. As above, blocked material cannot be ordered or re-used in the MPP/GPP. And (as above) in special circumstances also the viewing option

can be removed from the MPP, or the program can be completely hidden from MPP/GPP-search.

Any material designated as "license required" will be checked for suitability for viewing within the catalogue and any request for reuse of this material will be checked for context by customer service and/or the Legal department. This is particular relevant for material in public domain, and material that is requested by a S&V-account, being "own" material.

9.3.3 Terms of use

The public has access to metadata of the collections via the online catalogue on the General Public Portal (GPP), where they are permitted to play footage and sound that is in the public domain, or that have been explicitly released by the rights holders. For the Media Professional Portal (MPP), any individual media professional may request a personal account with specific rights. Based on these rights, the user can browse, preview, request/grant a license and download material. The specific terms of use (EULA) aimed at designated users[54] (MPP) or the general public[55] (GPP) can be found on both portals.

In case a user has any questions, contact details for the S&V customer service department are available on the portals as well. If needed, S&V customer service will consult the S&V legal department before answering a question.

The website of Sound & Vision has a section containing general information on the conditions and instructions for the usage of audio-visual material, including how to contact the customer service regarding more complicated copyright issues and requests[56]. Specific information on AV copyright regulation is also 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[57].

Where appropriate, Sound & Vision further makes use of the standard licensing model by Creative Commons, which allows re-use of material as long as the user complies with the conditions stated in the applicable Creative Commons license. A pop up will be shown when material on GPP/MPP is licensed under CC a specific Creative Commons license. This pop up contains a link to the definition of the applicable license on the Creative Commons website, where

54 userconditions and terms of service
55 https://zoeken.beeldengeluid.nl/licenses
56 https://www.beeldengeluid.nl/collectie
57 http://www.beeldengeluid.nl/auteursrechten

the terms and conditions of the license are explained. The user then has to agree to comply with the license and will then be able to download the material. If the user cannot comply with the terms and conditions of the Creative Commons license, the user will be referred to the licensing procedure for the material from the pop-up. For the GPP this means that the user can create an account and send a form to the customer service to apply for a license. For the MPP the user will have to obtain a license through the standard procedure that applies to all material that is available for licensing via the licensing module.

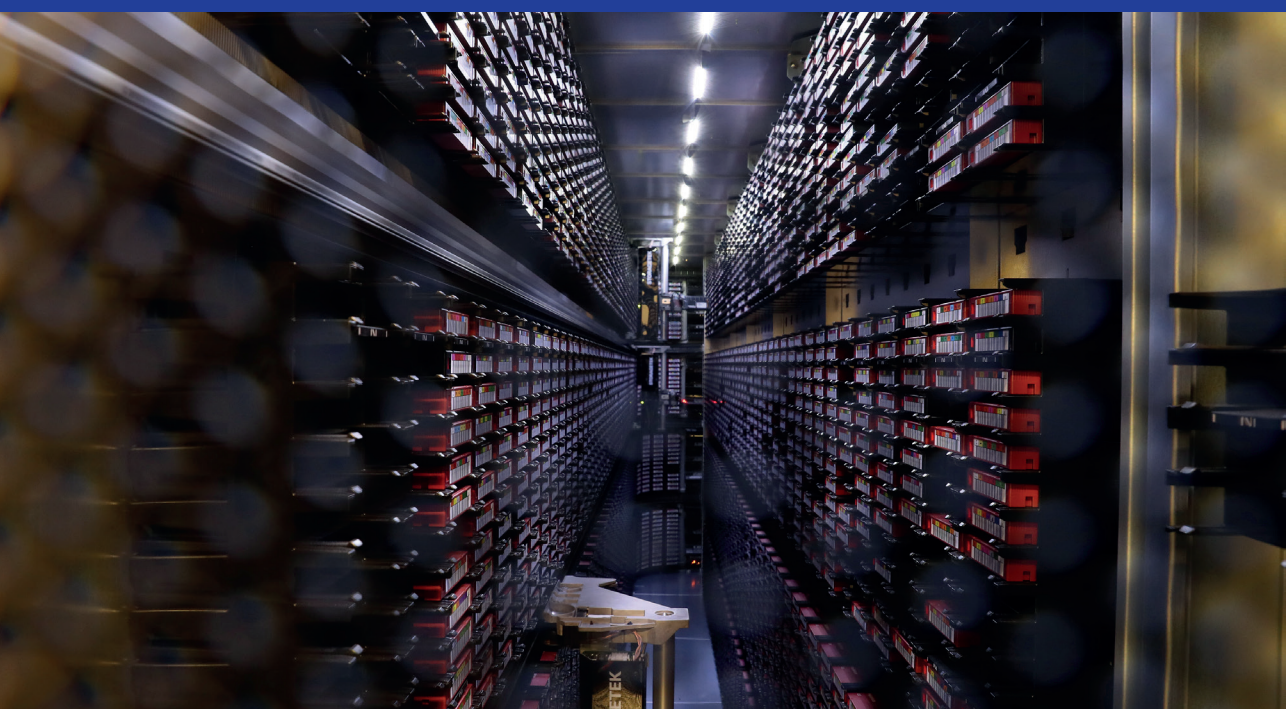
As regards Sound and Vision At School, S&V's, integrated audio-visual portal for primary, lower, middle and senior secondary schools, information on all the aspects of access and usage and the codes of conduct, can be found in the General Conditions on the site[58].

The use of material that is published on external platforms (MediaSuite, Peertube, Europeana) is governed by the tems and conditions of the portal. S&V makes sure that the permitted re-use is stated with each program (using a creative commons license).

58 https://beeldengeluidopschool.nl/#/algemene-voorwaarden

Technical infrastructure 10

Secure services at the core of the digital archive



10.1 The IT environment

At the core of S&V'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 S&V's digital materials and all supporting tasks and secondary activities. Sound & 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 terms 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. 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.

S&V 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 S&V 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 S&V 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 S&V (for example 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 & 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 S&V museum.
The Office environment, with its office applications and facility systems, supports all these processes. The S&V websites and portals are managed in the Hosting environment.

For almost all environments, SaaS is an integral part of the infrastructure. Only for the production environment does it (currently) hold true that all software and data run on-premises. The new public portal will utilize mezzanines of archive content to offer a streaming service via an Amazon Web Services (AWS), based on Content Delivery Network (CND), fully operating in the cloud. This new service will be part of the Hosting environment.

The introduction of Kubernetes as the infrastructure for our Service Layer, the possibility to scale up our on-premises resources to the cloud for peak traffic (hybrid cloud approach) has been introduced. The service layer acts as middleware between the MAM and the portal(s) to implement search indexes, metadata interfaces (OAI-PMH), and business logic. While it is not a production component of the core digital archive itself, it is essential for providing access to our target audiences. Further in the future, it is conceivable that IaaS will be used for disaster recovery, including a backup of archive material

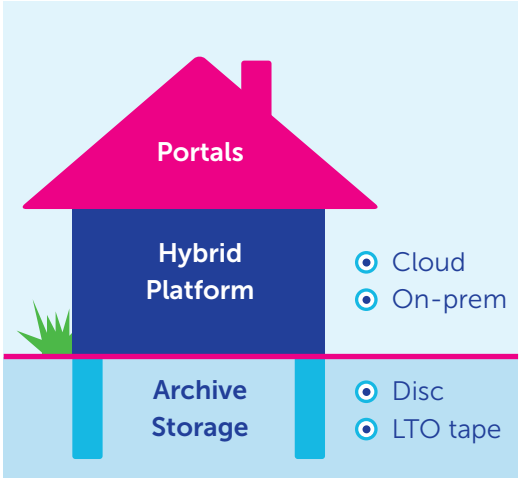


Fig. 17: Layered infrastructure

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. S&V'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. S&V provides the possibility of a VPN connection to allow internal and external users to access the internal network.

10.2 Workflows and Media Asset Management

At the heart of the Production environment is the MAM system DAAN (Digital Audio-visual 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 S&V'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 QC-software (BATON) guarantee both the quality and the completeness of the files and the metadata, during every ingestion.
For broadcast material the integrity of metadata is also detected after updates that are automatically performed from various sources as new descriptive metadata become available at set times after broadcasting. New versions of a programme (re-offered by the broadcaster after small corrections

have been made) will be detected and the files will be added to the original entry. The older versions will be blocked from the portals.

For al workflows the MAM system has rejection criteria for metadata built in (such as mandatory fields), as well as a workflow for repairing, logging and documenting errors.

In the rare case that a file needs to be repaired, the new version is regarded a new ingest with the archive date reset. A remark will be made that points to the archive date of the original.

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 facilitate that technical properties of the analogue material are shown in DAAN.

The systems are directly connected to the Digital Provision (DDV), the central infrastructure of the broadcast production environment. Separate additional import facilities for files and metadata are in place for radio, music recordings broadcasters and a few tenants. On top of that a Generic Importer is available for all other material.

To digitize material, a digitization batch must be created in the MAM system. Each batch is linked to a project. Within the project, it is defined for which user group the digitization is intended and to which tape group the material must go. This batch is picked up by a workflow manager application (using separate workflows for audio, video, and film). The workflow tracks the status per carrier through the different phases: digitization, audio, synchronization, and quality control (QC), leading to the final product. The files (wav, mxf or dpx+wav+mxf) are automatically ingested via the Generic Importer along with an XML file containing all technical data.

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.

Metadata about the provenance of 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 of ingestion is shown in the ingest task and as creation date of the item. The route of ingestion can be derived from the project for all files ingested via pre-ingest or digitization. For automated flows the tasklist indicates which flow the file went through. The QC profile is visible both in the QC task and at item level in the field Baton profile used. The archive management system DivArchive logs and stores basic provenance data with the AV-file, such as creation date.
By way of a three lettercode, data producers are identified in the

various ingestion workflows, and a unique internal id is generated by the MAM-system and passed on to the Media asset management system, establishing a link between metadata and files. The process of extracting and recording technical file characteristics will ensure that all files can be searched for and analysed on these characteristics. The output of all processes is logged and structured in the MAM-system database and as such can be used as an overview of the life cycle of the files from ingestion through to storage and access.

10.3 Storage management

Files arrive on a disk cache where the MAM-system applies basic business rules to determine what storage location is appropriate. This can be either a specific disk or the tape storage. The business rules use technical criteria (resolution) as well as file type to determine the appropriate storage location.

Digital archive masters are stored primarily on LTO tape. S&V employs DivArchive (a Telestream product) as archive management system for the tape storage. Tapes are loaded into a tape robot. The administration of master files takes place in DivArchive (DIVA), which is responsible for ingestion, integrity monitoring, copying actions, access, logging, management of tape groups (three lettercodes), 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 & 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 file name, archive date and objectname. Each ingest flow is recorded onto its own tape group. This is done to ensure maximum control over the storage policy for each tenant or collection source. 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.

Storage plan

The storage management system DivArchive, is responsible for creating copies according to the applicable storage plan, and ensuring that these are identical to the primary archival copy. A disk cache has been created to enable recently ingested material to be delivered faster (on a temporary basis) than is possible from tape. Files that arrive on the disk cache can only be removed after at least one successful copy has been made on tape. This process includes a check that copies have been produced in time. The underlying principle is that, as soon as resources become available, the

copy is created as quickly as possible following ingestion. Recovery is necessary whenever a tape fails to work (it may occure that a tape is no longer readable during a migration action, for example, or when an order is received). The copy can then be retrieved for restore.

Storage plans depend on the preservation level agreed with the depositor/data producer and/or the appraisal of the collection that justifies specific extra copies or locations. As soon as the ingestion process has been successfully completed, responsibility for the master files is accepted by the Digital Archive, and the predetermined storage plan.

Storage media

In response to errors on reading or writing files on a tape, operators will change bad tapes for new ones. In order to prevent this as much as possible, the quality of the tapes and the tape drives of the tape robots is proactively monitored. To do this, the software tool called 'Tape Analytics' from Oracle is used. 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. After all, 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 (for the tape robots on the Mediapark) this cannot occur.

Some digital archive masters are stored on disc, instead of tape. On these systems the technique of Erasure Coding is active. This is a method of data protection in which data is broken into fragments, expanded and encoded with redundant data pieces, and stored across a set of different locations or storage media. By this technique redundancy as well as file integrity are being monitored.

Tape migration

Every five to seven years all data tapes are rewritten to a new generation of LTO. A migration plan is drawn up for this process, which is monitored by the Preservation Board. The plan also includes an estimate of future volumes. This planning is carried out to ensure that the capacity and performance of 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'),

when three months old, 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 of a removed file) is possible in emergency situations, on condition the tape has not been deleted and no migration has taken place in the meantime. To find the correct location of the removed file the event table (see above) is used.

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. The entire production environment has a daily backup schedule. From other environments only the essential data is backuped. The nightly backup includes all the database files that contain the catalogued metadata of the MAM-system. In the event of data loss or database defects, backup copies can be restored.
The backup procedures are managed by the backup system called Rubrik. This process is strictly monitored. In case of a deviation, an alarm goes off (e.g., if the backup failed).
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. 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.

Disc management

Files that are kept on file servers are mirrored; in order to have two copies of these archival masters too, like on tape. The mirrorrule is processed at ingest per item. This concerns still images, low-resolution AV files, and text files that are not stored on tape due to the small size of the files; the file management system for the tapes is not designed for this. Monthly, the influx of masterfiles and proxy's on disc (mp4, mp3, tiff, and pdf) files is archived as objects in Diva; this also includes the low-res/ small files belonging to the archive masters, which are stored on Diva. Retroactively, the legacy data on these servers is also being archived in Diva in this manner.

Browse files of all preservation masters are also stored on disc and mirrored. The aim of these mirrored copies 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.

Fig. 18: Storage volumes per medium (2026)

Offsite disaster recovery

A disaster recovery approach is used to spread the geographical risk. Most storage plans define two copies, each on a different location. A second tape robot is placed with a professional commercial partner at the Media Park. A third tape robot has been implemented, situated at a secure distance. This third location functions primarily as a disaster recovery facility for the S&V's Production environment. Two key parts of the collection (digitized film and public broadcast video) have a copy stored in this facility. Also, the database of the storage management system and the database of the MAM-system are copied to this tape robot every night.

10.5 Security

Sound & Vision's Security policy[59] establishes which security measures and procedures are implemented to secure the entire digital infrastructure against all (deliberate or unintentional) internal and external threats, human error and incompetence, and deliberate sabotage. As a derivation of the themes that, according to OAIS, require security, S&V has structured its security policy into several topics. The main security measures per topic are mentioned below.

Physical security

Physical security is structured in several tiers, defining distinct areas, with security increasing as the information becomes more sensitive. The innermost tier provides the highest level of protection, e.g.:

- At the two data centres on the Mediapark, a theft and fire alarm system has been fitted. This is linked to a security organization and/or the police and fire brigade.
- The data centres have 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 u nauthorized persons, the data centres are secured with turnstile gates, a second factor or biometrics for authentication.

Strict procedures apply to the issuing of access passes. External visitors are registered and collected, as well as escorted upon departure. The S&V information security contingency plan describes scenarios such as fire or power outages and the procedures that apply in those cases.

Authentication and authorization

To secure internal access to devices and 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.

Authentication for access to servers in the infrastructure is based on user names and passwords. Password management protects against the misuse or exposure of passwords. Specific software is used for storing passwords. Access to systems that can bypass security measures, such as backup software and VPN connections, is additionally secured. Users are not granted any rights beyond those necessary for their tasks (for example, rights to change or delete data).
Fixed procedures apply for granting and revoking rights when employees join and leave. The rights are regularly monitored by management.

Access control

For a large part of the collection publication is restricted due to copyright considerations. Access to this material is secured by having a separate network and portal in place, accessible only via connections from IP addresses on the S&V whitelist.

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, via a dedicated instance of the media professional-portal, based on user and password authentication. Access is not permitted to anyone else. The internal MAM system remains accessible solely to administrators. Tenant access level, rights and restrictions are specified in their respective agreements.

Data integrity

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 collection. Alarms and events that entail significant risks are automatically sent via email 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. This is done by an independent organization. It 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). And other penetration tests detect less common vulnerabilities or combinations of vulnerabilities. Security improvements are made on the outcome of these testst.
Sound & Vision makes a daily backup of its production systems (web applications, meta databases and indexes, search software, operating systems and virtual machines).
Sound and Vision monitors whether planned backup procedures have been correctly implemented.

Data confidentiality

In the processing register, S&V records per application which personal data are processed and which retention periods apply. With regard to the archive, technical and organizational measures have been taken to guarantee the principle of minimal data processing. S&V handles data subject access requests or requests for rectification of certain data in the archive according to a fixed procedure. Privacy statements have been drawn up for the benefit of data subjects whose personal data the organization processes.

Employee awareness regarding incidents is actively maintained and regularly updated. The Data Protection Officer prioritizes incidents and ensures appropriate follow-up actions are taken, including additional or alternative measures to prevent similar incidents in the future.

Appendix 1

NISV 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 software packages. Of all ingested materials in this category, standard viewing versions are produced in MPEG4. Standard Definition (SD) Material must be encoded as MXF OP1a, D10-30 or D10-50, the standard for Digital Provision of the public broadcasters, on the basis of the SMPTE guidelines. High Definition (HD) material must be encoded as XDCAM HD422/50 Mbps.

ProRes files are supported in a number of variants and combinations of frame rates and resolutions. This format is lossy, but there is no limitation on the bitrate and various resolutions are possible, allowing a relatively high quality to be achieved.

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. This format is lossy and is widely used for Internet videos where the bitrate must be relatively low. MP4 files are supported in a limited number of variants and combinations of frame rates and resolutions.

Audio

The preservation format for audio is BWF. This format consists of the lossless WAV format, supplemented by additional metadata fields. For born-digital content that is not available otherwise, the MP3 format is archived provided certain specifications are met. This format is lossy, and therefore the recommendation is to apply the highest possible bitrate.

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.

Text

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

Photographs/Paper

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.

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.

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.

Appendix 2

Required knowledge and skills of Digital Archive staff

Job profiles at Sound and Vision include, wherever relevant, the competences required for digital archiving. This relates in particular to the job profiles of 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/playout 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.


SOUND&
VISION