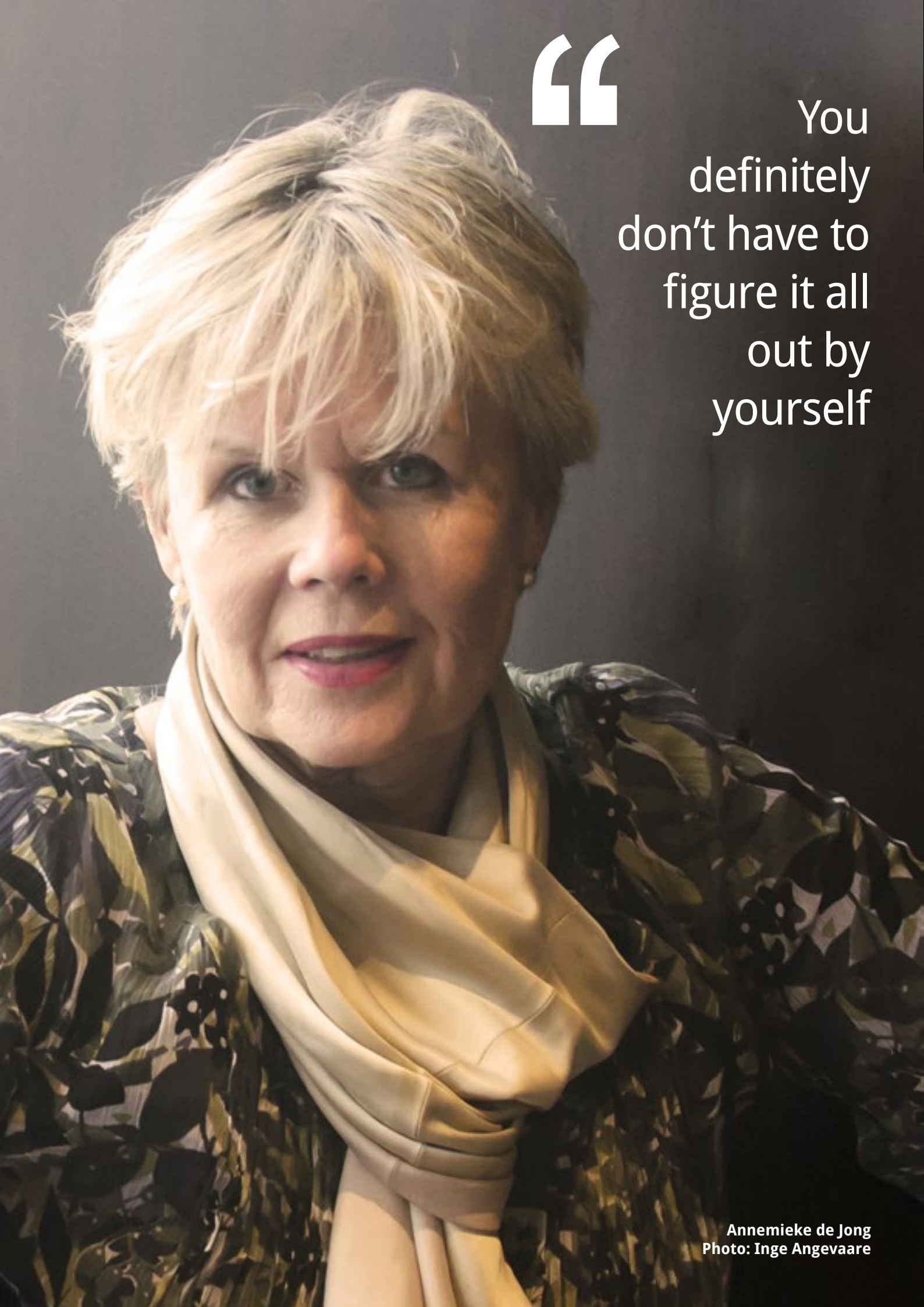
The challenges of becoming a Trusted Digital Repository

Annemieke de Jong is Preservation Officer at the Netherlands Institute for Sound and Vision (NISV) in Hilversum. She is responsible for setting out strategic policies for preserving and managing the institute's digital collections. Currently de Jong and her team are setting up a full set of requirements for the huge archive of NISV to become a Trustworthy Digital Repository (TDR). This involves a remodelling of the archival processes from ingest to access through to storage, as well as developing new administrative and organizational policies.

The Netherlands Institute for Sound and Vision is the central production archive for the Dutch public broadcasters. In this role the archive is responsible for storing and providing access to broadcasted television and radio. At the same time Sound and Vision functions as the Dutch national audiovisual archive. Thus, its collection also includes cultural photos and objects, amateur film and AV collections from businesses and social organizations. In addition, Sound and Vision plays a central coordinating role in the Dutch broadcasting and AV cultural heritage landscape, gathering and disseminating knowledge in the preservation and access domains. Currently Sound and Vision manages a collection that includes over 800,000 hours of radio, television and film. Every day, fragments from the repository are delivered to hundreds of users in the professional broadcast domain, at home, in educational institutions, in businesses and visiting Sound and Vision's museum. Recently, Sound and Vision has made it a strategic goal to become a TDR for Dutch audiovisual cultural heritage collections.



Annemieke de Jong
Photo: Inge Angevaare

How much of Sound and Vision's collections today are in digital form?

Since 2007, we ingest the complete broadcasted radio and TV programming of all twenty Dutch public broadcasters as well as its related metadata via an automated digital workflow system. This year, 2015, we will finish our 7-year project "Images for the Future" that has digitized more than half of our analogue film, video and audio collections. Our digital repository system to date holds approximately 7 petabytes of digital born and digitised material, which equals around 450.000 hours. On an yearly basis this is added to by some ten thousand hours of digital born broadcast programmes as well as the by the collection materials we store and preserve for organizations in the heritage field.

What is your strategy to keep all this material safe and accessible in the long run?

Our workflows, storage infrastructure and metadata conventions are traditionally closely related to the demands of broadcast environment, where access is the primary goal. When the first IT integration with production was designed in 2007, long-term preservation was no major consideration. In the initial infrastructure design, preservation processes were not explicitly incorporated. But obviously, whether it is public radio, television or other audiovisual material, we need to store it safely and keep it made permanently available for those that want to use it.

Managing our fast growing digital repository and the increasingly complex processes around it in a rational and responsible way, has therefore become one of our biggest challenges, leading to important questions: How to control the life cycle of the numerous broadcast programmes that are ingested daily? How to manage all the different formats, locations and size of the heritage collections that are entrusted to us, and, most important: how to guarantee the delivery of up-to-date formats to all of our depositors and users in the long term.

Trusted OAIS compliant archiving is an important strategic goal of the organization as a whole

We have become aware that in order to control and preserve our digital archive, more processes, more policy, more procedures and more metadata would need to be incorporated into our systems and workflows. Additionally, the role and responsibility of all parties involved, be it depositors, user groups as well as our own staff would need better definition. We basically would need to know precisely what 'digital preservation' signifies to us, being a national AV-archive, operating in a dynamic production environment. And also what common preservation concepts like 'authenticity' and 'integrity' actually imply for the management of our files.

To get a grip on the domain and find out how digital preservation would fit into our environment, we set up a project in 2013, to start 'optimizing' our digital archive. Our first goal was to deliver a set of normative policy documents covering the organizational, technical and data management aspects of our repository system. These documents would then have to function as the main reference model for the re-design of our workflows and systems, implementation being our second goal. We figured that trying to get formally certified as a TDR could well support the work, by providing direction, quality

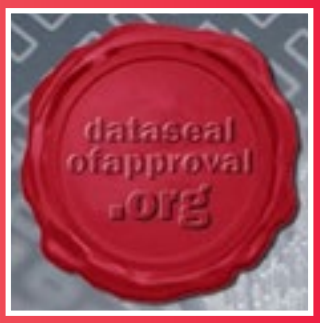
criteria and concrete guidelines. Becoming an acknowledged TDR, by acquiring a 'Data Seal of Approval', thus became our third goal, to be achieved in 2016.

What standards and reference models did you find useful?

OAIS, our main source of inspiration, made us more aware of the broad scope of the preservation domain, covering not only workflows, data objects and technology, but also the financial, legal and organizational aspects of trusted archiving, like the need to set up formal contracts with your depositors and users, the development of methods for risk management, training plans for the staff and so on.
One of the problems we encountered using OAIS, was that the model to date has seen wide take-up primarily in digital libraries and traditional archives. This can also be said for PREMIS, the most important preservation metadata standard. Their applicability and implementation in the media archive domain, where the emphasis lies on access and re-use – is still scarce. This meant we had to modify OAIS and PREMIS based processes and metadata to fit the specific situations and needs of AV files managed in our own dynamic production environment.
Besides OAIS and related standards and documents, we researched a lot of the Presto Prime deliverables that were written over the last couple of years. These documents were very useful in making us understand and define the issues and solutions for our own specific AV-archive production environment, and were all excellent stuff. Deliverable D2.2.1, on Modelling Preservation Processes, even became somewhat like our 'bible'!

What have you accomplished to date?

We now have a large set of normative policy documents in place, outlining OAIS compliant requirements for our archive, that we based on



The Data Seal of Approval : a certificate for trustworthy archives

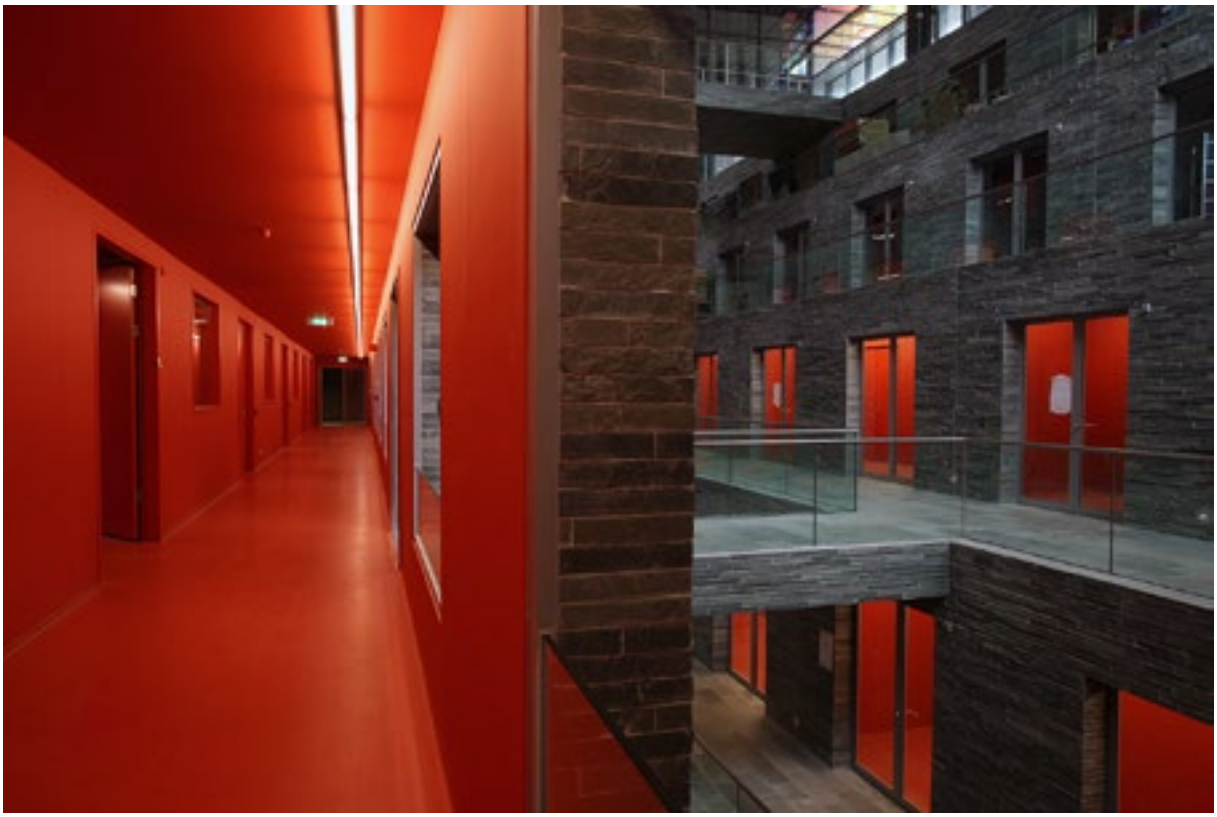
Trust is the basis of storing and sharing data. That trust must be present in various stakeholders. The data depositors want the assurance that their data in the digital archive are safe and will remain accessible, usable and meaningful. Data users have questions like: have the data been well kept, have they retained their authenticity and integrity, are the data of good quality, do the identifiers refer to the appropriate objects? The funders of digital archives have other concerns. They want to be certain that their investment in data production yields optimum returns, i.e. that the data will be available for long term reuse. What characteristics make a digital archive reliable? First, a digital archive's mission should be to give reliable long-term access to the digital data under their care, now and in the future. Second, there should be permanent monitoring, planning and maintenance. The threats and risks within their systems must be understood. Finally, there should be a regular audit and certification cycle in place. Reliability is not something you achieve once and can then take for granted. Certification can make an important contribution to the confidence of various stakeholders.

The Data Seal of Approval:

- Gives depositors and users the assurance that their data will be stored in a reliable manner and can be reused;
- Provides funding bodies with the confidence that data will remain available for reuse;
- Enables users to assess in a reliable manner the repositories that hold the data which they want to reuse;
- Supports data repositories in the efficient archiving and distribution of data.

the high level digital AV-Archive quality criteria we listed. We also worked on practical guidelines for creating submission and order agreements with our depositors and users. Starting from an analysis of the current technical demands of these groups as to the accessibility to our collections, we're we are now working to establish mechanisms to permanently monitor them, in order to connect their needs to our preservation planning on a structural basis. For the preservation services itself we have developed a definition of the different preservation and access guarantee levels that we can offer, based on the demands of the depositors and our own collection policy as a national archive. The core of our work has been the Information Model that defines normative workflows and preservation metadata. Two essential things were required to ensure both the integrity and the authenticity of our files: the defined preservation business processes, that ensure that preservation related events take place; and a mechanism by which an audit trail can be generated and maintained, allowing the archive to demonstrate the outcomes of these events. By tracking and registering the events can demonstrate the authenticity of the ingested object. Thus we fulfil the basic requirement of trustworthiness.

The model describes not only the actions an object undergoes during various workflows, but also the properties of the objects themselves. These properties are all defined in our AV-Preservation Metadata Dictionary, that includes both essential technical characteristics of audiovisual files as well as preservation metadata, which



In Sound and Vision's vaults, approximately 7 petabytes of digital born and digitalized material is stored to date. How to keep it safe and accessible for the long term?

focuses primarily on digital provenance: that is, metadata documenting the creation, chain of custody and change history over time. Rights related metadata, strictly referring to the rights to preserve, also form an essential element. For the technical metadata, we made a study of a variety of AV specific metadata schemas. PREMIS was chosen as the standard to use for the digital provenance.
In general our documents provide Sound and Vision with an important reference framework against which we can measure how far our current operations consciously reflect preservation lifecycle management in our own environment. We feel we now have a solid, theoretical basis for establishing a dedicated preservation structure that can be considered OAIS compliant. We have identified all those who play a role in the preservation process. Sound and Vision now knows how to operate according to the standards and will thus be able to demonstrate how these standards have been implemented. The implementation of the basic requirements of a 'trusted' repository allows us to prove to our depositors/producer and users, how we operate responsibly.

How was the archive staff involved in the work?

We realized from the start that the digital archive is not confined to the IT department and policy-development. The entire organization, especially the cataloguers, the access services staff and the acquisition and selection staff, plays a role in providing long term preservation and access guarantees. In order to broaden knowledge and awareness among the staff, we choose for a project structure, involving staff members from different departments. Getting everyone to better understand their own role would strengthen the responsibility they feel in preservation. Although it proved to be difficult for most project members to make time for researching and translating theoretical standards and processes into policy alongside their daily work

There are 16 guidelines that together determine whether a digital archive qualifies for the Data Seal of Approval. The guidelines for applying and verifying quality aspects concern the creation, storage, use and reuse of digital data. They have been designed with a focus on scientific materials, but they can be applied to all types of digital information, including audiovisual materials. Fundamental to the guidelines are five principles that together determine whether or not the digital data may be considered as sustainably archived:

1. The data can be found on the Internet.
2. The data are accessible, while taking into account relevant legislation with regard to personal information and intellectual property.
3. The data are available in a usable format.
4. The data are reliable.
5. The data can be referred to (persistent identifiers).

These principles are integral to the 16 guidelines. The DSA guidelines remain valid for a period of two calendar years: the Seal period. After this periode the DSA had to renewed. Reviews and assessment of evidence is being done by DSA peer reviewers. An online self-assessment tool has been developed to make the DSA application process easy and transparent.

The Data Seal of Approval offers the possibility of basic certification. The DIN standard provides a second set of guidelines. The 34 criteria were developed by the German organisation NESTOR (a consortium of museums, archives and libraries) and formalized as the DIN 31644 standard. This DIN standard is essentially a catalogue of criteria which digital archives should satisfy. In 2014 the first DIN-based audits have been conducted. The third way to evaluate a digital archive is provided by ISO standard 16363. This standard is based on the OAIS model that provides a framework for understanding archival concepts needed for the preservation of and long-term access to digital information.

responsibilities, the project approach did succeed in raising the level of consciousness about the meaning of the concepts 'digital preservation' and 'lifecycle management' and their importance for digital collection management. We have achieved that OAIS, its terminology and its process framework are no longer foreign to the majority of our organization. Certain technical processes such as validation and fixity checking have gained prominence and have now actually been implemented.

Three basic conditions to become a TDR, according to Annemieke:

1. Have your management committed;
2. Involve the rest of your organisation;
3. Get documents, standards and advice from other archives & knowledge centres.

How do TDR-requirements relate to other IT-developments in the archive?

It's important that all the knowledge gained during the preservation requirement trajectory remains the reference framework for any development in and around the digital archive. Our requirements demonstrate to IT staff that preservation processes and workflows first need mapping to functional areas in the IT architecture system as a whole. Only then can applications for carrying out the needed functionality be identified. In the meantime, the IT department needs to start finding solutions to support the newly developed business processes.
Good synchronization of the various developments within the archive also remains crucial. For example, at Sound and Vision, our requirements project was running at the same time that the process to acquire a new MAM system began. In order to enable that a new MAM would contribute to OAIS compliancy goals, steps in the Information Model workflows were quickly translated into detailed and concrete MAM requirements. This also meant that some requirements outside of the workflow needed to be analysed. It was at this point that it became clear that OAIS compliancy requirements themselves offer little concrete or technical solutions. In the end, those OAIS compliant lifecycle management requirements that became part of the overall MAM requirements actually represents a sub-set of the overall OAIS requirements.

What are the challenges for the near future?

Besides working out where and how within the IT architecture the preservation functions need to land, serious plans will have to be made to train the staff in digital life cycle management and preservation planning. At the same time we will need to find answers to the financial questions: what are the actual cost implications of implementing the full preservation scenario we worked out? How to connect costs and budgets to the different preservation levels we have developed? And who will pay for preservation in the long run: the archive itself as TDR, the producer/depositor or the user?
Another burning question concerns all the previously ingested files and metadata: how, for example, to ensure that what 'dark metadata', generated during earlier digitization processes, is brought into the preservation workflow and data management system? What does one do with the approximately 400.000 hours that were ingested into a non-OAIS compliant repository, and thus never underwent the fixity checking and validation processes?
Furthermore, the preservation business processes need to incorporate answers to new

questions: are we going to apply the full preservation workflow and metadata processes to all the collections and types of content (metadata, photos, written sources) that are ingested, or only to some? If the latter, how are we going to define different preservation levels for broadcast production material, cultural heritage material and contextual material?

Any recommendations for other AV-archives that consider starting an OAIS-TDR trajectory?

First and foremost: make sure your management is committed to the task! Help them understand that preservation is their primary business being a digital archive, by convincing them of the benefits: safe, accessible and well organized digital collections, skilled staff and happy stakeholders. In this way they will hopefully appoint trusted OAIS compliant archiving - and preferably becoming a certified TDR - to be an important strategic goal of the organization as a whole. This will guarantee you the necessary acknowledgement, the funding and the personnel resources. Secondly: share and communicate! Make sure all departments in your organization are somehow involved in the process. Include the IT people, so they understand that preservation is more than storing and migrating the bits, convince your ingest department of the importance of solid contracts with depositors and make sure cataloguers and archivists start to become familiar with the principles of digital lifecycle management as part of their daily work. And lastly: make use of anything that's

out there and that can help you understand what you need and what you want for your own organization: preservation standards, models, guidelines and so on. Ask your peer archives for their policies and their best practices. After all: every audiovisual archive now is struggling to learn how to best control their increasing digital collections, how to cope, what solutions to find and how to work in trusted ways. You definitely don't have to figure it all out by yourself.

Further reading

- **ISO 14721:2012. The Open Archival Information System Model**
http://public.ccsds.org/publications/archive/650x0m2.pdf

- **DRAMBORA, Digital Repository Audit Method Based on Risk Assessment**
https://www.prestocentre.org/tools-catalogue/drambora-digital-repository-audit-method-based-risk-assessment

- **CCSDS, Audit and Certification of Trustworthy Digital Repositories Recommended Practice**
Magenta Book, September 2011. http://public.ccsds.org/publications/archive/652x0m1.pdf

- **PREMIS**
https://www.prestocentre.org/standards/premis/2.2

- **CCSDS Producer-Archive Interface specification**
https://mail.google.com/mail/u/0/?shva=1#apps/pais+2012/n50/14b0c2303e05d3ac?projector=1

- **Presto Prime Deliverables**
http://www.prestoprime.org/project/public.en.html

- **DSA**
http://www.datasealofapproval.org

- **ISO**
http://www.iso16363.org/

- **DIN**
http://www.langzeitarchivierung.de

Contact the author
Annemieke de Jong, Preservation Officer Netherlands Institute for Sound and Vision: adjong@beeldengeluid.nl

This article was written for AV Insider #6, a periodical published by PrestoCentre.
For more information, visit www.prestocentre.org

PRESTO CENTRE