Choosing and implementing a MAM system

Peter Bubestinger-Steindl

(`Peter @ ArkThis.com`)

MAM? DAM!

- **DAM:** Digital Asset Management
- **MAM:** Media Asset Management
- **CMS:** Collection Management System
- aka: “Our Archival Catalogue”

Speaker notes

Disclaimer: DAMs or MAMs may be used for preservation, but are sometimes/often not designed or intended for use in a preservation context, but merely to store and handle “digital assets”: from regular office files (documents, images, etc) to managing in-house assets of larger companies. And some of these systems were then “also” used by archives.

The term “Collection Management System” usually indicates that it was more likely intended to be used in a preservation context, such as museums for example - where it may be used beyond digital: To handle physical collections even, like books, chairs, or anything. And files ;)

You may see the term “Collection Management System” being used interchangeably with DAM or MAM by the preservation community.

(Note: In the IT world, the abbreviation “CMS” is more commonly known as: “**Content** Management System” which is something completely different. In order to save some screen space, I will use the abbreviation “CMS” in these slides however instead of typing “Collection Management System”)

A DAM is usually the generic version of MAM - and sometimes the borders between “is it a MAM? is it a DAM?” are fuzzy and unclear, because they are so closely related.

Typical for “classic” DAMs: Often trimmed and designed for handling “2D material” (documents, images). When it comes to audiovisual, they’re mostly inadequate or not suitable for archive-suitable quality media handling.

MAMs are usually better suited for handling media. For example:

- auto-generating access/preview copies suitable for low-bandwidth/internet/browser access.
- image area annotation
- time based annotation (“markers”)

But watch out! Even systems designed to handle AV media, are often not taking too much care about preservation/archival aspects. It’s not uncommon that “looks/sounds good enough!” is exactly where you’re at.

So please:

- Try *before* you buy!
- Don’t trust sales. Ever.

A short list

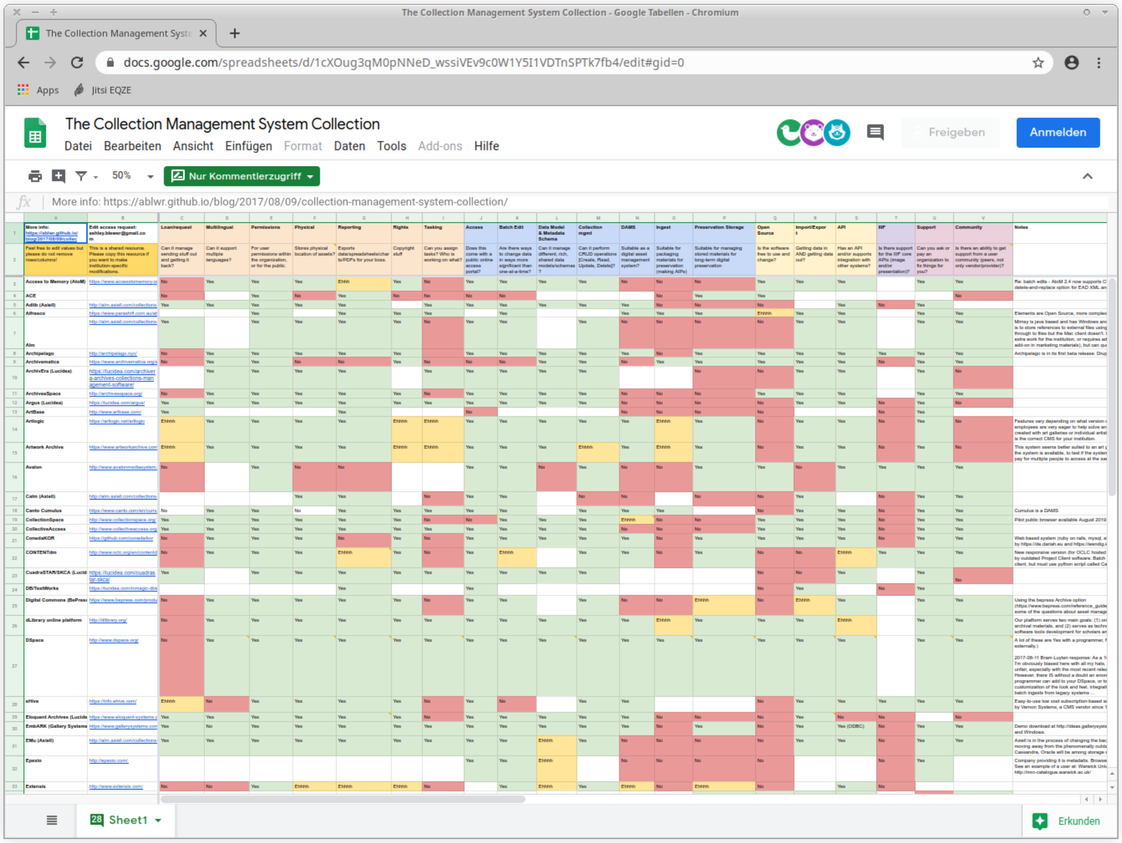
of some popular OpenSource CMS:

- AtoM (Access To Memory)
- Omeka
- CollectionSpace
- ResourceSpace

A loooong list...

of not only OpenSource CMS:

https://bits.ashleyblewer.com/blog/2017/08/09/collection-management-system-collection/


The Collection Management System Collection - Google Tabellen - Chromium
The Collection Management Syst
docs.google.com/spreadsheets/d/1cXOug3qM0pNNeD_wssiVEv9c0W1Y5I1VDTnSPTk7fb4/edit#gid=0
Apps
Jitsi EQZE
Datei
Bearbeiten
Ansicht
Einfügen
Format
Daten
Tools
Add-ons
Hilfe
Freigeben
Anmelden
50%
Nur Kommentierzugriff
Sheet1
Erkunden

The Collection Management System Collection

More info: https://ablwr.github.io/blog/2017/08/09/collection-management-system-collection/

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	
1	More info: https://ablwr.github.io/blog/2017/08/09/collec	Edit access request: ashley.blewer@gmail.com	Loan/request	Multilingual	Permissions	Physical	Reporting	Rights	Tasking	Access	Batch Edit	Data Model & Metadata Schema	Collection mgmt	DAMS	Ingest	Preservation Storage	Open Source	Import/Export	API	IIIF	Support	Community	Notes
2	Feel free to edit values but please do not remove rows/columns!	This is a shared resource. Please copy this resource if you want to make institution-specific modifications.	Can it manage sending stuff out and getting it back?	Can it support multiple languages?	For user permissions within the organization, or for the public.	Stores physical location of assets?	Exports data/spreadsheets/charts/PDFs for your boss.	Copyright stuff	Can you assign tasks? Who is working on what?	Does this come with a public online access portal?	Are there ways to change data in ways more significant than one-at-a-time?	Can it manage different, rich, shared data models/schemas?	Can it perform CRUD operations [Create, Read, Update, Delete]?	Suitable as a digital asset management system?	Suitable for packaging materials for preservation (making AIPs)	Suitable for managing stored materials for long-term digital preservation	Is the software free to use and change?	Getting data in AND getting data out?	Has an API and/or supports integration with other systems?	Is there support for the IIIF core APIs (image and/or presentation)?	Can you ask or pay an organization to fix things for you?	Is there an ability to get support from a user community (peers, not only vendor/provider)?	
3	Access to Memory (AtoM)	https://www.accesstomemory.o	No	Yes	Yes	Yes	Ehhh	Yes	No	Yes	Yes	Yes	Yes	No	No	No	Yes	Yes	Yes		Yes	Yes	Re: batch edits - AtoM 2.4 now supports C delete-and-replace option for EAD XML an
4	ACE		No		Yes	No	Yes	No	No	No	No				No	Yes	Yes					No	
5	Adlib (Axiell)	http://alm.axiell.com/collections-	Yes	Yes	Yes	Yes	Yes		Yes	Yes	Yes	Yes	Yes	Yes	No	No	No		Yes	No	Yes		
6	Alfresco	https://www.parashift.com.au/al			Yes		Yes	Yes	Yes		Yes			Yes	Yes	Yes	Ehhh	Yes	Yes		Yes	Yes	Elements are Open Source, more complex
7	Aim	http://alm.axiell.com/collections-	Yes		Yes	Yes	Yes	Yes	No	Yes	Yes		Yes	No	No	No	No	Yes	Yes		Yes	Yes	Mimsy is java based and has Windows and is to store references to external files using through to files but the Mac client doesn't. I extra work for the institution, or requires ad add-on in marketing materials), but can qu
8	Archipelago	http://archipelago.nyc/	No	Yes	Yes	Yes	Yes	Yes		Yes	Yes	Yes	Yes	Yes	No	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Archipelago is in its first beta release. Drup
9	Archivematica	https://www.archivematica.org/e	No	Yes	Yes	No	No	Yes	No	No	No	Yes	Yes	No	Yes	Yes	Yes	Yes	Yes	No	Yes	Yes	
10	ArchivEra (Lucidea)	https://lucidea.com/archivera-archives-collections-management-software/		Yes	Yes	Yes	Yes		Yes	Yes	Yes	Yes	Yes			No	No	Yes	Yes		Yes	No	
11	ArchivesSpace	http://archivesspace.org/	No	Yes	Yes	Yes	Yes	Yes	No	Yes	Yes	Yes	Yes	No	No	No	Yes	Yes	Yes		Yes	Yes	
12	Argus (Lucidea)	https://lucidea.com/argus/	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	No	No	No	No	Yes	Yes	No	Yes	No	
13	ArtBase	http://www.artbase.com/	Yes				Yes			No				No	No	No	No			No	Yes		
14	Artlogic	https://artlogic.net/artlogic	Ehhhh	Yes	Yes	Yes	Yes	Ehhhh	Ehhhh	Yes	Yes	Yes	Yes	Yes	Ehhhh	Yes	No	Yes	Yes	No	Yes	No	Features vary depending on what version c employees are very eager to help solve an created with art galleries or individual artist is the correct CMS for your institution.
15	Artwork Archive	https://www.artworkarchive.com	Ehhhh	Yes	Yes	Yes	Yes	Ehhhh	Ehhhh	Yes	Yes	Yes	Ehhhh	Yes	Ehhhh	Yes	No	Yes	Yes	No	Yes	No	This system seems better suited to an art g the system is available, to test if the system pay for multiple people to access at the sa
16	Avalon	http://www.avalonmediasystem.	No		Yes	No	No			Yes	Yes	No	Yes	No	No	Yes	Yes	No	Yes	Yes	Yes	Yes	
17	Calm (Axiell)	http://alm.axiell.com/collections-				Yes	Yes		No	Yes	Yes		No	No		No	No	Yes		No	Yes	Yes	
18	Canto Cumulus	https://www.canto.com/en/cum	No	Yes	Yes	No	Yes	Yes	Yes	Yes	Yes		Yes	Yes	Yes		No	Yes	Yes	No	Yes	Yes	Cumulus is a DAMS
19	CollectionSpace	http://www.collectionspace.org/	Yes	Yes	Yes	Yes	Yes	Yes	No	No	Yes	Yes	Yes	Ehhhh	No	No	Yes	Yes	Yes	No	Yes	Yes	Pilot public browser available August 2019
20	CollectiveAccess	http://www.collectiveaccess.org	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	No	No	No	Yes	Yes	Yes	No	Yes	Yes	
21	ConedaKOR	https://github.com/coneda/kor	No	Yes	Yes	Yes	No	Yes	No	Yes	Yes	Yes	Yes	Yes	No	Yes	Yes	Yes	Yes	No	Yes	Yes	Web based system (ruby on rails, mysql, e by https://de.dariah.eu and https://wendig.i
22	CONTENTdm	http://www.oclc.org/en/contentd	No	Yes	Yes	Yes	Ehhhh	Yes	No	Yes	Ehhhh		Yes	Yes	No	Yes	No	No	Ehhhh	Yes	Yes	Yes	New responsive version (for OCLC hosted by outdated Project Client software. Batch client, but must use python script called Ca
23	CuadraSTAR/SKCA (Lucid	https://lucidea.com/cuadrastar-skca/	Yes		Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes				No	No	Yes		Yes	No	
24	DB/TextWorks	https://lucidea.com/inmagic-dbt					Yes			Yes							No	Yes		No	Yes		
25	Digital Commons (BePress	https://www.bepress.com/produ	No	Yes	Yes	Yes	Yes	Yes	No	Yes	Yes		Yes	No	No	Ehhhh	No	Ehhhh	Yes		Yes	Yes	Using the bepress Archive option (https://www.bepress.com/reference_guide some of the questions about asset manage
26	dLibrary online platform	http://dlibrary.org/	No	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Ehhhh	Yes	Yes	Yes	Ehhhh		Yes	Yes	Our platform serves two main goals: (1) on archival materials, and (2) serves as techn software tools development for scholars an
27	DSpace	http://www.dspace.org/	No	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes		Yes	Yes	A lot of these are Yes with a programmer, externally.) 2017-08-11 Bram Luyten response: As a I'm obviously biased here with all my hats, unfair, especially with the most recent relea However, there IS without a doubt an enorm programmer can add to your DSpace, or to customization of the look and feel, integrati batch ingests from legacy systems ...
28	eHive	https://info.ehive.com/	Ehhhh	No	Yes	Yes	Yes	Yes	No	Yes	No		Yes	Yes			No	Yes	Yes	Yes	Yes	Yes	Easy-to-use low cost subscription-based w by Vernon Systems, a CMS vendor since 1
29	Eloquent Archives (Lucide	https://www.eloquent-systems.c	Yes	Yes	Yes	Yes	Yes	Yes	No	Yes	Yes	Yes	Yes	Yes	No	No	No	Yes	No	No		No	
30	EmbARK (Gallery Systems	https://www.gallerysystems.com	Yes	No	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	No	Yes	No	Yes	Yes (ODBC)	No	Yes	Yes	Demo download at http://ideas.gallerysyste and Windows.
31	EMu (Axiell)	http://alm.axiell.com/collections-	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Ehhhh	Yes	No	No	No	No	Yes	Yes	No	Yes	Yes	Axiell is in the process of changing the bac moving away from the phenomenally outd Cassandra, Oracle will be among storage
32	Epexio	http://epexio.com/								Yes	Yes	Ehhhh		No	No	No	No			No	Yes		Company providing it is metadatis. Browse See an example of a user at: Warwick Univ http://mrc-catalogue.warwick.ac.uk/
33	Extensis	http://www.extensis.com/	No	No	Yes	Ehhhh	Ehhhh	Ehhhh	No		Yes	Ehhhh	Yes	Ehhhh	No	Ehhhh	No		Yes	No			

Speaker notes

There is no one size fits all. As usual :P

Criteria Overview

- Controlled vocabulary lists
- Import/Export abilities
- Vendor/license dependency
- Metadata Standards compliance
- “Classic” vs “Entity relationships”
- Does it do Web?
- Access control
- Failsafe options
- Does it scale?

Your Wishes and Experiences?

- Feature Wishlist?
- Bad experiences (=avoid in the future)?
- Good experiences (=want this again)?
- Questions you’d like answered?

Speaker notes

Split into n groups: Each one writes down and presents about their experiences and wishes from a MAM:

- Feature Wishlist
- Bad experiences (=avoid in the future)
- Good experiences (=want this again)
- Questions you’d like answered

Controlled Vocabulary

> *“Controlled vocabulary schemes mandate the use of predefined, authorised terms that have been preselected by the designers of the schemes, in contrast to natural language vocabularies, which have no such restriction.”*

Source: “Controlled Vocabulary” (Wikipedia)"

CoVoc Examples

Same-same 😇, but different… 😭

- **35mm** = 35 mm = 35 millimètre
- **dup pos** = duplicate positive
- **de** = deu = german = German = alemán
- **yuv422p10le** = YUV, 4:2:2, 10 bpc
- **Director** = Directed by

Speaker notes

Controlled vocabulary: Be very greedy with “free text” fields, as they lead to chaos, disorder and mayhem!

On the other hand: It’s always good to have “some” field where to put the “doesn’t go anywhere else” stuff. And it may also be useful/necessary to be able to store/preserve the original terms “as-is”, since mapping different sources to a vocabulary may not be as straight forward or exact as one would hope.

Oh and other mapping fun: * Typos do happen! “35mn” anyone? * 1967-06-07-12 (Date with a “sort index” hack added in Freetext field)

Handling CoVocs

- Choice / definition of terms?
- Standard terms (e.g. ISO mapping) for: Language, Country, Region, Tech-Terms, etc.
- Maintaining / augmenting terms?
- Import / Export lists?
- Identifier vs Label

Identifier & Label

> *Can the MAM store & handle this properly?*

Label (for humans)	**Identifier**(for machines)
Albanian	`sq`, `sqi`, `alb`, …
German	`de`, `deu`, `ger`, …
Undetermined	`und`

Examples: ISO 639 (Languages), ISO 3166 (Countries),…

Speaker notes

Separating label from identifier, has the following benefits:

- Label can be translated arbitrarily.
- The meaning of the data stored is (more) “clear”. Less options for mis-interpretation errors.
- Can even be (more easily) upgraded to shared CoVocs and Linked Open Data in the future.
- Import/Export and exchange with others (and systems) greatly improved.

Shared CoVocs

- Same vocabulary terms/lists can be used across different systems/domains.
- Use existing CoVocs where possible.
- Share/publish yours.
- Be in contact/exchange with others.
- Be ready for embracing trade-offs.
- Can your MAM handle “dynamic” CoVocs?

Speaker notes

As great as sharing and using common CoVocs is, it brings the “problem” that you need to agree on (a) common term(s).

This can sometimes be trivial, and otherwise be non-trivial - and even very emotional. Practically, a major issue is often that different domains have different terms in their professional vocabularies already - which are (or may be) mis-interpreted or confusing for other domains.

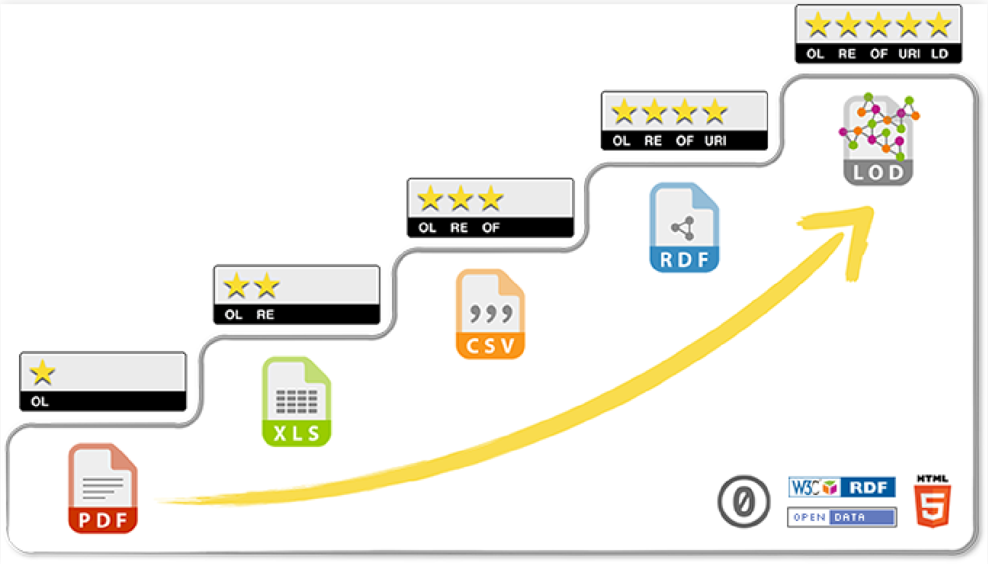
> *“But we call it ‘…’ - and calling it ‘…’ may completely confuse our operators”*

Example: The term “Agent” ;)

Import / Export

- Lists / CoVocs?
- Cataloguing data? $^{\text{(DB)}}$
- Your files? $^{\text{(storage)}}$
- Value/term/schema mapping options?
- Programming language & skills?

Access & Exchange: Data Formats



Speaker notes

The options for accessing, exchanging - basically working with your data - play a very important role in the choice of any catalogue / MAM.

Linked (Open) Data

> *“linked data (often capitalized as Linked Data) is structured data which is interlinked with other data so it becomes more useful through semantic queries”*

Source: Linked data (Wikipedia)

Speaker notes

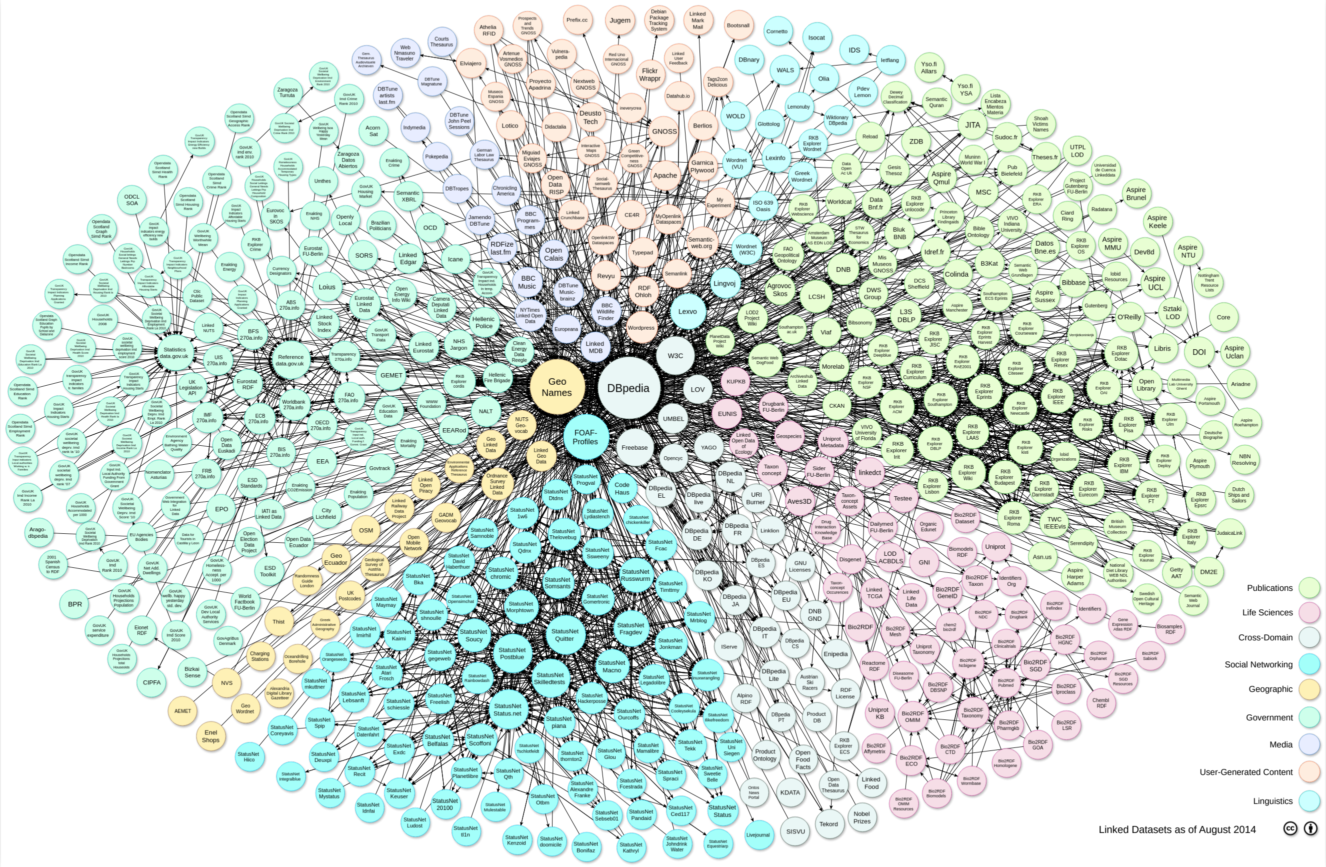
LoD is still a bit “new”, and there are many many many things to be considered, tested, found out, implemented, improved, etc. Yet, having isolated catalogues and maintaining all entitites individually, does not seem like a long-term sustainable approach.

However, many archives are still using conventional “flat” data models, local CoVocs - and are slowly transitioning towards shared CoVocs and Entity-Relationship models.

If you are interested in more hands-on practical experience with LoD and interoperability of data, you may want to get in touch with the FIAF cataloging commission, who is very actively charting the new territory of archives and LoD.

There’s a short summary from a FIAF LoD Workshop in 2019, which contains very valuable overviews and insights in a quick read :)

Linked (Open) Data



Wikidata

> *“[…] is a free and open knowledge base that can be read and edited by both humans and machines.”*

Example: Earth (Q2) vs Earth (Q83697636)

MD-Standards: Some of Them

- Dublin Core: Core fields (& their names)
- METS: (MD container) descriptive, administrative, structural
- PREMIS: A metadata framework
- EBUCore: descriptive & technical (broadcast use case focus)
- CEN EN 15907: Comprehensive description of cinematographic works
- FRBR ("furbur"): Comprehensive description of bibliographic works
- Mediainfo XML: Technical metadata (AV)

Speaker notes

So you see: Metadata itself needs to be stored in "a format" too. Same rules and properties as for the other data formats apply (open, accessible, etc)

PREMIS: PREservation Metadata: Implementation Strategies METS: Metadata Encoding and Transmission Standard EBUCore: European Broadcast Union (EBU) - based on Dublin Core EN 15907: Metadata standard for Cinematographic Works FRBR: Functional Requirements for Bibliographic Records

MAM Requirements for MD-Standards

- Store ID/ISO values (or just label)?
- Relationships between different schema?
- Import/export schema support?
- Field labels in UI vs “The Standard Terms”?
- Watch out for “EEE”:
  Embrace, Extend, and Extinguish

Speaker notes

EEE: Although the phrase “Embrace, Extend, and Extinguish” was invented by Microsoft, the principle is neither new, nor limited to their company:

With any standard, it may be so that someone implements it, but then feels like adding an additional field (for whatever reason). This may seem fine at first, but may lead to interoperability issues pretty fast.

If possible, pay attention to how “proper” someone/some system implements a certain Standard if they claim to do so. Sometimes this happens by best intentions, by accident or by strategic intentions.

In any case, you may have to deal/suffer from the consequences.

Data Structure Matters

> *“classical” catalogue = flat, table/cell thinking.*

- Better: Identify **Entities** and **Attributes**.
- Then: Identify **Relationships** (and their “`@`”)
- Entity Relationship Model (ERM)
- Think in “Semantic Triples”:

  `Subject - Predicate - Object`

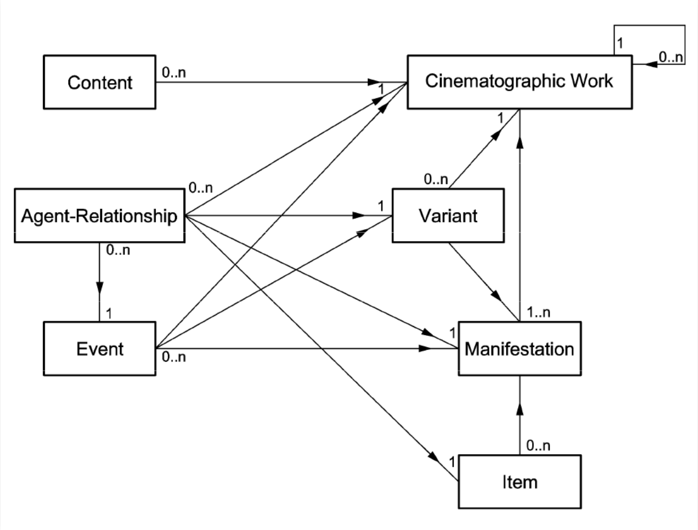
Example elgrito.witness.org

Speaker notes

Inspiration for terms, but IMO not useful because the relationships combine more than 1 thing, therefore their approach is by definition built to lead to thousands of “terms”, which actually describe too much - and cannot be combined and re-used properly: rda-registry.info

Good idea to put terms in a version control system - and publish them. Like this: RDA registry on Github

Another Example: EN 15909



Great! Now what about their “film-related materials” (posters, documents, etc)?

Relationships

- **Forward**: Agent1 `is parent` of Agent2
- **Reverse**: Agent2 `is child` of Agent1

Speaker notes

How (and if) a MAM handles relationships, their terms, directions - and their attributes - plays a major role, not only for modelling your data, but especially for search/retrieval and other future usage of whatever you've put in there.

Different “entites”

Relationships: @Attributes

> *(How well) can the system handle/search relationship-attributes?*

Examples:

- Which Work(s) feature which Agent(s) under the alias of …?
- List all … related to a work, in order of their …?

Oh, btw: “Says who?”

- Declare information sources?
- At which level?
- What if you have ≥ 1 sources?
- Who edited which field when and why?
- Could a source be “linked” (as relationship)?

Content annotations

- Regular descriptive metadata
- Image annotations?
- Time-based annotations?
- Import/Export of content annotations?
- International Image Interoperability Framework (IIIF) support?

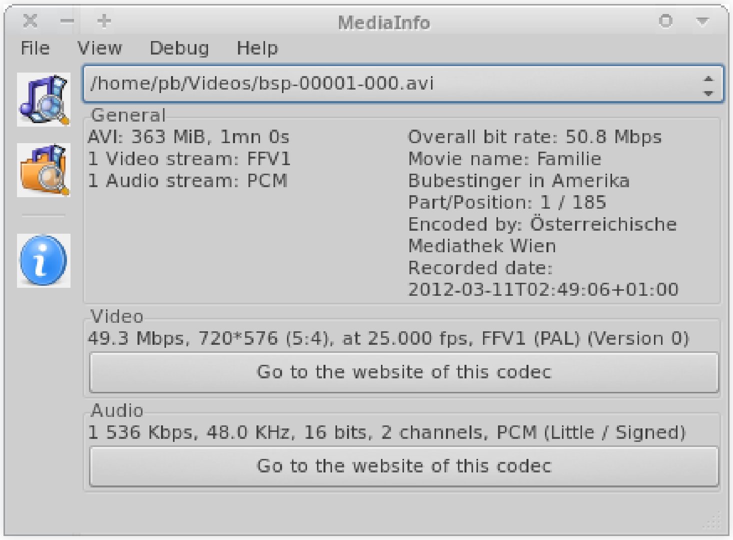
Reference between Catalogue and Files

- How are catalogue entries related to files (and vice-versa)?
- How (in)dependent of the catalogue can files be used/modified?

Files intact, but MAM gone?

Needle in a haystack?

Store identifying metadata


MediaInfo
File
View
Debug
Help
Go to the website of this codec
Go to the website of this codec

```
/home/pb/Videos/bsp-00001-000.avi
General
AVI: 363 MiB, 1mn 0s                Overall bit rate: 50.8 Mbps
1 Video stream: FFV1                Movie name: Familie
1 Audio stream: PCM                 Bubestinger in Amerika
                                    Part/Position: 1 / 185
                                    Encoded by: Österreichische
                                    Mediathek Wien
                                    Recorded date:
                                    2012-03-11T02:49:06+01:00
Video
49.3 Mbps, 720*576 (5:4), at 25.000 fps, FFV1 (PAL) (Version 0)
Audio
1 536 Kbps, 48.0 KHz, 16 bits, 2 channels, PCM (Little / Signed)
```

Speaker notes

The following may improve finding and/or re-organizing your files independent of the asset management system that has been used to allocate them before:

Store some identifying metadata in+next to your files:

- Minimalistic inside the file: Textual data that is most likely *not* to change over the lifetime of the file. Good practice: title, identifier, institution name, digitization date
- More metadata next to the content files (sidecar).
    - Store images as images.
    - Store textual data as structured text (XML is text). Makes it easier to modify, augment and migrate this data, but leave the large files in place.
- As simple as possible, as complicated as necessary.

Failsafe options

- Backup database.
- Backup storage.
- Setup MAM as Virtual Machine.
- Test export options.
- Documented export data format.
- Documentation & Source Code.

Does it do Web?

- Is the client browser-based?
- How hard/easy/possible to serve a web frontend?
  (APIs, components, programming language, …)
- Which operating system(s) are used/required?
- Does it provide eg a "REST API"?

Speaker notes

A classic example:

If the MAM client is browser-based, and the server engine runs on GNU/Linux OS with standard (=well known, well documented, well supported, FOSS-licensed) components, and provides proper and documented interfaces, having or implementing a web-interface option is way easier than if it requires a less-web-thinking OS like Windows or proprietary licenses or proprietary interfaces.

Some MAMs provide a “backend” for in-house use, and a separate “frontend” for web-access already.

From Local Catalogue to Web Access

> *Which level of separated/interconnected do you want/need?*

- 1 common database (DB)?
- 2 DBs, synchronized: how? + how often?
- What about the content (files)?
- Who may access (read/write/edit) what?

Access Control

Users, Roles, Rights, etc.

- How important for your collection/setup?
- How seamless? (SSO, LDAP, other?)
- How secure?
- How hard to maintain?
  (SysAdmin effort/needs/overhead/headaches?)

Does it scale?

- Speed? (search, retrieve, data, files, etc)
- Cache & tuning options?
- Synchronize multiple instances?
- Ongoing import/export? (in-house, external)
- License conditions and impact?
- Multi data schema in parallel?
- Multi language support?
- Migration to a “better scaling” system when needed?

Got Support?

- Whom to call in case of …?
- Option to switch/choose “your mechanic”?
- Which conditions?
- Support contracts?
- Which license model?

Summary of Criteria / Considerations

- CoVocs: Ability to have ID and label?
- Multiple MD-schemas in parallel? e.g. Different entry types: film, image, documents, physical objects, etc
- Can it be used to implement a certain Standard?
- Import/Export abilities (data and files)?
- Field mapping/translation?
- Content (file) storage layout? (naming, folder structure, links)
- Complexity of hardware/software setup? (For setup, maintenance and migration)
- What if your MAM is gone/broken?
- Artificial restrictions for future support/migrations?
- Access levels: Who may access/edit/etc which parts?
- Browser-based vs local client application
- Got support?

In practice: Try it!

- Get access to a demo installation.
- Try to import/export lists/CoVocs.
- Consider import/export with *your* data (layout).
- *Before* you make a final choice.
- Get cataloguers & technicians together.

The End

Questions? Comments?

Peter Bubestinger-Steindl

`Peter @ ArkThis.com`

Links

- Semantic Triple
- Resource Description Framework (RDF)
- Metadata & ERM Intro (witness.org)
- Entity Relationship Model (ERM)