Topic 4 - Ingest

Scope of submission

What is a SIP?

"An Information Package that is delivered by the Producer to the OAIS for use in the construction or update of one or more AIPs and/or the associated Descriptive Information." — OAIS, 1-15

Speaker notes

Meaning: The stuff someone gives you to preserve (for them).

For example:

- 1-or-more content files
- Metadata
- Other (relevant) information

Shape of SIPs to come

- File
- Folder(-structure)
- ZIP, TAR, etc.
- BagIt Bag
- Digital objects + database entry
- A video tape, film, image, tape, ...
- ...anything! :D

Mom... Where to SIPs come from?

- Creators
- Digitization vendors
- Staff
- Passionate collectors
- All of the above!
- and more...

How to decide...

what is accepted/required?

- Source (e.g., film, video)
- Agreements with producers
- Internal capabilities
- Metadata only available from producer
- Policies: collection, format
- etc...

Speaker notes

Source: Example: If it's a physical medium "film" it goes to a film archive, if it's a magnetic tape it may go somewhere else: Mediathek.

This may also be a(n unwritten) policy.

What is...?

Ingest activities

Speaker notes

These are as different as the world is colorful and amazing! ;)

They depend on things like: * media type * focus, capacities * policies * etc

Typical ones

- Prepare & record (analogue) source
- Generate unique ID
- Validate or generate fixity/hash data
- Format policy checks
- Format normalization
- Create derivatives
- Create metadata
- Logging
- Virus check
- etc...

Speaker notes

Questions: * Which of these activities applies to which kind of data formats? * Which steps could be "source preparation"? And for which kinds of materials?

The Unique Identifer

A "must have"!

Also known as:

- Object ID
- Archive signature
- Item ID
- UID, UUID
- ...

Speaker notes

Questions: * For what reason(s) would you want/need an identifier? * Why is it desired that it's unique? * What is "Linked Open Data", and what's its relation to

The Unique Identifer

Examples

- V-00815
- W/S #00034
- FBW002984
- 38AF2EC1A13494B9DF6FD6E75960307
- 111-ADC-4319
- VHS-0317
- adBDwKf_aSE
- Q83697636
- ...

Format Normalization

- Popular SIP to AIP use case
- Improve preservation properties
- By switching to a "better" format
- Cleaning/normalizing data (dialects)

Format Normalization

Examples

- Rewrap container (eg MKV, MOV, MXF)
- Audio to PCM
- Convert video to FFV1, V210, H.264, etc

Format Policy Checks

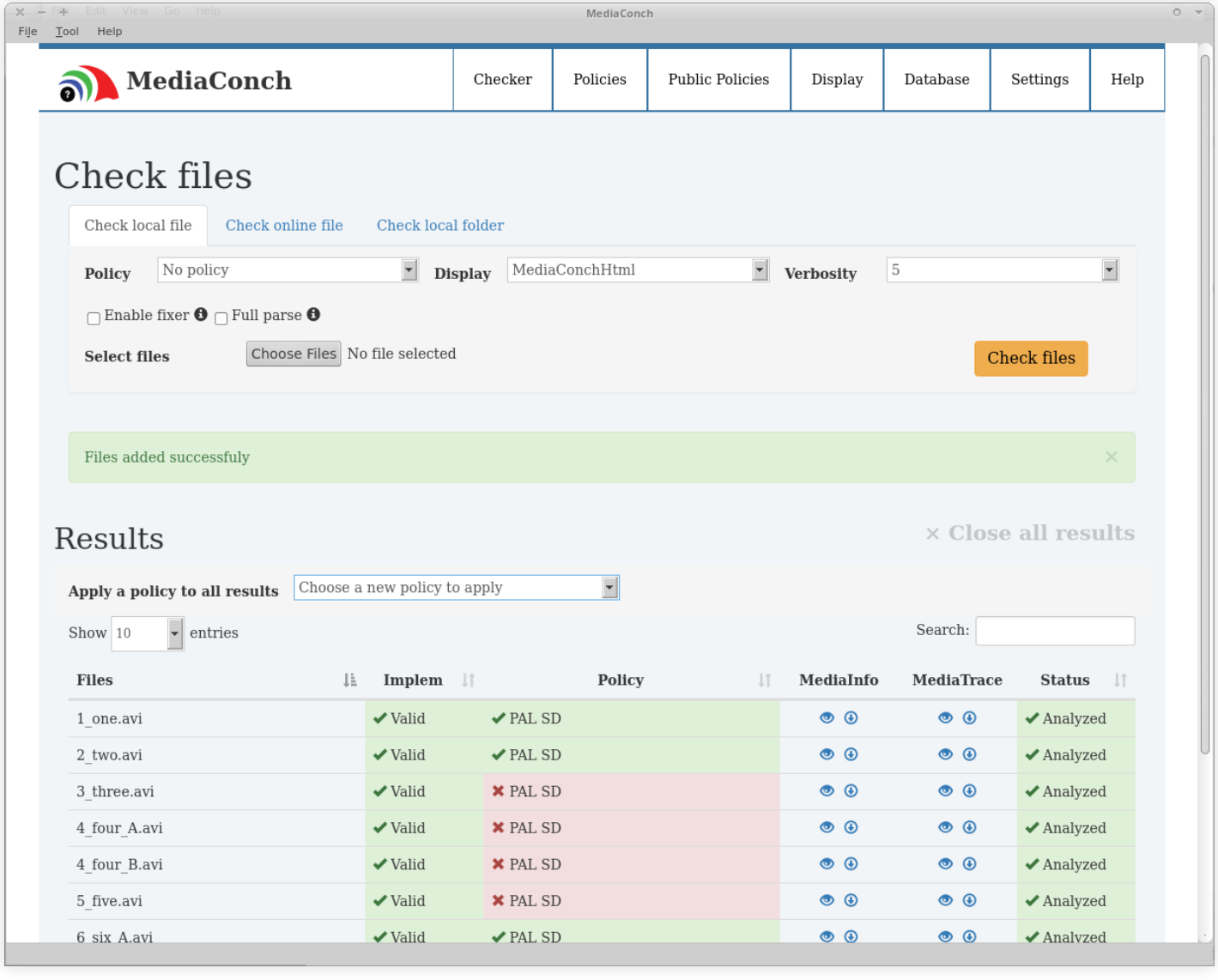
- Define conditions for tech-MD properties
- "whitelist" formats
- Spot irregularities

Speaker notes

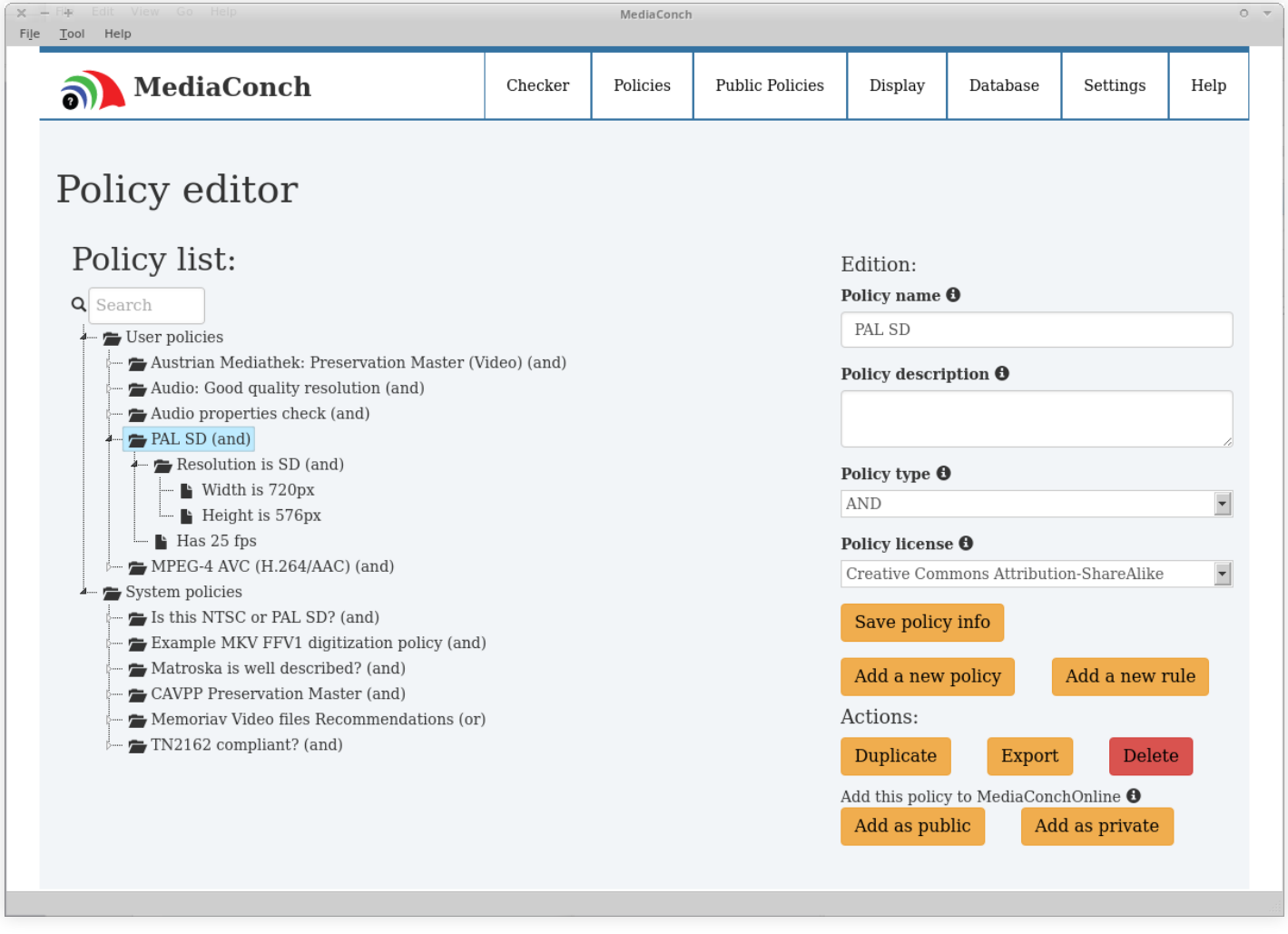
In order to decide which files can be ingested without format normalization, or need other special attention, it is good to define so called rules which properties a media file shall have. These rules are (also) called "policies".

These policies can then be used to decide further steps to be taken for certain objects.

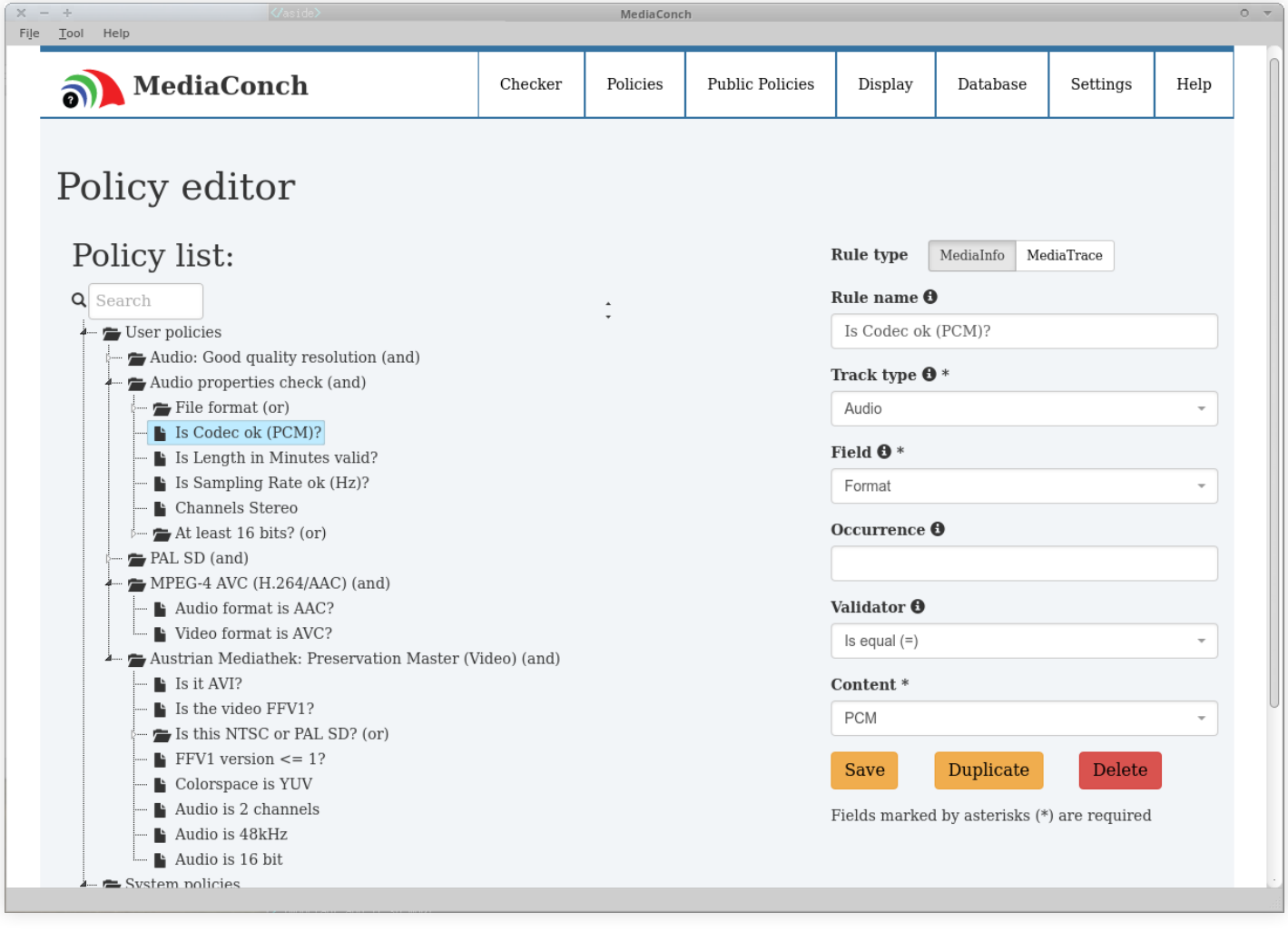
MediaConch


MediaConch
File Tool Help
MediaConch
Checker
Policies
Public Policies
Display
Database
Settings
Help
Check files
Check local file
Check online file
Check local folder
Policy
No policy
Display
MediaConchHtml
Verbosity
5
Enable fixer
Full parse
Select files
Choose Files
No file selected
Check files
Files added successfuly
Results
× Close all results
Apply a policy to all results
Choose a new policy to apply
Show 10 entries
Search:
Files
Implem
Policy
MediaInfo
MediaTrace
Status
1_one.avi ✔ Valid ✔ PAL SD ✔ Analyzed
2_two.avi ✔ Valid ✔ PAL SD ✔ Analyzed
3_three.avi ✔ Valid ✖ PAL SD ✔ Analyzed
4_four_A.avi ✔ Valid ✖ PAL SD ✔ Analyzed
4_four_B.avi ✔ Valid ✖ PAL SD ✔ Analyzed
5_five.avi ✔ Valid ✖ PAL SD ✔ Analyzed
6_six_A.avi ✔ Valid ✔ PAL SD ✔ Analyzed

MediaConch


MediaConch
File Tool Help
MediaConch
Checker
Policies
Public Policies
Display
Database
Settings
Help
Policy editor
Policy list:
Search
User policies
Austrian Mediathek: Preservation Master (Video) (and)
Audio: Good quality resolution (and)
Audio properties check (and)
PAL SD (and)
Resolution is SD (and)
Width is 720px
Height is 576px
Has 25 fps
MPEG-4 AVC (H.264/AAC) (and)
System policies
Is this NTSC or PAL SD? (and)
Example MKV FFV1 digitization policy (and)
Matroska is well described? (and)
CAVPP Preservation Master (and)
Memoriav Video files Recommendations (or)
TN2162 compliant? (and)
Edition:
Policy name
PAL SD
Policy description
Policy type
AND
Policy license
Creative Commons Attribution-ShareAlike
Save policy info
Add a new policy
Add a new rule
Actions:
Duplicate
Export
Delete
Add this policy to MediaConchOnline
Add as public
Add as private

MediaConch


MediaConch
File Tool Help
MediaConch
Checker
Policies
Public Policies
Display
Database
Settings
Help
Policy editor
Policy list:
Search
User policies
Audio: Good quality resolution (and)
Audio properties check (and)
File format (or)
Is Codec ok (PCM)?
Is Length in Minutes valid?
Is Sampling Rate ok (Hz)?
Channels Stereo
At least 16 bits? (or)
PAL SD (and)
MPEG-4 AVC (H.264/AAC) (and)
Audio format is AAC?
Video format is AVC?
Austrian Mediathek: Preservation Master (Video) (and)
Is it AVI?
Is the video FFV1?
Is this NTSC or PAL SD? (or)
FFV1 version <= 1?
Colorspace is YUV
Audio is 2 channels
Audio is 48kHz
Audio is 16 bit
System policies
Rule type
MediaInfo
MediaTrace
Rule name
Is Codec ok (PCM)?
Track type *
Audio
Field *
Format
Occurrence
Validator
Is equal (=)
Content *
PCM
Save
Duplicate
Delete
Fields marked by asterisks (*) are required

Possible Bottleneck

"Ingest can be a dangerous bottleneck. Don't let the perfect be the enemy of the good"

Comments?

Questions?