Metadatamodel NISV

Scope

The NISV archive holds a variety of material, such as audio, video, images, documents and physical objects like cameras, costumes and papers. The metadata model contains the descriptive metadata for all these different asset types.

The MAM-system is called ‘DAAN’. Its internal database holds all descriptive metadata as well as technical metadata of the digital assets of NISV. Analogue material has placeholders, that enable to add some technical metadata of the carrier and to relate to digitised versions of that material (provided that that is available), as well as to descriptive metadata.

DAAN is composed of two proprietary systems: Viz-One and Mayam. As Viz-One is the core MAM-system. Mayam is the workflow engine, that handles workflow-metadata under the bonnet.

DAAN keeps metadata of all files that are stored in the archive. But the asset management (storagemanagement, filemanagement) is done by DivArchive (known as DIVA). For instance: the way copies of a file are managed over different locations is not visible in DAAN.

This document gives an overview of the NISV metadatamodel. It is restricted to the metadata in Viz-One. It covers all descriptive metadata, as well as most technical metadata of all assets.

Datamodel hierarchy

The vertical hierarchy of the datamodel consists of the levels *series, season, program, package, item, logtrack and logtrack item, file* (see figure 2).

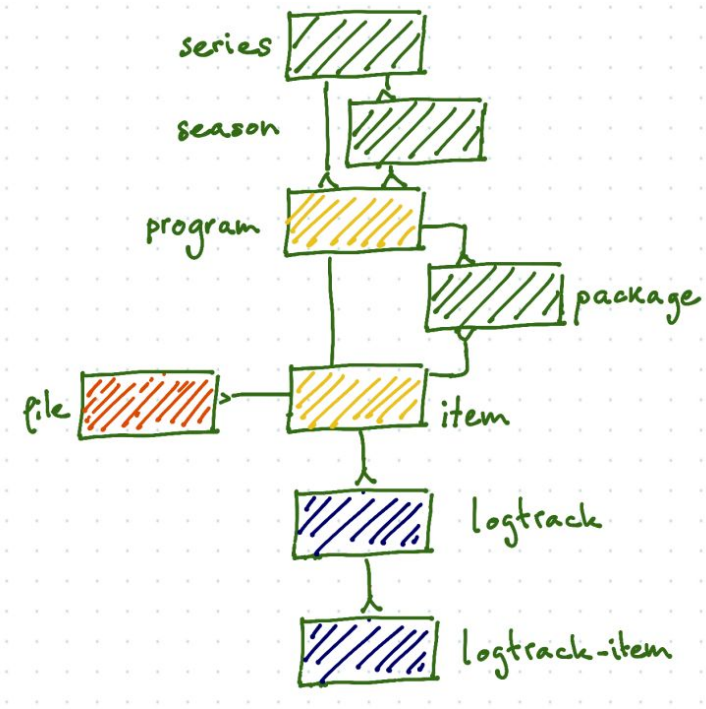


Figure 2: hierarchy of levels

Annotation about the colors:
green: only used when applicable
yellow: core entities
blue: several types, including the technical logtrack that defines the relation between the two core entities
red: automatically generated metadata at ingest/upload files

File

At *file* level only technical metadata is available, that will be automatically generated at ingest or upload of the file. The files are categorised by mediatype: highres, proxy, auxiliary and other. One file will be regarded as the main file; this will usually be the highres, e.g. MXF, WAV, Tiff or (if not available) the first auxiliary file.

Item

The *item* refers to the asset-item: metadata regarding either digital media or analogue material. When the item regards digital media the metadata will represent only the main file. When the item regards analogue material, no files will be attached; the item is a placeholder and as such represents a specific analogue carrier.

The items are differentiated by Item type. Each type has its own set of attributes. In addition to the digital media and analogue material, a few specific types have been created:

- Contract (IPR contracts)
- Licensecontract
- Order Items

These items are not related to any program or logtrack. They can only have auxiliary files.

Logtrack

The *logtrack* is a timeline on the streaming video or audio. It has a color and a type. Current types are:

- Marker (technical logtracks)
- Baton (remarks and errors)
- Scene descriptions
- Custom
    - subtitles deaf and hearing impaired provided by the broadcasters (TT888)
    - subtitle translations of non-Dutch subs (Cavena)
    - speaker labels (thesaurus labels of speakers with an in- and outpoint)
    - extracted terms (thesaurus labels of corporations, geographical names, person names and subject terms derived from subtitles with in- and outpoints that refer to the scene description or program.)

Logtrack item

The *logtrack item* is defined by the timecode in and timecode out on a specific logtrack. The logtrack items have a set of attributes, depending on the (sub)type of logtrack and/or other typical types of the program or item.

The technical logtrack relates items to programs. It is used to define the start and end position of the program on the digital file and also to specify different versions of the same program (e.g. clean feeds). This information is needed to publish the item on the portal; all digital and analogue items that are to be published must have a technical logtrack.

Package

A *package* is needed to relate items to programs in some corner cases:

- to put several member items (digital) belonging to one program in playing order, necessary for playout
- to relate multiple programs to a single item (digital or analogue)

Program, series, season, (scene descriptions)

There are three entities and one specific logtrack that hold the descriptive metadata of the asset.
The *program*, *series and season* all have very similar attributes. This means that these levels can be described at the same detail.

- the program is the single production or may be the episode in a series
- the series has the overarching information of a multi production
- the season holds specific information that only belongs to a subset of episodes

The logtrack items of the scene-description logtrack also have very similar attributes. These logtrack items describe (timecoded) segments of the program.

Program is the main entrance for search. Results of a search by external users are always presented at program level. Even though in some cases most information is found at a

higher level (series or season). Some attributes are propagated downwards to have essential information directly available in the member programs. This facilitates performance during search and access.

FRBR

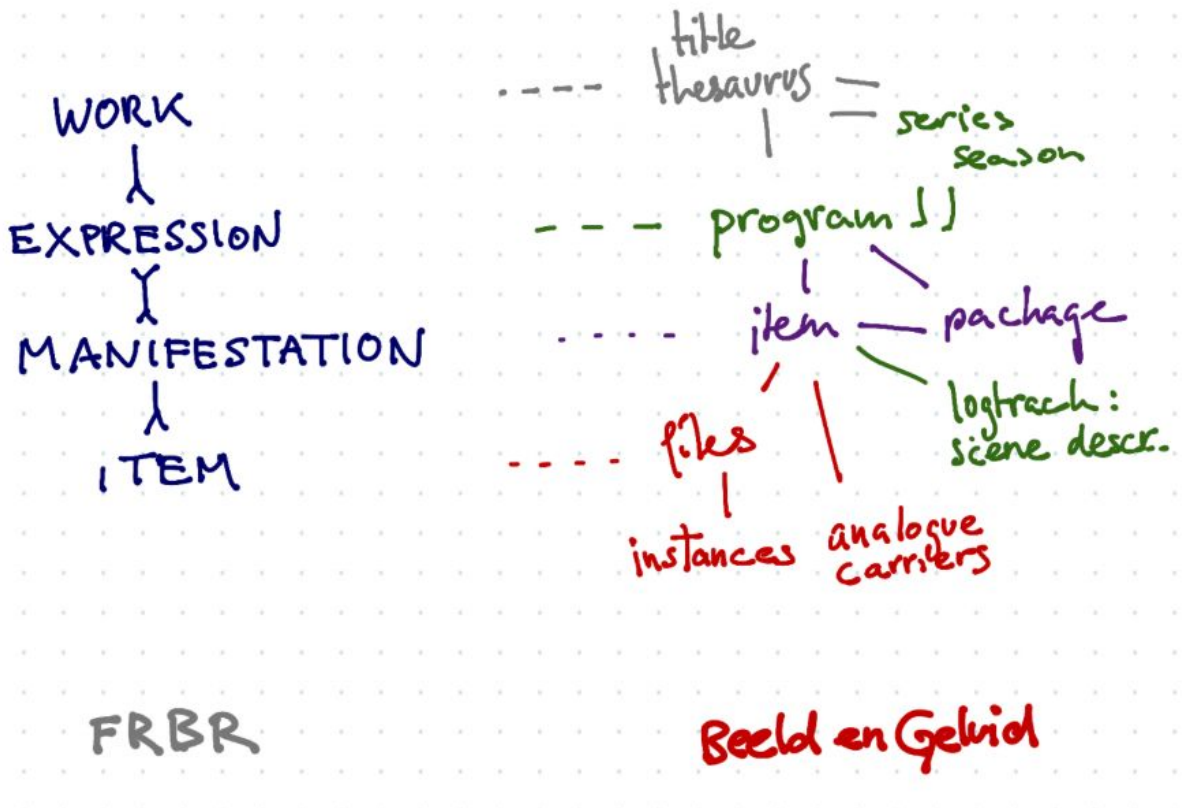
The hierarchy as described is set-up similar to the FRBR-model. Except for the *Work*; this is not modelled as a separate level, although assets belonging to the same work can be grouped via the title thesaurus attribute which provides for grouping different assets under a uniform program title.

Programs (together with series and seasons, as well as scene descriptions) describe the *Expression* of some artistic content.

Items (as named in our hierarchy), describe the physical entity or *Manifestation* of the program. An item in our MAM-system, together with the package when needed, represent the files that are kept for playout or viewing, and for preserving the content in our archive.

At an even more concrete level these files are managed in a filesystem, where several copies are kept. These copies or instances kan be interpreted as the *Items* of the FRBR.

So far for the digital born material. When it comes to digitized material, the analogue carriers will also be described as items in our hierarchy, with a pointer to our physical asset management system. The same holds for analogue material that is not yet digitised.



Attributes descriptive metadata

Non repeatable/Repeatable

Some values are non repeatable such as

- Main Title
- Summary, Long summary, Museum summary
- Episode number
- Material type
- Production Country

Other values are repeatable such as

- Thematic group
- Genre
- Subject terms

Other values are presented in a table with multiple rows, such as

- Crew (Name, Role, Annotation)
- Cast (Name, Character, Annotation)
- IPR Metadata (IPR contract ID, Maker, Current Rightsholder…)
- Publication (Production ID, Publication date, Date annotation, Repetition…)
- Recording Information (Date, End date, Production year…)

Directory

In some cases an attribute represents information that is structured as a new entity. These are called directories. A directory can be referred to at any level of the hierarchy.
For instance:

- *Persons* is used in the IPR metadata at all four descriptive levels. It refers to information of the person being a maker and has Name, Date of birth, Date of death and Contact Information
- *Organisations* is used when the maker is not a person but an organisation
- *Collections* is used at item level and refers to information of the NISV-collection this material is acquired
- *Projects* is also used at item level and refers to the digitisation project

Thesaurus/Dictionary

Some attributes are linked to a thesaurus-list to manage consistency of the values of these attributes. Examples of thesaurus-lists are:

- Persons
- Corporations
- Titles
- Geographical names
- Subject terms

Other attributes are managed by a dictionary, to make sure only controlled values are used

- Role
- Title-type

Attributes technical metadata

File

Some basic technical data on each file is automatically represented at file level.

- General (Filename; Location; Size)
- Details (MIN ID; Container; Date Created; Created by)
- Picture (Format; Bitrate; Resolution; Aspect Ratio)
- Sound (Format; Bitrate; Sample[rate]; Track)

This information is read from the header of the file or created by the system during ingest.

Additionally, depending on the ingest settings, all header information of a file can be extracted (labels and values) and represented as physical metadata.

Technical logtrack item

The technical logtrack item holds metadata on how program and item (digital/analogue) are related:

- In, Out, Duration
- Sequence number
- Carrier type (Program, Clean feed, Photo)
- Program ID

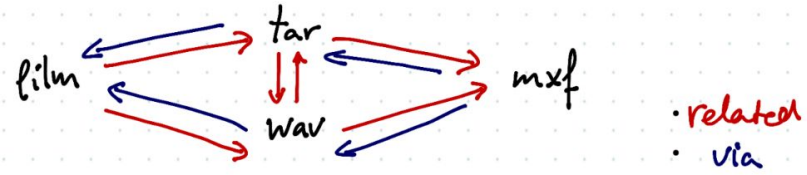
This information is used by the package in case a package is needed to support the relationship between program and item. The package itself has no original metadata.

Relations

The item is related to program. This vertical relationship is called a ‘collections’ relationship. Additional horizontal relationships can be:

- Via: relationship to the source item(s); from child to parent
- Related/Via: relationship to derived item(s); from parent to child[1]
- Related: relationship between sibblings

Figure 3: example horizontal relationships



Item

During ingest, depending on settings, some technical metadata from the header of the main file may be stored at file level. Other technical metadata will be created by the system or part of the ingested metadata. Examples:

- Carrier number
- Technical annotation
- File size

[1] for new ingest this will also get type Via instead of Related

- Playing time

For analogue items a selection of metadata is copied from the physical asset management system (PAM).

Access Rights

To be able to manage the access to the metadata and distribute correct rights to Read/Write/Delete, two concepts are relevant:

- tenants
- publishing points

Tenants are set at program level and propagated vertically through the hierarchy. The tenant is the owner of the asset (metadata) and media (files). All entities of the hierarchy (except files) will have a set of attributes depending on the tennant.

Publishing points are also set at program level and propagated vertically through the hierarchy. Publishing points refer to tenants and to access points such as a portal. For each program it is defined whether other tenants and/or portal have access to the assets (metadata) and media (files) in terms of read/write/delete.

Each user of the system belongs to only one tennant. For each program the combination of tenant and settings for publishing points defines the access rights of the user.
Additionally these rights may be further restricted by the roll of the user.

2018/Marjolein Steeman